

Natural Language as the Basis for Meaning Representation and Inference

Ido Dagan¹, Roy Bar-Haim¹, Idan Szpektor¹, Iddo Greental², and Eyal Shnarch¹

¹ Bar Ilan University, Ramat Gan 52900, Israel,
dagan@cs.biu.ac.il

² Tel Aviv University, Tel Aviv 69978, Israel

Abstract. Semantic inference is an important component in many natural language understanding applications. Classical approaches to semantic inference rely on logical representations for meaning, which may be viewed as being “external” to the natural language itself. However, practical applications usually adopt shallower lexical or lexical-syntactic representations, which correspond closely to language structure. In many cases, such approaches lack a principled meaning representation and inference framework. We describe a generic semantic inference framework that operates directly on language-based structures, particularly syntactic trees. New trees are inferred by applying entailment rules, which provide a unified representation for varying types of inferences. Rules were generated by manual and automatic methods, covering generic linguistic structures as well as specific lexical-based inferences. Initial empirical evaluation in a Relation Extraction setting supports the validity and potential of our approach. Additionally, such inference is shown to improve the critical step of unsupervised learning of entailment rules, which in turn enhances the scope of the inference system.

This paper corresponds to the invited talk of the first author at CICLING 2008.

1 Introduction

It has been a common assumption that the structure of natural language is not suitable to formally represent meanings and to conduct inferences over them. Indeed, according to the traditional formal semantics approach inference is conducted at a logical level. Texts are first translated, or *interpreted*, into some logical form and then new propositions are inferred from interpreted texts by a logical theorem prover. Meaning and inference are thus captured by representations that are “external” to the language itself, and are typically independent of the structure of any particular natural language.

However, practical text understanding systems usually employ shallower lexical and lexical-syntactic representations, which clearly correspond to the structure of the particular natural language being processed. Such representations are sometimes augmented with partial semantic annotations like word senses,