

Report: Short Term Scientific Mission

COST Office Science Officer : MR RALPH STUEBNER, ralph.stuebner@cost.eu

COST MC Chair : DR AGATA SAVARY, agata.savary@univ-tours.fr

COST STSM Manager : Pr. CVETANA KRSTEV, cvetana@matf.bg.ac.rs

COST STSM Reference Number : COST-STSM-IC1207-20132

Period : 2014-05-19 to 2014-05-23

COST Action : IC1207

STSM type : Regular (from France to Germany)

STSM Applicant : Dr Yannick PARMENTIER, University of Orléans, Orléans(FR),
yannick.parmentier@univ-orleans.fr

STSM Topic : Metagrammars and multi-word expressions

Host : Laura KALLMEYER, Heinrich-Heine-Universität Düsseldorf, Düsseldorf(DE),
kallmeyer@phil.uni-duesseldorf.de

1 Purpose of the visit

In a previous STSM by Simon Petitjean (January 2014), a collaborative work concerning formal grammar description (so-called metagrammar) and multi-word expressions has been started. More concretely, Simon Petitjean's STSM allowed for the extension of the XMG language [Crabbé et al., 2013] in order to describe semantic representations in the Frame Semantics formalism (see e.g. [Kallmeyer and Osswald, 2012]).

This STSM proposal can be seen as a follow-up of Simon Petitjean's STSM, with some additional objectives. Prof. Laura Kallmeyer's group has been working extensively on the use of both Lexicalised Tree Adjoining Grammar (LTAG) and Role and Reference Grammar (RRG) to represent natural language. While LTAG offers a formally appealing framework (LTAG languages are polynomially parsable), RRG offers a linguistically-rich framework where one can express both syntactic and semantic constraints. While offering a rich framework, RRG lacks some formal specification. It has been shown [Kallmeyer et al., 2013] that such a specification can be obtained by converting RRG instances into equivalent Tree Wrapping Grammars (TWG), which could in turn be described via metagrammars. The objectives of this STSM is to extend the concept of metagrammar to RRG (potentially by using TWG as a pivot formalism). While doing this, a particular attention will be paid to MWEs.

2 Description of the work carried out

Several meetings took place between the beneficiary of the STSM and members of the host institute (namely, Laura Kallmeyer, Timm Lichte, Wolfgang Maier, and Rainer Osswald). These allowed to start ongoing discussions on the syntactic and semantic representation of MWEs within LTAG and RRG.

As an example of a syntactic-semantic representation of the MWE *to kick the bucket*, see Figure 1 below. From this Figure, one can see that the syntax / semantics interface consists of unification variables, which allow to bind syntactic and semantic information (namely syntactic nodes with semantic arguments).

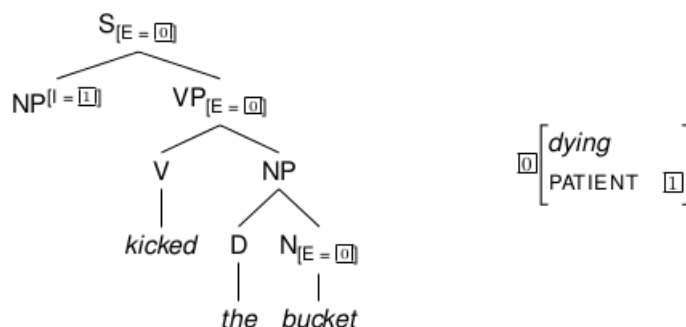


Figure 1: LTAG representation of *kicked the bucket* using frame semantics

We first looked at those idiomatizations, which exhibits some syntactic variability, such as the famous example *to spill the beans*. A representation of such expressions in LTAG has been proposed by [Abeillé and Schabes, 1989]. In their work, the authors propose to associate such an expression with a specific elementary LTAG tree, and to use adjunction constraints to prevent unwanted modifications. From a semantic point of view, the proposed representation of idiomatizations is somehow problematic. Indeed, it leads to atomic semantic representations where modifiers cannot have a narrower scope than the full expression. In other words, in Figure 1, the semantic frame does not allow for the further semantic modification of the internal NP.

To illustrate this issue, consider the following usages of *to spill the beans*:

- a) John spilled the beans.
- b) John spilled the proverbial beans.
- c) John spilled the political beans.
- d) John spilled the juicy beans.

In example a) above, one can find the two classical readings, the action of throwing beans, and the revealing of a secret. Both readings could be represented via distinct elementary LTAG trees. b) allows for the disambiguation of a), making only the second reading valid. Interestingly, when looking at c), one may want to express the fact that the revealed secret has something to do with

politics. How to do this with the classical LTAG approach ? Example d) is even trickier, since the adjective juicy usually applies on food, here it probably refers to some sensitive secret.

This example illustrates the very interesting discussions led during the STSM, and which took most of the time. The work around metagrammars has thus been delayed and will take place via other means (e.g. skype meetings).

3 Description of the main results obtained

Among the outcomes of this one-week STSM, one may cite:

- a better understanding of the semantic issues arising when representing MWEs in LTAG / RRG,
- concrete ideas for a future and ongoing collaboration (e.g. further extensions of the XMG language, implementations of MWE-aware metagrammars for French and German),
- a concrete introduction to the several members of Prof. Laura Kallmeyer's group, who may contribute to this work around the metagrammar-based representation of MWE syntactic and semantic properties.

4 Future collaboration with host institution

During this STSM, a meeting has been specifically dedicated to the opportunity to propose a joint-project for next year's ANR-DFG call for project proposals.¹

A first draft of a project proposal around metagrammar-based representations of syntax and semantics and their parsing, has been designed. Further work on this proposal will be carried out remotely in the coming months.

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

Publications involving co-authors from Prof. Kallmeyer's group in Düsseldorf and from Dr. Yannick Parmentier's group in Orléans are expected in a medium term (especially if the above-mentioned joint-project is selected for funding).

6 Confirmation by the host of the successful execution of the mission

Prof. Laura Kallmeyer, Universität Düsseldorf:

The STSM can be considered as very successful, as it made it possible for both institutions to reinforce existing contacts via fruitful discussions, and concrete plans for the future. Yannick Parmentier's visit was a chance for us to present

¹The French and German Research Agencies (named respectively ANR and DFG) support a common program to promote collaborative research in Humanities, see <http://www.agence-nationale-recherche.fr/suivi-bilan/sciences-humaines-et-sociales/fral/>.

our formal syntactic and semantic representation of MWEs to Yannick while benefiting from his feedback on a high-level encoding of these representations within XMG-based metagrammars. This visit also made it possible to define the outline of a future joint project proposal, where the research interests of our respective groups would meet.

7 Financial summary

Travel expenses : 276.30 EUR (266.90 EUR for return train tickets from Orléans to Düsseldorf, and 9.40 EUR for local public transportation)

Living expenses : 468 EUR (368 EUR for 4 nights –single room– including breakfast, and 20 EUR per day for meals)

The estimated cost of the STSM was 700 EUR, the real cost is 744.30 EUR.

8 Other comments

The beneficiary is grateful to the members of the host institution for the very enriching discussions.

References

- [Abeillé and Schabes, 1989] Abeillé, A. and Schabes, Y. (1989). Parsing Idioms in Lexicalized TAGs. In *Proc. of the 4th EACL*, pages 1–9, Manchester, UK.
- [Crabbé et al., 2013] Crabbé, B., Duchier, D., Gardent, C., Le Roux, J., and Parmentier, Y. (2013). Xmg: extensible metagrammar. *Computational Linguistics*, 39(3):591–629.
- [Kallmeyer and Osswald, 2012] Kallmeyer, L. and Osswald, R. (2012). A frame-based semantics of the dative alternation in lexicalized tree adjoining grammars. *Empirical Issues in Syntax and Semantics*, 9:167–184.
- [Kallmeyer et al., 2013] Kallmeyer, L., Osswald, R., and Van Valin, Jr., R. D. (2013). Tree wrapping for role and reference grammar. In Morrill, G. and Nederhof, M.-J., editors, *Formal Grammar. 17th and 18th International Conferences, FG 2012 Opole, Poland, August 2012, Revised Selected Papers*, number 8036 in Lecture Notes in Computer Sciences, pages 175–190. Springer.