

## *On the semantics of noun compounds*

STAN SZPAKOWICZ<sup>1,2</sup>, FRANCIS BOND<sup>3</sup>,  
PRESLAV NAKOV<sup>4</sup> and SU NAM KIM<sup>5</sup>

<sup>1</sup>*School of Electrical Engineering & Computer Science, University of Ottawa, Canada*

<sup>2</sup>*Institute of Computer Science, Polish Academy of Sciences, Poland*

<sup>3</sup>*Division of Linguistics & Multilingual Studies, Nanyang Technological University, Singapore*

<sup>4</sup>*Qatar Computing Research Institute, Qatar Foundation, Qatar*

<sup>5</sup>*Faculty of Information Technology, Monash University, Australia*

The noun compound – a sequence of nouns which functions as a single noun – is very common in English texts. No language processing system should ignore expressions like *steel soup pot cover* if it wants to be serious about such high-end applications of computational linguistics as question answering, information extraction, text summarization, machine translation – the list goes on. Processing noun compounds, however, is far from trouble-free. For one thing, they can be bracketed in various ways: is it *steel soup*, *steel pot*, or *steel cover*? Then there are relations inside a compound, annoyingly not signalled by any words: does *pot* CONTAIN *soup* or is it FOR COOKING *soup*? These and many other research challenges are the subject of this special issue.

The volume opens with Preslav Nakov's survey paper on the interpretation of noun compounds.<sup>1</sup> It serves as an excellent, thorough introduction to the whole business of studying noun compounds computationally. Both theoretical and computational linguistics consider various formal definitions of the compound, its creation, its types and properties, its applications, and its approximation by paraphrases. The discussion is also illustrated by a range of languages other than English. Next, the problem of bracketing is given a few typical solutions. There follows a detailed look at noun compound semantics, including coarse-grained and very fine-grained inventories of relations among nouns in a compound. Finally, a "capstone" project is presented: textual entailment, a tool which can be immensely helpful in many high-end applications.

Diarmuid Ó Séaghdha and Ann Copestake tell us how to interpret compound nouns by classifying their relations with kernel methods. The kernels implement intuitive notions of lexical and relational similarity which are computed using distributional information extracted from large text corpora. The classification is tested at three different levels of specificity. Impressively, in all cases a combination of both lexical and relational information improves upon either source taken alone.

<sup>1</sup> The first authors of all four papers cut their teeth as scientists on noun compounds. ☺

Paul Nulty and Fintan Costello's work introduces techniques for ranking paraphrases of some of the vast number of possible semantic relations which can hold between nouns. A semi-supervised probabilistic method of ranking candidate paraphrases of relations is combined with a new method which selects plausible relational paraphrases at different levels of semantic specificity. These methods are motivated by the observation that existing relation classification schemes often exhibit a highly skewed class distribution, and that lexical paraphrases of semantic relations can vary widely in semantic precision: a single relation can be described in many ways.

A lexical-semantic method of interpreting and bracketing compounds is the subject of the paper by Su Nam Kim and Tim Baldwin. They develop an automatic method based on the use of semantic relations. It interprets the unseen compounds by measuring the lexical-semantic similarity – derived from *WordNet* – with known tagged compounds. The effectiveness of this method is demonstrated on the interpretation of two-term and three-term compounds. Finally, the paper shows that the interpretation method can boost the coverage and accuracy of noun compound bracketing.

So there you have it: A special issue on noun compound interpretation, with a comprehensive survey of the field and three papers which delve into a few demanding problems. Just dive in. You will not regret it.

Happy reading!

Stan Szpakowicz, Francis Bond, Preslav Nakov, and Su Nam Kim are guest editors.