Report: Short Term Scientific Mission

COST-STSM-ECOST-STSM-IC1207-130114-039056

Simon Petitjean

February 21, 2014

1 Purpose of the STSM

The goal of the mission was to formalize and implement the integration of a frame semantics module into the XMG compiler. eXtensible MetaGrammar is a tool used for grammar engineering, which allows to reason about the language at a more abstract level.

One of the main concepts of XMG is the concept of dimensions. They are used to separate the different levels of linguistic description (syntax, semantics...). The purpose of the STSM was to integrate a frame semantics dimension, which would allow to generate a Tree Adjoining Grammar where each tree is paired with a frame.

2 Description of the work carried out during the STSM

The extensibility of XMG allows to add new levels of linguistic representation, with dedicated description languages and compiling modules (like solvers). We were concerned with the design of a dimension for frame representations, i.e. typed feature structures. The work therefore consisted of the following steps:

The first step was to choose a description language for typed feature structures, adapted to describe frame fragments like the ones proposed in [1]. The unification of two typed feature structures implies the unification of their types, according to general type constraints or a given type hierarchy. The type unification rules used in [1] are given by the means of general feature structure constraints.

The second step was to implement these constraints into the XMG processor, compute the set of valid types, and handle type unification.

The last step was to connect these compiling phases together and write a toy metagrammar to illustrate some examples of frame fragments combination, including multi word expressions.

This work was done thanks to constant collaboration with Timm Lichte and regular meetings with Laura Kallmeyer and Rainer Osswald (about one meeting every two days). We also met Christof Rumpf, who gave us some useful pieces of advice about type encoding in Prolog.

3   Description of the main results obtained

The main result is the integration of a set of type unification constraints and of a frame description language into XMG. Until now, these modules have been incorporated into an operative prototype which was validated on a toy metagrammar.

4   Future collaboration with host institution

Collaboration remains active. A further visit is planned for the end of 2014. Future collaboration will also address the treatment of MWE in metagrammars. We have already begun with preparatory work in this area.

5   Foreseen publications/articles resulting or to result from the STSM

The integration of frame semantics into a metagrammar compiler is the subject of an article which will be submitted to a special issue of the Journal of Language Modelling on High-level Methodologies for Grammar Engineering.

6   Confirmation by the host institution of the successful execution of the STSM

Timm Lichte: I happily confirm that the execution of the STMS was successful from our point of view. The collaboration with Simon Petijean was

very focused and fruitful over the whole two weeks of his stay. I doubt very much that similar results could have been achieved in the same period of time under remote collaboration.

References

[1] Laura Kallmeyer and Rainer Osswald. Syntax-driven semantic frame composition in Lexicalized Tree Adjoining Grammar. *Journal of Language Modelling*, 1:267–330, 2013.