COGNITIVE ASPECTS OF NATURAL LANGUAGE PROCESSING :

Wheels for the mind of the language producer

Michael Zock
C.N.R.S. – LIF (Marseille, France)
http://pageperso.lif.univ-mrs.fr/~michael.zock/

Languages are not only means of expression, but also vehicles of thought, allowing us to discover new ideas (brainstorming) or clarify existing ones by refining, expanding, illustrating more or less well specified thoughts. Of course, all this must be learned, and to this end we need resources, tools and knowledge on how to use them.

Knowledge can be encoded at various levels of abstractions, considering different units (words, sentences, texts). While *semantic maps* represent words and their relations at a micro-level, *schematic maps* (tree banks, pattern libraries) represent them combined in larger chunks (macro-level). We all are familiar with *microscopes*, *maps*, and *navigational tools*, and we normally associate them with professions having little to do with NLP. I will argue during my talk that this does not need to be so. Methaphorically speaking, we do use the very same tools to process language, regardless of the task (analysis vs. generation) and the processor (machine vs. human brain).

Dictionaries are resources, but they can also be seen as ***microscopes*** as they reveal in more detail the hidden meanings, nutshelled in a word. This kind of information display can be achieved nowadays by a simple mouse-click, even for languages whose script we cannot read (e.g. oriental languages for most Europeans). A corpus query system like *Sketch Engine* can reveal additionally very precious information: a word's grammatical and collocational behaviour in texts.

Unlike inverted spyglasses, which reduce only size, ***macroscopes*** are tools that allow us to get the great picture. Even though badly needed, they are not yet available in hardware stores, but they do exist in some scientists' minds. They are known under the headings of pattern recognition, feature detectors, etc. The resulting abstractions, models or blueprints (frames, scripts, patterns) are useful for a great number of tasks. I will illustrate this point for patterns via two examples related to real-time *language production* and *foreign language learning* (acquisition of fluency via a self-extending speakable phrasebook).

Semantic ***maps*** (wordnets, thesauri, ontologies, encyclopedias) are excellent tools for organizing words and knowledge in a huge multidimensional meaning space. Nevertheless, in order to be truly useful, i.e. to guarantee access to the stored and desired information, maps are insufficient — we also need some navigational tool(s). To illustrate this point I will present some of my ongoing work devoted to the building of a lexical ***compass***. The assumption is that (a) people always have some knowledge when looking for a word (more or less direct neighbors ; relationship holding between the source- and target word); and (b) that their knowledge-states are variable and hard to predict. This being so, we must build bridges between what is known or available (source word) and the target. I also assume that people have a highly connected conceptual-lexical network in their mind. Finding a word amounts thus to entering the network at any point by giving a related word (*source word*) and to follow then the links (associations) until one has reached the *target word*.

To allow for this kind of navigation, I believe that we need to do three things : (a) build an association network, (b) cluster the set of words, i.e. the associated terms we get in response to the input (word coming to the user's mind while trying to access the target; *tip of the tongue problem*), and (c) give meaningful names to the clusters. While the first step consists in building the *semantic map* within search takes place, the role of the next two steps is to support navigation. The role of the resulting *categorial tree* is to organize the set of words triggered by some input. Since any input is likely to yield many outputs (all words being associated with many other words) it is important to organize the resulting set of words, as otherwise we will drown the user.