

UNL'98 European Workshop

Universidad Politécnica de Madrid

- ◆ 1st Year Organization
- ◆ Technical Tasks - 1st Year
- ◆ 2nd Year Objectives
- ◆ 2nd Year Plan
- ◆ Current State
- ◆ Development lines
- ◆ Conclusions

1st Year Organization of the Project

- Technical Tasks:

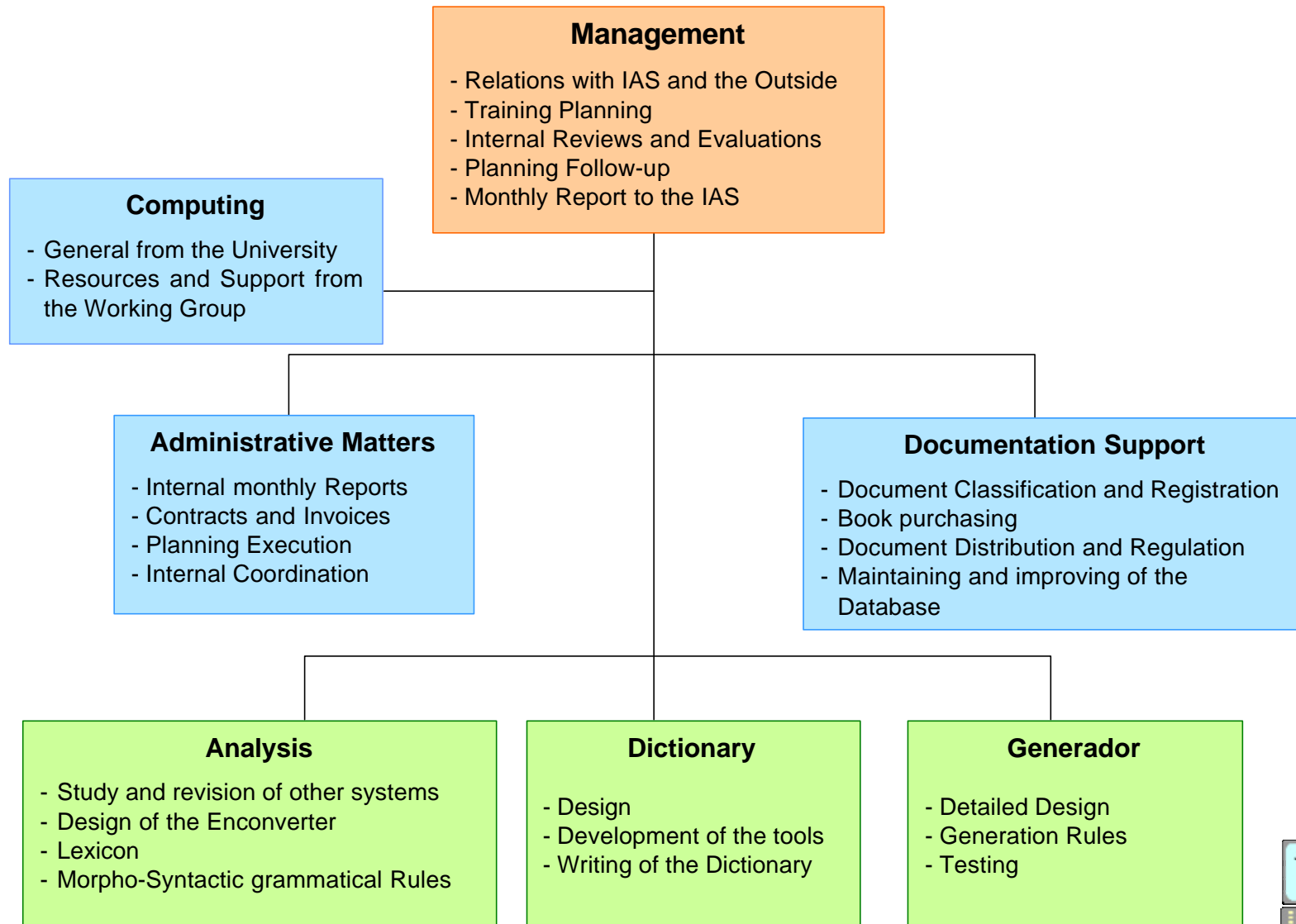
- ◆ Analyzer
- ◆ Dictionary
- ◆ Generator

- Management Tasks:

- ◆ Medium/long-term project.
- ◆ Medium-sized and multidisciplinary team.



1st Year Organizational Chart



Technical Tasks: 1st Year

- Analyzer:
 - ◆ Definition of the global computational architecture.
 - ◆ Compilation of a core of morphosyntactic rules.
- Dictionary:
 - ◆ Balanced and not specialized organization of the contents.
 - ◆ Implementation as a relational Database supported by several tools specifically developed by UPM team.
 - ◆ Introduction of 101.952 UW/Spanish word pairs.
 - ◆ Number of semantic blocks: 9.807
 - ◆ Number of morphologic blocks: 11.328
- Generator:
 - ◆ Writing of grammatical reports.
 - ◆ Design of the Generator
 - ◆ Implementation of the Generator design with the DeConverter rules.
 - ◆ Evaluation for the Final Report with the 1997 Corpus.



2nd Year Objectives

- Consolidate what was done in 1st year. (Improvement and expansion).
 - ↪ Define the functionalities of the UNL conversion environment.
 - ↪ Implement a UNL conversion visual environment.
 - ↪ Robust and scalable system.
- Orientate the development towards useful products.
 - ↪ Refining of the Dictionary contents.
 - ↪ Increase the contents.
 - ↪ Expand the system application coverage.
 - ↪ Improve the Generation functioning.
- Demonstrate the viability of the UNL system.
 - ↪ Define a reliable demonstration framework.

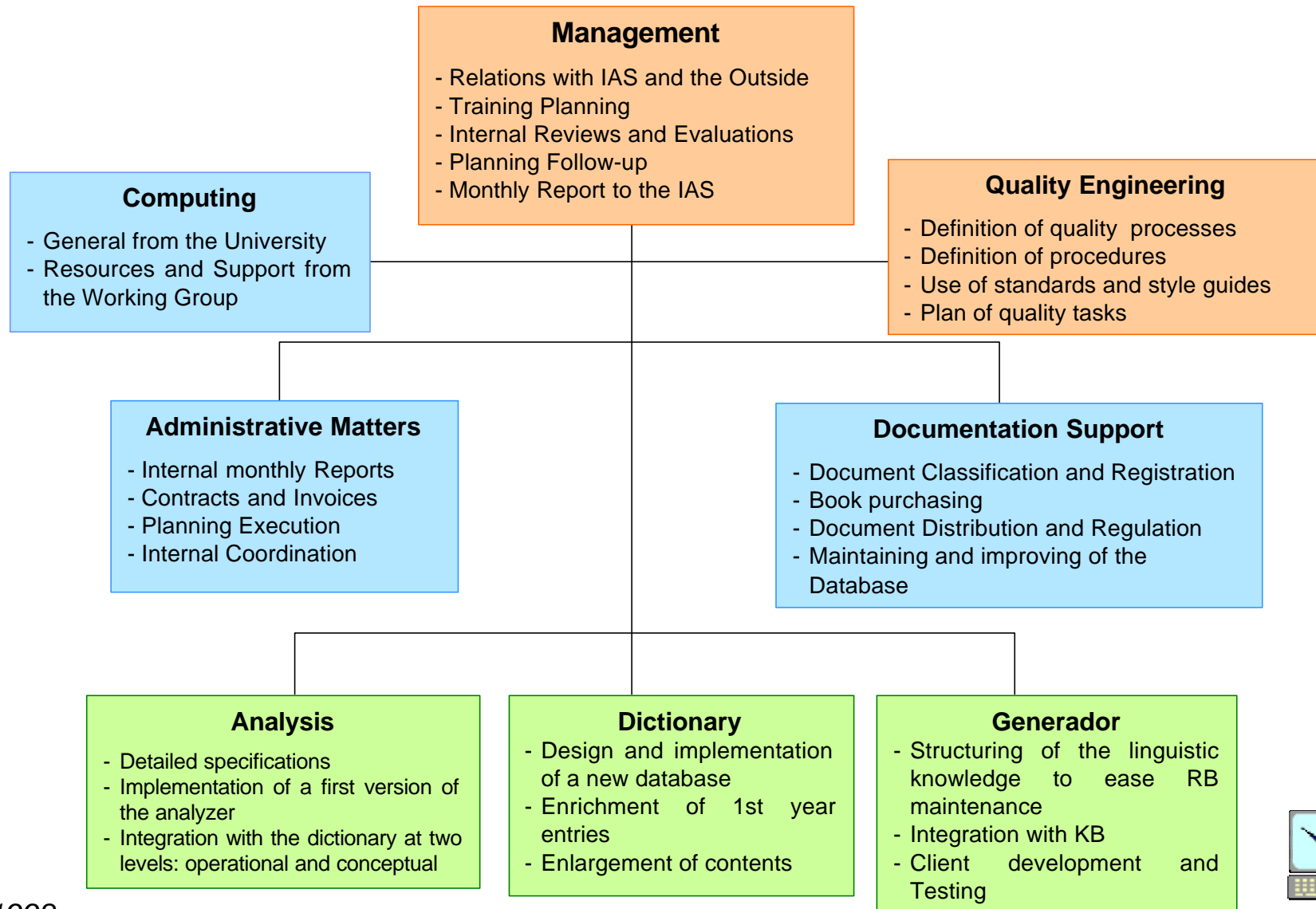


2nd Year Plan

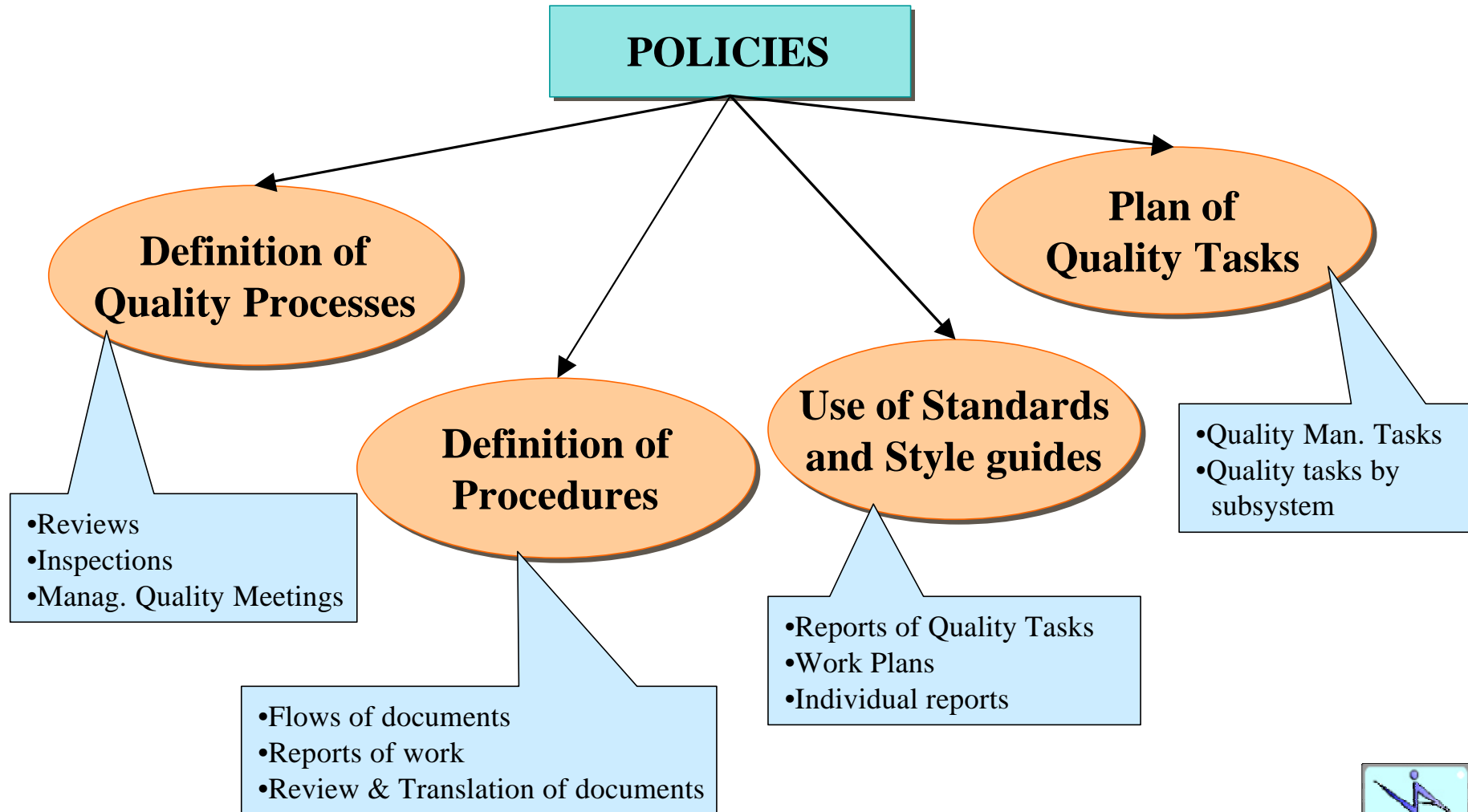
- Management:
 - ◆ Quality Engineering
- Analyzer
 - ◆ Detailed specifications of the Analyzer tool: functionalities, scope of automatic capabilities, interface layout, etc.
 - ◆ Implementation of a first version of the Analyzer.
 - ◆ Integration with the Dictionary at two levels: operational and conceptual.
- Dictionary
 - ◆ Design and implementation of a new database that meets the requirements expounded in the objectives.
 - ◆ Improvement of 1st year entries (New features, refining).
 - ◆ Enlargement of contents, doubling the size of the Dictionary.
- Generator
 - ◆ Structure the linguistic knowledge to ease the Rule Base maintenance.
 - ◆ Integration with Knowledge Base.
 - ◆ Client development and Testing.



2nd Year Organizational Chart



Quality Plan Goals



Current State: Analyzer

Tasks	1 st Year	Currently
Especification & Design	Global architecture	<ul style="list-style-type: none"> - Lexical analysis module - UNL assisted environment User interface
Implementation	Not started	<ul style="list-style-type: none"> - Lexical analysis (Prolog) - "Blue prints" User Interface (C++/Prolog) – (A)
Rules	Morphosyntactic analysis of the phrase syntagms	Pending (B)

○ Comments:

(A) The software architecture consists of a Prolog engine embedded into a C++ Windows application.

(B) We are currently pounding the scope of the automatic capacity versus the user contribution.



Current State: Dictionary

Tasks	1 st Year	Currently
Design of the lexical DB	Definition of a set of lexical categories (32) and grammatical features (52)	Refining and expansion of the lexical information (+20 grammatical features) (A)
Dictionary Implementation	<ul style="list-style-type: none"> - <i>mysql</i> prototypes - Several dedicated tools 	<ul style="list-style-type: none"> - Microsoft SQL ServerTM (B) - Several visual tools (Ms Access views, queries) (C)
N° of entries (pairs)	101.952	About 36.000 more

○ Comments:

(A) Refining with morphologic and semantic features as well as argument structures.

(B) Orientated to support a future commercial product.

(C) More reliable and easier maintenance.



Current State: Generator

Tasks	1 st Year	Currently
Generation Rules	About 3.000	About 4.000
Rules Coverage	Not applicable	No application yet (A)
UNL'97 Corpus sentences	About 50% (free of errors)	About 60% (free of errors)

- Comments:

(A) In order to evaluate the coverage, it is necessary to have a controlled corpus.



Generator: development lines

- Equivalence of Design Modules and Groups of Rules:
 - ◆ Allows Rule Base maintenance.
 - ◆ Extension of new linguistic phenomena within the Design: Evolutionary development.
 - ◆ Easy to change partial mechanisms (for example: Agreement)
- Testing Strategy:
 - ◆ Unit Test Corpus for each Design Module, also with simple but effective Logic Verification Tools.
 - ◆ Definition of an Integration Test Corpus covering different Modules.
 - ◆ Independent Acceptance Corpus (?).



Analyzer: development lines

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



Dictionary: development lines

- **Maintaining** 1st year criteria regarding domain and part-of-speech balance.
- **Increasing** the use of available public resources and automatic cataloging tools.
- **Improving** the semantic information attached to an entry applying KB's concepts.



Conclusions

- Every task of the project obeys a structured **methodology of production**.
- Clear **definition of the UNL system** and different modules.
- Clear **definition of the utility** of the sub-products.
- Definition of a **demonstration framework** for the end of the 2nd year in the UNL'98 Demonstration (New York).
- It is essential to reach an agreement on the definition of a **controlled test corpus** as soon as possible.
- UPM Team commitment to assure the **quality** of the processes as well as of the software product.
(Introduction of Quality Engineering methods)



Documents produced (1)

- ◆ [UPM/UNL/WP/T4.2/V1.0/62]; *Rules for the definition of sub-sentential phrases*; Luis Iraola; 30/11/97
- ◆ [UPM/UNL/WP/T4.2/V1.0/ 57]; *The Verb: Core of the Verbal Syntagm*; Christèle Legeard; 1/12/97
- ◆ [UPM/UNL/WP/TAnaliz/V1.0/58]; *Notes on the Strategies of Search in the Dictionary*. Luis Iraola; 12/12/97
- ◆ [UPM/UNL/WP/TAnaliz/V1.0/59]; *Features Structures as Directed Acyclic Graphs*; Luis Iraola; 18/12/97
- ◆ [UPM/UNL/WP/T2.2/V1.0/64]; *Demonstration Dictionary of the Generator*; David Escorial; 19/1/98
- ◆ [UPM/UNL/WP/T2.2/V1.0/65]; *Demonstration Corpus: Input and Output of hte Generator Demonstration*; Francisco Astudillo, David Escorial; 19/1/98
- ◆ [UPM/UNL/WP/T2.2/V1.0/66]; *Demonstration Rules in the DeCo format*; Francisco Astudillo, David Escorial; 19/1/98 [UPM/UNL/Del/V1.0/91]; *First Year Final Report*; J. Cardeñosa, F. Astudillo, D. Escorial, L. Iraola, F. Moscoso. 19/01/98 (Written in English)
- ◆ [UPM/UNL/WP/T2.1/V1.0/ 92], *Description of the DeConverter*, David Escorial, 20/5/97
- ◆ [UPM/UNL/WP/T2.2/V1.0/ 93], *Verbal Phrase*, David Escorial 5/5/97



Documents produced (2)

- ◆ [UNL/UNL/WP/T1.0/V1.0/ 94], *Collaboration for UNL specification*, Nikolai Puntikov, 21/1/98
- ◆ [UPM/UNL/WP/T1.0/V1.0/ 95], *Dictionary review with regard the generation of the new corpus*, Luis Iraola, 29/1/98
- ◆ [UPM/UNL/WP/TAnaliz/V1.0/ 96], *Simple sentence classification*, Christèle Legeard, 9/2/98
- ◆ [UPM/UNL/WP/TRev/V1.0/ 97], *Dictionary 98: Work plan*, Luis Iraola; Fermin Moscoso, 3/2/98
- ◆ [UPM/UNL/WP/T1/V2.1/ 98], *Conceptual design of the database*, Fermin Moscoso; Luis Iraola, 3/3/98
- ◆ [UPM/UNL/WP/Tanaliz/V1.0/ 98], *Enconverter 98: goals and work plan*, Luis Iraola, 23/2/98
- ◆ [UPM/UNL/WP/Tplan/V1.0/ 99], *Dictionary 98: goals*, Luis Iraola, 1/4/98
- ◆ [UPM/UNL/WP/Tplan/V2.0/ 102], *Enconverter 98: work plan*, Luis Iraola, 1/4/98
- ◆ [UPM/UNL/WP/Tplan/V3.0/ 103], *Dictionary 98: work plan*, Luis Iraola; Fermin Moscoso, 1/4/98
- ◆ [UPM/UNL/WP/Tplan/V1.0/ 104], *Quality plan for the project UNL*, Edmundo Tobar, 14/4/98
- ◆ [UPM/UNL/WP/T1.0/V1.0/ 105], *Comments to the UNL specifications for the specifications task force meeting*, David Escorial, Luis Iraola, 13/5/98



Documents received

- ◆ [UNL/UNL/WP/T1.0/V1.0/85]; *Decol v1.1*; UNL Center; 17/11/97
- ◆ [UNL/UNL/WP/T1.0/V1.0/86]; *Mailing list formation*; UNL Center; 26/11/97
- ◆ [UNL/UNL/WP/T1.0/V1.0/87]; *Anomaly SCOPE solution report*; UNL Center; 1/12/97
- ◆ [UNL/UNL/WP/T1.0/V1.0/88]; *Decol v1.1*; UNL Center; 15/12/97
- ◆ [UNL/UNL/WP/T1.0/V1.0/89]; *UNL 97 corpus*; UNL Center; 15/12/97
- ◆ [UNL/UNL/WP/T1.0/V1.0/89]; *UNL 97 corpus*; UNL Center; 15/12/97
- ◆ [UNL/UNL/WP/T1.0/V1.0/106]; *Decol v1.4*; UNL Center; 13/5/98
- ◆ [UNL/UNL/WP/T1.0/V1.0/ 107], *Rules of the prototype of generation Italian*, Irina Prodanoff, Nicoletta Calzonari, 13/5/98.
- ◆ [UNL/UNL/WP/T1.0/V1.0/ 108]; *Universal Networking Language System*, UNL Center, 19/5/98.

