



# UNL

## Spanish Analyzer

- Summary of the 1<sup>st</sup> year
- Progress lines for the 2<sup>nd</sup> year
- Goals for the 2<sup>nd</sup> year
- Current state

Madrid - May 98

## Summary of the 1<sup>st</sup> year

- Setting up the guidelines of the future Enconverter:
  - functionality of the tool
  - operation mode
  - user profile
- Global architecture of the sub-system:
  - division into four functional modules
  - isolation of the functional modules from the user interface
  - setting up the interaction with the Dictionary



The Enconverter is a software tool designed to provide:

- a) syntactic analysis of the input sentence
- b) semantic representation of sentence's meaning
- c) 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:

- a) supervises the results proposed by the sub-system, and
- b) 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:

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

## Summary of the 1<sup>st</sup> year

- Developing of groundwork for future application into the Enconverter:
  - definition of a tree of lexical categories and a set of grammatical features for them
  - compilation of a set of morpho-syntactic rules for parsing sub-sentential phrases
  - selection of computational formalisms for manipulating lexical and syntactic information



## Rule example

Infinitives can perform as nuclei of nominal phrases:

*Fumar es peligroso; Sigue jugando hasta ganar*

Antecedent:

the constituent is a verb  
the tense of the constituent is infinitive

Consequent:

a nominal phrase is constituted  
the verb is the core of the nominal phrase

```
nominal_phrase --> verb
  <verb:tense> = infinitive,
  <nominal_phrase:core> = verb
```



## Progress lines for the 2<sup>nd</sup> year

- Continuation of the groundwork done in the 1<sup>st</sup> year:
  - enhancement and refinement of grammar's coverage
  - application of the semantic abstractions of UNL to the Spanish language
- Detailed design and implementation of a first version of the Spanish Enconverter



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.

## Goals for the 2<sup>nd</sup> year

- Enhancement and validation of the grammar with a corpus of Spanish sentences
- Creation of a functional Enconverter providing:
  - a visual interface to the task of enconverting the input sentence
  - on-line retrieval of lexical information
  - automation of the lexical analysis
  - automation of the syntactic analysis
  - guidance through the semantic analysis
  - automatic generation of the UNL text



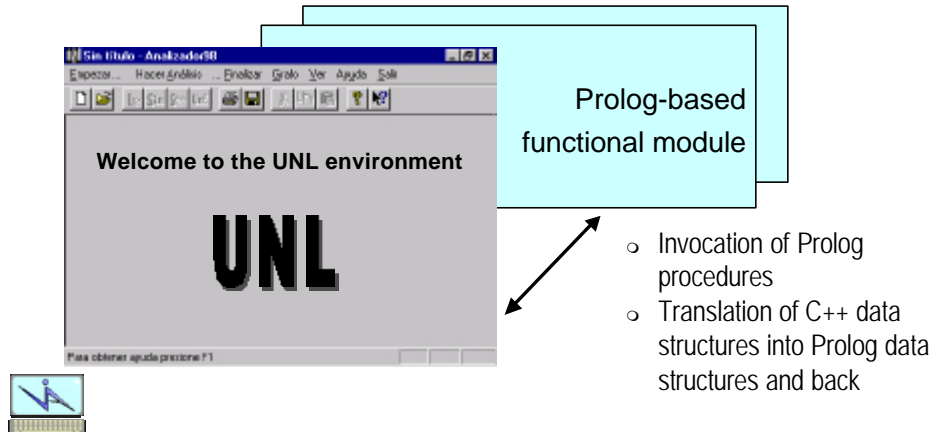
The implementation of an Enconverter along these guide-lines will have two major benefits.

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

## Analyzer: development lines

Software architecture that combines:

- use of state-of-the-art tools for the user interface
- use of highly productive tools for the functional modules



## Analyzer: development lines

- Analyzer environment:
  - Analyzer's usability and productivity
  - Integration of automatic and manual processes
  - Extensibility of automatic capabilities
- Cooperation with the Dictionary for:
  - achieving an operational interface
  - providing enough information for the enconverting process.



## Current state

- Specification of new requirements for the Dictionary, concerning:
  - richer classification of lexical entries
  - on-line interface between both sub-systems
- Requirement specifications and detailed design of the User Interface
- Requirement specifications and detailed design of the Lexical Analysis Module



## Current state

- Lexical Analyzer implemented functionality:
  - segmentation of the initial character string
  - processing of multi-word entries
  - processing of verbal particles
  - interaction with the Dictionary
- User Interface implemented functionality:
  - selection and editing of input sentence
  - presentation of the lexical analysis
  - presentation of the feature structures associated with each elementary constituent



## Current state

**Introduce o Selecciona una Oración**

Oración Seleccionada:  
Nuestros padres tienen un jale.

Oraciones disponibles:  
existe un lugar en la mancha de cuyo nombre no quiero acordarme.  
Juan tenía una botas con las que andaba siete leguas.  
**Nuestros padres tienen un jale.**  
pepe también, pero él no andaba tanto.

Ayuda (F1)

Análisis Léxico  
.....>

Cancelar



# Current state

**Análisis Léxico** [Minimizar] [Maximizar] [Cerrar]

Selecciona Elemento Constituyente

nuestros
padres
tienen
un
yate

Selecciona Categoría Gramatical

Sustantivo       Preposición

Adjetivo       Determinante

Adverbio       Verbo

Ver Rasgos Asociados

Ver Grafo Asociado

Ayuda      Cancelar      -----> [Análisis Sintáctico]



## Current state

Ver Matriz de Rasgos Asociada

Categoría

Concordancia

Genero

Número

Persona

Tipo

Número





**UNL**

## **Analizador**

Christèle Legeard  
Fermín Moscoso del Prado  
Luis Iraola

Madrid - Noviembre 98

## Analizador: cumplimiento de tareas

Tarea	Plazo	Productos
Prueba del Analizador Léxico	25/05/98 – <b>24/07/98</b>	Analizador Léxico <i>Plan de Pruebas del Analizador Léxico<sup>0</sup></i>
Especificación de requisitos y compleción de la gramática	13/07/98 - <b>28/08/98</b>	Reglas de análisis gramatical <i>Especificaciones del Analizador Sintáctico</i>
Diseño e implementación del Analizador Sintáctico	20/07/98 - <b>28/08/98</b>	Analizador Sintáctico <i>Diseño e implementación del Analizador Sintáctico<sup>*</sup></i>
Depuración del Analizador Sintáctico	03/08/98- <b>28/08/98</b>	<i>Plan de pruebas del Analizador Sintáctico<sup>*</sup></i>
Especificación de la representación semántica	29/06/98 - <b>31/07/98</b>	<i>Especificación de la representación del significado y de las heurísticas asociadas</i>
Diseño e implementación de heurísticas	3/08/98 - 11/09/98 <b>25/09/98</b>	Heurísticas de selección de atributo y relación semántica. Validación de RSUs. <i>Diseño e implementación del procedimiento de validación<sup>0</sup> de representaciones semánticas UNL</i>



## Analizador: cumplimiento de tareas

Tarea	Plazo	Productos
Especificación requisitos del conversor a UNL	31/08/98 - 4/09/98	<i>Especificaciones del Conversor a UNL</i>
Diseño e implementación del conversor a UNL	7/09/98 – 18/09/98 <b>15/09/98</b>	Conversor a UNL <i>Diseño e implementación del Conversor a UNL</i>
Depuración del conversor a UNL	21/09/98 - 2/10/98 <b>28/09/98</b>	<i>Plan de pruebas del conversor a UNL</i>
Integración	5/10/98 – 30/10/98 <b>en curso</b>	<i>Plan de pruebas de integración*</i>



## Herramientas auxiliares

- Conjugador nominal y verbal
- Generador de *scripts* SQL
- *Parser* de ficheros UNL
- Extractor de UWs desde ficheros UNL

