Wide-coverage NLP with Linguistically Expressive Grammars

Julia Hockenmaier
Department of Computer Science,
University of Illinois
`juliahmr@illinois.edu`

Yusuke Miyao
National Institute of Informatics
`yusuke@nii.ac.jp`

Josef van Genabith
Centre for Next Generation Localisation,
School of Computing,
Dublin City University
`josef@computing.dcu.ie`

1 Introduction

In recent years, there has been a lot of research on wide-coverage statistical natural language processing with linguistically expressive grammars such as Combinatory Categorial Grammars (CCG), Head-driven Phrase-Structure Grammars (HPSG), Lexical-Functional Grammars (LFG) and Tree-Adjoining Grammars (TAG). But although many young researchers in natural language processing are very well trained in machine learning and statistical methods, they often lack the necessary background to understand the linguistic motivation behind these formalisms. Furthermore, in many linguistics departments, syntax is still taught from a purely Chomskian perspective. Additionally, research on these formalisms often takes place within tightly-knit, formalism-specific subcommunities. It is therefore often difficult for outsiders as well as experts to grasp the commonalities of and differences between these formalisms.

2 Content Overview

This tutorial overviews basic ideas of TAG/CCG/LFG/HPSG, and provides attendees with a comparison of these formalisms from a linguistic and computational point of view. We start from stating the motivation behind using these expressive grammar formalisms for NLP, contrasting them with shallow formalisms like context-free grammars. We introduce a common set of examples illustrating various linguistic constructions that elude context-free grammars, and reuse them when introducing each formalism: bounded and unbounded non-local dependencies that arise through extraction and coordination, scrambling, mappings to meaning representations, etc. In the second half of the tutorial, we explain two key technologies for wide-coverage NLP with these grammar formalisms: grammar acquisition and parsing models. Finally, we show NLP applications where these expressive grammar formalisms provide additional benefits.

3 Tutorial Outline

1. Introduction: Why expressive grammars

2. Introduction to TAG

3. Introduction to CCG

4. Introduction to LFG

5. Introduction to HPSG

6. Inducing expressive grammars from corpora

7. Wide-coverage parsing with expressive grammars

8. Applications

9. Summary

References

Aoife Cahill, Michael Burke, Ruth O'Donovan, Stefan Riezler, Josef van Genabith and Andy Way. 2008. Wide-Coverage Deep Statistical Parsing using Automatic Dependency Structure Annotation. *Computational Linguistics*, 34(1). pp.81-124, MIT Press.

Yusuke Miyao and Jun'ichi Tsujii. 2008. Feature Forest Models for Probabilistic HPSG Parsing. *Computational Linguistics*, 34(1). pp.35-80, MIT Press.

Julia Hockenmaier and Mark Steedman. 2007. CCGbank: A Corpus of CCG Derivations and Dependency Structures Extracted from the Penn Treebank. *Computational Linguistics*, 33(3). pp.355-396, MIT Press.