



---

# ALLC BULLETIN

---

Association for Literary and Linguistic Computing  
Association de Littérature et de Linguistique Computationelles

Editor: Gordon Dixon

Volume 13 Number 1 1985

## Italian Texts

The enrichment of the Italian corpora recorded in machine readable form has contained, in the main part following the standards and procedures of the textual archive of the Istituto di Linguistica Computazionale (ILC). Experimentation in the lemmatization of Italian texts using the Italian Machine Dictionary has also continued. There has been a pause in the text processing work of the Accademia della Crusca, which in 1983 celebrated the 400th anniversary of its foundation, as it is now restructuring the program and the procedures for its 'Tesoro delle Origini'. The results of analyses and processing of Italian texts are now being applied to some new sectors. For example, the concordances and the indexes of a corpus of contemporary Italian texts, containing approximately one and a half million occurrences, have been given to Italian Institutes of Culture abroad. In this way, advanced students of Italian who do not have the chance to spend much time in Italy, are able to become acquainted with 'real' Italian 'linguistic' usage (as distinct from linguistic 'norm') through the evaluation of the frequencies and distribution, in the corpus, of the different linguistic phenomena: grammatical constructions, different communicative contexts and types of texts, idiomatic expressions, etc.

An interesting development is the growing interest in the electronic text-processing of the language of the mass media which can, in this way, complete the very extensive synchronic and diachronic corpora of literary texts already in machine-readable form.

At the present moment, feasibility studies are being made to collect and process extensive samples of 'regional Italian', (written texts, recorded conversations and interviews, questionnaires), to be compared both with standard Italian, and with the dialects of nearby areas. The use of new technologies for the preparation of specific computational linguistic analyses will be organised in the form of a data base which will be accessible on-line for an Italian university network.

Important progress has also been made in researches on syntactic and semantic parsers for Italian. These studies interest both linguistic and computational linguistic research, and are connected with knowledge representation systems. Their connection with practical objectives (i.e. the typical interests of natural language processing systems: man-machine communication, data base interrogation, speech synthesis, etc.), must also be underlined.

*A. Zampolli  
LLC, Piza*