A. ZAMPOLLI
Towards Reusable Linguistic Resources
Abstract

1. The paper will try to identify the various concurrent factors which, in the last years, have contributed to the increasing attention paid to the issue of reusable linguistic resources for natural language processing. The role of various major Organisations (EEC, Council of Europe, DARPA, NEH, etc.) will be examined, taking into account the reasons which seem to have stimulated their active participation. The recognition of the need of reusable linguistic resources by various sectors of "automated linguistic data processing" will also be considered as a catalyzer for the recent trend towards the cooperation of various scientific and professional Associations (ACL, ALLC, ACH, etc.).

2. Various interpretations of the term "reusability" have been proposed. We shall consider the relationships and differences between:

    -   the feasibility of the utilization of existing repositories of linguistic information, as potential sources for the construction of NLP components;

    -   the feasibility of constructing new large linguistic knowledge bases, explicitly designed for multifunctional uses, in various research frameworks and development activities.

    In particular, we shall examine the connections between various core aspects of the construction of reusable linguistic resources, and

    -   the general problems of the "evaluation" in the field of natural language processing, both at the general level of shareability of the state-of-the-art knowledge, and at the level of specific NLP systems or linguistic data collection;

    -   the design of "standards" at the level of representation formalisms and of the identification of linguistic units and description categories.

3. Examples will be taken by some of the various international projects nowadays underway in the field of lexica (ET7, Acquilex, Multilex, Genelex, The lexical Consortium, etc.), corpora (European Network of Textual Corpora, British National Corpus, Data Collection Initiative, etc.), and standards (Text Encoding Initiative).

4. The paper will review and compare some of the results already acquired, and will state some problems, both at the scientific and organizational level which, in the opinion of the author, are still to be solved, and represent major controversial issues. It is expected that the round table, which will be held during the Conference with the participation of the representatives of some of those projects, will discuss these problems and indicate priorities for research, practicable ways towards their solution, and possible convergence of efforts.