

## **Introduction to this special issue on *Cognitive Aspects of Natural Language Processing***

Realising that no single discipline will ever suffice to reveal the functioning of the human mind, cognitive science was born. The aim from the outset was to analyse complex behaviour (speaking, thinking, problem-solving) from different angles and to build models that could account for this complexity.

This ambitious goal has seemed promising enough to attract the attention of numerous researchers. Given the inherent multifaceted nature of the object studied there are many fields concerned, but as the results are generally accessible only via the journals of the respective fields (linguistics, psychology, ergonomics, etc.), they are not always visible to the rest of the cognitive science community. Some researchers are not even fully aware of the cognitive relevance of their work. Natural Language Processing (NLP) is one of these fields. It has existed for decades and has produced many interesting results. Unfortunately, most researchers working in this field take a very narrow, engineering-centred point of view, focusing exclusively on applications, admitting sometimes not to be interested at all in cognition or the inner workings of the human brain.

Interestingly enough these same people are strongly inclined to evaluate their systems. Yet the ultimate measure for evaluating NLP systems can only be humans (how adequate is it to meet users' needs?). In other words, true optimization must take (at least partially, or at some point) the human dimension into account. So, even if researchers claim not to be interested in cognition, their optimized algorithms are in the end substitutes of cognitive processes (e.g. in machine translation).

Of course one can argue that there are a number of tasks which machines do perfectly well (arithmetics, chess playing, etc.) without having or needing any information concerning the human mind or the end user. But

why deal only with fully automated behavior? Why not build tools likely to assist people in their tasks? If one agrees that this latter goal is worth achieving, then one must take into account the needs and habits of the end user and the functioning of his/her mind.

While there is no doubt that evaluation is required at some point, it does not guarantee progress or insights per se. We might optimize algorithms at a local level, but this hardly brings us any closer to our goal. So a broader perspective is needed, and this includes the study of human cognition. To get the necessary insights and develop a bigger picture, we should adopt and integrate multiple viewpoints. In other words, we should take the approach of cognitive science. This is precisely the goal of the papers presented in this volume.

Nearly all of the contributors to this volume participated in a symposium entitled “Enhancing electronic dictionaries with the help of lexicographers, computer scientists and cognitive psychologists” which took place in Beijing in 2010 as part of the 7<sup>th</sup> International Conference of Cognitive Science. Thus, several of the papers in this volume deal with the lexicon, its structure (some kind of a semantic map), and usage (word access, i.e. navigation). Sierra and Hernandez present a methodology for creating a Lexical Knowledge Base to be used in an onomasiological dictionary, that is, a dictionary that allows meaning-based access to words. Their goal is to automate the creation of such dictionaries in order to help people overcome the *tip of the tongue problem*.

A similar goal is pursued by Zock and Schwab, who present various search methods rather than a single strategy solely based on the words occurring in a term’s definition. Having shown that storage does not guarantee access, they sketch a roadmap of how to build a resource (association-based index drawn from a corpus) likely to help users find the words they are looking for.

Okamoto and Ishizaki also rely on associations. They describe the principles of building and using an *Associative Concept Dictionary* in order to summarize texts or to disambiguate words. They compare the quality of the summarization with that of human participants as well as those based on conventional methods such as term frequencies. The results show that their method is better than the conventional one using term frequencies. Their

results for word sense disambiguation are also promising.

Chen et al. present work on automated affect recognition. Their annotation schema allows one not only to annotate explicit and implicit expressions of emotion, but also to encode their respective emotional levels. Taking the theories of cognitive psychologists into account, they present a system capable of automatically creating large, high-quality emotion corpora able to recognize emotions in written texts of Chinese or English. Moreover, as their annotation schema is generic, it may be applied to other languages as well.

Finally, Rapp proposes the hypothesis that languages are acquired by discovering, memorizing, and reproducing statistical patterns found in perceived language. He provides evidence for this with respect to word familiarities, word associations, and some aspects of meaning. As research on this progresses, it will be interesting to see the extent to which this claim holds for other aspects of language (e.g. syntax), for different types of languages, and whether it applies to both first and second language acquisition.

The articles included in this volume do not cover the entire spectrum of topics related to NLP pursued within the framework of cognitive science. No single volume could do that. Nevertheless, it is hoped that these articles will illustrate the potential benefits of this approach, which rests on a valid premise, namely, the need to integrate different viewpoints in order to do justice to the complexities of the phenomena or problems at hand.

It is our hope that other scholars will proceed further along this path and deepen our understanding of the functioning of the human mind and that such insights, will allow us to create applications that either simulate language processing or assist humans in their task of processing language, be it in their mother tongue or a foreign language.

Co-Editors:

Michael Zock  
LIF-CNRS  
163, Avenue de Luminy,  
13288 Marseille, France  
michael.zock@lif.univ-mrs.fr

Reinhard Rapp  
Universities of Mainz and Leeds  
Weiherstr. 61  
72074 Tübingen, Germany  
reinhardrapp@gmx.de

