



Consiglio Nazionale delle Ricerche

# **GREY LITERATURE AND COMPUTATIONAL LINGUISTICS: FROM PAPER TO NET**

**Claudia Marzi, Gabriella Pardelli, Manuela Sassi**  
**Istituto di Linguistica Computazionale (ILC)**  
**Consiglio Nazionale delle Ricerche (CNR) - Italy**

# computational linguistics (CL) and language inquiry

- Computational Linguistics (CL) has changed the way we look at human language as a subject of scientific inquiry , shifting emphasis from abstract knowledge to real usage
- in CL, text understanding is “doing things with words” and requires the ability to master a heterogeneous system of skills based on the processing of complex information structures (e.g. reading, marking-up, summarizing, retrieving, classifying etc.)

# CL and the web

- the growing popularity of the web as an unbounded repository of unstructured text information led to an increasing interest in the application of CL methodologies to document access and retrieval
- web users can take advantage of CL tools and methodologies to get intelligent and selective access to on-line text documents
  - minimizing problems of information overflow
  - avoiding the strictures of pre-indexed document repositories

# CL and the web in Italy

- over the last five years, more and more Italian Universities have introduced CL courses into their Humanities curricula
- CL courses have the potential of addressing a much wider demand for a more aware use of the web than a purely academic one

# the role of grey CL literature

- to make it up for comparative shortage of white CL literature in Italian, CL courses have sprouted dedicated web sites providing tutorials, exercises, power-point presentations and other teaching materials
- on-line materials offer introductory information for a better understanding of:
  - *aspects of computer architecture and functioning*
  - *issues of digital text encoding and document representation*
  - *aspects of text browsing with personalized search patterns*
  - *issues of document mark-up and classification*
  - *fundamentals of document content indexing*

# the role of grey CL literature (II)

- provide a meeting point between academic information providers and non academic information consumers
- provide remote on-line access to actual course materials
- modify the general public attitude towards computer-based information access
- prompt more personalized ways of accessing web-based information

# case study:

## “Informatica Umanistica”

- **course overview**
  - goals
  - prerequisites
- **full set of course slides**
- **power point**
- **full set of teaching material offered during the course**
- **on-line exercises**
- **downloadable documents**
- **links to websites of interest and downloadable software**
- **access to on-line tools for Italian text processing**

# concluding remarks

- **CL → access & retrieval of web-based info**
  - vast majority of web-based knowledge available in huge on-line repositories of electronic text documents
  - automated, intelligent access of such repositories is precondition to their existence
  - web users want to access this info in an increasingly more dynamic, goal-oriented & flexible way
  - such demands will be met through integration of knowledge-rich and language-intelligent technology
  - prompts more aware ways of searching and accessing information on the web
  - sets high standards for information dissemination and sharing