



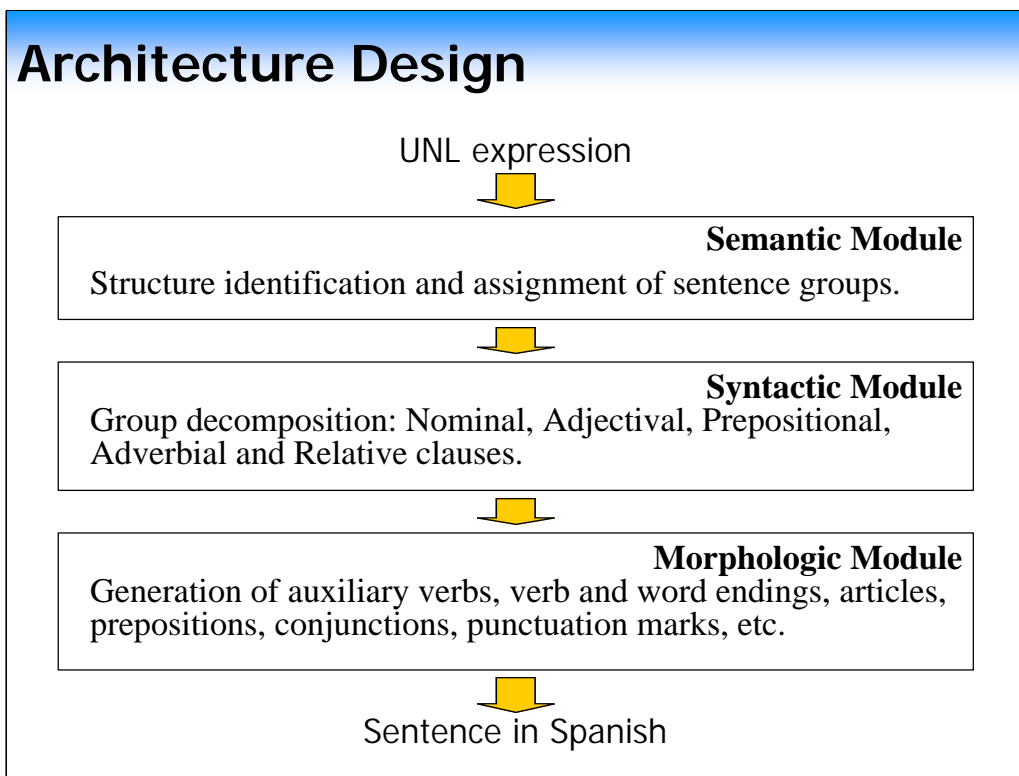
# UNL - Spanish Deconverter

Architecture Design  
Demonstration  
Trace analysis  
Conclusions

Paris - November 1997

## Presentation:

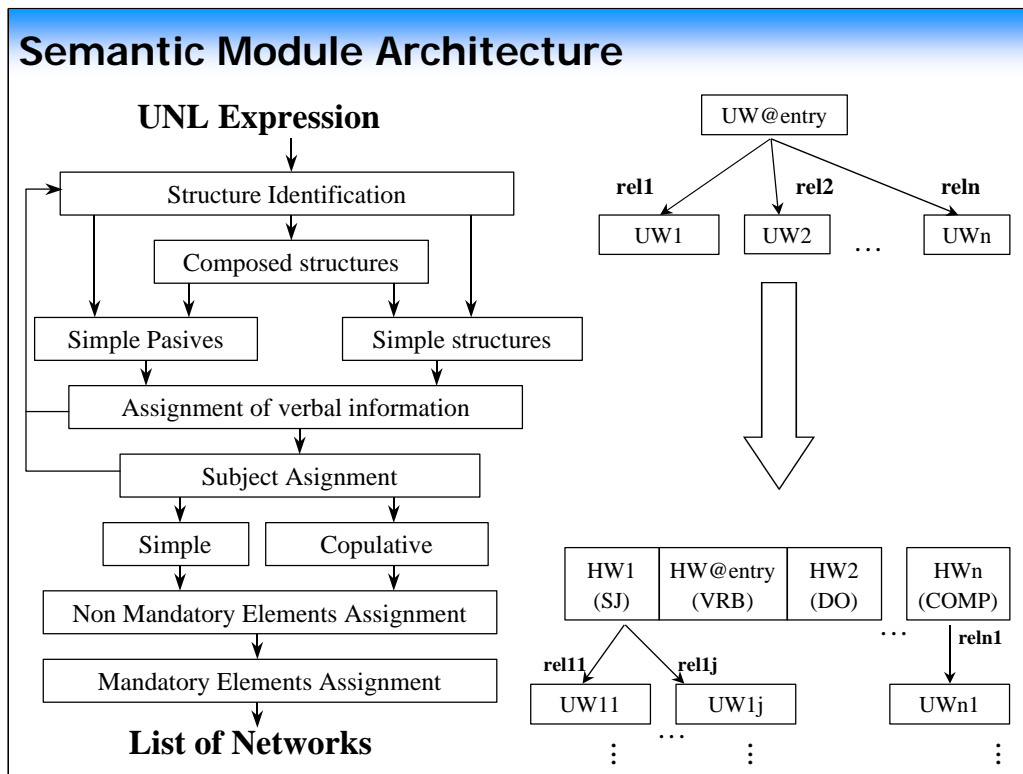
- Architecture Design of the complete System.
- Guidelines of the Implementation.
- Presentation of the Demonstration.
- Tracking of an example through the Design.
- Conclusions and Suggestions.



This full architecture description is an overview of the one shown in the Pisa Workshop on May 1997.

The main points are:

- To know what is the high level structure (co-ordination, conditional, passive voice, copulative, ...).
- Assignment of groups around the verb, describing their grammatical and syntactical functions.
- To explode the groups formed after the description, increasing their morphological information, taking into account their position respect the nucleus of the group.
- Generation at word level (prepositions, auxiliary verbs, etc.)



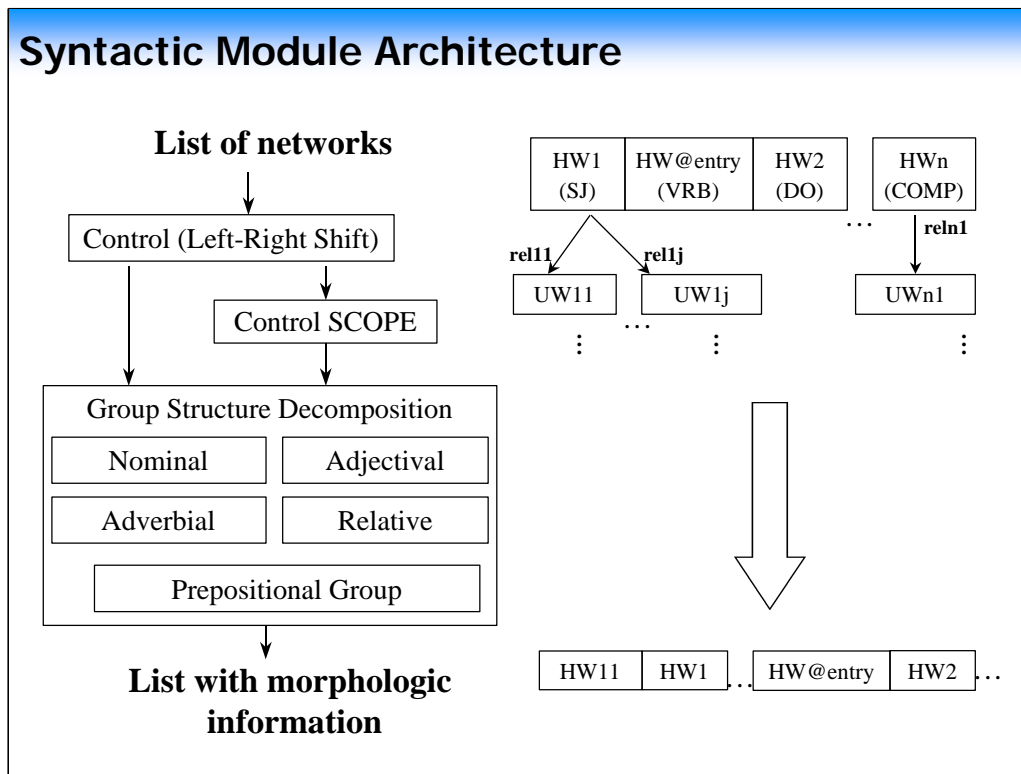
This module is the first one to deal with the UNL expression, so it must contain the higher level checks to classify and identify which kind of sentence it is, that is the *Structure Identification*. The decisions are to determine if the sentence is copulative, passive, active, conditional, coordinate and so forth.

After this first check, we are forced to assign and compare the verbal root, if this verbal root is not the one the DeCo select (comparing the aspect, intention, mode and tense) at the top of the deconversion, we have to backtrack until the beginning to select a new one). We select and assign the subject of the sentence and again compare with the verbal root, in this case about the number and person (if the comparative fail we backtrack to the top).

Note that if the verb is regular, the two checks will be correct and no backtrack is produced.

After that we *select the mandatory positions* (basically: Direct Object and Indirect Object) from the relationships. Then we *assign* (i.e. rise the node to the list) the *optional elements* before the *mandatory* ones are assigned.

We finally obtain a list of networks, where the roots of these networks (or grammatical trees) are the nuclei of the groups (nominal, adjectival, adverbial,...). This list will be the entry for the Syntactic Module.



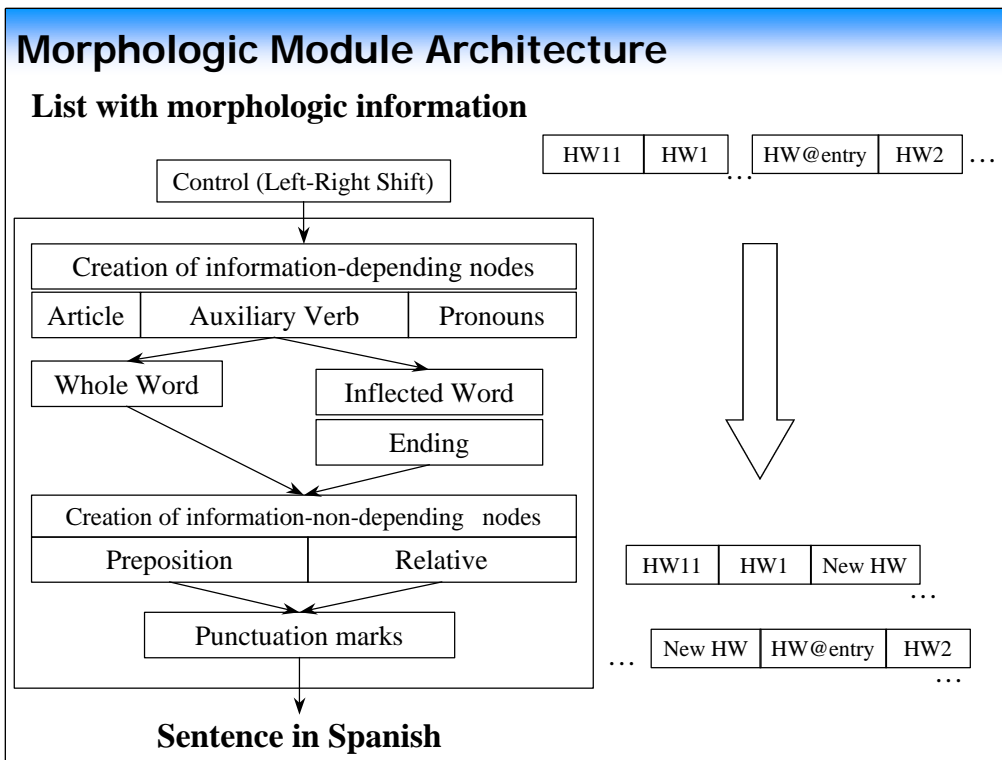
The input for this module is a list of networks (trees) as output from the generation at semantic level.

Each network in the list is a group that should be expanded. The nucleus of each one could be a noun, an adjective, an adverb, a noun affected by a preposition or a relative particle introducing a new clause.

The control rules must parse the list from left to right until there will not be any relation hanging down the list, taking into consideration the SCOPE or different kinds of clustering (like co-ordination inside the different groups).

The rules inside the module have to deal with the position of the adjectives respect to the noun, use of pronouns (personal, possessive, ...), choose of the right adverbs, and so forth.

Note that the position of these groups respect the kernel of the sentence, the verb, had to be determined in the previous module.



This last module complete the generation of the sentences adding the endings of the words, specially the verbs (according to the person, number, tense and mode). This module also tackle with the insertion of auxiliary verbs for aspects and intentions of the sentence.

The generation of prepositions, relative pronouns and articles is embedded with the Syntactic module due to the DeCo constraints. This is not really a problem: the morphological rules will be fired when they could be, completing the words of the sentence.

To end the whole process: the rules add the punctuation marks.

## Implementation guidelines

- Extremely simple Inference Engine (DeCo).
  - Implantation of our three-level-Design and Subdesigns in the DeCo formalism.
  - Use of the backtracking mechanism as the basic (and unique) tool for implementing complex operations.
- Integration of the Syntactic and Morphologic Modules with the same control rules.
  - "Update and Back" DeCo policy.
  - Reduce to one pass generation control (equivalent with the conceptual double pass).
  - Split frequent backtracking module (Semantic) from isolated backtracking modules (Syntactic and Morphologic).

The implementation guidelines are centred in the simplicity of the DeCo, and the application of our designs to this "programming language".

This simplicity only allows one effective mechanism for performing complex operations: backtracking.

Other point in our implementation is the integration of the syntactic and morphologic modules, separating them from the Semantic or high level module. The later constitutes non local backtracking application, while in the former the application is local, and does not affect the main structure.

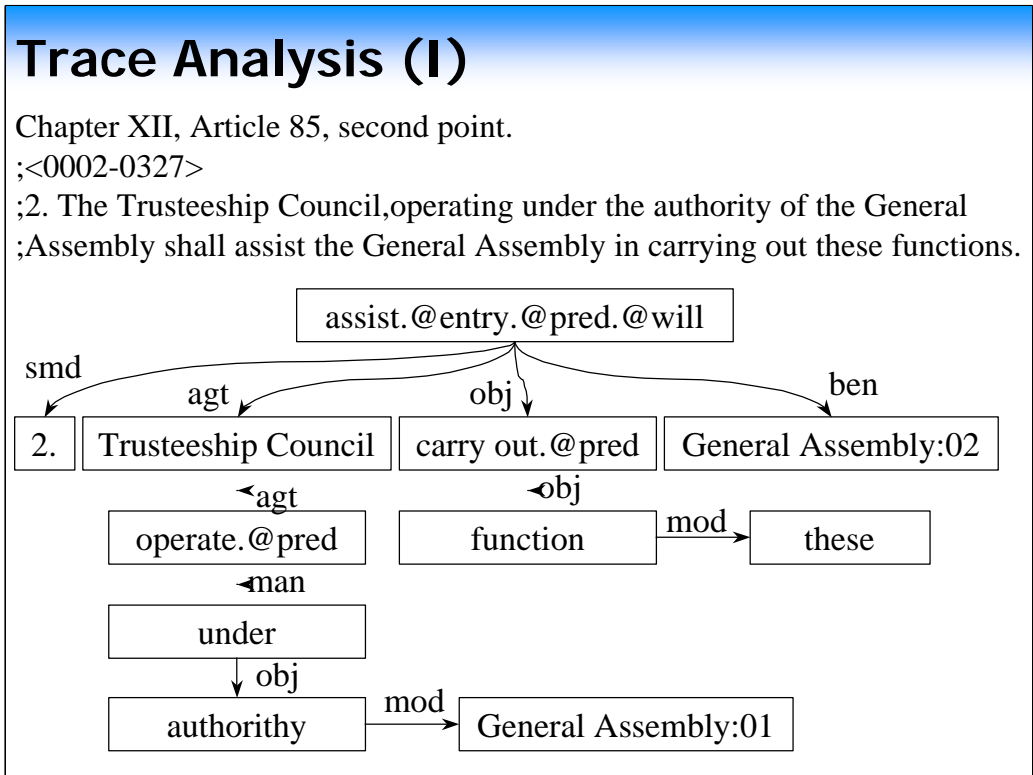
## Demonstration

- Prototype evolved from a Spanish General Purpose Deconverter:
  - Based on the Corpus (4 chapters of the UN Charter).
  - UNL Specifications (April'97, August'97 and new features from the Corpus).
- Demo Statistics:
  - Rules: around 1,000.
  - Demo dictionary: around 800 entries.
- Selection of expressions from the Corpus:
  - Around twenty expressions from the Corpus (corresponding to eight articles of the UN Charter).

The rules for this Demo are an evolution from a Spanish General Purpose Deconverter. This evolution is based on the Corpus established during September-October 1997, Chapters 8, 9, 12 and 13 of the U.N. Charter.

The rules are about 1,000; while the dictionary entries for the expressions expressions selected from the Corpus are nearly 800.

The selection of the expression have been taken from the 4 chapters, more than twenty expression (excluding titles and subtitles), in groups of articles (total number of articles: eight).



The sentence is a very good representative expression from the Corpus, and it is enough simple to follow the generation process. The structure is: SUJ(agt) + VRB(entry) + D.O.(obj) + I.O.(ben)

The Subject Noun group is affected by another verbal group with a reverse agt relation. There is also a preposition that head a prepositional group (using the man relationship).

The Direct Object introduce another verbal nucleus.

## Trace Analysis (II)

### First step: Structure determination, verbal form assignment and subject selection

State of the NODE LIST:00

```
/ << / 2.|uw=2. / [ ] / [Consejo de Administración
Fiduciaria|uw=Trusteeship Council] / ayud|uw=assist(icl>event) /
cumpl|uw=carry out / Asamblea General|uw=General Assembly /
>> /
```

- “Ayudar” (to assist) is a regular verb, so it is not necessary to backtrack for the correct verbal root (depending on tense, aspect and intention).
- Simple structure: Subject + Verb + Direct Object + Indirect Object
- Direct Object as a predicate node with no specifications about tense, aspect or intention: the verb should be in ing-form (gerund).

Structure identification for the expression, insertion of the smd (non semantic apposition) before the whole sentence. Simple subject assignment, no needing of passive or other forms (e.g. inclusion of “se” pronoun).

Identification of the final forms (headwords) for the inserted Universal Words: verbs (main verb and direct object verb).

## Trace Analysis (III)

### Second step: Decomposition of Sentence Group and morphologic fulfilment (list single pass)

State of the NODE LIST:00

```

/ << / 2.|uw=2. / / e| / / Consejo de Administración
Fiduciaria|uw=Trusteeship Council / / oper|uw=operate(obj>organization) /
ando / / bajo|uw=under(icl>manner) / / la / /
autoridad|uw=authority(icl>power) / (null) / / de / / la / / Asamblea
General|uw=General Assembly / [ ] / [ayud|uw=assist(icl>event)] /
cumpl|uw=carry out / funci|uw=function(icl>role) / Asamblea
General|uw=General Assembly / >> /

```

- Reverse relation (agt) “to operate” with no specifications on tense, aspect or intention; solved with the gerund form. This node is a new verbal kernel, so the rules for structure are again applied.
- Selection of the appropriate endings for verbs and words.
- Generation of prepositions, articles and punctuation marks.

## Trace Analysis (IV)

**Final Result: Completion of the list pass and end of the generation**

State of the NODE LIST:00

```
/ << / 2. | uw=2. / / el / / Consejo de Administración Fiduciaria | uw=Trusteeship
Council / / oper | uw=operate(obj>organization) / ando / /
bajo | uw=under(icl>manner) / / la / / autoridad | uw=authority(icl>power) / (null)
/ / de / / la / / Asamblea General | uw=General Assembly / /
ayud | uw=assist(icl>event) / ará / / cumpl | uw=carry out / iendo / /
estas | uw=these / / funci | uw=function(icl>role) / ones / / a / / la / / Asamblea
General | uw=General Assembly / . / [>>] /
```

- Final presentation of the sentence:

2. el Consejo de Administración Fiduciaria operando bajo la autoridad de la Asamblea General ayudará cumpliendo estas funciones a la Asamblea General.

;Time 9 Sec

## Conclusions

- Spanish Deconverter:
  - Improvement of the Morphologic level (specially generating the correct articles and prepositions).
  - Exhaustive testing and debugging.
- DeCo Inference Engine
  - Two new operations for improving computation time:
    - Parametric headword substitution (vs. the current proposal).
    - Movement of subnetworks within the local network.
  - Integration in a better Development Environment.
- Global project
  - Unification of the different "styles" of writing UNL.
  - Production of enough testing and development Corpora (Importance of the Analysis-Automatization).