

6 • 5 Italy

Antonio Zampolli
University of Pisa, Italy

The general objectives of the national strategic project of the CNR: “natural language processing” and its successor “methods and tools for the language industries” can be summarized as follows:

- To create and analyze a textual reference corpus of written and spoken Italian, connected with corpora of specialized technical sublanguages and with a multilingual corpus of “contrasted” texts and their translation.
- To design a set of components for analysis and generation of Italian, including a formalized grammar of a large subset of written Italian, which could be used in different computational frameworks.
- To create an Italian multifunctional lexical database. Particular attention will be given to the problem of the reusability of existing lexicographical resources. A feasibility study is underway to evaluate how far the lexical information can be represented, according to a polytheoretical model which should ensure their usability in different linguistic frameworks. A very rich set of semantic relations among the lexical entries will also be specified including a conceptual taxonomy. The

Italian lexical database will be connected to similar lexical DBs which are being constructed in other countries. Bilingual dictionaries in machine readable form are used as a basis for this task.

- To ensure and coordinate the Italian participation in European and International projects in natural language processing and language industries, including machine translation.

The research lines, besides their specific goals, produce knowledge, information, data, methods, tools, competences directly usable in machine translation. From the national projects viewpoint, the participation in machine translation European projects is expected, in addition to its specific results, to provide elements helping to define the possibilities and limits of the state-of-the-art in machine translation, to evaluate the consequences of unresolved linguistic phenomena on the usefulness of machine translation, to suggest priorities for research and development in the near future. We are particularly interested in the definition of strategies, activities, objectives for research at the lexical and semantic level, which seem to be of crucial relevance in machine translation.