# Introduction

The Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-5) was held on 23 June 2011 following the ACL HLT 2011 conference in Portland, Oregon. Like the first four SSST workshops in 2007, 2008, 2009, and 2010, it aimed to bring together researchers from different communities working in the rapidly growing field of structured statistical models of natural language translation.

During these past five years, statistical machine translation research has seen a movement toward not only tree-structured and syntactic models incorporating stochastic synchronous/transduction grammars, but also increasingly semantic models. There is no doubt that issues of deep syntax and shallow semantics are closely linked, and this encouraging trend has been reflected at recent SSST workshops. Semantic SMT research now includes context-dependent WSD (word sense disambiguation) for SMT (Carpuat and Wu 2007, 2008; Chan, Ng and Chiang 2007; Giménez and Màrquez 2007); SRL (semantic role labeling) for SMT (Wu and Fung 2009); and SRL for MT evaluation (Lo and Wu 2010, 2011).

In order to emphasize structure and representation at semantic and not only syntactic levels, "Semantics" has been explicitly added to the name of this year's Workshop (the acronym remains SSST), and is a special workshop theme.

We selected 15 papers for this year's workshop. Many either directly fall under the special theme of Semantics in SMT, or span the area between deep syntax and shallow semantics, illustrating the variety of semantic representations and models that are relevant to current statistical MT.

SRL predicate-argument structure clearly emerges as a useful representation for many aspects of SMT and MT evaluation. Wu and Palmer show that it is possible to automatically learn accurate cross-lingual SRL mappings between Chinese and English SRL annotated bitext. Input-side SRL is used to define reordering rules for Chinese-English word alignment (Meyers, Kosaka, Liao and Xue), and to improve pairwise translation hypothesis ranking (Pighin and Màrquez). Output-side SRL informs rule extraction in hierarchichal phrase-based SMT (Gao and Vogel), and provides structure for meaningfully comparing translation hypotheses and references in MT evaluation (Lo and Wu).

WSD also emerges as a prominent research direction with semantically richer SMT models designed to address ambiguity in translation lexical choice. Banchs and Costa-jussa use Latent Semantic Indexing to build a context-dependent phrase-based SMT model. Jiang, Du and Way integrate input paraphrases into SMT via confusion networks. Lefever and Hoste show that dedicated classifiers learned on parallel corpora outperform phrase-based SMT on a cross-lingual WSD task. SMT can also be seen as a tool to enrich semantic resources: McCrae, Espinoza, Ponsoda, Aguado-de-Cea and Cimiano propose several strategies for automatically translating ontologies and taxonomies, leveraging their rich semantic structure to compensate for the weakness of standard text translation methods.

A rich range of syntactic and tree-based approaches for learning translation rules is also seen. Attardi, Chanev and Miceli Barone learn reordering rules for a decoding approach drivenby a input-side dependency parser to guide reordering. Hanneman and Lavie describe a method for inducing nonterminals in synchronous/transduction grammars, by clustering nonterminal-pairs across input and output languages. Na and Lee propose a method for encoding alternative binarizations of a single input-side dependency tree into a forest by merging vertices before extracting translation rules. Hanneman,

Burroughs and Lavie extract synchronous/transduction grammar rules combining input-side and output-side parse tree information with the highly lexicalized approach of hierarchical phrase-based methods. Input-side parse features are incorporated within a maximum-entropy reordering approach by Xiang, Ge and Ittycheriah. On the formal side, Saers and Wu show how to simplify calculation of rule expectations for expectation-maximization training of transduction grammars as well as monolingual grammars, by reifying rules directly into the hypergraph representation of a deductive system so that a rule becomes an extra child rather than meta-information of a hyperedge.

Thanks once again this year are due to our authors and our Program Committee for making the SSST workshop another success.

Dekai Wu, Marianna Apidianaki, Marine Carpuat, and Lucia Specia