



Research Networking Programmes

Short Visit Grant or Exchange Visit Grant

(please tick the relevant box)

Scientific Report

The scientific report (WORD or PDF file – maximum of eight A4 pages) should be submitted online within one month of the event. It will be published on the ESF website.

Proposal Title: Complex words and semantic transparency

Application Reference N°: 4677

1) Purpose of the visit

The aim of this visit was to build on the success of two previous short visits made by Martin Schäfer (the applicant) to Melanie Bell (the host), and to further develop our joint work on semantic transparency. Semantic transparency is known to play an important role in the storage and processing of complex words (e.g. Marslen-Wilson et al 1994) and human raters of transparency achieve high levels of agreement (e.g. Frisson et al 2008). The aim of our work is to understand the factors that contribute to this perceived transparency. Bell & Schäfer (2013) [funded by Networks grant 4677] modelled the transparency both of compound words and of individual compound constituents, and showed that shifted word senses reduce perceived transparency, while certain semantic relations between constituents increase it. However, this finding is problematic in at least two ways. Firstly, it is not clear whether there is a solid basis for establishing whether a specific word sense is shifted or not. Secondly, work on conceptual combination by Gagné and collaborators has shown that relational information in compounds is accessed via the concepts associated with individual modifiers and heads, rather than independently of them (e.g. Spalding et al 2010 for an overview). This led us to the hypothesis that it is not whether a specific word sense is etymologically shifted, nor whether a specific semantic relation is used

per se, that makes a compound constituent more or less transparent; rather, it is the degree of expectedness of a particular word sense and a particular relation for a given constituent. Bell & Schäfer (2014) [funded by Networks grant 6520] provided evidence for this hypothesis, at least for the modifier constituents (N1) of English compound nouns. To model the expectedness of word senses and semantic relations for a given modifier, we used the constituent families of the compounds in our dataset. We calculated the proportion of compound types in each constituent family with each semantic relation (after Levi 1978) and each WordNet sense of the constituent in question (Princeton 2010). We take these two measures to reflect the expectedness of the respective relations and WordNet senses of the constituents: if a relation or sense occurs in a high proportion of the constituent family, it is more expected. In our final model for N1 transparency, the expectedness of specific word senses and semantic relations for N1 are significant predictors: the more expected the relation or sense, the more transparent the constituent is perceived to be. More intriguingly, they are also significant predictors for N2 transparency, suggesting that humans rating the transparency of individual compound constituents make use of the local compound context. We are currently working on the final piece of our modelling: the incorporation of the corresponding predictors for the head noun (N2). We hypothesise that the inclusion of these predictors will improve our model of whole compound transparency. Furthermore, we want to test whether N2 predictors affect judgements of N1 transparency, since this will have implications for our understanding of how compounds are interpreted.

The short trip had two immediate aims: to finalise our new model for semantic transparency in order to incorporate it in Schäfer & Bell (in prep.) and to write an abstract for the 2015 NetWordS conference.

2) Description of the work carried out during the visit

At the beginning of the visit, Martin Schäfer finished the coding of the N2 families in our dataset. We had now the fully coded N1 and N2 constituent families of the dataset described in Reddy et al (2011) at our disposal, so our first step was to calculate the missing proportional variables for the N2 data. In order to make our findings more easily reproducible, we decided to carry out the necessary calculations with the help of an R-script, and not, as previously, using MS Excel. After writing the script, we tested it against our previous excel-based calculations for the N1 constituent families. We then revisited all the methodological steps we had taken so far, in particular focusing on the problem of dealing with plural and singular wordforms in our data. In extracting compound families, we took all strings of exactly two nouns that follow an article in the British National Corpus and which also occur four times or more

in the USENET corpus (Shaoul & Westbury 2010). We did not lemmatize the forms in the process, with the result that in our final data we had frequencies for singular and plural wordforms of a given noun noun combination only if these forms already occurred in the BNC. We decided that it would be advantageous to have data from the USENET corpus on all singular and plural wordform permutations, independent of their occurrence in the BNC. The inclusion of this further data necessitated some additional steps, in particular the preparation of these wordform permutations on the basis of the BNC data, and furthermore, getting the frequencies for this data from the USENET corpus. While these two latter steps could not be done in the only 4 full days of the visit, we agreed on an ambitious timetable to get this done as soon as possible.

3) Description of the main results obtained

The main outcomes of the visit were

- Completion of data coding
- Revision of methodology to be both more rigorous and more easily reproducible
- Production of R code for lemmatization
- Initial draft of methodology section of our paper to be submitted to morphology
- Abstract submitted for the final Networks conference

4) Future collaboration with host institution (if applicable)

We submitted an abstract to the Networks 2015 conference, hoping to be able to present our findings there. In addition, we are currently working on a journal article, reporting our findings. This article will be submitted to a special issue of *Morphology*.

In the longer term, we plan to extend the work to investigate contextual effects in the interpretation of compounds.

5) Projected publications / articles resulting or to result from the grant (*ESF must be acknowledged in publications resulting from the grantee's work in relation with the grant*)

Schäfer, Martin & Melanie J. Bell. Modelling semantic transparency. *Morphology*.

6) Other comments (if any)

We are extremely grateful for the funding through Networks - The European Network on Word Structure. This is the third time that a short visit grant has allowed Martin Schäfer time in Cambridge, and this has really benefitted us. Both the fruitful face-to-face discussions, and also the blocking of time for one dedicated purpose, have led to important advances in our work and our thinking.

References

Bell, M. J. & M. Schäfer 2013. Semantic transparency: challenges for distributional semantics. In Herbelot, A., R. Zamparelli & G. Boleda (eds.). Proc. of the IWCS 2013 Workshop, Potsdam: ACL, 1-10

Bell, M. J. & M. Schäfer 2014. Semantic transparency and constituent informativity in English compound nouns. Paper presented at LAGB 2014, Oxford

Frisson, S., E. Niswander-Klement, & A. Pollatsek (2008). The role of semantic transparency in the processing of english compound words. *Brit. J. of Psychology* 99(1), 87-107

Levi, J. N. 1978. The syntax and semantics of complex nominals. New York: Academic Press

Marslen-Wilson, W., L. K. Tyler, R. Waksler, & L. Older 1994. Morphology and meaning in the English mental lexicon. *Psychological review* 101(1), 3-33

Princeton University (2010) "About WordNet." WordNet. Princeton University. 2010

Schäfer, M. & Bell, M. J. (in prep.) Modelling semantic transparency. *Morphology*

Spalding, T. L., C. L. Gagné, Allison C. M. & H. Ji. 2010. Relation-based interpretation of noun-noun phrases: A new theoretical approach. In Olsen, S. (ed.). *New Impulses in Word-Formation*, 283-315