

A Modular Account of Information Structure in Extensible Dependency Grammar

Ralph Debusmann¹, Oana Postolache², and Maarika Traat³

¹ Programming Systems Lab, Saarland University, Saarbrücken, Germany
rade@ps.uni-sb.de

² Computational Linguistics, Saarland University, Saarbrücken, Germany
Al. I. Cuza University, Iași, Romania
oana@coli.uni-sb.de

³ Informatics, University of Edinburgh, Scotland
M.Traat@sms.ed.ac.uk

Abstract. We introduce a modular, dependency-based formalization of Information Structure (IS) based on Steedman’s prosodic account [1, 2]. We state it in terms of Extensible Dependency Grammar (XDG) [3], introducing two new dimensions modeling 1) prosodic structure, and 2) theme/rheme and focus/background partitionings. The approach goes without a non-standard syntactic notion of constituency and can be straightforwardly extended to model interactions between IS and other dimensions such as word order.

1 Introduction

Information Structure (IS) is the way in which people organize their utterances. Usually, in an utterance there is a part that links the content to the context, and another that advances the discourse by adding or modifying some information. IS is an important factor in determining the felicity of an utterance in a given context. Among the many applications where IS is of crucial importance are content-to-speech systems (CTS), where IS helps to improve the quality of the speech output [4], and machine translation (MT), where IS improves target word order, especially that of free word order languages [5].

In this paper we present a modular, dependency-based account of IS based on Steedman’s prosodic account of IS for Combinatory Categorical Grammar (CCG) [1, 2]. Similarly to Steedman, we establish a bi-directional correspondence between IS and prosodic structure, i.e. when the IS is known, we can determine the prosodic structure (e.g. in CTS systems), and when we have the prosodic information, we can extract the IS (e.g. to augment dialog transcripts).

We state our approach in terms of Extensible Dependency Grammar (XDG) [3], which allows us to take a modular perspective on IS. We distinguish three notions of constituency: syntactic, prosodic, and information structural, which are related, but not identical. Thus, differently from Steedman, we can decouple syntax from information structure, and do not assume non-standard syntactic constituents. By this, we can monotonically add IS to existing XDG grammars.