

Artificial Intelligence Master (UPC-URV-UB), course 2010-2011

Complementary Artificial Intelligence Seminar 2 (SEM-IA2), 2010-2011

Topic: Syntax, Semantics and Structure in Statistical Translation

Lecturer: Prof. Dekai Wu

Human Language Technology Center

Department of Computer Science and Engineering

Hong Kong University of Science and Technology

<http://www.cs.ust.hk/~dekai/>

Dates: May 2-6, 2011, from 10-13

Place: Building XXXXX Room YYYYY. Universitat Politècnica de Catalunya-BarcelonaTech, Campus Nord, Jordi Girona 1-3, 08034 Barcelona

Contents/Abstract: Over the past twenty years, we have attacked the historical methodological barriers between statistical machine translation and traditional models of syntax, semantics, and structure. In this series of lectures, we will survey some of the central issues and techniques from each of these aspects, with an emphasis on 'deeply theoretically integrated' models, rather than hybrid approaches such as superficial statistical aggregation or system combination of outputs produced by traditional symbolic components. On syntactic SMT, we will explore the trade-offs for SMT between learnability and representational expressiveness. After establishing a foundation in the theory and practice of stochastic transduction grammars, we will examine very recent new approaches to automatic unsupervised induction of various classes of transduction grammars. We will show why stochastic linear transduction grammars (LTGs and LITGs) and their preterminalized variants (PLITGs) are proving to be particularly intriguing models for the bootstrapping of inducing full-fledged stochastic inversion transduction grammars (ITGs). On semantic SMT, we will explore the trade-offs for SMT involved in applying various lexical semantics models. We will first examine word sense disambiguation, and discuss why traditional WSD models that are not deeply integrated within the SMT model tend, surprisingly, to fail. In contrast, we will show how a deeply embedded phrase sense disambiguation (PSD) approach succeeds where traditional WSD does not. We will then turn to semantic role labeling, and discuss the challenges of early approaches of applying SRL models to SMT. Finally, on semantic MT evaluation, we will explore some very new human and semi-automatic metrics based on semantic frame agreement. We show that by keeping the metrics deeply grounded within the theoretical framework of semantic frames, the new HMEANT and MEANT metrics can significantly outperform even the state-of-the-art expensive HTER and TER metrics used by DARPA, while at the same time maintaining the desirable characteristics of simplicity, inexpensiveness, and representational transparency.