

Natural Language Access to Electronic Archives

Leonardo Lesmo

Dipartimento di Informatica and Centro di Scienze Cognitive,

Università di Torino, Italy

Email: lesmo@di.unito.it

MICHAEL INTERNATIONAL CONFERENCE

INTERNATIONAL PROJECTS

1. NATURAL LANGUAGE

- **Interfaces:** Access via spoken or written language to stored data for the standard user in her/his own language

MAIN PROBLEMS

- **Availability of linguistic tools:** dictionaries, syntactic analysers, ontologies
- **Multilingual alignment:** Mapping of data and descriptions among different languages

2. TOOLS

Understanding languages implies having knowledge about **lexicon** (dictionary), **syntax** (grammar) and **semantics** (meaning).

For understanding a sentence, procedures are needed for accessing the dictionary (**morphology**), analyzing the structure (**parsing**) and extracting meaning (**interpretation**)

University of Torino:

- large dictionaries for Italian (30.000 entries), English (50.000), Spanish (5.000) and Catalan (5.000)
- A parser for four languages, plus partial tests on French
- A general procedure for semantic interpretation
- Procedures for managing ontologies (i.e. repositories of the meaning of words)

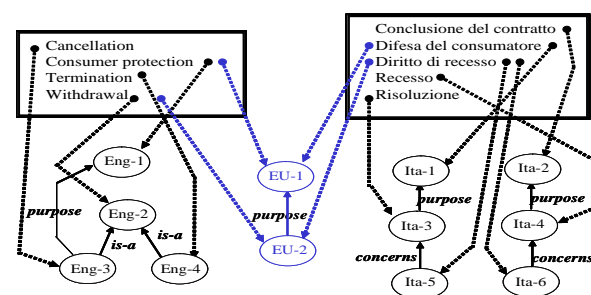
3. Alignment (Ex 1): The SYLLABUS Project

<http://www.uniformterminology.unito.it>

- **GOAL:** Establish a Uniform Terminology for the European Laws and provide interlingual access to stored documents
- **PARTICIPANTS:** Torino, Oxford, Barcelona, Lyon, Warsaw, Nijmegen, Münster
- UE Directives must be translated in the different European languages and adopted in the National Legal Codes
- Problem 1: Not all terms have an exact correspondence in all languages
Ex. Comerciante (Spanish), Trader (English), Commerçant (French), Commmerciant (Italian) has no correspondence in German
- Problem 2: The same meaning is expressed with different terms (in the same language) in the EU directives wrt. the national code.
Ex: "Reasonably" is translated as "Ragionevolmente" in the EU directive 99/44/EC, while is transposed to "Con ordinaria diligenza" in the Italian National laws.
- Problem 3: The manual management of data is complex and prone to errors.

4: SYLLABUS: Adopted Solution

- **Use of ONTOLOGIES**
(an Ontology is a conceptualization of the domain)



- **Implementation of a graphical interface**

Syllabus						
Umeanings of the european term "clear and comprehensible"						
Actions	umeaning id	Juridic language	Term	Description	References	Attachments Associations
Show ontology	eu2	english	clear and comprehensible	*...shall be provided in clear and comprehensible manner	dir 97/7: art. 4.2; dir 2002/65: art. 3.2	

5. Interfaces (Ex 2): the HOPS Project

<http://www.hops-fp6.org>

GOAL: enabling a citizen (tourist) to get information about the services provided by a municipality

PARTICIPANTS: Barcelona, Camden (London), Torino. Four languages: Castilian, Catalan, English, Italian

SOURCES: The backend databases of the existing web portals

TEST DOMAIN: Cultural events occurring in the three cities (concerts, museums, movies, theatre plays, ...)

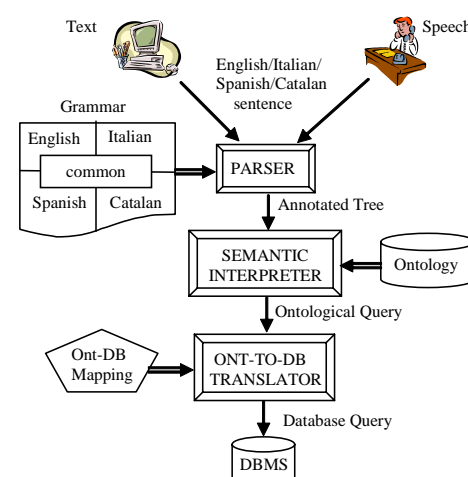
Example:

User: Could you tell me if there are any exhibitions of contemporary art in Torino this weekend?

System: There are four of them. Do you want a list?

6. HOPS: Adopted Solution

- Multilingual access via text or voice
- Semantic interpretation based on ontological knowledge



7. CONCLUSIONS

- Natural Language Processing enables to face practical problems in real domains
- Ontologies are a pervasive source of knowledge