

The UNL'98 Symposium

Universidad Politécnica de Madrid

18-22 November 1998

New York



UNL Project

Spanish Point of View - End of 1998

OBJECTIVES

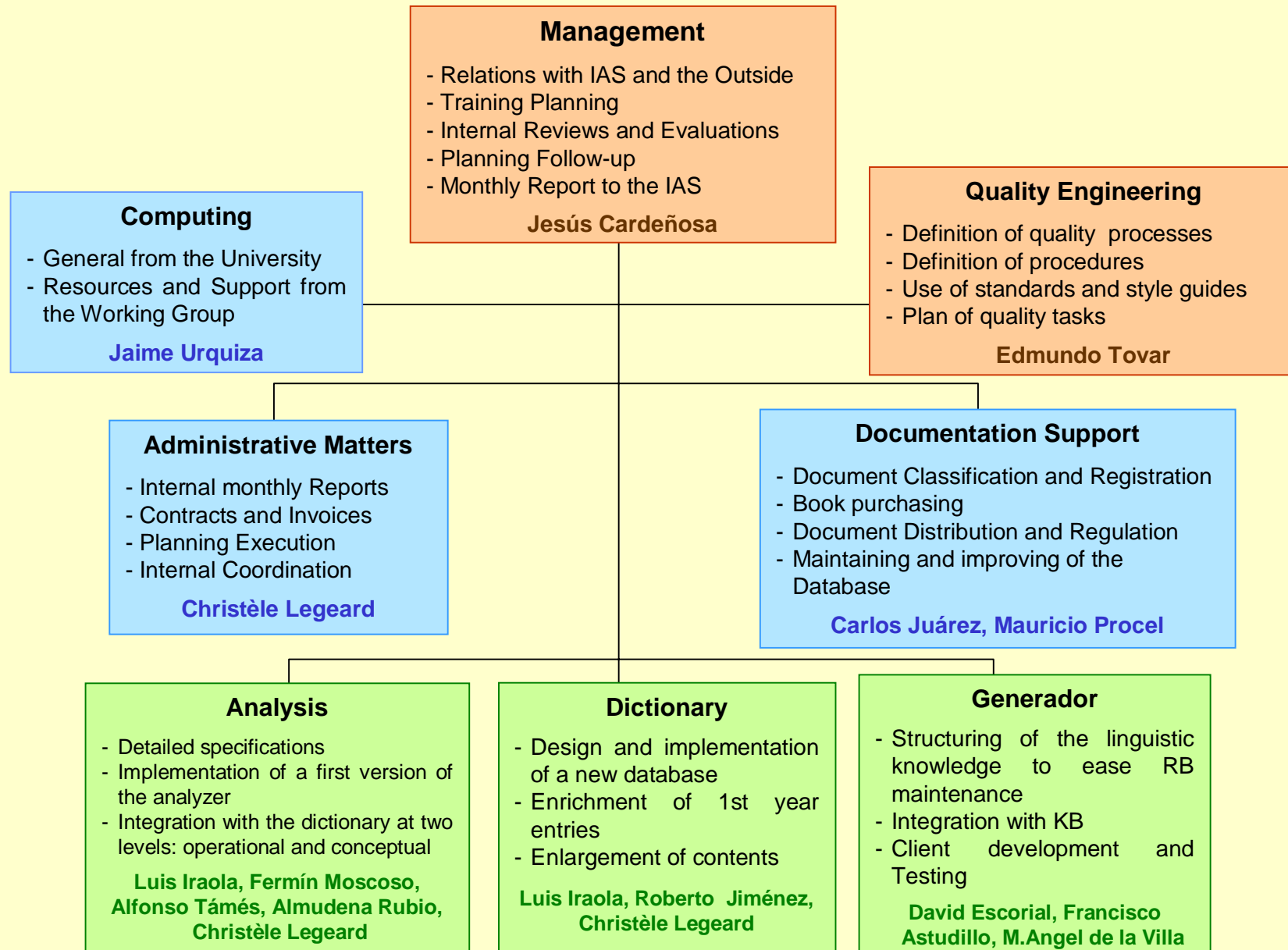
- Product Orientation.
- Quality of the components.
- Detection of gaps on the market.

PLAN

- Rigorous Development Plans.
- Quality Plan according to ISO 9001.
- Strategic Plan → Future of UNL as a System and Organization.
- Rigorous Documentation Control.



UPM Organization





Documentation and Archive Service

- Standardization of Documents.
- Standardization of Registration.
- Rapid service of search and maintenance.





Document Cover Page

Grupo de Validación y aplicaciones industriales

Departamento de Inteligencia Artificial
Facultad de Informática
Universidad Politécnica de Madrid



PROYECTO UNL

DeConverter specifications

Autor(es) : Mellyng Zita; Francisco Artidillo; UNL center
Institución(es) : UPM
Fecha : 11-10-98
Tarea : de.co
Tipo : Working Paper
Versión : 2.0
Referencia : UPM/UNL/WP/TdecoA/2.0/142
Path :
Acceso : Confidencial

Resumen



Documentation Data Base

Formul_Documentos

Título DeConverter specifications

Autores
Meiying Zhu
Francisco Astudillo
UNL center

Proyecto UNL
Tipo Working Paper
Tarea deco
Versión 2.0
Nº Doc. 142

Resumen

Institución UPM
Referencia UPM/UNL/MP/Tdeco/V2.0/142
Archivo
Acceso Confidencial

Observaciones

Fecha producción 11-nov-98

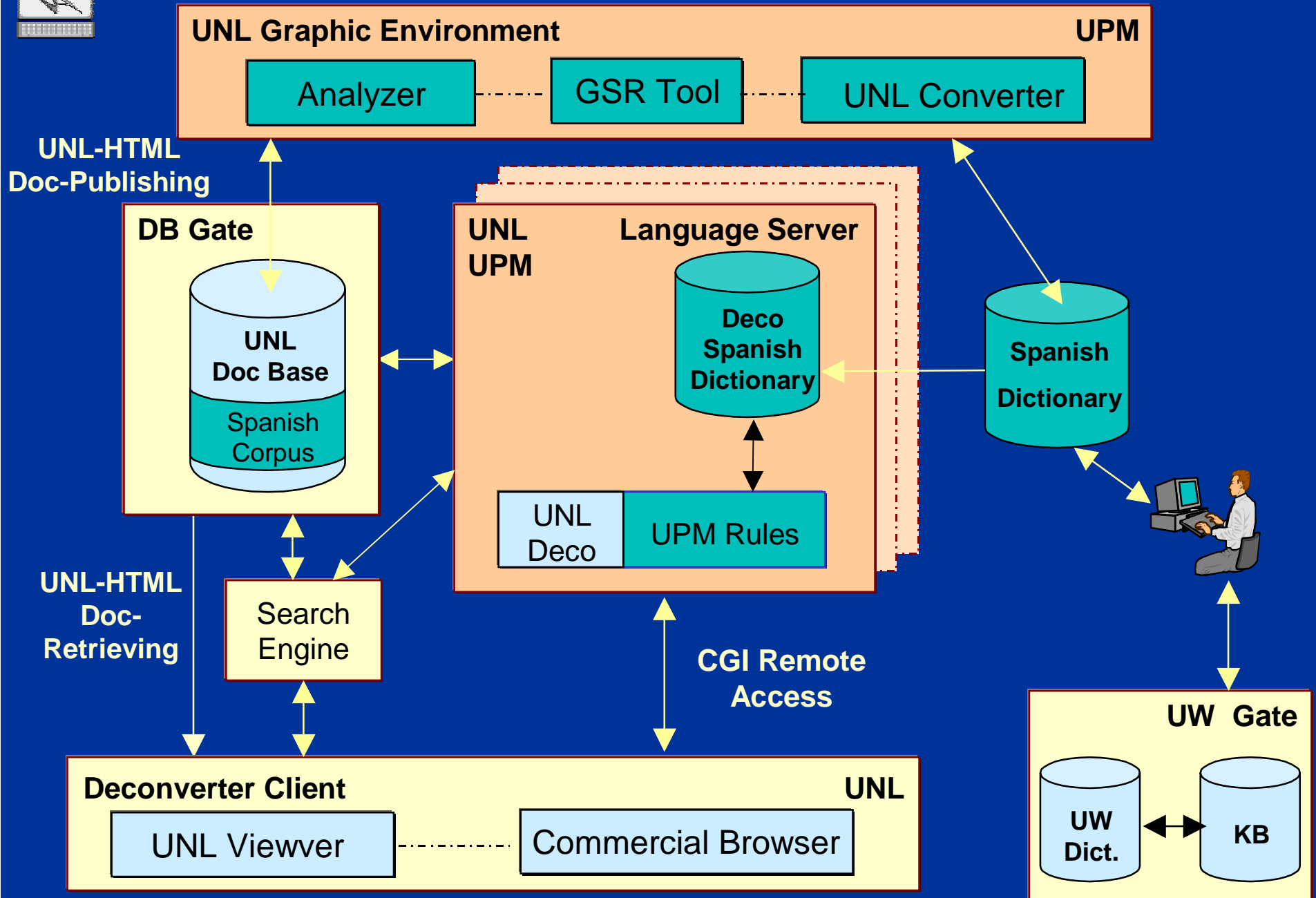
Fecha SAD 13-nov-98

Distribución

Nombre	Fecha	Medio de entrega
David Escorial	13-nov-98	SAD

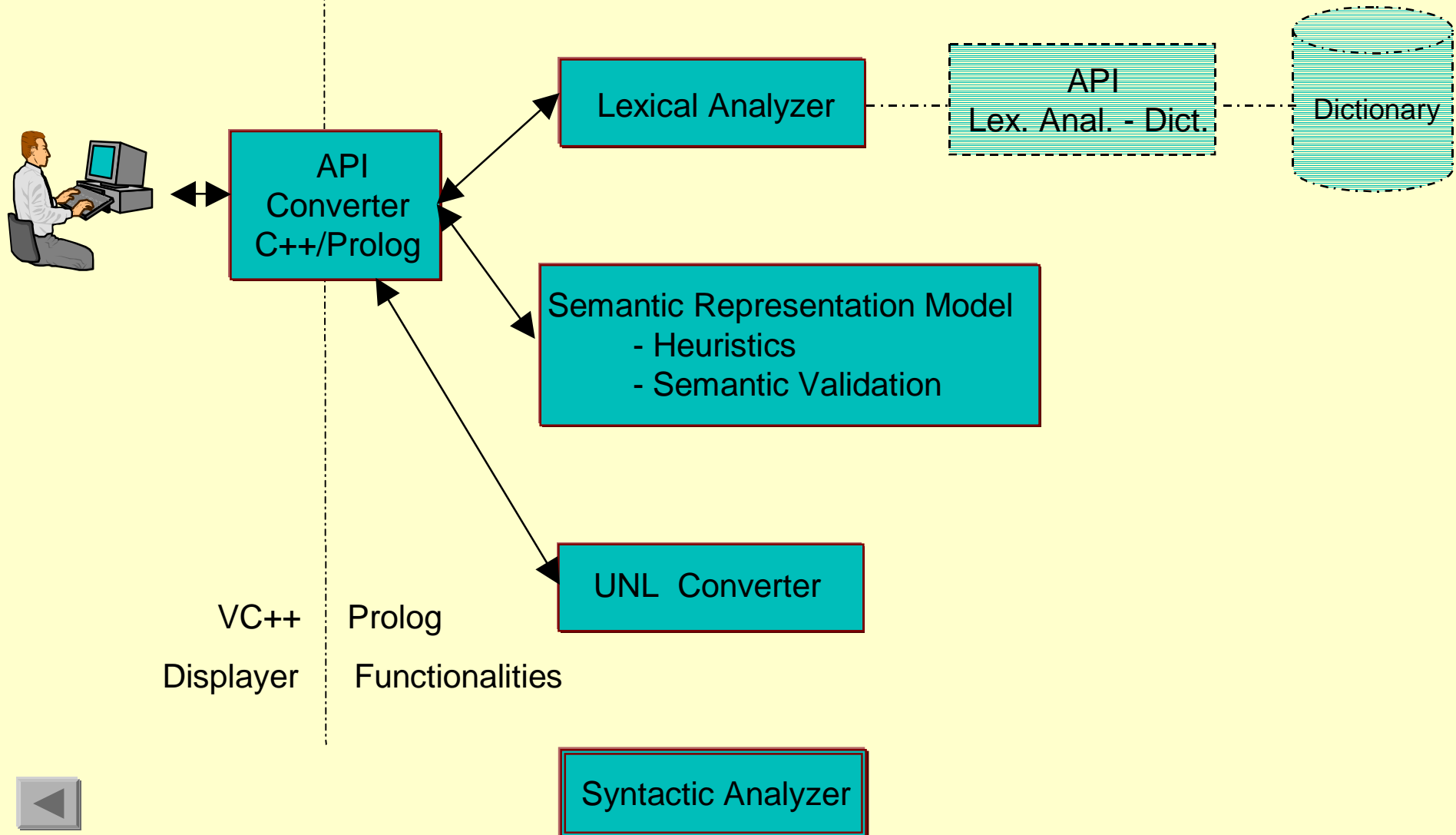
Registro: 1 de 1

Registro: 276 de 276



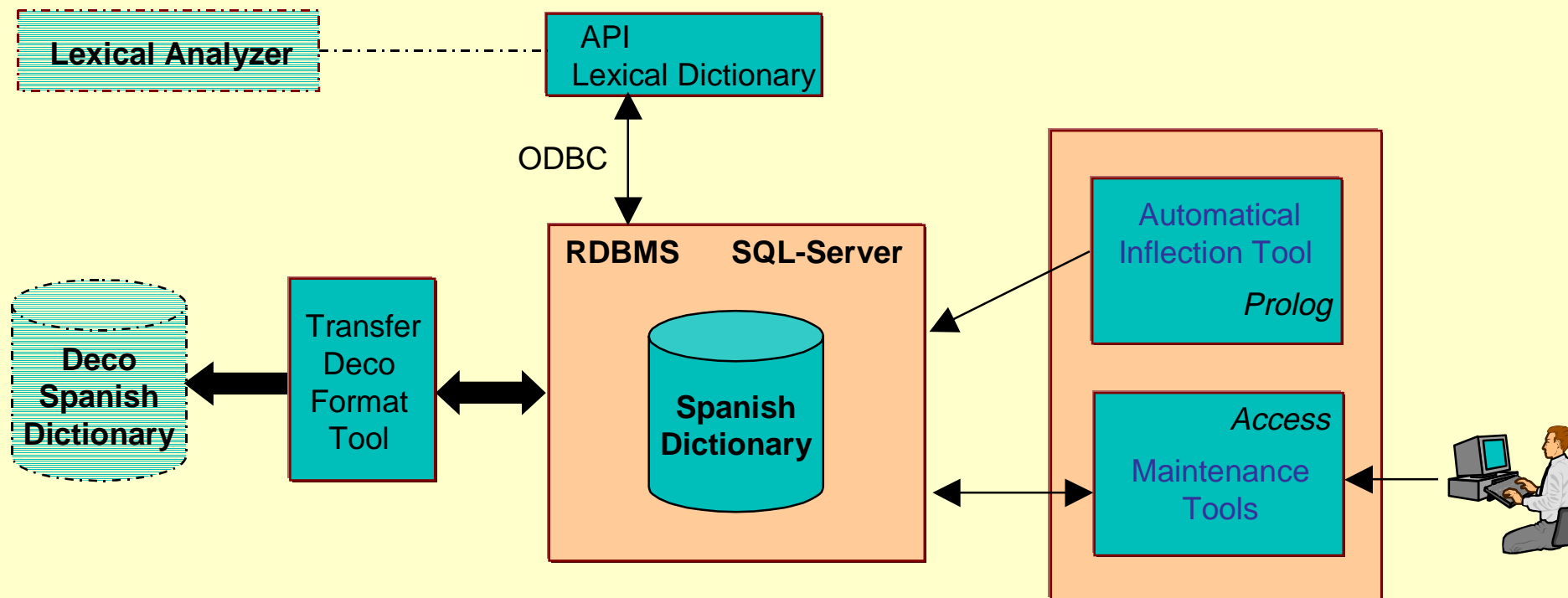


UNL Graphic Environment



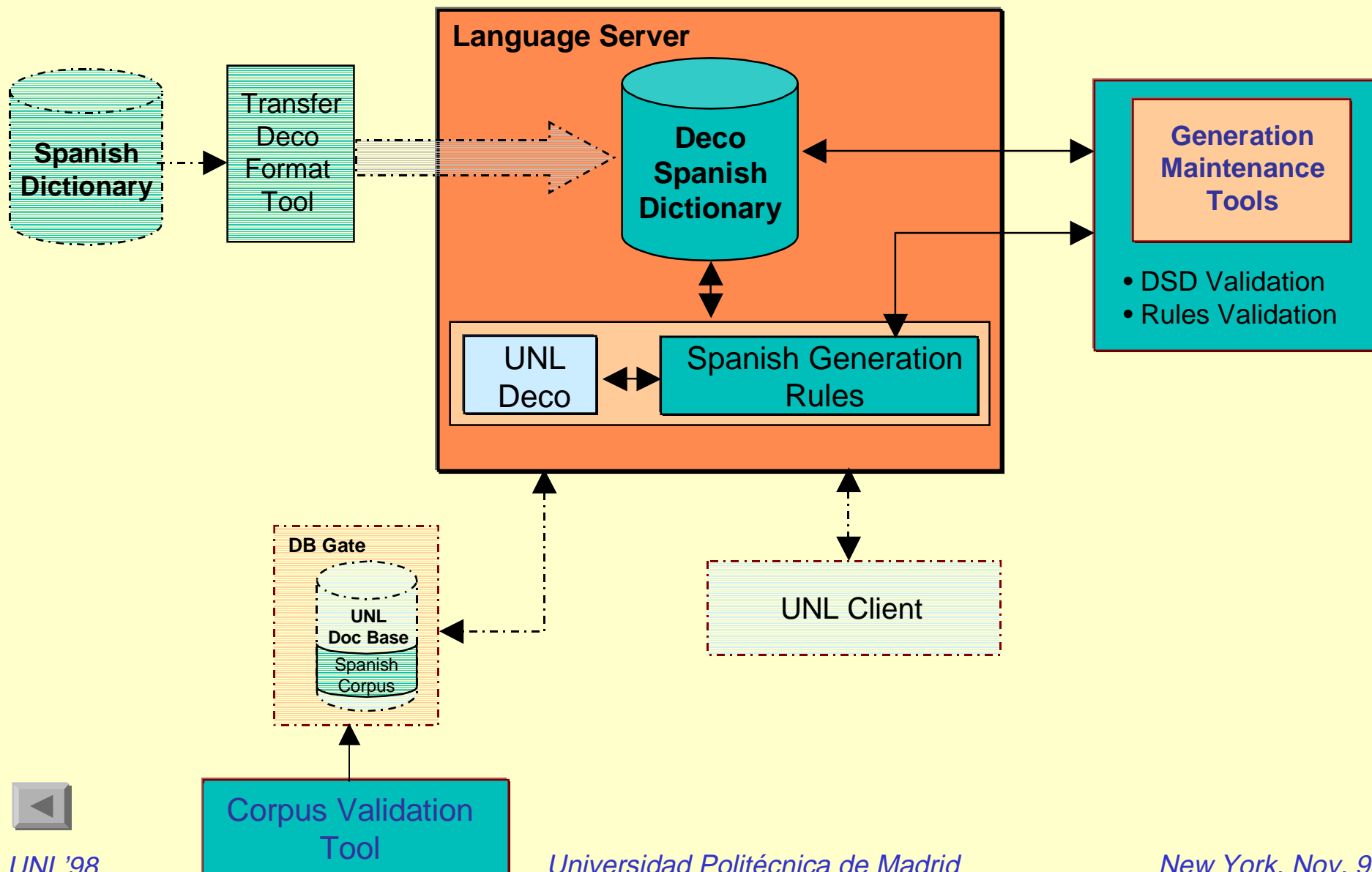


Dictionary





Deconverter





Lexical Analyzer

- Completely terminated.
- Developed in SWIProlog.
- Some functionalities are:
 - It segments sentences into lexical units.
 - It separates contractions.
 - It groups together locutions.
 - Others...





UNL Editor Environment

UNL ENTORNO DE ANÁLISIS Y EDICIÓN UNL - [prueba.txt]

Archivo Editar Ver ANALIZADOR LEXICO GRÁFICO CODIGO UNL Ayuda

TEXTO A ANALIZAR

Zagalo mete presión a Francia ante la final
■
ÁNGELS PIÑOL, París

"Perfecto. ■ Brasil-Francia. ■ La final soñ técnico de Brasil, prefirió ver el Francia-Crc atendió en el hotel de Ozoir, donde reside l periodistas que le aguardaban. ■ Mientras festejaban enloquecidos el éxito histórico d su propia euforia. ■ Más seguro que nunca escapar la quinta Copa del Mundo, el entre comer el campo, el césped, la pelota, la To si es necesario. ■ No vamos a dejar escap luchar segundo a segundo". ■ Y fue aún m mucho su primera final. ■ Ya nos vemos c

■
Aimé Jacquet, seleccionador francés, adm "Son los maestros del fútbol", dijo, pero ap la ventaja de jugar en casa. ■ El factor can más mínimo a Brasil.

■
Faltan dos días para la final y nadie duda y: críticas que ha recibido en este Mundial. ■ que encabeza Sukoer (5 goles), empatado c (Argentina), ambos desde hace tiempo de tienen la convicción de que puede romper e

EDITOR GRAFICO

```
graph TD; A(mete presión) -- agente --> B(Zagalo); A -- experimentador --> C(Francia); A -- manera --> D(ante); D -- tema --> E(final);
```

CODIGO UNL

```
agt(put pressure on(icl>event).@present.@pred.@entry, Zagalo;  
exp(put pressure on(icl>event).@present.@pred.@entry, France  
man(put pressure on(icl>event).@present.@pred.@entry, in fron  
obj(in front of, final round(fld>soccer).@sg)
```

ANALIZADOR LÉXICO

| Zagalo || mete presión || a || Francia || ante || la || final |

Preparado





Semantic Representation Model

- **Heuristics**
 - They limit the search space of the grammatical categories associated to a node.
- **Validation**
 - It assures that the graphic representation fits to UNL semantics.
 - It generates a series of warning and error messages that guide the user for UNL Conversion.
 - The system prevents the generation of UNL sentences that do not fulfill UNL syntax and semantics.





Syntactic Analyzer

- 119 very compressed rules.
- Syntactic coverage of 90% of phrase syntagms and simple sentences.
- Not determinist (it offers a range of possible solutions to a determined query).





Spanish Dictionary

QUANTITATIVE ISSUES

- Actual number of pairs (Spanish inflected words/UW) : 302.000 aprox.
- They are distributed as follows :
 - Substantives : 11.558
 - Verbs : 2.414
 - Adjectives : 1.738
 - Adverbs : 272
- More than 50 features for each word.
- Two Software environments: SQL Server and Access.

TOOLS

- Automatic verb inflection tool.
- Maintenance tool.
 - Insertion of new elements.
 - Search and Inspection for detecting errors.
 - Modification of contents.
- Transfer DeCo Format tool.





Dictionary Maintenance Tool

The screenshot displays two windows from the Dictionary Maintenance Tool. The left window, titled "MORFOLOGIA", contains a search section with fields for "Identificador" (288) and "Palabra Cita" (comer). Below this are fields for "Raíz" (com), "Primeras letras", "Forma de cita" (comer), "Longitud mínima" (4), and "Longitud máxima" (10). A table lists morphological entries with columns for "Identif." and "Desinencia".

Identif.	Desinencia	Rasgos
10039	e	Rasgos
10041	emos	Rasgos
10119	o	Rasgos
10193	a	Rasgos
10194	a	Rasgos

The right window, titled "PALABRAS UNIVERSAL...", shows a search for "eat" with "Uw:" and "275" fields. It features a "Bloques semánticos" section with entries like "40413" and "Entradas".





UPM Deconverter

QUANTITATIVE ISSUES

- 4.209 rules.
- Coverage of 90% of the Spanish Grammar.
- Corpus 98 based on :
 - 6 daily press papers of soccer.
 - 196 sentences.
- Two accesses (automatic and manual) to the Spanish Language Server.

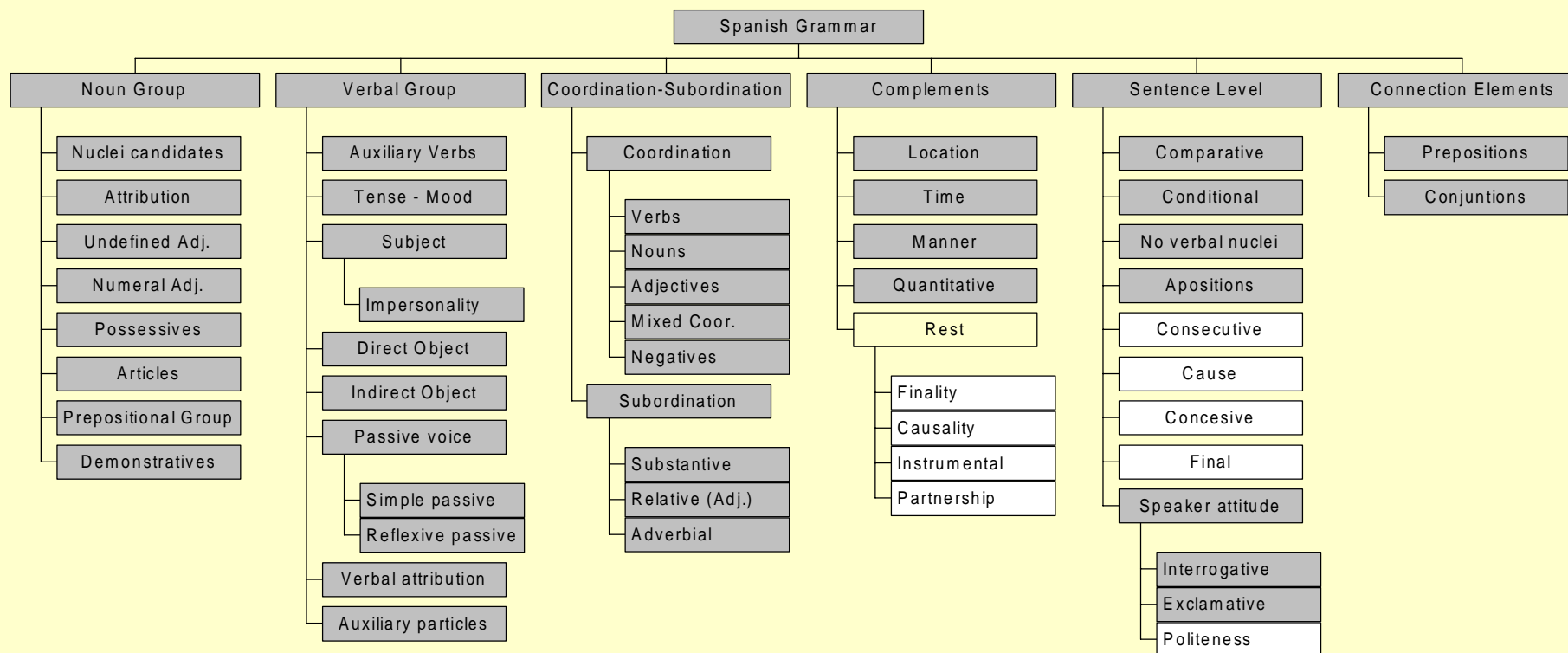
TOOLS

- DeCo Spanish Dictionary Validation tool.
 - Completeness Checker.
 - Insertion of new rules.
 - Logic Rule Verification.
- Tool for Syntactic Validation of UNL files (Corpus).





Grammar Coverage Estimation





Deco Maintenance Tool

The screenshot displays the Deco Maintenance Tool interface. The main window, titled "VLR [-Verificación Lógica de Reglas-]", shows the following statistics:

Resultados	
Número de líneas:	8298
Número de Reglas:	4208
Número de comentarios:	1.495
Número de errores:	

The "Herramientas" section includes buttons for "Clasificador de reglas" and "Verificador de Completitud".

The "Verificador de Completitud" dialog box provides detailed analysis:

Marcadores del diccionario	Marcadores premisa	Marcadores conclusión
CNT (no usado) 1PL 1SG 2PL 2SG 3PL	1pl 1sg 2pl 2sg 3pl 3sq	<ptn 1pl 1sg 2pl 2sg 3pl
Total: 586	Total premisas: 281	Total conclusiones: 270
Usados por reglas: 471	Premisas falsas: 65	Conclusiones inalcanzables: 52
Utilización: 80 %	Porcentaje: 23 %	Porcentaje: 19 %

The "Reglas" section shows a list of errors, including:

- {SHEAD_UNDEF,^blk:+blk} "[],BLK:+SHEAD"^(^PERIOD)P80: -- error -- Regla no disparable (45)
- "CNJ::man:" {@entry,@pred:+INTROD_CONJ} P100: -- aviso -- Marcadores redundantes (72)
- {SHEAD_UNDEF,^blk:+blk} "[],BLK:+SHEAD"^(^PERIOD)P80: -- error -- Regla no disparable (155)

Summary statistics for rules:

Reglas no disparables: 148	Porcentaje: 3 %
Total reglas: 4208	

The interface also includes a "Aceptar" button and a status bar at the bottom indicating "Para obtener ayuda presione F1" and "ln 1, col 1".