



# UNL

## Spanish Enconverter

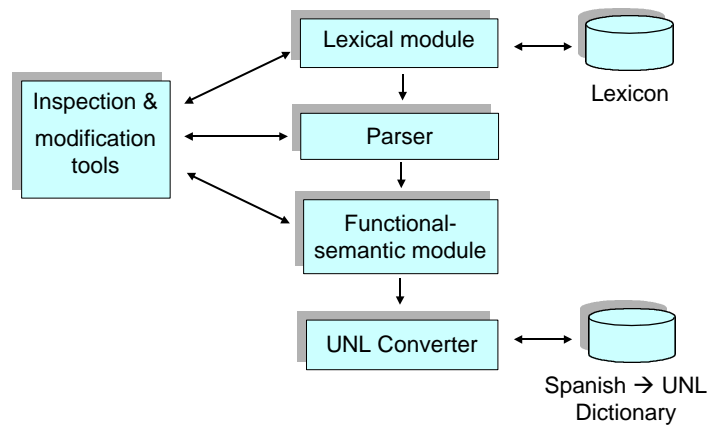
- architecture
- achievements
- benefits

Paris - November 97

This brief presentation of the Spanish Enconverter offers a general outline of the whole system, and of the work already done on the different modules that make up the Enconverter.

We will finish stressing some of the benefits expected from such a system for the whole UNL project.

# Architecture



The Enconverter is a software tool designed to provide:

- syntactic analysis of the input sentence
- semantic representation of sentence's meaning
- automatic conversion of that representation into UNL.

The Enconverter's functionality has been divided into four modules. Each one completes a major step towards the translation of the input sentence into UNL.

The Enconverter is a software assistant for its user, who:

- supervises the results proposed by the sub-system, and
- modifies and completes those results, if necessary.

The intended user is not an NLP expert, but perhaps a person with a bachelor's degree in Linguistics or just a person with a sound training in Spanish grammar.

As an interactive tool, several usability issues should be met:

- efficient analysis of the input sentence and quick response to user's command.
- employment of terminology and representation techniques well-known in Linguistics.

## Achievements

- Lexical module:
  - a) definition of a formalism of grammatical features using DAGs
  - b) definition of a set of basic lexical categories
- Parser:
  - a) definition of a declarative rule formalism
  - b) compilation of rules for most of the sentential phrases
- Functional-semantic analysis:
  - a) definition of a semantic model for Spanish based on the same abstractions than UNL.



We have established a set of nine basic categories (noun, adjective, verb, etc), most of them subdivided into more specific groups. For each category (and sub-category), a set of grammatical features have been defined in order to represent its grammatical properties.

The grammatical features belong to three different classes:

- a) morphological: gender, number ...
- b) syntactical: transitive, countable
- c) semantic: animate, human

At the moment, we have written rules for nominal, determinative and adjectival phrases. Adverbial, pronominal and prepositional phrases are being currently defined, and we will turn to simple and complex sentences after that.

The semantic model we plan to apply is based on:

- a) a set of primitive semantic objects (actions, entities, qualities, manners, propositions, etc.)
- b) a set of primitive semantic attributes or features that characterize these objects (animate, instrument, etc.)
- c) a set of primitive semantic relations (agent, theme, beneficiary, cause, etc.) that can be established between objects.

This representation is the base for the automatic conversion of sentence's meaning (achieved in this module) into UNL.

## Benefits

- A successful Enconverter will highly facilitate the production of high-quality UNL documents from Spanish texts.
- The Enconverter's interaction with the two other sub-systems (Dictionary and Deconverter) will help to improve both of them:
  - a) it imposes new requirements over the Dictionary, forcing to enrich and validate its contents
  - b) it provides newly generated *corpora* for testing and demonstrating the Deconverter.



The implementation of an Enconverter along these guide-lines will have two major benefits.

Point b) is regarded as very important during the development phase for an extensive and systematic testing of the Deconverter.