

Constructing Semantic Tagsets

Alain Polguère
Department of linguistics and translation
University of Montreal
C.P. 6128, Succ. Centre-Ville
Montreal (Quebec) H3C 3J7
Canada
polguera@ere.umontreal.ca

1 Goal of the Session

This session will be devoted to the problem of building language-specific semantic tagsets. The concept of semantic tagset will be taken as referring to systems of core semantic units that give structure to the lexicons of natural languages. Because they are language-specific, semantic tagsets are not constructed based on the same methodologies as general/universal conceptual ontologies.

When building semantic tagsets, many questions need to be answered; for example:

- To what extent are semantic tagsets language-specific?
- What should be the relationship between lexical meanings of the language and semantic tags?
- Are semantic tags necessarily hierarchically organized? If so, what should tag hierarchies look like?
- What are the main methodologies available for the construction of semantic tagsets?
- Can semantic tagsets, which are language-specific, be used in multilingual Natural Language Processing to establish correspondences between lexical units of different languages?

2 Organization of the session

In order to “fuel” the debate, I will first make available to participants a list of lexical units of a number of languages (at least English and French) whose semantic tagging seems to me to raise interesting problems and questions. Each unit will be accompanied by examples (to make clear the senses we are talking about) and by a short summary of the potential tagging problems I see it posing. Participants are invited to study this list before the session starts and prepare any contribution they may wish to make to the discussion, based on the following questions:

- What would be your own particular solution to the problems posed?
- Can a parallel be drawn with the tagging of corresponding lexical units of other languages you know?
- What would be your solution, if any, for dealing with all the potential problems within one single coherent system of semantic tagging?

All participants with a “message to deliver” regarding the above-mentioned problems are invited to make a very short presentation of the semantic tagset they are using or developing and how it offers answers to those problems (or why it fails to do so). The purpose is not so much to decide who has got the right solution. I believe we should postulate that nobody has got it, and that we precisely need to meet and discuss the problem of semantic tagset construction in order to get closer to the most suitable solution. My own participation in this aspect of the session will be the presentation of the work we are doing in Montreal, on the building of a semantic tagset for a French dictionary of collocations.

3 Outcome of the session

If the session proves fruitful, and I am sure it will, I will propose to those interested in the enterprise to later work on the writing of a summary of all the questions we raised, all solutions we came up with, and all conclusions we drew.