Review of Master Thesis

Author: Max Jakob
Title: Mapping the Prague Dependency Treebank Annotation Scheme onto Robust Minimal Recursion Semantics
Supervisor: RNDr. Markéta Lopatková, Ph.D.

The goal of the thesis is to investigate a correspondence between two semantic formalisms, the dependency-based tectogrammatical layer of the Prague Dependency Treebank (PDT) and the logic-based Robust Minimal Recursion Semantics (RMRS). The author has developed a mapping algorithm that converts PDT tectogrammatical trees into RMRS structures. For each node of the dependency tree, a *node-RMRS* is created. These node-RMRSs are combined into final RMRS structure on the basis of compositional roles formulated by the author.
The thesis consists of five chapters; it includes a rich list of references, list of figures and list of tables. The attached CD contains text of the thesis, detailed documentation, tools used by the author and scripts written by the author.

After a short introduction, the first chapter briefly mentions related works, namely English Resource Grammar and Verbmobil project, a project focusing on mapping generic logical forms onto RMRs structures, and investigation of correspondence of PDT and categorical grammars.
The second chapter provides a reader with necessary information on PDT (pp. 5-16). The author manages to clearly describe basic aspects of PDT formalism (stratificational annotation; fundamental features of tectogrammatical layer – concepts of dependency relation and effective relation as well as coordination; valency dictionary). Then an essential introduction to (R)MRS follows (pp. 16-25). The author offers a comprehensive and clear explanation of basic concepts of this formalism (esp. elementary predications EPs, their configurations, MRS graphs, hooks and constraints); this introductory description of RNRS can be considered as one of the interesting gains of the project.
The extensive third chapter dealing with a correspondence between the two formalisms represents a core part of the thesis (pp. 26-62). The author describes step by step the construction of a final RMRS: creating RMRS nodes for each of tectogrammatical nodes, assigning functional nodes, constructing lexical EPs and finally combining node-RMRSs into a final RMRS structure. Special attention is devoted to valency and free modifications and to coordination, especially with respect to the effective relation – elaborated rules are formulated for adequate translation of PDT-like dependency structures onto RMRS structures. The description is accompanied with a number of illustrative examples, which allow a reader to follow a non-trivial correspondence of the involved formal structures. The author also carefully lists all phenomena skipped during the conversion as well as lost information. The chapter ends with description of an algorithm designed for the conversion.
The fourth chapter focuses on the evaluation of output RMRS structures. Two basic structural criteria are used: so called net criterion and criterion of existence of (at least one) configuration for a valid MRS structure. The author discusses the gained results; he proves his deep insight into the problem. (A manual checking was not possible as there are no Czech data for RMRS available.)
The last chapter concludes the text and mentions possible future work.

The thesis is clearly written, with no gaps in argumentation. It is written in good English, I have not detected any serious errors or other formal imperfections.

Conclusion

The reported thesis brings new interesting results concerning the correspondence between the two semantic formalisms. It proves the author's ability to solve independently and creatively assigned tasks in the area of NLP and to clearly formulate his goals, used methods and the gained results.

The thesis complies with the requirements for Master Thesis at MFF. I recommend to accept the thesis for the defense.

Prague, January 21, 2010
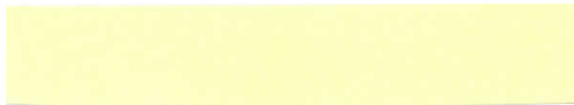
RNDr. Markéta Lopatková, Ph.D.
Institute of Formal and Applied Linguistics
Charles University in Prague