



Universitat d'Alacant  
Universidad de Alicante

**Esta tesis doctoral contiene un índice que enlaza a cada uno de los capítulos de la misma.**

**Existen asimismo botones de retorno al índice al principio y final de cada uno de los capítulos.**

**[Ir directamente al índice](#)**

**Para una correcta visualización del texto es necesaria la versión de [Adobe Acrobat Reader 7.0](#) o posteriores**

**Aquesta tesi doctoral conté un índex que enllaça a cadascun dels capítols. Existeixen així mateix botons de retorn a l'índex al principi i final de cadascun dels capítols .**

**[Anar directament a l'índex](#)**

**Per a una correcta visualització del text és necessària la versió d' [Adobe Acrobat Reader 7.0](#) o posteriors.**



Universitat d'Alacant  
Universidad de Alicante

Departamento de Lenguajes y Sistemas Informáticos  
Universidad de Alicante



## Resolución de información temporal y su aplicación a la Búsqueda de Respuestas

Estela Saquete Boró

Memoria para optar al grado de Doctor en Informática bajo la dirección  
de

Dr. Patricio Martínez Barco  
Dr. Rafael Muñoz Guillena

Alicante, 28 de febrero de 2005



Universitat d'Alacant  
Universidad de Alicante



Departamento de Lenguajes y Sistemas Informáticos  
Universidad de Alicante

Universitat d'Alacant  
Universidad de Alicante

## Resolución de información temporal y su aplicación a la Búsqueda de Respuestas

Estela Saquete Boró

Memoria para optar al grado de Doctor en Informática  
bajo la dirección de

Dr. Patricio Martínez Barco  
Dr. Rafael Muñoz Guillena

Alicante, 28 de febrero de 2005

Cofinanciado por el Gobierno de España (CICyT) con los proyectos número TIC2000-0664-C02-02 y número TIC2003-07158-C04-01, y el Gobierno de la Comunidad Valenciana (OCyT) con el proyecto número CTIDIB-2002-151.



Universitat d'Alacant  
Universidad de Alicante



Universitat d'Alacant  
Universidad de Alicante

## Prefacio

El trabajo que aquí se presenta supone la culminación de mi trayectoria investigadora desde sus orígenes en el año 1998, fecha en la que comencé mis estudios doctorales hasta la actualidad. Introduzco a continuación este trabajo de Tesis con un breve resumen de mi trayectoria investigadora.

Finalizados los estudios de Ingeniería en Informática Superior en la Universidad de Alicante, mi actividad investigadora se inicia en el año 1998 con la realización de los cursos del programa de doctorado **“INGENIERÍA LINGÜÍSTICA, APRENDIZAJE AUTOMÁTICO Y RECONOCIMIENTO DE FORMAS”** perteneciente al Departamento de Lenguajes y Sistemas Informáticos (DLSI) de la Universidad de Alicante (UA), compaginando estos estudios con mi trabajo como becaria en el Laboratorio Multimedia de la Universidad de Alicante y posteriormente en el departamento de Informática de la Caja de Ahorros del Mediterráneo (CAM), finalizando la tesina en 2002 y obteniendo de esta forma la Suficiencia Investigadora.

Ese mismo año, en Noviembre, ingresé en el Departamento de Lenguajes y Sistemas Informáticos en calidad de Ayudante (LOU). A su vez, también entro a formar parte del Grupo de Investigación de Procesamiento del Lenguaje y Sistemas de Información perteneciente al DLSI y liderado por el Dr. Manuel Palomar Sanz donde continúo mi investigación en el campo del Procesamiento del Lenguaje Natural, en concreto, dedicándome profundamente al estudio de como reconocer y resolver la información temporal en textos escritos, usando diferentes técnicas, hasta el momento no probadas, en este campo.



Universitat d'Alacant  
de Alicante

## II Prefacio

Durante este periodo de realización de Tesis se han producido diversas publicaciones y ponencias relevantes que se destacan a lo largo de la tesis. Sin embargo me gustaría resaltar aquí dos de ellas que considero especialmente importantes dada la trascendencia científica del marco en el que se producen. Se trata por una parte, de la presentación de parte de este trabajo en el último congreso celebrado de la Association of Computational Linguistics (ACL) en julio de 2004, que constituye uno de los congresos sobre Lingüística Computacional de mayor reconocimiento internacional y de los más competitivos por su altísimo índice de rechazos durante el proceso de selección. Y por otra parte, la contribución a la revista "Data and Knowledge Engineering Journal" (indizada en Science Citation Index) donde se sintetiza el trabajo de esta tesis y que se encuentra actualmente aceptado y pendiente de su publicación final.

Finalmente, me gustaría destacar un último factor que considero importante en la conclusión de esta Tesis y es la estancia de investigación que realicé en la Universidad de Edimburgo (United Kingdom) en el periodo comprendido entre mayo y agosto de 2004, bajo la supervisión de la Dra. Bonnie Webber y el Dr. Johan Bos, donde colaboré con su equipo de investigación en diferentes tareas relacionadas con la Búsqueda de Respuestas en inglés, y que me permitieron completar el trabajo aquí presentado.



Universitat d'Alacant  
Universidad de Alicante

## Agradecimientos

Al finalizar este trabajo, quiero mostrar mi profundo agradecimiento a todos aquellos que me han ayudado a que finalmente sea una realidad.

En primer lugar me gustaría destacar todo el apoyo y ayuda recibida por mis directores de Tesis: Patricio Martínez Barco y Rafael Muñoz Guillena. Creo que sin unos directores de Tesis con esta calidad profesional y personal hubiera sido mucho más difícil llegar hasta donde he llegado. Muchos han sido los consejos, las ideas y las ayudas que me han proporcionado y que siempre he valorado enormemente, pero sobre todo, ha sido su apoyo incondicional el que me ha alentado a seguir siempre hacia delante.

También estoy particularmente agradecida a José Luis Vicedo, compañero con el que he colaborado en múltiples artículos, y que ha ayudado en gran medida a enfocar mi Tesis por un camino innovador, dándole una mayor calidad al trabajo realizado. Por otro lado, tampoco puedo olvidarme de agradecer a Manuel Palomar ese primer enfoque que me dio, cuando yo apenas sabía que era el Procesamiento de Lenguaje Natural, de lo que podría ser mi investigación, estando siempre pendiente de mi trabajo y mi persona.

Me gustaría también agradecer la labor de Antonio Ferrández, quien ha actuado como supervisor final del trabajo, resolviéndome también cualquier inquietud que me pudiera surgir.

Tampoco puedo pasar por alto el apoyo de mis compañeros del Grupo de Procesamiento del Lenguaje y Sistemas de Información: Andrés Montoyo, Armando Suárez, Paloma Moreda, Borja Navarro, Cristina Cachero, Fernando Llopis, Jaime Gómez, Jesús Peral, Juan Carlos Trujillo, Maxi Saiz, Rafa Romero, Irene Ga-



Universitat d'Alacant  
de Alicante

#### IV Agradecimientos

rrigós, Santiago Meliá, Sonia Vázquez, Rafa Junior, además de los ya nombrados Patricio Martínez Barco, Rafael Muñoz, Jose Luis Vicedo, Manuel Palomar y Antonio Ferrández, así como al resto de colaboradores del grupo que sería muy largo nombrar aquí. Su ayuda ha estado presente tanto en el plano profesional como en el personal.

A todos los compañeros del Departamento de Lenguajes y Sistemas Informáticos que me han apoyado firmemente en mi periodo de investigación.

A todos los revisores anónimos de congresos y revistas a los que hemos sometido nuestra investigación, ya que con sus orientaciones y críticas nos han permitido centrar nuestros objetivos.

Finalmente quiero hacer un agradecimiento a toda mi familia, haciendo una mención especial a mi hermana Desireé y mis padres, y por supuesto a mi perro Bubú, que ha estado ahí desde el principio y no se ha quejado nunca. También agradecer su apoyo a mis amigos y especialmente a Estefanía, que siempre ha sido muy comprensiva. Y por último, pero no menos importante, agradecer con mucho cariño a Adrián su último aliento.



## Índice general

<b>1.. Introducción</b> .....	1
1.1. Motivación .....	4
1.2. Objetivos .....	12
1.3. Organización de la Tesis .....	14
<b>2.. Trabajo previo</b> .....	19
2.1. Ontologías temporales .....	22
2.1.1. Ontologías basadas en la información verbal ..	22
2.1.2. Ontologías basadas en las expresiones tempo- rales .....	26
2.2. Tareas para el tratamiento de la información temporal	28
2.2.1. Anotación de expresiones temporales .....	30
2.2.2. Ubicación de eventos en la línea temporal ....	31
2.2.3. Establecimiento de relaciones y orden entre eventos .....	32
2.3. Esquemas de Anotación Temporal .....	32
2.4. Estado del arte en sistemas de Anotación Temporal.	39
2.5. Estado del arte en sistemas de Ubicación de Eventos	45
2.6. Estado del arte en sistemas de Anotación de Rela- ciones y Ordenación de Eventos .....	48
2.7. Conclusiones del capítulo .....	50
<b>3.. El modelo temporal</b> .....	55
3.1. Ontología Temporal .....	56
3.1.1. Conceptos temporales .....	56
3.1.2. Relaciones entre conceptos .....	60
3.2. Reglas del modelo temporal .....	60
3.3. Conclusiones del capítulo .....	63



## VI Índice general

<b>4.. TERSEO Monolingüe: Sistema de Resolución de Expresiones Temporales aplicado a la Ordenación de Eventos</b> .....	65
4.1. Esquema de Anotación de las expresiones temporales	66
4.2. Descripción del sistema .....	68
4.2.1. POS TAGGER .....	70
4.2.2. Analizador Temporal .....	71
4.2.3. Unidad de Resolución de Expresiones Temporales .....	72
4.2.4. Unidad de Ordenación de eventos .....	73
4.3. Conclusiones del capítulo .....	75
<b>5.. Extensión del sistema: Adquisición automática de conocimiento para otras lenguas</b> .....	77
5.1. Multilingüidad automática .....	78
5.1.1. Unidad de Traducción .....	81
5.1.2. Unidad de Depuración de Expresiones Temporales .....	82
5.1.3. Unidad de Generación Automática de la Gramática Destino .....	82
5.1.4. Unidad de Palabras Clave .....	83
5.1.5. Motor de búsqueda de Nuevas Expresiones Temporales .....	85
5.1.6. Unidad de Asignación de Reglas de Resolución	86
5.2. Conclusiones del capítulo .....	87
<b>6.. Aplicaciones: Búsqueda de Respuestas Temporal</b> ..	89
6.1. Propuesta de una taxonomía de preguntas temporales	92
6.2. Arquitectura multicapa de un sistema de Búsqueda de Respuestas .....	94
6.3. Arquitectura de una sistema de Búsqueda de Respuestas Temporal .....	96
6.4. Unidad de Descomposición de Preguntas .....	98
6.4.1. Identificación de Tipo .....	99
6.4.2. Reconocimiento y Resolución de Expresiones Temporales .....	100
6.4.3. División de la Pregunta .....	101

6.4.4. Ejemplo de funcionamiento de la unidad de Descomposición .....	104
6.5. Unidad de Recomposición de la Respuesta.....	105
6.5.1. Preprocesamiento de la salida del sistema ge- neral de Búsqueda de Respuestas .....	106
6.5.2. Filtrado de las respuestas individuales .....	107
6.5.3. Comparación y composición de la respuesta final .....	108
6.6. Adaptación del sistema al castellano .....	109
6.6.1. Ejemplo de funcionamiento de la unidad de Descomposición para el castellano .....	111
6.7. Conclusiones del capítulo .....	112
<b>7.. Evaluación del sistema.....</b>	<b>115</b>
7.1. Evaluación del sistema TERSEO monolingüe .....	116
7.2. Evaluación del sistema TERSEO multilingüe .....	120
7.3. Evaluación TERN 2004 .....	123
7.3.1. Transformación de las etiquetas TERSEO a las etiquetas TIMEX2 .....	127
7.3.2. Resultados Evaluación TERN 2004 .....	133
7.3.3. Análisis de errores .....	136
7.3.4. Evaluación alternativa .....	142
7.4. Evaluación del sistema de Búsqueda de Respuestas Temporal en inglés .....	144
7.4.1. Evaluación de la descomposición de la pregunta.....	144
7.4.2. Comparativa del sistema de BR Temporal con un sistema general de BR .....	147
7.5. Evaluación CLEF2004 .....	151
7.6. Enriquecimiento del sistema TERSEO .....	153
7.7. Conclusiones .....	154
<b>8.. Herramientas .....</b>	<b>157</b>
8.1. Analizador temporal.....	158
8.2. Sistema TERSEO .....	159
8.3. Sistema Búsqueda de Respuestas Temporal .....	162
8.3.1. Descomposición de la pregunta .....	162
8.3.2. Recomposición de la pregunta .....	164



Universitat d'Alacant  
Universidad de Alicante

VIII Índice general

8.4. Conclusiones .....	166
<b>9.. Conclusiones .....</b>	<b>169</b>
9.1. Trabajos en progreso .....	176
9.2. Trayectoria investigadora .....	178
9.2.1. Estudio inicial (1998-2000) .....	179
9.2.2. Consolidación y especialización (2000-2003) ..	179
9.2.3. Expansión y aplicación (2004) .....	180
<b>A.. Anexo Formalización de las reglas de resolución del modelo temporal .....</b>	<b>183</b>
A.1. Reglas de resolución desde la perspectiva de la lógi- ca de 1er orden .....	183
<b>B.. Anexo Modelo Temporal castellano .....</b>	<b>193</b>
B.1. Gramática en castellano .....	193
B.1.1. Reglas para la identificación de Expresiones Temporales Explícitas .....	193
B.1.2. Reglas para la identificación de Expresiones Temporales Implícitas .....	194
B.2. Expresiones temporales en castellano clasificadas según la ontología temporal .....	194
<b>C.. Anexo Modelo Temporal inglés .....</b>	<b>205</b>
C.1. Gramática en inglés .....	205
C.1.1. Reglas para la identificación de Expresiones Temporales Explícitas .....	205
C.1.2. Reglas para la identificación de Expresiones Temporales Implícitas .....	206
C.2. Expresiones temporales en inglés clasificadas según la ontología temporal .....	206
<b>Referencias .....</b>	<b>221</b>



Universitat d'Alacant  
Universidad de Alicante

## Índice de cuadros

2.1. Resumen de esquemas de anotación . . . . .	52
2.2. Cronología de las aproximaciones existentes . . . . .	53
3.1. Funciones primitivas definidas por Hobbs y usadas en el modelo temporal propuesto . . . . .	62
3.2. Funciones primitivas nuevas . . . . .	62
3.3. Reglas resolución DIA CONCRETO . . . . .	63
4.1. Ejemplo de reglas de identificación de expresiones explí- tas e implícitas . . . . .	72
4.2. Ejemplo de algunas reglas de resolución del modelo temporal . . . . .	73
5.1. Ejemplo de traducciones para los tres traductores . . . . .	82
5.2. Ejemplo de algunas palabras clave temporales . . . . .	84
6.1. Claves de ordenación para cada señal . . . . .	102
7.1. Evaluación del sistema TERSEO monolingüe . . . . .	119
7.2. Evaluación del sistema TERSEO multilingüe . . . . .	121
7.3. Conjunto de atributos del etiquetado TIMEX2 . . . . .	125
7.4. Significado de cada elemento del formato para duraciones	129
7.5. Posibles valores del atributo ANCHOR_DIR . . . . .	129
7.6. Transformación realizada para duraciones . . . . .	130
7.7. Elementos utilizados para representar expresiones difusas	131
7.8. Transformación para expresiones difusas . . . . .	131
7.9. Posibles valores atributo MOD . . . . .	132
7.10. Resultados entrenamiento TERSEO desglosado por atri- butos (Broadcast News) (I) . . . . .	133



## X Índice de cuadros

7.11. Resultados entrenamiento TERSEO desglosado por atributos (Broadcast News) (II) .....	134
7.12. Tamaño del corpus de test .....	134
7.13. Resultados de TERSEO en TERN 2004 desglosado por atributos (I) .....	135
7.14. Resultados de TERSEO en TERN 2004 desglosado por atributos (II) .....	135
7.15. Resultados de TERSEO en evaluación alternativa TERN 2004 desglosado por atributos (I) .....	143
7.16. Resultados de TERSEO en evaluación alternativa TERN 2004 desglosado por atributos (II) .....	143
7.17. Evaluación de la descomposición de la pregunta en inglés	146
7.18. Evaluación del sistema de BR Temporal completo .....	150
7.19. Resultados Tarea Piloto CLEF 2004 .....	152
7.20. Evaluación del sistema TERSEO monolingüe enriquecido	154
7.21. Comparación del sistema TERSEO monolingüe y multilingüe .....	154
8.1. Resultados obtenidos por el analizador .....	160
B.1. Simbología utilizada en el anexo .....	193
B.2. Reglas para la identificación de expresiones explícitas en castellano .....	194
B.3. Reglas de la gramática en castellano con nomenclatura MACO .....	195
B.4. Listado de palabras clave temporales en castellano .....	196
B.5. Expresiones temporales en castellano (I) .....	197
B.6. Expresiones temporales en castellano (II) .....	198
B.7. Expresiones temporales en castellano (III) .....	199
B.8. Expresiones temporales en castellano (IV) .....	200
B.9. Expresiones temporales en castellano (V) .....	201
B.10. Expresiones temporales en castellano (VI) .....	202
B.11. Expresiones temporales en castellano (VII) .....	203
B.12. Expresiones temporales en castellano (VIII) .....	204
C.1. Simbología utilizada en el anexo .....	205
C.2. Reglas para la identificación de expresiones explícitas en inglés .....	206



Índice de cuadros XI

C.3. Reglas de la gramática en inglés con nomenclatura MA-CO(I) .....	207
C.4. Reglas de la gramática en inglés con nomenclatura MA-CO (II) .....	208
C.5. Reglas de la gramática en inglés con nomenclatura MA-CO (III) .....	209
C.6. Listado de palabras clave temporales en inglés.....	209
C.7. Expresiones temporales en inglés (I) .....	210
C.8. Expresiones temporales en inglés (II).....	211
C.9. Expresiones temporales en inglés (III) .....	212
C.10Expresiones temporales en inglés (IV) .....	213
C.11Expresiones temporales en inglés (V).....	214
C.12Expresiones temporales en inglés (VI) .....	215
C.13Expresiones temporales en inglés (VII) .....	216
C.14Expresiones temporales en inglés (VIII) .....	217
C.15Expresiones temporales en inglés (IX) .....	218
C.16Expresiones temporales en inglés (X).....	219



Universitat d'Alacant  
Universidad de Alicante



## Índice de figuras

2.1. Representación gráfica de la ontología temporal en Moens <i>et al.</i> 1988 .....	24
4.1. Representación gráfica de TERSEO .....	69
4.2. Ordenación de los eventos en la línea temporal .....	76
5.1. Representación gráfica de la unidad TERILI .....	79
5.2. Representación gráfica de la extensión automática multilingüe de TERSEO .....	80
6.1. Representación gráfica de un sistema de Búsqueda de Respuestas multicapa .....	95
6.2. Representación gráfica del sistema de Búsqueda de Respuestas temporal .....	96
6.3. Árbol de decisión para Identificación de Tipo .....	100
6.4. Árbol de decisión para la división de la pregunta .....	103
6.5. Árbol de decisión para la división de la pregunta en castellano .....	110
7.1. Representación gráfica de las medidas de precisión y cobertura .....	119
7.2. Comparativa F-Medida de los diferentes sistemas para atributo VAL .....	136
7.3. Comparativa F-Medida de los diferentes sistemas para atributo TEXT .....	136
7.4. Comparativa F-Medida de los diferentes sistemas para atributo SET .....	137
7.5. Comparativa F-Medida de los diferentes sistemas para atributo MOD .....	137



Universitat d'Alacant  
Universidad de Alicante

#### XIV Índice de figuras

<b>7.6. Comparativa F-Medida de los diferentes sistemas para atributo ANCHOR_VAL</b> .....	138
<b>7.7. Comparativa F-Medida de los diferentes sistemas para atributo ANCHOR_DIR</b> .....	138
<b>7.8. Comparativa F-Medida de los diferentes sistemas para sólo detección</b> .....	139
8.1. Analizador temporal .....	159
8.2. Pantalla principal sistema TERSEO .....	161
8.3. Pantalla resultado sólo resolución sistema TERSEO ....	162
8.4. Pantalla resultado Reconocimiento y Resolución sistema TERSEO .....	163
8.5. Pantalla resultado Ordenación Eventos sistema TERSEO	164
8.6. Pantalla principal sistema BR Temporal(Descomposición)	164
8.7. Pantalla principal sistema BR Temporal(Recomposición)	165



Universitat d'Alacant  
Universidad de Alicante

## 1. Introducción

*“Pero, ¿qué es el tiempo? ¿Quién podrá fácil y brevemente explicarlo? ¿Quién puede formar idea clara del tiempo para explicarlo después con palabras? Por otra parte, que cosa más familiar y manida en nuestras conversaciones es el tiempo. Entendemos muy bien lo que significa esta palabra cuando la empleamos nosotros y también cuando la oímos pronunciar a otros. ¿Qué es, pues el tiempo? Sé muy bien lo que es, si no se me pregunta. Pero cuando quiero explicárselo al que me pregunta, no lo sé.”* (San Agustín, *Confesiones*, XI, 14)

Desde que hace más de seiscientos años San Agustín se plantea todas estas cuestiones, muchos han sido los intentos de la humanidad por encontrar una definición exacta del concepto del tiempo. Desde *“tiempo”* inmemorial, los humanos tratamos de medir el paso del tiempo para organizar nuestra vida y ordenar nuestro destino. Quien les iba a decir, hace 4000 años a los egipcios, cuando inventaron el primer reloj solar, que este invento evolucionaría tanto como para acabar siendo lo suficientemente pequeño para ser llevado en la muñeca midiendo el tiempo con una precisión casi meridiana. Y no sólo se miden las horas, también los días, meses, años,... con sistemas como los calendarios, que además adoptan diferentes formas a lo largo de la historia y para las diferentes culturas.

La humanidad ha considerado fundamental la medida del tiempo para poder organizar sus actividades, sincronizar sus relaciones sociales, establecer predicciones y situar su pasado. Pensad si no, que sería de nuestra vida cotidiana si el tiempo no tuviera en ella



2 1. Introducción

un papel fundamental. Además, la distinción entre presente, pasado y futuro es esencial para la temporalidad. Un determinado evento, que ahora es presente, fue futuro y será pasado. Tres situaciones que aunque incompatibles están interrelacionadas. Además, el pasado sigue vivo en el presente, en los recuerdos y de algún modo también el futuro está contenido en el presente a través de la esperanza. Si el futuro no existe, el presente pierde sentido.

*“Mal acabará quien pretenda adentrarse en el **futuro**, ignorando lo que sucedió en el **pasado** porque entonces no vivirá el **presente**”.* (Proverbio oriental)

*“El **instante** es la continuidad del **tiempo**, pues une el **tiempo pasado** con el **tiempo futuro**”.* (Aristóteles)

Además, es importante destacar que no sólo se habla de temporalidad en cifras, sino también usando expresiones que denominaremos expresiones temporales y que además, por las características del lenguaje en general, pueden ser un conjunto muy amplio puesto que no hay una única manera de expresar un determinado concepto. La lengua es rica, y por tanto, nos permite expresar el tiempo, no sólo en fechas y horas sino también en expresiones que se refieren a estas fechas pero que no las nombran explícitamente.

*“Siguen pasando **los días**  
y sólo tu voz me sana,  
escuchando la armonía,  
que ojalá te vea **mañana**.*

***Hoy es un día más,**  
ya no se qué hacer,  
no creo en verte jamás,  
aunque sufro más que **ayer**.  
No te quito de mi mente,  
ya no sé quién soy,  
esperando **aquel instante**,  
que tal vez sea **hoy**.”*



(Mario de la Cuadra G.)

Y no es únicamente una cuestión de la riqueza del lenguaje, las expresiones temporales también implican el poder relacionar unos elementos con otros en un determinado texto. Imaginemos que únicamente utilizáramos fechas concretas para referirnos a los eventos, por ejemplo:

*“El 25 de diciembre se produjo un accidente. El 26 de diciembre ya se habían detectado las causas”.*

Sin embargo, si se utilizan expresiones temporales para narrar el texto, el resultado sería el siguiente:

*“El 25 de diciembre se produjo un accidente. Al día siguiente ya se habían detectado las causas”.*

Como se puede observar comparando estos dos ejemplos, a pesar de que los dos se están refiriendo a la misma información, en el segundo caso, los diferentes eventos del texto se encuentran interrelacionados entre ellos, existe una conexión entre los diferentes elementos del texto.

Un humano es capaz de interpretar todas estas expresiones y tener un concepto claro del momento en el tiempo al que dicha expresión se están refiriendo y la relación entre los diferentes eventos ocurridos. El problema surge cuando se requiere el manejo de una gran cantidad de información y la interpretación del contenido de la misma. Esta tarea es costosa desde el punto de vista de su realización manual por parte de un humano y es ahí donde las máquinas y los procesos automáticos cobran importancia. Pero las máquinas no eran capaces por sí solas de pensar como un humano. Diversos investigadores centraron sus esfuerzos en crear sistemas que sean capaces de procesar el lenguaje natural, y en concreto para este caso la información temporal, proporcionando al sistema automático esa parte de conocimiento que no posee. Posteriormente, una vez que el sistema es capaz de asociar una fecha concreta a una expresión temporal, también es capaz de or-



denar los eventos asociados a esas fechas o extraer información relacionada con un evento concreto. Esto facilita en gran medida, e incluso soluciona en muchos casos, la tarea de extraer esta información que necesita ser buscada entre cantidades ingentes de documentos.

## 1.1 Motivación

El Procesamiento del Lenguaje Natural cobra cada vez más importancia debido al enorme número de documentos escritos en formato digital existentes y sobre todo, debido al auge de Internet. El Procesamiento del Lenguaje Natural (PLN) (Moreno *et al.*, 1999) es una disciplina que forma una parte esencial dentro de la Inteligencia Artificial y que se encarga de investigar y formular mecanismos computacionalmente efectivos que faciliten la interrelación hombre/máquina y permitan una comunicación menos rígida que los lenguajes formales. En esta disciplina se distinguen dos tipos de tareas: tareas finales e intermedias.

Dentro de las tareas finales destacan diversas aplicaciones relacionadas con Internet como son la recuperación de información, los sistemas de búsqueda de respuestas y la extracción de información. Existen a su vez, otras aplicaciones más relacionadas con el tratamiento o manejo de datos en textos digitalizados como son la traducción automática, la generación de resúmenes, los sistemas de diálogo, etc. A continuación se detallan un poco más cada una de estas aplicaciones:

- *Recuperación de información* (RI), una buena definición del término recuperación de información se encuentra en Lancaster (1968): “Un sistema de recuperación de información no informa a un usuario con respecto a una pregunta que éste realiza, sino que meramente informa de la existencia de documentos relacionados con dicha pregunta”. Por tanto, los sistemas de recuperación de información procesan colecciones de textos según una condición de búsqueda y proporcionan como resultado una lista ordenada de todos aquellos textos en los que dicha información ha sido encontrada. Estas tareas alcanzan mayor importancia

teniendo en cuenta que, según Hausser (1999), se calcula que en 10 años se habrá duplicado la información escrita durante toda la historia y por tanto cada vez se hace más necesario poder extraer y recuperar información a través de los contenidos semánticos de los textos. Muchos autores consideran la extracción de información como un proceso posterior a la recuperación de información. La recuperación de información ha sido tratada por diferentes autores y desde diferentes puntos de vista. En este sentido, cabe destacar el interés despertado por este tipo de sistemas en la comunidad investigadora y que ha llevado en los últimos años a la celebración de competiciones internacionales como TREC (2004) y CLEF (2004), dónde diferentes trabajos han aportado sus conclusiones a esta tarea a través de la participación de sus sistemas en estas competiciones, como, por ejemplo, Savoy (2004), Llopis *et al.* (2002; 2004) y Martínez Santiago *et al.* (2004).

- *Extracción de información (EI)*, los sistemas de extracción de información procesan textos que contienen una información relevante de forma no estructurada y los convierten de forma total o parcial en un conjunto de datos relevantes (Smeaton, 1997). La diferencia entre recuperación de información y extracción de información radica en que los primeros localizan documentos relevantes mientras que los segundos partiendo ya de los documentos relevantes, extraen la información relevante. En Wilks (1997) y Gaizauskas y Wilks (1998) se introduce el concepto de información pre-especificada o plantilla, donde el usuario especifica una serie de items de los cuales le interesa extraer información.

Un ejemplo de un sistema de extracción de información sería el sistema EXIT (Llopis *et al.*, 1998). Este sistema tiene como objetivo la extracción de información de compra-venta de textos notariales.

Además, al igual que ocurre con la recuperación de información, la tarea de extracción de información también se trata en conferencias como las conferencias Message Understanding



Conference (MUC)<sup>1</sup>, lo cual ha proporcionado la posibilidad de comparar y evaluar los diferentes sistemas dedicados a la extracción de información.

Puesto que la extracción de información es dependiente del dominio en el que trabaja el sistema, también se han realizado estudios sobre como adaptar automáticamente los sistemas utilizando técnicas de aprendizaje automático, reduciendo así el coste de dicha adaptación (Turmo, 2003).

- *Sistemas de búsquedas de respuestas (question answering)(QA)*, Se puede definir la búsqueda de respuestas (BR) como aquella tarea automática realizada por ordenadores que tiene como finalidad la de encontrar respuestas concretas a necesidades precisas y arbitrarias de información formuladas por los usuarios (Vicedo *et al.*, 2003).

Las primeras investigaciones en este campo utilizaron técnicas de recuperación de información para realizar el proceso de la búsqueda de respuestas (Cormack *et al.*, 1999; Fuller *et al.*, 1999; Allan *et al.*, 2000). Sin embargo, estas aproximaciones presentaron un pobre rendimiento en tareas en las que se requería una respuesta escueta y precisa como contestación a la pregunta.

Inmediatamente se empezó a experimentar con la aplicación de técnicas de PLN cada vez más complejas para mejorar la precisión a la hora de extraer una respuesta. Aunque son varios los sistemas que aplican técnicas enmarcadas en estos niveles, sólo el trabajo descrito en Moldovan *et al.* (2002) consigue un nivel de satisfacción que justifica el esfuerzo empleado en su aplicación.

En los últimos años, ha habido un cambio de tendencia orientado a la utilización de la recuperación de información basada en documentos o pasajes y posteriormente, a la aplicación sobre estos pasajes o documentos de técnicas superficiales de PLN en detrimento del uso de técnicas más complejas (Soubbotin & Soubbotin, 2002; Magnini *et al.*, 2002; Magnini *et al.*, 2003; Yang & Chua, 2002).

<sup>1</sup> [http://www.itl.nist.gov/iaui/894.02/related\\_projects/muc/](http://www.itl.nist.gov/iaui/894.02/related_projects/muc/)

- *Traducción automática* (TA), los sistemas de traducción automática se definen como aquellos sistemas que traducen textos escritos en un lenguaje origen a un lenguaje destino (Forcada & Pérez, 2004). En una sociedad multilingual como en la que hoy vivimos, sistemas de estas características son totalmente necesarios. Los pioneros en la creación de este tipo de sistemas surgieron en los años 50 y desde entonces se han ido mejorando las técnicas hasta la actualidad (Hutchins, 2001).

Las técnicas actuales de traducción automática se clasifican en:

- Traducción directa: Se trata de una traducción palabra a palabra sin representación intermedia de la información a traducir (Hutchins & Somers, 1992).
- Traducción indirecta: por transferencia o por interlingua. En la traducción indirecta primero se realiza una representación intermedia que será utilizada posteriormente para realizar la traducción (Arnold *et al.*, 1994).
- Aproximaciones inductivas a la traducción automática: Mientras que las técnicas nombradas anteriormente son de naturaleza deductiva, actualmente están aumentando los sistemas que utilizan aproximaciones inductivas también denominada “traducción automática basada en corpus”. Se destacan dos tendencias principales:
  - Traducción automática estadística: Usa modelos estadísticos para obtener las traducciones más probables de las oraciones (Brown *et al.*, 1990).
  - Traducción automática basada en ejemplos: Intenta construir plantillas de traducción generales a través de ejemplos (Carl *et al.*, 2001).
- *Generación de resúmenes automática*(*summarization*), los sistemas de generación de resúmenes procesan un texto proporcionando como salida una sinopsis de todo el texto procesado (Mani & Maybury, 2001). En concreto, en esta tarea, es muy importante la necesidad de analizar los diferentes factores de cohesión léxica del documento, como por ejemplo, la repetición léxica (Boguraev & Neff, 2000).



Cabe destacar las conferencias DUC<sup>2</sup> (Document Understanding Conference) enfocadas a sistemas de generación de resúmenes y a la evaluación de los mismos. Entre los autores que han presentado su sistema en alguna de las conferencias DUC con buenos resultados destacan Baldwin y Ross (2001).

Además, muchos sistemas de generación de resúmenes no son completamente automáticos, sino que permiten al usuario interactuar con el sistema para mejorar la salida del mismo (Orasan *et al.*, 2003).

- *Sistemas de diálogo (SD)*, son sistemas en los que se estudia el comportamiento del diálogo en su aspecto intelectual. En estos sistemas se produce una comunicación hombre/máquina en lenguaje natural, por lo que el sistema debe ser capaz de interpretar la petición del usuario, extraer la información que busca ese usuario y contestar en lenguaje natural a dicha petición. En el trabajo realizado por Fernández (Fernández, 2000) se muestran las tres tendencias de investigación actuales en el campo de los sistemas de diálogo visto desde las tres grandes perspectivas de la Inteligencia Artificial. Desde el punto de vista *metodológico*, dicha investigación pretende una obtención de modelos de diálogo hombre-máquina así como su formalización teórica. Desde el punto de vista *cognitivo* se busca la motivación de los modelos propuestos. Y por último, desde el punto de vista *computacional* lo que se pretende es la implementación de los modelos de gestión de diálogo, y su integración en otros componentes de procesamiento del lenguaje natural como el reconocimiento y generación de voz o la interpretación y generación del lenguaje natural. Estos componentes deben incluir la resolución de sus fenómenos lingüísticos. En concreto, autores como Martínez-Barco y Palomar (2000) (Martínez-Barco, 2001) (Palomar & Martínez-Barco, 2001) han estudiado las necesidades de la última línea de investigación, haciendo especial énfasis en la resolución de fenómenos lingüísticos como la anáfora. Dicho fenómeno será explicado posteriormente.

<sup>2</sup> <http://duc.nist.gov/>



En cuanto a la arquitectura de los sistemas de diálogo, según Bernsen *et al.* (1998), debe estar formada por los siguientes módulos:

- Procesamiento del habla (reconocimiento y síntesis).
- Procesamiento del lenguaje natural (comprensión y generación).
- Gestión del diálogo (control y contexto).

La información, según este autor, fluye secuencialmente de un nivel a otro. Sin embargo, trabajos recientes publicados por Allen *et al.* (2000), demuestran que el mantenimiento de una arquitectura con este flujo de información simple no es viable cuando se pretende construir un sistema capaz de gestionar diálogos de cierta complejidad. En este sentido, Allen *et al.* proporcionan una arquitectura genérica para los sistemas de diálogo por medio de la cual, existe un flujo principal de información desde la entrada de un enunciado hablado hasta la generación de la respuesta por parte de la máquina, pero además existe un flujo de información secundario entre los diferentes componentes que remarcan la necesidad de una interacción entre los distintos niveles más allá del flujo principal.

En las tareas intermedias se resuelven problemas característicos del lenguaje. Uno de los mayores problemas de la simulación del lenguaje humano a través de los ordenadores es el problema de la “ambigüedad”. Los diferentes tipos de ambigüedad que se presentan son:

- *Ambigüedad estructural*: La ambigüedad estructural aparece a nivel sintáctico. Se dice que una oración es ambigua estructuralmente cuando posee más de un árbol sintáctico de derivación. Por ejemplo: “Juan o Victor y Elena deberán ir”. El árbol sintáctico dependerá de la prioridad que le demos a uno de los dos nexos de unión. Este tipo de ambigüedad ha sido estudiada para diferentes idiomas, destacando el trabajo de Voutilainen y Järvinen (1995) para el inglés y los trabajos de autores como Atserias *et al.* (1998), Castellon *et al.* (1998), Martínez-Barco *et al.* (1998) para el castellano.



## 10 1. Introducción

- **Ambigüedad léxica:** Se presenta cuando una palabra tiene más de un significado. Se puede clasificar en:
  - **Pura:** Afecta sólo a nivel semántico. Las palabras son diferentes en contenido pero desempeñan un mismo papel sintáctico en una oración. Por ejemplo: “Me senté en el *banco*” y “Fui a cobrar al *banco*”.

Los métodos que tratan este tipo de ambigüedad pueden clasificarse en:

- **Métodos supervisados:** Utilizan un corpus anotado semánticamente (Mihalcea & Moldovan, 2001; Molina *et al.*, 2002; Suarez, 2004).
- **Métodos no supervisados:** Utilizan un corpus no anotado para resolver la ambigüedad (Rigau *et al.*, 1997; Rigau *et al.*, 1998; Montoyo & Palomar, 2000a; Montoyo & Palomar, 2000b; Magnini & Strapparava, 2000).

Otra posible clasificación de los métodos podría ser:

- **Métodos basados en conocimiento:** Utilizan conocimiento lingüístico previamente adquirido (Montoyo & Palomar, 2001).
- **Métodos basados en corpus:** Estos métodos usan técnicas estadísticas y de aprendizaje automático para inducir modelos de uso del lenguaje a partir de una gran cantidad de ejemplos de texto (Agirre & Martínez, 2004; Yarowsky, 1995; Suárez & Palomar, 2002).

Ambas clasificaciones se encuentran relacionadas puesto que en esta última clasificación, los métodos realizan aprendizaje supervisado si los ejemplos están etiquetados y aprendizaje no supervisado cuando no lo están.

- **Categorial:** una palabra presenta ambigüedad léxica categorial cuando aparte de tener diferentes significados, estos desempeñan una función sintáctica distinta en la oración (Brill, 1995). Por ejemplo: “El *cura* dio la misa”, “La *cura* será dolorosa” y “Esta enfermedad no se *cura*”.
- **Ambigüedad de ámbito de cuantificación:** Se presenta cuando aparecen en una misma oración un cuantificador existencial y un cuantificador universal. Por ejemplo: “El profesor recibió un regalo de todos los padres”. En esta frase “un regalo” es un



cuantificador existencial y “todos” es un cuantificador universal. Esta frase podría interpretarse como que el profesor ha recibido un regalo de cada padre o como que el profesor a recibido un único regalo de todos los padres (Allen, 1995).

- *Ambigüedad de función contextual*: Se define como las diversas interpretaciones que puede tener una frase dependiendo del contexto (Allen, 1995).
- *Ambigüedad referencial*: Se presenta cuando para un pronombre o sintagma nominal definido aparece más de una referencia. Surge entonces el concepto de anáfora. En el pasado, las referencias y descripciones fueron estudiadas por los filósofos Frege (1892) y Russell (1919). En la literatura actual se pueden encontrar diversas definiciones del fenómeno de la anáfora. Entre todas ellas destaca Hirst (1981) que define la anáfora como el mecanismo que nos permite hacer en un discurso una referencia abreviada a alguna entidad o entidades, con la confianza de que el receptor del discurso sea capaz de desabreviar la referencia y por consiguiente determinar la entidad a la que se alude.

Existen diferentes tipos de anáfora, según el tipo de antecedente del que se trate, entre las que habría que destacar:

- *Anáfora pronominal*: en este caso se utilizan pronombres para hacer referencias o incluso oraciones completas o situaciones anteriores. Para la resolución de este tipo de anáfora existen métodos que utilizan un análisis completo como Saiz-Noeda *et al.* (Saiz-Noeda *et al.*, 2001) y otros que realizan aproximaciones usando un conocimiento denominado pobre, mejorando así la tarea en tiempo y coste como, por ejemplo, los trabajos presentados en Mitkov (1995; 1997; 1998) y Ferrández (1998).
- *Descripciones definidas*: son expresiones anafóricas formadas por sintagmas nominales. Existen diversas técnicas para la resolución de descripciones definidas, entre las que destacan aquellas que están basadas en corpus, es decir, utilizan información obtenida a partir del corpus de trabajo (Poesio & Vieira, 1998; Cardie & Pierce, 1998). Destacar los trabajos de Muñoz y Ferrández (2000) y Muñoz *et al.* (2000) para descripciones definidas en castellano.



- Verbal: este tipo de anáfora se produce fundamentalmente en inglés mediante auxiliares (*do* o *have*). Esta anáfora puede referirse bien a un verbo o bien a una frase verbal (Webber, 1988).

Si se centra el problema en la ambigüedad referencial, en la mayoría de los sistemas, la resolución de la anáfora ayuda a mejorar la eficiencia de los mismos. Se destaca, por ejemplo, el trabajo presentado en Peral (2002) donde se aplica la resolución de la anáfora a la Traducción Automática. Además, otra tarea intermedia como es el reconocimiento de entidades (Muñoz *et al.*, 1998) también tiene un papel muy importante en esta mejora.

Desde el punto de vista del Procesamiento del Lenguaje Natural, el tratamiento de las expresiones temporales tiene mucha relación con el reconocimiento de entidades, puesto que una expresión temporal es una entidad y el reconocimiento de las mismas permite identificarlas dentro de un texto. A su vez, también tiene mucha relación con el concepto de anáfora y su resolución. Se entiende por anáfora el mecanismo que nos permite hacer en un discurso referencia a alguna entidad o entidades con la confianza de que el receptor del discurso sea capaz de interpretar la referencia y por consiguiente determinar la entidad a la que se alude. El caso de las expresiones temporales podría considerarse un tipo especial de anáfora, puesto que una expresión temporal hace referencia a un fecha concreta nombrada anteriormente y necesita esta fecha para poder ser interpretada de una manera precisa. Por ello, el tratamiento de las expresiones temporales cobra gran importancia dentro del Procesamiento del Lenguaje Natural.

## 1.2 Objetivos

El objetivo general que persigue este trabajo es el de desarrollar un sistema automático que permita identificar, clasificar y resolver las expresiones temporales que pueden aparecer en colecciones de textos digitales de noticias escritos en lenguaje natural. Además, también se pretende salvar la barrera de los diferentes idiomas existentes desarrollando un sistema capaz de trabajar en



un nivel multilingüe. Una vez resueltas las expresiones se pretende realizar una ordenación de los eventos relacionados con dichas expresiones con el fin de conseguir una cronología de dichos eventos. En la práctica, establecer una cronología implica que muchas tareas dentro del campo del Procesamiento de Lenguaje Natural, como por ejemplo, la realización de resúmenes, la búsqueda de respuestas a preguntas temporales, etc., se conviertan en retos fácilmente alcanzables.

Más concretamente, para poder conseguir el objetivo general de este trabajo serán necesarias las siguientes líneas de actuación:

- Estudio del estado del arte de todos aquellos sistemas que se dediquen al tratamiento de cualquier tipo de información temporal, ya sean verbos, expresiones temporales, etc. Además, se necesita conocer los modelos temporales existentes hasta el momento que sirven de base para los sistemas de tratamiento de información temporal.
- Estudio del ámbito de temporalidad que se desea para el sistema a desarrollar y a partir de ahí desarrollar un modelo temporal nuevo o utilizar uno ya existente dependiendo de las necesidades de dicho sistema.
- Definición de un esquema de anotación temporal nuevo o utilización de uno ya existente dependiendo de la aplicación posterior para la que va a ser usada dicha anotación. Interesa utilizar un esquema de anotación simple y concreto para una mayor facilidad en el manejo de la información anotada.
- Desarrollo de un sistema capaz de identificar, clasificar y resolver las expresiones temporales de textos en castellano, así como todos los recursos necesarios para el correcto funcionamiento del sistema, como por ejemplo, un analizador temporal parcial para el reconocimiento de las expresiones. El sistema, por tanto, será capaz de marcar las expresiones temporales con una etiqueta y en sus atributos aparecerá la información de las fechas concretas a las que se refieren estas expresiones. Se desea un sistema lo menos dependiente posible del idioma para su extensión a otras lenguas.



- Utilización del marcado de las expresiones temporales con fechas concretas para que el sistema pueda ser capaz de ordenar los eventos asociados con dichas expresiones anotadas.
- Realización de una extensión del sistema a otras lenguas de la manera más automática posible para que no se requiera conocimiento previo de la lengua a la que se desea extender el sistema.
- Aplicación del sistema de reconocimiento y resolución de expresiones temporales a tareas concretas del Procesamiento del Lenguaje Natural. Una de estas tareas puede ser la Búsqueda de Respuestas cuando las preguntas tienen un carácter temporal, puesto que actualmente no están siendo tratadas por los sistemas existentes.
- Evaluación de todos los sistemas propuestos, realizando tanto evaluaciones propias como a través de la participación en conferencias de evaluación internacionales con el objetivo de medir la eficiencia de los sistemas propuestos con respecto a los demás sistemas existentes. Con ello se consigue valorar si las novedades incluidas por los sistemas propuestos aportan mejores resultados.

Además, aunque la aplicación principal del sistema sea la Búsqueda de Respuestas Temporal, su aplicación no debe limitarse a este tipo de tarea de Procesamiento de Lenguaje Natural, sino que puede aplicarse a muchas otras tareas como son la Realización de Resúmenes o la Extracción de Información.

### 1.3 Organización de la Tesis

Este trabajo se ha estructurado en los siguientes capítulos:

En el capítulo 2 se presenta el estudio del estado del arte en relación a los sistemas que tratan información temporal. En primer lugar, se realiza un estudio de las diferentes ontologías temporales existentes hasta el momento. Posteriormente, los diferentes sistemas que realizan tratamiento de información temporal se agrupan en tres grupos, dependiendo de la tarea concreta que dichos sistemas son capaces de resolver. Las tareas que se pueden realizar son la anotación de las expresiones temporales, la ubicación en la línea



temporal de los eventos de un texto asociados a una fecha concreta y el establecimiento de las relaciones temporales que existen entre los eventos que no tienen una fecha concreta asociada, indicando la relación de orden entre los mismos.

El capítulo 3 se dedica a presentar la formalización del modelo temporal en el que se basará el sistema desarrollado en esta Tesis y que consta de una ontología de los conceptos que queremos tratar y las relaciones entre estos conceptos, un conjunto de reglas de identificación y resolución y su posterior formalización. La ontología propuesta se fundamenta sobre cinco clasificaciones diferentes en función de posibles aspectos temporales de las expresiones. La formalización de dicha ontología se basa en lógica de primer orden para especificar todos los conceptos y las relaciones que se desea que el sistema sea capaz de tratar.

El capítulo 4 detallará los fundamentos del sistema TERSEO propuesto en este trabajo. Este sistema está basado en Conocimiento y es capaz de reconocer y resolver expresiones temporales y en concreto, en un estado inicial de desarrollo, era capaz de trabajar únicamente en castellano. Una vez reconocidas y resueltas las expresiones, los eventos asociados con dichas expresiones son ordenados por el sistema según los resultados obtenidos, consiguiendo de esta forma una ordenación cronológica de los eventos en los textos.

En el capítulo 5 se presenta el desarrollo de la extensión automática del sistema TERSEO monolingüe, consiguiendo de esta forma que el sistema fuera capaz de reconocer y resolver expresiones temporales en cualquiera de los idiomas para los que el sistema hubiera sido previamente extendido. La extensión del sistema se realiza automáticamente utilizando un sistema de traducción automática de las reglas de identificación originales y un conjunto de textos no anotados de los que se extraen nuevas reglas de identificación para expresiones temporales en la lengua destino. Para conseguir estas nuevas expresiones se realiza una búsqueda de palabras clave temporales en la lengua destino. Una vez se ha conseguido las nuevas reglas para esta lengua, éstas serán utilizadas por TERSEO para reconocer y resolver las expresiones temporales de textos en dicha lengua.



El capítulo 6 se centra en una de las aplicaciones en las que el sistema TERSEO puede utilizarse dentro del campo del Procesamiento de Lenguaje Natural como es la Búsqueda de Respuestas. Más concretamente, la Búsqueda de Respuestas a preguntas temporales de diferentes grados de complejidad. Se presenta una arquitectura multicapa capaz de procesar estas preguntas temporales, transformándolas a preguntas que puedan ser tratadas por un sistema actual de Búsqueda de Respuestas sin implicar ningún cambio en dicho sistema. La capa de procesamiento de preguntas temporales tiene dos partes. Por un lado, un preprocesamiento de las preguntas complejas que permite dividir las preguntas simples y por otro lado, un postprocesamiento de las preguntas simples que permite formar la respuesta final de la pregunta compleja. El sistema se desarrolló en un principio para el inglés, pero adaptando únicamente el módulo de división de la pregunta pudo también utilizarse para el castellano.

El capítulo 7 muestra una amplia evaluación de cada una de las partes presentadas en los capítulos anteriores así como el razonamiento de los resultados obtenidos para cada caso. El sistema TERSEO monolingüe y multilingüe ha sido evaluado con un corpus de prueba formado por artículos digitales tanto en castellano como en inglés. Además, se incluye la participación del sistema en un foro de evaluación internacional dedicado a medir los resultados de sistemas que realizan reconocimiento y resolución de expresiones temporales en inglés. Posteriormente se realizó una evaluación del sistema multicapa de Búsqueda de Respuestas Temporal, comparando los resultados obtenidos por el mismo con respecto a un sistema de Búsqueda de Respuestas general. Finalmente, el sistema de Búsqueda de Respuestas temporal en castellano participó en la tarea piloto de la conferencia de evaluación internacional CLEF 2004. La finalidad de esta tarea era medir los resultados de los sistemas que realizaban Búsqueda de Respuestas Temporal en castellano.

En el capítulo 8 se muestran las diferentes herramientas desarrolladas para la consecución de este trabajo de Tesis. Estas herramientas son:



- Un analizador temporal que permite el análisis parcial de los textos para obtener más rápidamente sólo aquellas expresiones susceptibles de ser expresiones temporales.
- El sistema TERSEO para el reconocimiento, resolución, anotación y ordenación de expresiones temporales en textos, tanto para el castellano como para el inglés.
- El sistema de Búsqueda de Respuestas Temporal, desarrollando en concreto las unidades de Descomposición de la pregunta y de Recomposición de las respuestas y utilizando un sistema general de Búsqueda de Respuestas online.

Finalmente, en el capítulo 9 se establecen las conclusiones y los aspectos que hemos considerados importantes para el desarrollo de esta Tesis.



Universitat d'Alacant  
Universidad de Alicante



## 2. Trabajo previo

La extracción de información temporal en textos escritos no es una tarea nueva. Reichenbach (1947) ya consideró la importancia de la extracción de la información temporal que introducían los tiempos verbales en inglés. Para la realización de dicha extracción estableció tres puntos en el tiempo:

- Tiempo del Discurso o *Speech Time* (S): Momento en el que se realiza la proposición que detalla un evento.
- Tiempo del Evento o *Event Time* (E): Extensión temporal de la propia proposición.
- Tiempo de Referencia o *Reference Time* (R): Momento temporal dentro del propio evento al que se hace referencia y que viene determinado por una expresión temporal que acompaña al evento, como por ejemplo: ayer, ahora, mañana, etc.

Además, determinó que, para cada uno de los tiempos verbales existentes, estos puntos temporales se relacionan de una determinada manera:

- Presente simple  $S=E=R$
- Pasado simple  $(E=R)<S$
- Futuro simple  $S<(E=R)$
- Pretérito perfecto  $E<(S=R)$
- Pasado perfecto  $E<R<S$
- Futuro perfecto  $S<E<R$

Este trabajo estableció un punto de referencia en el área del análisis de la información temporal en los textos, surgiendo entonces el concepto de “anáfora temporal basada en los tiempos verbales”, entendiendo por anáfora el mecanismo que nos permite hacer en un discurso una referencia a alguna entidad o entidades,



con la confianza de que el receptor del discurso sea capaz de interpretar la referencia y por consiguiente determinar la entidad a la que se alude. Trabajos como Partee (1973), Chomsky (1980), Sag and Hankamer (1984), Hinrichs (1986) o Webber (1988) intentan establecer una analogía entre las anáforas estudiadas hasta el momento (pronominal, descripciones definidas,...) y la anáfora establecida por los tiempos verbales. Esta similitud se establecía porque la anáfora temporal, al igual que el resto de las anáforas, también necesita de un referente en el texto para poder ser interpretada correctamente.

Pero, no sólo los tiempos verbales tienen similitudes con la anáfora. También determinadas expresiones temporales establecen anáforas con respecto a otras expresiones temporales nombradas previamente. Por ejemplo, dado el texto: “La conferencia fue el 25 de febrero. Dos días después terminó”, la expresión temporal “dos días después” se refiere a la fecha nombrada anteriormente “25 de febrero”. Por tanto, si dichas anáforas existen, necesitan ser resueltas para poder extraer información más precisa de los textos. Para reconocer y resolver esta información anafórica es necesario establecer qué tipo de entidades van a ser reconocidas y resueltas. Esto ha llevado a muchos autores a establecer ontologías temporales basadas en diferentes aspectos de la información temporal, y que determinan las entidades o conceptos susceptibles de tratamiento.

El término ontología se usa en el ámbito de la ingeniería del conocimiento para referirse a un conjunto de conceptos organizados jerárquicamente y las posibles relaciones entre conceptos. Su utilidad es la de representar los conocimientos de forma lo suficientemente genérica y consensuada como para que la misma definición pueda ser reutilizada y compartida en distintos dominios y por diferentes aplicaciones. En este sentido, y centrándonos de nuevo en la información temporal, se pueden distinguir dos tipos de ontologías temporales:

- Ontologías temporales basadas en tiempos verbales.
- Ontologías temporales basadas en expresiones temporales.



Estas ontologías permiten definir claramente qué entidades o conceptos temporales serán posteriormente tratados por las aplicaciones y qué relaciones se establecerán entre ellos.

El establecimiento de una ontología temporal permitirá realizar las tres tareas necesarias para obtener la información temporal contenida en todo texto (Setzer & Gaizauskas, 2002):

- Anotación temporal precisa y detallada de las entidades temporales usando esquemas de anotación.
- Ubicación en la línea temporal de los eventos de un texto con expresiones temporales asociadas, asignándoles marcas de tiempo a dichas expresiones. Estas marcas de tiempo serán fechas del calendario y determinan la fecha exacta en la que ocurrió el evento.
- Establecimiento de las relaciones temporales que existen entre los eventos que no están asociados a una expresión temporal, y por tanto tampoco a una fecha concreta, ordenando dichos eventos según esas relaciones establecidas.

En la literatura existente, para abordar estas tareas existen dos tipos diferentes de enfoques o aproximaciones:

- Sistemas basados en Conocimiento: Son sistemas que utilizan una serie de reglas obtenidas a partir de un estudio del problema a resolver en los textos que van a ser tratados. Dichas reglas son almacenadas en una Base de Conocimiento.
- Sistemas basados en Corpus: Estos sistemas obtienen las reglas a partir de técnicas de aprendizaje automático mediante el uso de corpus anotados.

En las siguientes secciones se mostrará una cronología de autores y sistemas centrados en las tareas de extracción y comprensión de la información temporal que permitirá proporcionar una visión más clara de cómo ha evolucionado la investigación en este aspecto.

Además, se realizará una revisión en el capítulo 6 del estado del arte de aquellas aplicaciones que pueden nutrirse de esta información temporal resuelta con el fin de incrementar su rendimiento,



como por ejemplo la Recuperación de Información o la Búsqueda de Respuestas.

## 2.1 Ontologías temporales

Previamente a mostrar los diferentes algoritmos relacionados con la información temporal desarrollados por diversos autores, se van a presentar las diferentes ontologías temporales que permiten definir el problema que vamos a abordar de una manera clara y completa. En concreto, las ontologías temporales pueden estar basadas en la información verbal o en las expresiones temporales. A continuación se presentan diversas ontologías pertenecientes a ambos tipos.

### 2.1.1 Ontologías basadas en la información verbal

Una ontología basada en información verbal se caracteriza por clasificar los diferentes tipos de eventos en base a las propiedades del verbo que los define, como por ejemplo el aspecto o el tiempo verbal. Esta información sirve a su vez para determinar la relación de unos eventos con otros.

#### Moens y Steedman (1988)

Una de las ontologías más importantes en ese sentido, basándose en los conceptos de Reichenbach, es la que establecieron Moens y Steedman (1988), donde las entidades representadas son eventos (cosas que ocurren) y estados (condición actual de una cosa o persona). En concreto, los eventos se clasificaban en:

1. Puntuales:
  - Con Culminación: Es un evento puntual que proviene de un proceso previo ya finalizado, por ejemplo, “Juan alcanzó la meta”.
  - Sin Culminación: Es un evento puntual que no viene de un proceso previo, por ejemplo, “Juan tosió”.
2. Procesos:



- **Culminados:** indica un proceso que ha durado un determinado tiempo y ya ha finalizado, se combina con la preposición “en” a la cual se le asocia el tiempo de duración del proceso, por ejemplo: “Pedro acabó el soneto en 15 minutos”. Pretende dar importancia al resultado obtenido.
- **No culminados:** Indica un proceso que todavía continúa y se asocia con la preposición “durante” para indicar cuánto ha durado hasta el momento, por ejemplo: “Pedro ha estado escribiendo el soneto durante 15 minutos”. Enfatiza la importancia del proceso y no de su resultado.

Un estado es susceptible de cambiar pero se mantiene durante un periodo de tiempo y a menudo no tiene definido un principio y un final. Los principales tipos de estados definidos por los autores son los siguientes:

1. **Estados habituales:** Son los estados que denotan un evento puntual que se repite periódicamente en el tiempo.
2. **Estados consecuentes:** Son los estados provocados por la culminación de un evento.
3. **Estados progresivos:** Son estados que continúan en progreso durante el tiempo de referencia (R) al que nos estamos refiriendo.

Para definir su ontología los autores establecieron un grafo (figura 2.1) en el que situaron los diferentes objetos o entidades de la misma y las relaciones entre ellos. Dichas relaciones implican que un determinado objeto puede convertirse en otro aplicando algún cambio en su tiempo o aspecto verbal o mediante el uso de adverbios. A esta característica de cambio la denominaron *contingencia*.

El concepto de contingencia puede verse más claro en los siguientes ejemplos:

- Un evento puntual (con culminación o sin culminación) que se itera durante un tiempo determinado puede convertirse en un proceso (por ejemplo: “Juan tosió” es un evento puntual que puede convertirse en proceso al añadirle “Juan tosió durante 20 minutos”).

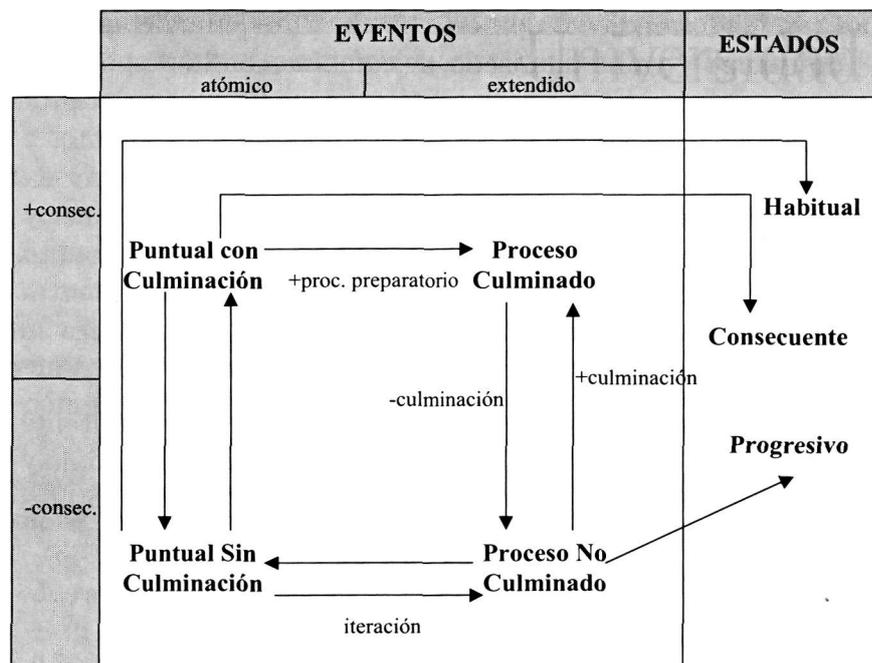


Figura 2.1. Representación gráfica de la ontología temporal en Moens *et al.* 1988

- Un proceso culminado puede pasar a ser un proceso no culminado si eliminamos la culminación del mismo (por ejemplo: “Juan acabó de escribir” se puede convertir en “Juan escribió”).

Los autores definen también el concepto de un evento como un núcleo cuya estructura abarca un proceso preparatorio, una culminación y un estado consecuente.

Cuando hablamos de relaciones temporales en esta ontología nos referimos a características como el aspecto, el tiempo verbal o la modalidad.

1. Dentro del aspecto verbal tenemos los siguientes tipos:
  - El progresivo: Es una situación que empezó en el pasado y que todavía está en proceso. Lógicamente, por tanto, un evento puntual no puede considerarse progresivo a menos que se itere en el tiempo. La entrada será un proceso y la salida un estado progresivo, por ejemplo, “Roger estuvo co-



rriendo una milla la semana pasada. Esta semana está corriendo 3”.

- El perfecto: Requiere una culminación como entrada y lo que se obtiene es un estado consecuente, es decir, la situación final es la consecuencia producida por esa culminación, por ejemplo, “Harry ha alcanzado la meta” es una culminación y su estado consecuente será que “Harry es el ganador de la carrera”.

2. Dentro del tiempo verbal, distinguimos entre:

- Presente
- Pasado
- Futuro

El tiempo verbal se considerará anafórico cuando sea necesario establecer un punto de referencia para poder interpretar el significado de dicho verbo. Los autores se basan en la idea de Reichenbach (1947), que establecía el tiempo de referencia, el tiempo del evento y el tiempo del discurso. Sin embargo, discrepan en la idea de que el tiempo de referencia es siempre un instante puntual, puesto que afirman que éste puede ser un periodo. Esto fue expuesto también por Dowty (1982) (1986), Hinrichs (1986), Partee (1973), y Moens y Steedman (1988).

3. Por último, cuando hablamos de modalidad nos estamos refiriendo a los diferentes sentidos que los verbos modales pueden dar a una determinada oración.

Steedman (1997) formalizó su ontología usando el cálculo de predicados, donde los eventos son guardados como primitivas, asociadas o no a un punto en el tiempo. Los eventos extendidos serán un par de eventos puntuales identificando su punto de inicio y su punto de final. En cuanto a las operaciones permitidas entre eventos tendremos:

- Precedencia: un evento ocurre antes que otro.
- Simultaneidad: la ocurrencia de los eventos se solapa en el tiempo.
- Contingencia: un evento de un tipo determinado puede convertirse en un evento de otro tipo aplicando algún cambio en su tiempo, aspecto verbal o mediante el uso de adverbios.



Las transiciones del grafo de la figura 2.1 son implementadas como procedimientos de inferencia.

### 2.1.2 Ontologías basadas en las expresiones temporales

Además de las ontologías que vienen establecidas por la información verbal en los textos, existen ontologías basadas en las expresiones temporales de dichos textos. En las siguientes subsecciones establecemos una cronología de las ontologías más destacadas basadas en expresiones temporales, como son Cycorp, KSL y DAML.

#### Cycorp<sup>1</sup>

Se trata de una ontología desarrollada por esta empresa privada, que contempla 6.000 conceptos de todo tipo, no exclusivamente relacionados con temporalidad.

Esta ontología está basada en lógica y su parte temporal contempla eventos y relaciones temporales entre dichos eventos. Posee un conjunto de precondiciones muy completo expresivamente hablando que son representados usando diferentes predicados.

Como resultado, la compañía Cycorp tiene desarrollado una Base de Conocimiento fundamentada en esta ontología que contiene una taxonomía de conceptos y relaciones entre dichos conceptos, así como la representación formal de las interconexiones entre los conceptos. Esta Base de Conocimiento está siendo utilizada en sistemas comerciales y académicos.

#### KSL (1991)

KSL (1991) es una ontología temporal definida por *The Stanford Knowledge Systems Lab*. En esta ontología se define una clase de *entidades temporales* y dentro de ella las subclases de *instantes* e *intervalos*. Los instantes y los intervalos se tratan como elementos primitivos en la línea temporal y sobre eso construye

<sup>1</sup> <http://www.cyc.com/> Visitada el 03/01/2005



una jerarquía de clases, relaciones, axiomas e instancias. Introduce ciertas mejoras a otras ontologías genéricas anteriores, como por ejemplo:

- Contempla intervalos abiertos y cerrados: La ontología anterior Cycorp por ejemplo, no diferencia entre intervalos abiertos o cerrados.
- Implementa la granularidad temporal para refinar más: Esto conlleva definir determinadas unidades de tiempo que permiten una mayor especificación en las medidas temporales.
- Se incluyen clases para representar los meses del año, los días del año y los días de la semana: Al definir estos elementos como clases se puede guardar mucha información en las mismas.

El lenguaje utilizado para el desarrollo de dicha ontología se denomina KIF<sup>2</sup> (*Standard Upper Ontology Knowledge Interchange Format*). KIF es un lenguaje que permite la representación de meta-información y el intercambio de conocimiento entre diferentes sistemas.

Existen esquemas de anotación temporal que se han basado en esta ontología para su desarrollo, como podrá comprobarse en posteriores subsecciones.

### DAML Time Ontology (2002)

Esta ontología (Hobbs, 2002) está desarrollada usando lógica temporal de primer orden. Tal y como ocurría en KSL, esta ontología define una clase principal denominada *entidades temporales* y dentro de ésta establece las subclases *instante* e *intervalo*, pero además, mediante el uso de predicados y funciones, añade la posibilidad de definir las relaciones entre las subclases. Para la representación de las unidades temporales usa aritmética ordinaria, pudiendo con todo esto representar gran parte de la información temporal existente. Algunos conceptos temporales, como secuencias de tiempo (cada miércoles, todos los días,...) o conceptos temporales vagos (pronto, recientemente,...) no se contemplan

<sup>2</sup> <http://logic.stanford.edu/kif/kif.html> Visitada el 03/01/2005



en esta ontología. Tampoco existen funciones para resolver muchas expresiones temporales implícitas más complejas.

## 2.2 Tareas para el tratamiento de la información temporal

Una vez definida la información temporal, mediante las ontologías anteriormente presentadas, se puede comenzar a trabajar en el tratamiento de dicha información. Actualmente, existen dos aproximaciones diferentes a la hora de tratar la problemática de anotar y resolver las entidades temporales en textos escritos: basadas en Conocimiento y basadas en Aprendizaje. A continuación se presenta cada una de ellas.

### Sistemas basados en Conocimiento

Un sistema basado en Conocimiento puede definirse como un sistema que resuelve problemas para un dominio concreto utilizando una representación simbólica del conocimiento humano. Las características principales de este tipo de sistemas son:

- Representación explícita del conocimiento
- Capacidad de razonamiento independiente de la aplicación específica
- Alto rendimiento en un dominio específico

Estas características se deben a la separación entre:

- Conocimiento específico del problema: Se trata de un conjunto de conocimientos aplicables al dominio concreto con que se está trabajando. Se pueden utilizar diferentes formalismos a la hora de representar dicho conocimiento aunque los sistemas, en el caso de la información temporal, definen el conocimiento como reglas. Es un elemento estático dentro de la arquitectura del sistema.
- Metodología para solucionar el problema: Proceso que efectúa el razonamiento a partir de los datos del problema que se intenta solucionar utilizando el conocimiento que posee. Esta meto-



dología es genérica, es decir, que se puede aplicar a diferentes dominios sólo cambiando el conocimiento que se utiliza.

El rendimiento de los sistemas basados en Conocimiento radica en la cantidad y la calidad del conocimiento de un dominio específico y no tanto de las técnicas de solución de problemas.

Por tanto, si los sistemas de tratamiento de información temporal utilizan el enfoque basado en Conocimiento, necesitarán establecer todas las reglas necesarias para poder resolver las expresiones temporales de los textos y almacenar estas reglas en una base de conocimiento.

### **Sistemas basados en Corpus**

Los sistemas basados en Corpus son un tipo de sistemas que desarrollan técnicas que aprenden a tratar o resolver una determinada tarea automáticamente. Estos sistemas suelen estar muy relacionados con métodos estadísticos ya que al igual que ellos se basan en el estudio de los datos.

Los algoritmos más comunes utilizados en este tipo de sistemas, que son aplicados también para el Procesamiento de Lenguaje Natural son:

- **Aprendizaje supervisado:** El sistema aprende de uno o varios corpus de ejemplo anotados previamente con la etiqueta correcta.
- **Aprendizaje no supervisado:** Se considera no supervisado cuando no existe tal anotación en los corpus y se utilizan otro tipo de recursos para aprender, como por ejemplo, diccionarios electrónicos, tesauros, etc.

En concreto, el Aprendizaje Automático que se usa en la mayoría de los sistemas de tratamiento de información temporal será un aprendizaje supervisado que utiliza corpus anotados manualmente y supervisados (Suarez, 2004) para entrenar al sistema y así aprender automáticamente como resolver las diferentes expresiones temporales que aparecen en los textos.



## 30 2. Trabajo previo

Una vez decidida la aproximación a utilizar, las tareas que pueden realizarse para tratar los diferentes aspectos de la información temporal son:

- Anotación de las expresiones temporales.
- Ubicación en la línea temporal de los eventos de un texto asociados a una fecha concreta.
- Establecimiento de las relaciones temporales que existen entre los eventos que no tienen una fecha concreta asociada, indicando la relación de orden entre los mismos.

A continuación se detallan cada una de estas tareas y los resultados que se pretenden conseguir con el uso de las mismas.

### 2.2.1 Anotación de expresiones temporales

Las ontologías temporales detalladas en la sección anterior se han usado como base para la creación de un lenguaje de anotación de información temporal que pueda ser usado por los usuarios, aunque estos no sean expertos en ontologías. Diferentes aplicaciones en el campo del Procesamiento del Lenguaje Natural se ven beneficiadas cuando utilizan corpus anotados temporalmente. Algunas de estas aplicaciones pueden ser:

- Extracción de Información, por ejemplo, normalizando referencias temporales para una entrada en base de datos.
- Búsqueda de Respuestas de preguntas relacionadas con temporalidad. La Búsqueda de Respuestas se vuelve más eficiente y precisa.
- Resúmenes, ordenando temporalmente la información en los mismos.
- Traducción automática, traduciendo y normalizando las referencias temporales.
- Visualización de información, para ver cronologías de eventos.

La anotación de los corpus requiere seguir una pauta de anotación que vendrá determinada por un esquema de anotación. El esquema de anotación indicará qué y cómo deben anotarse los elementos temporales, además de información adicional para cada



elemento anotado. Esta información adicional se almacenará en los atributos de las etiquetas del esquema. En un esquema de anotación temporal, la información susceptible de ser anotada es toda aquella que tenga que ver con la temporalidad, por ejemplo: expresiones temporales, fechas, eventos, etc. y que a su vez estará especificada en la ontología que el esquema de anotación use como base.

Un esquema de anotación debe contemplar las siguientes características básicas:

- Expresividad o completitud: Cantidad de propiedades que puede captar el esquema.
- Simplicidad: Esquemas fáciles de entender.
- Minimalidad: No expresar conceptos en función de otros.
- Formalidad: Cada concepto tiene una interpretación única, precisa y bien definida.

### 2.2.2 Ubicación de eventos en la línea temporal

Otra tarea a realizar con la información temporal sería la denominada asignación de marcas de tiempo a eventos (*Time-Stamping*). Esta tarea implica asignar fechas en el calendario a los eventos en los textos, gracias a las expresiones temporales que acompañan a dichos eventos. Una vez determinada la fecha concreta a la que se refiere la entidad temporal que acompaña al evento, ésta puede ser colocada en una línea temporal describiendo el desarrollo de una determinada situación. Al colocar los eventos en la línea temporal es muy posible que éstos aparezcan en un orden diferente al que aparecen en el texto. Además, al establecer esta estructura ordenada de los eventos, la comprensión del texto por parte del usuario es mucho más sencilla, puesto que intuitivamente un humano también realiza una recomposición de los diferentes eventos del discurso según un orden cronológico.

Para que esta ubicación temporal pueda realizarse es necesario resolver las expresiones temporales asociadas a los eventos, asignando así a estos eventos las marcas de tiempo que se corresponden con las fechas que denotan dichas expresiones temporales.



### 2.2.3 Establecimiento de relaciones y orden entre eventos

Las dos tareas presentadas en las dos subsecciones anteriores se encargan de marcar los eventos y las expresiones temporales en los textos con fechas en el tiempo, pero sin tener en cuenta las posibles relaciones temporales entre eventos a los que no se les puede asignar una fecha concreta en el calendario.

Definiremos relación temporal como una relación entre proposiciones que determina el orden en el tiempo de las entidades temporales que pretendemos relacionar. Por entidades temporales se entiende eventos y expresiones temporales, por tanto será posible determinar relaciones entre un evento y una expresión temporal o entre dos eventos. Las relaciones temporales vienen normalmente denotadas en los textos por preposiciones o adverbios temporales, como por ejemplo: *después, antes, durante, etc.* A partir de este momento, se denominarán señales temporales a todos aquellos elementos de una oración que denoten una relación temporal entre los elementos de la misma.

Además, la ordenación de eventos puede realizarse de dos maneras:

- Intra-oracional: Cuando el orden se establece dentro del contexto de cada una de las frases en el texto.
- Inter-oracional: Cuando el orden se establece sobre los eventos y expresiones temporales del texto completo.

Además, esta tarea de ordenación de eventos no necesita que obligatoriamente todos los eventos estén relacionados con una fecha concreta en el calendario, sino que en aquellos casos en que los eventos no tienen una marca temporal-concreta se utilizará la información de la señal temporal para establecer la ordenación adecuada.

## 2.3 Esquemas de Anotación Temporal

A continuación, se detallan tres de los esquemas de anotación más destacados y utilizados (TIDES, STAG, TIMEML), haciendo



especial hincapié en los elementos que cada uno de los esquemas es capaz de anotar así como los atributos asociados a cada uno de estos elementos. El resumen de esta información se presenta en el cuadro 2.1.

## TIDES

Durante la época de los noventa se hicieron grandes esfuerzos en la creación de corpus anotados que sirvieran de base para otras aplicaciones. Ello llevó a definir un esquema que marcara las pautas de anotación de dichos corpus. Uno de estos trabajos fue desarrollado bajo el apoyo de la organización DARPA en un programa de investigación que se denominó TIDES (Radev & Sundheim, 2002a). Este esquema ha ido evolucionando hasta nuestros días y su finalidad era su aplicación a sistemas de Búsqueda de Respuestas, Caracterización de Eventos, Resúmenes, etc. Este esquema de anotación está basado en la ontología temporal definida por KSL (1991). No es capaz de expresar todas las distinciones de esta ontología pero todo lo definido en el esquema de anotación forma parte de la ontología.

El proceso de anotación de TIDES se divide en dos pasos:

1. Marcar las expresiones temporales en un documento.
2. Identificar el valor temporal que la expresión está representando.

A lo hora de marcar las expresiones temporales se utilizan un conjunto de disparadores léxicos con sentido temporal. Toda expresión que contenga alguno de estos disparadores será considerada una expresión temporal.

El esquema de anotación TIDES únicamente marca expresiones temporales y la etiqueta que utiliza para ello se denomina *TIMEX2* (Wilson *et al.*, 2001). Además, esta etiqueta puede contener los siguientes atributos:

- **VAL**: es un atributo que se usa para representar el valor de una expresión ya sea como un punto en el tiempo o como una duración.



34 2. Trabajo previo

- **PERIODICITY:** Se utiliza para expresiones de tiempo que indiquen recurrencia regular. La recurrencia vendría expresada por palabras como “siempre”, “cada”, ...
- **MOD:** este atributo es usado para complementar a otros, cuando la expresión contiene algún modificador que puede cambiar su valor final, por ejemplo, la palabra “aproximadamente” antes de una expresión temporal.
- **SET:** es un atributo que indica si la expresión será un conjunto de valores puesto que la expresión indica una repetición en el tiempo, por ejemplo, la expresión “todos los jueves” tendría como valor del atributo SET un Sí.
- **GRANULARITY:** indica la unidad temporal en la que está representado cada miembro del conjunto
- **NONSPECIFIC:** es un valor de Sí/No y se usa para expresiones difusas y que por tanto no tienen una representación del atributo VAL concreta, por ejemplo: “a day on March”.
- **COMMENT:** Permite indicar comentarios para cada etiqueta.

La anotación se basa en dos principios básicos:

1. Si un humano puede determinar un valor para la expresión temporal, ésta debe ser marcada
2. El atributo VAL debe ser basado en la evidencia interna al conocimiento que se está anotando. Puede haber expresiones que sean anotados pero no tengan el atributo VAL porque este valor dependa del conocimiento del mundo.

Además, los valores del atributo VAL pueden ser clasificados en tres grupos básicos:

- **Puntos en el tiempo.** Todas aquellas expresiones que sean respuestas para preguntas de tipo “¿Cuándo?”
- **Duraciones.** Todas aquellas expresiones que sean respuestas para preguntas del tipo “¿Cuánto tiempo?”
- **Frecuencias.** Todas aquellas expresiones que sean respuestas para preguntas del tipo “¿Con qué frecuencia?”



## STAG

STAG (Setzer & Gaizauskas, 2000a) (Setzer & Gaizauskas, 2000b) es un esquema de anotación temporal desarrollado en la Universidad de Sheffield y que está orientado a la identificación de eventos en textos de noticias, a la relación de estos eventos con puntos en una línea temporal y además, a la anotación de señales temporales y relaciones de unos eventos con otros. La diferencia con TIDES es que TIDES únicamente anotaba expresiones temporales.

Los elementos temporales que serán anotados mediante este esquema son:

- Eventos (*EVENT*). Intuitivamente, un evento es algo que ocurre, algo que uno podría situar imaginariamente en una línea temporal. Los eventos pueden ser vistos como instantáneos o pueden ocurrir en un periodo de tiempo. Además los eventos pueden tener relaciones de correferencia entre ellos. Por eso es necesario anotar también la Identidad de Eventos. En los artículos de periódicos, muchas veces se hace referencia a eventos que fueron introducidos previamente en dicho artículo y es necesario poder identificarlos y así luego poder utilizarlos como referencia.
- Tiempos (*TIMEX*). Los tiempos se definen como expresiones temporales que acompañan a los eventos para determinar el momento en el que éstos ocurren.
- Relaciones Temporales (*SIGNAL*). Denotan posibles relaciones temporales existentes entre dos eventos o entre un evento y una expresión temporal, por ejemplo, relaciones de “antes” o “después”.

Se recomiendan 4 pasos a la hora de realizar la anotación:

1. Anotar todos los eventos y expresiones temporales sin tener en cuenta la relación entre ellos.
2. Relacionar aquellos eventos y expresiones temporales donde esté señalado explícitamente.
3. Relacionar todos los objetos temporales que no estén todavía emparejados con un evento.
4. Recorrer todos los eventos anotados y anotar la identidad de eventos.



## TIMEML

En 2002 aparece un nuevo esquema de anotación patrocinado por la institución ARDA y que se denominó TimeML (Radev & Sundheim, 2002b). TimeML es un esquema de anotación para marcar eventos, expresiones temporales y las relaciones temporales entre estos eventos que combina los dos esquemas de anotación temporales anteriores: TIDES y STAG, pero mejora estos esquemas al utilizar las mejores características de cada uno de ellos y añadiendo a su vez características nuevas, como la posibilidad de que un evento esté relacionado con más de un objeto indexado. Recordemos que en STAG sólo podía relacionarse con uno. También contempla que un evento implique dos o más acciones como veremos posteriormente. A continuación se especifican las características de este nuevo esquema.

Existen cuatro grupos principales de etiquetas:

- *EVENT*: Se denominan eventos a aquellas situaciones que ocurren.
- *TIMEX3*: Expresiones que denotan tiempo como por ejemplo: “12 de enero de 2004” o “el próximo año”.
- *SIGNAL*: Marca palabras como “en”, “durante”, etc. que denotan una relación temporal entre dos objetos.
- *LINK*: Sirve para marcar la relación entre dos objetos, ya sea una relación temporal (*TLINK*), de subordinación (*SLINK*) o aspectual (*ALINK*).

Vamos a ver a continuación, qué atributos pueden aparecer para cada una de las etiquetas anteriores. Para la etiqueta *EVENT* podemos tener los siguientes atributos:

- *ID*: es un atributo obligatorio que contiene un identificador para cada evento.
- *CLASS*: Los eventos se han clasificado en clases:
  - *REPORTING*: Eventos que son narrados por una persona u organización. Por ejemplo, eventos con el verbo “contar” o “explicar”.



- **PERCEPTION**: Eventos que conllevan la percepción física de otro evento, por ejemplo, eventos con el verbo “mirar” o “escuchar”.
- **ASPECTUAL**: Eventos que marcan el inicio de algo, el fin de algo, la culminación de algo o la continuación de algo, por ejemplo “empezar”, “proseguir”, etc.
- **LACTION**: Un evento de este tipo introduce un argumento de evento que describe una acción o situación de la que se puede inferir algo dada la relación que esta acción tiene con el evento **LACTION**. Por ejemplo, en la frase “Algunas compañías están intentado monopolizar el negocio”, “están intentando” sería un evento de tipo **LACTION** y “monopolizar” sería el argumento de evento que introduce.
- **LSTATE**: Se refiere a lo mismo que el anterior pero en lugar de eventos serían estados, usando verbos como “esperar que”, “opinar que”, etc.
- **STATE**: Describe estados.
- **OCCURRENCE**: Describe el resto de eventos que pueden ocurrir y que no se han contemplado en ninguno de los grupos anteriores.
- atributos adicionales: por ejemplo, el tiempo y el aspecto del verbo contenido en el evento.

En cuando a los atributos de la etiqueta *TIMEX3*, aunque se basa en la anotación realizada en *TIDES* para la expresiones temporales, existen algunas diferencias respecto a ésta. Por ejemplo, no se van a utilizar los atributos *SET*, *PERIODICITY*, *GRANULARITY* y *NON\_SPECIFIC*. Sin embargo, la información de estos tres primeros atributos aparece en una nueva etiqueta denominada *MAKEINSTANCE*, cuyo significado se explica posteriormente. Por tanto, los atributos de la etiqueta *TIMEX3* en este esquema serían:

- **ID**: es un atributo obligatorio para identificar la expresión temporal.
- **TYPE**: es un atributo obligatorio de indica el tipo de expresión a la que la etiqueta se refiera (fecha, hora o duración).



- **VALUE**: es equivalente al atributo VAL de *TIMEX2* y se especificará exactamente igual.
- **MOD**: es un atributo opcional con el mismo propósito que se usaba en *TIMEX2*.
- **temporalFunction**: es un atributo binario que expresa que el valor de la expresión temporal necesita ser determinado a través de la evaluación de una función.
- **anchorTimeID**: es un atributo opcional que indica el id de otra expresión temporal con la cual está relacionada temporalmente la expresión que estamos etiquetando en este momento.
- **valueFromFunction**: en caso de necesitar una función para resolver la expresión, el valor devuelto estaría en este atributo.
- **atributo functionInDocument**: este atributo indica la función de *TIMEX3* que devuelve la fecha con la que la expresión temporal está relacionada y que la necesita para poder resolverla, por ejemplo, la fecha de creación del documento, la fecha de modificación del documento, etc.

La etiqueta *SIGNAL* marca un elemento textual que denota la relación entre dos entidades temporales, ya sean éstas tiempos o eventos. Algunas de las señales que se pueden encontrar son:

- Preposiciones temporales como “en”, “antes”, “después”, ...
- Conjunciones temporales como “mientras”, “cuando”, ...
- Modificadores temporales como “cada”, “todos”, ...
- Expresiones negativas como “nunca”, ...
- Modales como “podría”, “debería”, ...
- Preposiciones que señalan modalidad como “a”.
- Caracteres especiales que pueden aparecer en fechas como “-”, “/” ...

La etiqueta *SIGNAL* únicamente tiene el atributo ID.

Como ya comentábamos en la introducción, se pueden dar casos en los que un evento denota dos o más acciones, como por ejemplo: “Juan imparte clases los lunes y los jueves”. Para estos eventos aparece una nueva etiqueta denominada *MAKEINSTANCE* que permite anotar este tipo de casos. La etiqueta *MAKEINSTANCE* aparece para cada evento al menos una vez y en caso de que un evento denote más de una acción, habrá tantas etiquetas



*MAKEINSTANCE* como acciones denote el evento. Los atributos para esta etiqueta serán:

- **ID**: un identificador de la instancia
- **eventID**: indica el identificador del evento con el que se relaciona
- **signalID**: indica el identificador de la señal que marca la repetición de la acción
- **CARDINALITY**: permite indicar la cardinalidad de la repetición.

Por último, para representar las relaciones entre entidades tenemos tres tipos de etiquetas *LINK*:

- *TLINK*: Representa una relación temporal entre dos eventos o entre un evento y un tiempo. La relación establecida indica el orden entre dichos eventos, por ejemplo, “simultáneo”, “antes”, “después”, etc.
- *SLINK*: El enlace de subordinación es usado para contextos que introducen relaciones entre dos eventos o un evento y una señal. Esta relación puede ser introducida por verbos modales, negativas, etc.
- *ALINK*: Representa la relación entre un evento aspectual y su argumento de evento. Suele expresar relaciones de “Iniciación”, “Culminación”, “Terminación” y “Continuación”.

## 2.4 Estado del arte en sistemas de Anotación Temporal

Cuando hablamos de la tarea de la anotación temporal existen dos fases a abordar:

1. Reconocimiento de las expresiones, que consiste únicamente en anotar dichas expresiones como expresiones temporales
2. Normalización del significado temporal de dichas expresiones en una etiqueta perteneciente al esquema de anotación que se esté utilizando en este momento.

Los sistemas y conferencias dedicados a abordar estas tareas se presentan ordenados cronológicamente a continuación:



### Conferencias MUC (1987-1998)

Entre 1987 y 1998, las conferencias de entendimiento de mensajes (*Message Understanding Conferences, MUC*)<sup>3</sup> han desarrollado una evaluación cuantitativa de sistemas que ahora se les llama generalmente Sistemas de Extracción de Información. Mientras que las evaluaciones del MUC típicamente definían muchas tareas de evaluación, la tarea más relevante que surgió en MUC-5 fue la del reconocimiento de entidades de nombres. Esta tarea requería el reconocimiento y la clasificación de entidades de nombres como personas, localizaciones, organizaciones, cantidades monetarias y expresiones de tiempo. El objetivo del reconocimiento de las expresiones era marcarlas usando un conjunto de etiquetas SGML y clasificarlas usando un atributo de tipo. Los tipos DATE se referían a expresiones completas o parciales de fecha y los tipos TIME se referían a expresiones completas o parciales de hora. Tanto las expresiones de tiempo absolutas como relativas tenían que ser marcadas, aunque estos dos tipos no se distinguían en su anotación.

Los participantes de MUC-5 y MUC-7 debían asignar fechas en el calendario a determinados tipos de eventos, además esta relación no era explícita, lo que obligaba a inferir información para conseguirla. En la evaluación del MUC-7, los mejores sistemas obtuvieron una F-Medida de aproximadamente el 94 % de esta tarea.

### Sistema REES (2000)<sup>4</sup>

Es un sistema enfocado a extraer eventos y las relaciones entre ellos, siendo capaz de determinar 100 tipos de relaciones y eventos. Una vez que esta información ha sido extraída, se generan unas plantillas con la misma. La arquitectura del sistema está dividida en tres componentes principales:

1. Un anotador de eventos, nombre comunes y nombres propios
2. Un módulo de resolución de correferencias basado en reglas

<sup>3</sup> [http://www.itl.nist.gov/iaui/894.02/related\\_projects/muc/](http://www.itl.nist.gov/iaui/894.02/related_projects/muc/)

<sup>4</sup> <http://www.cs.brandeis.edu/jamesp/arda/time/readings/anlp2000SRA.doc> Visitada el 03/01/2005



### 3. Un generador de plantillas.

#### Sistema TEMPEX (2002)<sup>5</sup>

Tempex (Wilson *et al.*, 2000) es un sistema que reconoce y anota expresiones temporales en textos utilizando el esquema de anotación TIDES. El sistema fue desarrollado por la corporación MITRE y está implementado en Perl. Maneja muchos formatos de fecha pero algunos elementos del esquema estándar no están completamente implementados, como por ejemplo las expresiones embebidas o las expresiones temporales con valor no identificado.

#### Conferencia TERN (2004)

Con el objetivo de evaluar diferentes sistemas que realicen dichas tareas, y patrocinado por el programa ACE (*Automatic Content Extraction Program*) surge la evaluación internacional TERN2004<sup>6</sup> (*Time Expression Recognition and Normalization*). Dicha evaluación está basada en un trabajo que comenzó en 1999 estableciendo un conjunto de guías para la anotación de información temporal en los textos y que terminó convirtiéndose en el esquema de anotación TIDES explicado en la sección 2.3. Esta evaluación pretende llevarse a cabo anualmente, mejorando así el esquema de anotación a medida que se vayan detectando posibles mejoras en el mismo.

La tarea del TERN consistía en hacer un tratamiento de las expresiones temporales realizando un reconocimiento y normalización de dichas expresiones o únicamente reconocimiento. Además, este tratamiento de las expresiones temporales podía realizarse tanto para el inglés como para el chino. A continuación detallamos los sistemas que participaron en “Sólo Reconocimiento” y los que participaron en “Reconocimiento y Normalización”, explicando brevemente los fundamentos en los que se basan cada uno de estos sistemas. Como podrá observarse, los sistemas que realizan “Sólo Reconocimiento” suelen estar basados en Aprendizaje Au-

<sup>5</sup> [http://www.cs.brandeis.edu/jamesp/arda/time/readings/GVW\\_TempExTut.ppt](http://www.cs.brandeis.edu/jamesp/arda/time/readings/GVW_TempExTut.ppt)  
Visitada el 03/01/2005

<sup>6</sup> <http://timex2.mitre.org/tern.html> Visitada el 01/08/2004



tomático mientras que los que también resuelven las expresiones temporales están en su mayoría basados en Conocimiento.

Las universidades o corporaciones actuales que realizan *únicamente anotación* de las expresiones temporales sin determinar su valor temporal son las siguientes:

### *Alias-I Inc.*<sup>7</sup>

El sistema desarrollado por Alias-I Inc. se denomina LingPipe (Carpenter, 2004) y está basado en Aprendizaje Automático. Utiliza métodos estadísticos y heurísticos para el reconocimiento de las entidades entre los que se encuentran:

- Un detector de entidades de nombre estadístico
- Un detector de sentencias heurístico
- Un motor de resolución de correferencia heurístico

El sistema está diseñado para el reconocimiento de entidades en inglés y para su extensión a otras lenguas debe ser entrenado nuevamente.

### *CLSR Universidad de Colorado*

El sistema desarrollado por esta Universidad (CLSR, 2004) utiliza un algoritmo de Aprendizaje Automático para el reconocimiento de las expresiones temporales. Se anotan las expresiones temporales a nivel de token, lo cual permite anotar también expresiones temporales embebidas. El autor usa un esquema de clasificación de izquierda a derecha, token a token, discriminativo y determinístico para determinar las etiquetas asignadas a cada token. Con el fin de anotar cada expresión se definen una serie de características para cada token, que a su vez son complementadas con un anotador basado en reglas y un anotador basado en estadísticas.

<sup>7</sup> <http://www.alias-i.com/lingpipe/> Visitada el 03/01/2005



### **IBM<sup>8</sup>**

El sistema desarrollado por IBM (Ittycheriah *et al.*, 2003) también está basado en Aprendizaje Automático. En concreto, se trata de una sistema estadístico que utiliza un clasificador basado en Máxima Entropía.

### **METACARTA Inc.**

El sistema desarrollado por METACARTA (2004) utiliza un método de Aprendizaje Automático que está basado en Bigramas HMM (*Hidden Markov's Model*). El sistema fue entrenado utilizando únicamente el corpus de entrenamiento del TERN y es capaz de reconocer expresiones temporales en inglés. Según los autores, el sistema podría ser mejorado significativamente utilizando un corpus de entrenamiento más grande.

### **Universidad de Sheffield<sup>9</sup>**

El sistema desarrollado por esta Universidad (Setzer & Gaizauskas, 2001) consiste en un sistema que utiliza el análisis morfológico y la confrontación de patrones para el reconocimiento de las expresiones temporales. El sistema es, actualmente, un prototipo y está basado en Conocimiento, siendo capaz de reconocer expresiones temporales en inglés y en chino. En concreto, el sistema tiene tres componentes principales:

- Tokenizador y analizador de parte del discurso
- Identificador de expresiones basado en patrones de reglas
- Sistema basado en Conocimiento

Por otro lado, las universidades o corporaciones que actualmente realizan *reconocimiento y normalización* de expresiones temporales se detallan a continuación:

<sup>8</sup> <http://www.watson.ibm.com/> Visitada el 03/01/2005

<sup>9</sup> <http://www.dcs.shef.ac.uk> Visitada el 03/01/2005



### ***Universidad de Amsterdam***

El sistema desarrollado por la Universidad de Amsterdam (2004) está basado en Conocimiento y utiliza un método basado en reglas para el reconocimiento y normalización de las expresiones temporales. Muchos de los problemas de este sistema se deben al hecho de que la cobertura del mismo no es muy amplia, es decir, no hay suficientes patrones para obtener buenos resultados de reconocimiento y resolución. Como trabajo futuro, pretenden incorporar un método de Aprendizaje Automático para la tarea de reconocimiento de expresiones.

### ***CLAC, Universidad Concordia***

La Universidad de Concordia (CLAC, 2004) ha desarrollado un sistema que realiza el reconocimiento y la normalización de las expresiones temporales en inglés. Para la detección de expresiones se basa en un autómata y para la resolución utiliza como posibles fechas referentes o bien la fecha de publicación del documento o la expresión temporal más recientemente reconocida en dicho texto. En este sistema, el reconocimiento de algunas expresiones no explícitas funciona incorrectamente por la necesidad de mayor información contextual, sintáctica y semántica.

### ***Cymfony Inc.***

Este sistema (Cymfony, 2004) es un sistema basado en Conocimiento que utiliza reglas gramaticales para el reconocimiento de expresiones y fórmulas temporales para la resolución de las mismas. Este sistema genera un conjunto propio de etiquetas que fueron traducidas para la participación en TERN. Información de contexto, como por ejemplo, los tiempos verbales, no se están teniendo en cuenta en este sistema.



### *ITC-irst*

El sistema desarrollado por este centro de investigación (ITC-irst, 2004) se denomina Chronos y es capaz de reconocer expresiones temporales tanto en inglés como en italiano, utilizando un sistema de Reconocimiento de Entidades de Nombre basado en reglas. Este sistema realiza primero una detección de las expresiones temporales usando un conjunto de reglas y anota dichas expresiones. Seguidamente, y haciendo uso de esa anotación intermedia, utiliza una serie de heurísticas para resolver el valor de las expresiones anotadas.

### *Lockheed Martin*

El sistema desarrollado por esta empresa (Aerotext, 2004) para el reconocimiento y resolución de expresiones temporales se denomina AeroText y es un sistema basado en Conocimiento capaz de extraer expresiones de tiempo específicas y convertir las expresiones de tiempo en un intervalo. El Motor de Inferencia que trabaja con la Base de Conocimiento utiliza un mecanismo simple de patrones que podría a su vez ser utilizado para cualquier tipo de análisis lingüístico realizando únicamente modificaciones en la Base de Conocimiento. Este sistema tampoco utiliza información del contexto para mejorar la resolución de las expresiones.

## **2.5 Estado del arte en sistemas de Ubicación de Eventos**

La tarea de la ubicación temporal de eventos se centra en reconocer los eventos en el texto y ubicar dichos eventos en la línea temporal. Esto implica, por tanto, resolver las expresiones temporales asociadas a los eventos, asignando a estos eventos marcas de tiempo que se corresponden con las fechas que denotan dichas expresiones temporales. Las aportaciones más destacadas en esta tarea se detallan a continuación.



### **Marcado de eventos en MUC-5 (1993) y MUC-7 (1998)**

Además de la tarea de reconocimiento de entidades de nombre, en MUC-5 y MUC-7 también era necesario determinar la relación existente entre las expresiones temporales y los eventos encontrados como parte de la información que se necesitaba extraer en la tarea de recuperación de información. Los participantes tenían que determinar fechas para los eventos encontrados. Los resultados obtenidos en esta tarea concreta no fueron demasiado altos debido a la dificultad que representa.

### **Filatova y Hovy (2001)**

Los autores Filatova y Hovy (2001) describen un método para dividir una noticia en frases, siendo cada frase un evento, y asignar a estos eventos marcas de tiempo (ya sean puntos en el tiempo o intervalos). Es un sistema basado en Conocimiento capaz de resolver tanto las referencias de tiempo implícitas como las explícitas. Se asume que las unidades sintácticas que denotan eventos son oraciones simples y éstas son identificadas con un analizador que produce árboles sintácticos anotados semánticamente. Algunos problemas son ignorados, como por ejemplo, las oraciones que tienen varios verbos conjugados en diferentes tiempos verbales.

El método de anotación utiliza dos puntos temporales como referencia para resolver las expresiones temporales. Por un lado, la fecha del artículo (teniendo en cuenta sólo el día, no la hora) y por otro lado, el último punto temporal asignado dentro de la misma frase. El procedimiento de anotación sería el siguiente:

1. El texto se divide en frases que denotan eventos.
2. Todas las expresiones de fecha en el texto son extraídas.
3. Se asigna una fecha a cada evento basándose bien en:
  - la expresión de fecha más reciente en la misma frase, o
  - si no está definida, la fecha del artículo
4. Si no hay ninguna expresión de fecha presente en la frase se usa la información del tiempo verbal de la oración para asignar



a ese evento un intervalo de tiempo relativo a la fecha del artículo.

Con el fin de asignar fechas a los eventos se aplican una serie de reglas de resolución. Cuando la expresión de fecha está presente en la frase, estas reglas tienen en cuenta las preposiciones que la rodean, como por ejemplo, “*después*”, “*antes*”, ... Eso permite llevar a cabo una especificación más completa.

Una vez que todos los eventos están marcados con una fecha, estos eventos son organizados en orden cronológico. Los autores indican que hay un 77.85 % de asignaciones temporales correctas a los eventos en un pequeño corpus de entrenamiento.

### Schilder y Habel (2001)

El sistema desarrollado por Schilder y Habel (2001) es un sistema también basado en Conocimiento, en el cual las expresiones temporales son clasificadas como:

- Expresiones que denotan tiempos
- Expresiones que denotan eventos

Su objetivo es situar las expresiones temporales en una línea temporal absoluta, para producir un conjunto de entidades temporales ordenadas linealmente. Para las expresiones que denotan tiempo, esto implica resolver expresiones que denomina indexadas como “ayer”, “ahora”, ... para poder especificar completamente una fecha en el calendario. Para las expresiones que denotan eventos, se asocia una fecha a un evento, extrayendo las relaciones temporales señaladas por sintagmas preposicionales. Se propone un conjunto de relaciones temporales que son: *before* (antes), *after* (después), *incl* (incluido), *at* (en), *starts* (empieza), *finishes* (termina) y *excl* (excluido), equivalentes a las relaciones propuestas por Allen (1983). Es importante destacar que las relaciones temporales en este caso no se anotan sino que simplemente se utilizan para asociar una marca temporal a un evento.

Para realizar la anotación semántica, los autores han desarrollado un sistema cuya parte principal es un Transductor de Estados Finitos (*Finite State Transducer, FST*) basado en reglas



escritas manualmente. La lengua destino es el alemán. Las etiquetas FST pueden:

- Anotar expresiones de tiempo.
- Anotar eventos que se corresponden con verbos.
- Anotar eventos que son expresiones nominales con una versión experimental de etiqueta.

Una representación semántica es entonces propuesta, basándose en las inferencias anotadas, especialmente aquellas que denotan expresiones temporales. En su estado actual, el FST establece relaciones entre eventos y tiempos. El anotador fue evaluado en un pequeño corpus de 10 artículos y la precisión obtenida fue del 84.49 %.

## 2.6 Estado del arte en sistemas de Anotación de Relaciones y Ordenación de Eventos

Una última tarea a considerar en el procesamiento de la información temporal es la obtención y anotación de las relaciones entre eventos. Los trabajos más destacados se presentan a continuación.

### **Katz y Arosio(2001)**

Uno de los sistemas más referenciados en el tratamiento de las relaciones temporales entre eventos es el de Katz y Arosio (2001). Este sistema está basado en Aprendizaje Automático y usa un gran corpus multilingüe, anotado no sólo con características morfológicas y sintácticas, sino también con relaciones temporales, y de esta forma adquiere automáticamente el conocimiento léxico que es necesario para determinar la interpretación temporal en un discurso narrado. Las relaciones temporales tratadas en este sistema son intra-oracionales, es decir las relaciones temporales internas a la oración.

El proceso que se sigue comienza por asociar un intervalo temporal con cada verbo de la frase, siendo las relaciones temporales



entre estos eventos lo que realmente se pretende establecer. La interpretación temporal suele ir muy ligada al contexto sintáctico. Esta relación es necesaria para detectar tanto a las relaciones semánticas entre las fechas como a las relaciones sintácticas entre las palabras de las frases que se refieren a esas fechas.

Los autores añaden una capa de anotación semántica a un texto ya anotado sintácticamente. La relación entre los verbos en la frase es etiquetada mediante una relación temporal. La precedencia, la inclusión y sus equivalentes son las posibles relaciones temporales que pueden aparecer. Un árbol treebank de búsqueda anotado y multilíngüe ha sido creado y cada frase es almacenada en una base de datos relacional con la información sintáctica y las anotaciones temporales.

Aunque es un trabajo muy válido, solamente tiene en cuenta las relaciones temporales entre elementos de la misma frase, lo que lo hace algo limitado.

### Setzer y Gaizauskas (2002)

Basándose en éste y utilizando también Aprendizaje Automático, Setzer y Gaizauskas (2002) plantean una representación de los eventos ordenados pero en un grafo. Esta técnica proporciona más información que si únicamente éstos se representan en la línea temporal.

En el grafo, los nodos son tiempos o eventos y los arcos son las relaciones temporales. La ventaja de utilizar el grafo es que no obliga a relacionar un evento con un tiempo concreto, sino únicamente relacionarlo con otro evento encontrado previamente, permitiendo, por tanto, establecer a su vez relaciones evento a evento.

Los elementos que se anotan son:

- Eventos, utilizando la etiqueta *EVENT*
- Tiempos, utilizando la etiqueta *TIMEX*
- Relaciones temporales. Si un evento está relacionado con otro, uno de los eventos lleva el ID del otro. Si el evento se relaciona con un tiempo, entonces el evento lleva el ID del tiempo y de



la relación temporal. Para indicar la relación entre más de dos eventos o eventos con tiempos se utiliza una etiqueta extra denominada LINK, que no anota ninguna expresión del documento sino que únicamente sirve para indicar este tipo de relaciones múltiples.

El proceso de anotación se divide en dos pasos:

1. Se anotan eventos, tiempos y señales temporales. Las relaciones temporales explícitas son establecidas y guardadas como atributos de los eventos.
2. Basándose en la información conseguida en el paso 1, todas las inferencias posibles son dibujadas de acuerdo a las relaciones *included* (incluido), *includes* (incluye), *after* (después), *before* (antes) y *simultaneous* (simultáneo).

## 2.7 Conclusiones del capítulo

A lo largo de la historia, han sido muchos los autores que han intentado tratar algún aspecto de la información temporal de los textos. Inicialmente, y como era necesario saber qué tipo de entidades se iban a reconocer, se establecieron una serie de ontologías que definían tanto las entidades susceptibles de estudio como las relaciones que podrían llegar a establecerse entre estas entidades. Estas ontologías servirían de base para poder realizar un tratamiento más sofisticado de la información temporal abordando tres tipos principales de tareas:

- Anotación temporal de textos.
- Ubicación de eventos en la línea temporal.
- Establecimiento de relaciones y orden entre eventos.

Se ha realizado una profunda revisión de los sistemas existentes que usan tanto técnicas basadas en Conocimiento como en Corpus. Un resumen de la cronología de los diferentes sistemas para cada una de estas tres tareas así como de los esquemas de anotación existentes se muestra en el cuadro 2.2.



Por tanto, y teniendo en cuenta el trabajo previo, la propuesta realizada en esta Tesis intentará resolver las tres tareas relacionadas con la extracción de información temporal, siendo para ello primero fundamental definir una ontología temporal. Para definir la ontología se usa como base la definida por Hobbs (2002), pero ha sido necesario definir una ampliación de la misma puesto que el esquema de anotación propuesto en este trabajo pretendía cubrir un rango de expresiones más amplio que lo que la ontología definida por Hobbs permitía. Una vez especificada la ontología se obtiene un esquema de anotación de las expresiones temporales contempladas en dicha ontología. Los esquemas de anotación existentes hasta el momento, presentados en este capítulo, contemplan demasiada información que no es necesaria para cubrir las necesidades del sistema aquí propuesto, por lo que se decidió establecer un esquema propio para anotar los resultados obtenidos por el sistema. No obstante, el sistema planteado será a su vez capaz de anotar los textos usando las etiquetas del esquema TIMEML (Radev & Sundheim, 2002b), como se presenta en el capítulo 7.

El sistema propuesto en este trabajo será un sistema basado en Conocimiento, pero diferenciándose de los sistemas existentes en la posibilidad de obtener además automáticamente las reglas de identificación de expresiones temporales para otras lenguas. A la hora de ubicar los eventos encontrados en los textos en la línea temporal, el sistema se basa en los trabajos desarrollados por Filatova y Hovy (2001) y Schilder y Habel (2001), pero ampliando el número de expresiones temporales que son reconocidas y resueltas por estos autores.

La ordenación de eventos realizada por el sistema presentado en esta Tesis se basa en las fechas obtenidas en la resolución de las expresiones temporales de los textos, teniendo en cuenta tanto las relaciones temporales intra-oracionales tratadas ya por Katz y Arosio (2001), como las relaciones inter-oracionales, tratadas previamente por autores como Setzer y Gaizauskas (2002), pero estableciendo el orden en una línea temporal en lugar de un grafo y consiguiendo con menos complejidad buenos resultados de ordenación.



	<b>TIDES</b>	<b>STAG</b>	<b>TIMEML</b>
<b>ET</b>	<i>Etiqueta:</i> TIMEX2	<i>Etiqueta:</i> TIME	<i>Etiqueta:</i> TIMEX3
	<i>Atributos:</i>	<i>Atributos:</i>	<i>Atributos:</i>
	VAL	CalDate	VALUE
	PERIODICITY	-	-
	MOD	-	-
	SET	-	-
	GRANULARITY	-	-
	NONSPECIFIC	-	-
	COMMENT	-	-
	-	tid	ID
	-	type	TYPE
	-	-	temporalFunction
	-	-	anchorTimeID
-	-	valueFromFunction	
-	-	functionInDocument	
<b>Eventos</b>	-	<i>Etiqueta:</i> EVENT	<i>Etiqueta:</i> EVENT
	-	<i>Atributos:</i>	<i>Atributos:</i>
	-	eid	ID
	-	class	CLASS
	-	argid	-
	-	tense	-
	-	aspect	-
	-	relatedToEvent	-
	-	eventRelType	-
	-	relatedToTime	-
	-	timeRelType	-
	-	signalID	-
	-	-	<i>Etiqueta:</i> MAKEINSTANCE
	-	-	<i>Atributos:</i>
	-	-	ID
-	-	eventID	
-	-	signalID	
-	-	Cardinality	
<b>Señales</b>	-	<i>Etiqueta:</i> SIGNAL	<i>Etiqueta:</i> SIGNAL
	-	<i>Atributos:</i>	<i>Atributos:</i>
	-	Sid	ID
<b>Relación</b>	-	-	<i>Etiqueta:</i> LINK
	-	-	<i>Atributos:</i>
	-	-	EventInstanceID
	-	-	TimeID
	-	-	SignalID
	-	-	RelatedToEvent
	-	-	RelatedToTime
	-	-	RelType
-	-	magnitude	

Cuadro 2.1. Resumen de esquemas de anotación



Año	Esquemas anotación	Anotación temporal	Ubicación eventos	Ordenación eventos
1987-1998		MUC	MUC-5 y MUC-7	
2000		REES		
2001	TIDES		Filatova y Hovy	Katz y Arosio
	STAG		Schilder y Habel	
2002	TIMEML	TEMPEX		Setzer y Gaizauskas
2004		Sistemas en TERN		

Cuadro 2.2. Cronología de las aproximaciones existentes



Universitat d'Alacant  
Universidad de Alicante



Universitat d'Alacant  
Universidad de Alicante

### 3. El modelo temporal

Los modelos, desde el punto de vista informático, proporcionan los medios necesarios para conseguir abstraer la realidad asociada a cualquier problema. Un modelo general es un conjunto de herramientas para representar un determinado tipo de información y el conjunto de operaciones que se pueden realizar con dichas herramientas. En concreto, el modelo temporal sirve para establecer una representación temporal y, por tanto, la ubicación de las expresiones temporales dentro de esta representación. El modelo temporal estará formado por un lenguaje (conjunto de símbolos y sintaxis) más una interpretación para ese lenguaje (semántica) que permite la identificación, clasificación y resolución de expresiones temporales en cualquier lengua.

El modelo temporal propuesto en esta Tesis y sobre el que se basa el sistema posteriormente desarrollado consta de dos partes principales:

- Un ontología temporal, en la que se indican los conceptos temporales relevantes en el dominio en el que nos encontramos (artículos de periódicos) así como las relaciones existentes entre dichos conceptos.
- Un conjunto de reglas de dos tipos: reglas de identificación o reconocimiento que permiten detectar una expresión temporal dentro de una lengua (dependientes de la lengua) y definen la parte sintáctica del modelo y reglas de interpretación o resolución que permiten calcular el valor absoluto asociado a la expresión temporal dentro del modelo y definen, por tanto, la parte semántica del modelo (independientes de la lengua).

Además, es necesaria la representación formal de este modelo temporal para que pueda ser utilizado como base de un sistema.



En nuestro caso se ha utilizado como base la ontología DAML (Hobbs, 2002) cuyo objetivo fue desarrollar una ontología temporal representativa que expresara conceptos temporales simples y propiedades comunes a cualquier formalización de tiempo.

Para poder obtener tanto la sintaxis como la semántica de los diferentes conceptos representados en el modelo temporal propuesto en este trabajo, ha sido necesario realizar un estudio de todas las expresiones temporales que podían encontrarse en un corpus no anotado de periódicos digitales.

A continuación se muestra tanto la definición de la ontología temporal que se propone como el conjunto de reglas de identificación y resolución asociadas a dicha ontología.

### 3.1 Ontología Temporal

La ontología temporal propuesta en este trabajo contempla todos los conceptos relevantes a tener en cuenta en el dominio en el que nos encontramos (artículos de periódicos digitales) y sus relaciones entre sí. Para establecer la ontología se define una taxonomía de expresiones temporales que contemple todos los posibles aspectos temporales de una determinada expresión.

#### 3.1.1 Conceptos temporales

Todos los conceptos temporales de la ontología propuesta se contemplan en una taxonomía temporal que se basa en cinco clasificaciones posibles para las expresiones temporales atendiendo a cinco propiedades básicas que puede contemplar cada una de ellas. Las cinco clasificaciones propuestas se ha realizado según:

- El tipo de resolución
- La fecha de referencia
- La unidad temporal
- El carácter temporal
- El valor temporal



Por tanto, toda expresión temporal tendrá cinco aspectos temporales que la ubican dentro de la ontología propuesta. Cada aspecto temporal se corresponde con un elemento de cada una de las clasificaciones propuestas que se encuentran detalladas en la subsecciones siguientes. Es necesario determinar los aspectos temporales de una determinada expresión, no sólo para reconocerla sino también para saber como resolverla.

### **Taxonomía de expresiones según el tipo de resolución**

Se proponen dos tipos de expresiones temporales básicas cuando hacemos referencia a la forma en que dichas expresiones son resueltas:

- **Expresiones Temporales Explícitas:** Expresiones que directamente denotan una fecha o periodo de fechas y no necesitan ser resueltas usando una regla de resolución. Estas expresiones pueden ser de varios tipos:
  - *Fechas completas con o sin expresión de hora:* “11/01/2002”, “el 4 de enero de 2002”, ...
  - *Expresiones referidas a eventos:* “el curso 2002-2003”, “Navidad”, “Semana Santa”...
- **Expresiones Temporales Implícitas:** Expresiones que necesitan ser resueltas para poder inferir la fecha o periodo de fechas al que se refieren. Esto implica tener una regla de resolución que permita resolver la expresión. Se consideran expresiones temporales implícitas, por ejemplo: “ayer”, “la semana pasada”, ...

### **Taxonomía de expresiones según la fecha de referencia**

En el proceso de resolución, una determinada expresión temporal implícita puede hacer referencia a dos tipos de fecha:

- **Expresiones que se refieren a la fecha del documento:** Expresiones que necesitan la fecha del documento para poder ser interpretadas y además siempre es así, independientemente del contexto en el que se encuentre la expresión. Para obtener la fecha explícita asociada a una expresión como “ayer” o “el



próximo mes” es necesario conocer la fecha en la que fueron usadas, es decir, la fecha en la que se escribió el documento.

- **Expresiones que se refieren a una fecha nombrada previamente:** Expresiones que, para ser resueltas, necesitan una fecha nombrada anteriormente en el texto como referencia. Estas expresiones sí serán dependientes del contexto en el que se encuentran, puesto que si las cambiáramos de contexto también cambiaría su interpretación. Dentro de este grupo tenemos expresiones como: “un mes después”, “un año antes”, ...

### Taxonomía de expresiones según la unidad temporal

Las expresiones temporales deben ser clasificadas en función de la unidad temporal a la que hacen referencia puesto que la regla de resolución necesita conocer dicha unidad temporal para poder transformar esa expresión a una fecha concreta o periodo de fechas. Las unidades temporales contempladas en esta taxonomía son las siguientes:

- **DIA:** Expresiones como “el día siguiente”, “ayer”, “mañana”
- **MES:** Expresiones como “el mes pasado”
- **AÑO:** Expresiones como “el año que viene”
- **SEMANA:** Expresiones como “la semana pasada”
- **FINSEMANA:** Expresiones como “este fin de semana”
- **DIASEMANA:** Expresiones como “el lunes pasado”
- **EVENTO:** Expresiones como “la próxima Navidad”

Las operaciones necesarias para resolver una expresión que se refiere a la unidad temporal DIA serán diferentes a las necesarias para resolver una expresión temporal que se refiere a la unidad temporal SEMANA y es por esto que se necesita ubicar cada expresión temporal susceptible de ser resuelta en uno de estos grupos.

### Taxonomía de expresiones según el carácter temporal

Se define carácter temporal como el momento en el tiempo al que la expresión hace referencia respecto al momento presen-



te. Los caracteres temporales básicos que se contemplan en esta taxonomía son:

- **PASADO:** Expresiones que indican un momento pasado con respecto a la fecha de referencia, por ejemplo: “la semana pasada”
- **PRESENTE:** Expresiones que se ubican en el mismo momento que la fecha de referencia, por ejemplo: “este año”
- **FUTURO:** Expresiones que se ubican en un momento posterior al de la fecha de referencia, por ejemplo: “la próxima semana”

Además, estos caracteres básicos pueden combinarse para obtener otros caracteres más complejos.

- **PASADO/PRESENTE:** Expresiones que denotan un periodo que comienza en un momento anterior a la fecha de referencia pero que continúan en el momento actual, por ejemplo: “desde hace cinco días”
- **PRESENTE/FUTURO:** Expresiones que denotan un periodo de tiempo que comienza en el momento actual y termina en un momento futuro, por ejemplo: “durante los próximos cinco meses”

### Taxonomía de expresiones según el valor temporal

Por último, también será necesario conocer qué tipo de resultado obtenemos una vez que la expresión temporal ha sido resuelta. Esto es lo que denominamos representación del valor temporal de la expresión y puede dividirse en dos subclasificaciones principales:

- Según el grado de definición de la expresión:
  - **Expresión concreta:** Expresiones cuyo valor temporal se puede especificar de manera concreta con una fecha o periodo de fechas, por ejemplo: “ayer”, “hace dos meses”
  - **Expresión difusa:** Expresiones cuyo valor temporal sólo se puede aproximar puesto que la expresión no determina una fecha o periodo concreto de fechas, por ejemplo: “hace unos



días”. Esta expresión no indica cuantos días y por tanto no se puede dar un valor exacto de resolución.

- Según el grado de amplitud de la expresión:
  - **Expresión simple:** Expresiones que se refieren a un único valor de fecha, ya sea concreto o difuso, por ejemplo: “mañana”, “hace unos días”
  - **Expresión periodo:** Expresiones que se refieren a un intervalo de fechas ya sea concreto o difuso, por ejemplo: “los dos años siguientes”, “durante semanas”.

Tanto el grado de definición como el grado de amplitud se combina para conseguir definir el tipo de valor temporal que tendrá una determinada expresión, por ejemplo, la expresión “ayer” será clasificada como “concreta-simple”, y así sucesivamente.

### 3.1.2 Relaciones entre conceptos

Existen dos relaciones entre los conceptos temporales definidos previamente:

- **Relación es un:** es la relación que definen las diferentes taxonomías, es decir, “ayer” es un **DIA** o “hace unos días” es una **Expresión Difusa**. La relación **es un** define las dependencias verticales entre los conceptos de la jerarquía ontológica.
- **Relación de precedencia:** es la relación que define un orden temporal entre los diferentes conceptos de la ontología y que posteriormente se utilizará para realizar la ordenación de eventos, por ejemplo: “ayer” precede a “hoy” y a su vez “hoy” precede a “mañana”. La relación de **precedencia** define las dependencias horizontales entre los conceptos de la ontología.

## 3.2 Reglas del modelo temporal

Una vez definida la ontología temporal, es decir, los conceptos temporales y las relaciones que se pueden establecer entre dichos conceptos, es necesario establecer el conjunto de reglas asociadas a dicha ontología.



Estas reglas se clasifican en dos tipos: Reglas de identificación o reconocimiento y reglas de interpretación o resolución.

### **Reglas de identificación**

Engloba el conjunto de reglas a través de las cuales se identifican las expresiones temporales ubicadas dentro de los conceptos temporales definidos en la ontología. Estas reglas son dependientes del idioma y por tanto será necesario formalizar un conjunto de reglas de identificación para cada lengua. La formalización de estas reglas para el modelo aquí propuesto se realizará usando gramáticas temporales parciales (Ver anexo B).

### **Reglas de resolución**

Estas reglas permiten resolver o interpretar los conceptos de la ontología, lo que implica la resolución de las expresiones temporales identificadas. La formalización de estas reglas se ha realizado tomando como base la lógica de primer orden. La lógica de primer orden es un sistema formal que permite razonar sobre un universo de discurso o dominio. Para poder hacer esta tarea, la lógica proporciona dos instrumentos: un lenguaje con el que poder hacer aserciones o afirmaciones de interés y unas reglas con las cuales determinar el valor de verdad de las afirmaciones o aserciones realizadas. En concreto, el lenguaje aplicado a las expresiones temporales permitirá realizar afirmaciones sobre el universo de dichas expresiones temporales y las reglas determinarán el valor de las afirmaciones o aserciones temporales realizadas. El hecho de que la lógica de primer orden sea un sistema formal permite que el modelo temporal creado a partir de ella sea perfectamente formal y preciso. Por tanto, en el trabajo que se presenta en esta Tesis se define un conjunto de reglas de resolución para los diferentes tipos de expresiones temporales de la ontología presentada en la sección 3.1 usando como base lógica de primer orden y las funciones primitivas definidas por Hobbs (Cuadro 3.1). Además, la resolución de todas las expresiones está basada, o bien en la fecha



## 62 3. El modelo temporal

del documento, o bien en una fecha nombrada previamente. Para ello se definen dos nuevas primitivas “*documentDate*” y “*previousDate*” respectivamente. Por último, necesitamos una función que nos devuelva el número de semana dentro del año para una fecha concreta que se pasa como argumento (“*weekOfYear*”) (Cuadro 3.2).

$interval(x)$	Indica que $x$ es un intervalo.
$day(x)$	$x$ es un día del calendario.
$start-of(x)$	Indica el día de inicio del intervalo $x$ .
$end-of(x)$	Indica el día de fin del intervalo $x$ .
$Hath(S, n, u, x)$	Un conjunto $S$ de $n$ intervalos de calendario de tipo $u$ (*Day*, *Month*, ...) forman el intervalo de calendario $x$ .
$interval-between(z, x, y)$	$z$ es un intervalo que comienza en el instante $x$ y acaba en el instante $y$ .
$in-interval(y, x)$	El instante $y$ pertenece al intervalo $x$ .
$duration(z, u) = n$	El intervalo $z$ está formado por $n$ unidades temporales de tipo $u$ .
$before(x, y)$	La entidad temporal $x$ ocurre antes que la entidad temporal $y$ .
$after(x, y)$	La entidad temporal $x$ ocurre después de la entidad temporal $y$ .
$Monday(x, y), Tuesday(x, y), \dots, Sunday(x, y)$	Indica que $x$ es el día de la semana (Lunes, martes, ...) de la semana $y$ .

**Cuadro 3.1.** Funciones primitivas definidas por Hobbs y usadas en el modelo temporal propuesto

$documentDate(T)$	Fecha del documento
$previousDate(T)$	Fecha nombrada anteriormente
$weekOfYear(T)$	Número de semana dentro del año para una fecha

**Cuadro 3.2.** Funciones primitivas nuevas

Una vez establecidas las funciones primitivas en las que se basa el modelo temporal, se procede a definir las reglas necesarias para resolver todo tipo de expresión temporal. En el Cuadro 3.3

se muestra un ejemplo de las reglas de resolución para aquellas expresiones cuya unidad de tiempo es el DIA y que serán interpretadas como un fecha concreta en el tiempo. El conjunto de todas las reglas de resolución definidas pueden ser consultadas en el anexo A.

$yesterday(x) \equiv$	$\exists y, z [documentDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = 1 \wedge day(x)]$
$daysAgo(x) \equiv$	$\exists y, z [documentDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = n \wedge day(x)]$
$dayBefore(x) \equiv$	$\exists y, z [previousDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = 1 \wedge day(x)]$
$daysBefore(x) \equiv$	$\exists y, z [previousDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = n \wedge day(x)]$
$today(x) \equiv$	$documentDate(x)$
$tomorrow(x) \equiv$	$\exists y, z [documentDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = 1 \wedge day(x)]$
$daysWithin(x) \equiv$	$\exists y, z [documentDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = n \wedge day(x)]$
$dayLater(x) \equiv$	$\exists y, z [previousDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = 1 \wedge day(x)]$
$daysLater(x) \equiv$	$\exists y, z [previousDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = n \wedge day(x)]$

**Cuadro 3.3.** Reglas resolución DIA CONCRETO

### 3.3 Conclusiones del capítulo

En este capítulo se ha definido un modelo temporal que consta de una ontología temporal y un conjunto de reglas asociadas a los conceptos representados en dicha ontología.

La ontología definida consta de un conjunto de conceptos temporales representados bajo cinco diferentes taxonomías interrelacionadas entre ellas y que permiten expresar todos los posibles



64 3. El modelo temporal

aspectos temporales de una determinada expresión. Una vez que la expresión ha sido ubicada en la ontología, la regla de resolución de la expresión es directa pues depende de dichos aspectos temporales. Se definen a su vez dos relaciones básicas que se pueden establecer entre los conceptos representados en la ontología, estas relaciones son: **es un** y **precedencia**. Con todo esto, la ontología temporal permite establecer el ámbito de identificación de las expresiones que el sistema propuesto en esta Tesis será capaz de tratar.

Por último, se definen un conjunto de reglas formales asociadas a la ontología definida previamente y cuya formalización se basa en lógica de primer orden pudiendo de esta forma ser usadas por el sistema de reconocimiento y resolución de expresiones temporales.



Universitat d'Alacant  
Universidad de Alicante

## 4. TERSEO Monolingüe: Sistema de Resolución de Expresiones Temporales aplicado a la Ordenación de Eventos

El modelo temporal definido en el capítulo 3 constituye la base para el trabajo que se desarrolla en esta Tesis. Dicho modelo representa el conjunto de objetos y las interrelaciones entre ellos permitiendo la construcción de los diferentes conjuntos de expresiones temporales en cualquier lengua, además de un criterio que permitirá realizar ordenaciones cronológicas de eventos.

La definición del modelo permite a su vez la especificación de un esquema de anotación de expresiones temporales en el que se contempla tanto la anotación de los textos con la interpretación temporal de las expresiones reconocidas, como también la información cronológica de dichas fechas en el texto. Por tanto, para obtener esta información cronológica será necesaria una estrategia de resolución de expresiones temporales implícitas, convirtiéndolas en explícitas. Finalmente, mediante la incorporación de técnicas de ordenación de eventos se facilita la creación de un sistema capaz de comprender la información temporal y las interrelaciones cronológicas existentes en cualquier documento de lenguaje natural. Una aproximación empírica al tratamiento de las expresiones temporales fue expuesta por Pustejovsky *et al.* (2002) en el dominio de los diálogos de planificación, en el que las expresiones temporales son limitadas. Sin embargo, el dominio de los artículos de periódico, sobre el que se trabaja en esta Tesis, es mucho más rico en este tipo de expresiones.

Con esta base se ha construido el sistema TERSEO, cuyas siglas significan en inglés *Temporal Expression Recognition System applied to Event Ordering* (Sistema de Reconocimiento de Expresiones Temporales aplicado a la Ordenación de Eventos) (Saquete *et al.*, 2003b), que es capaz de reconocer y resolver las expresio-



nes temporales que aparecen en documentos en castellano y posteriormente realizar una ordenación de los eventos relacionados con dichas expresiones temporales. La descripción y el proceso de construcción de este sistema se presenta en las siguientes secciones.

#### 4.1 Esquema de Anotación de las expresiones temporales

Los esquemas de anotación definen una serie de pautas para anotar en textos los elementos definidos en la ontología en la que se basan así como las etiquetas en las que se almacena la información que se va a extraer para cada uno de esos elementos. Actualmente existen diferentes esquemas de anotación, tal y como se describía en el capítulo 2, como por ejemplo, TIDES (Radev & Sundheim, 2002a), STAG (Setzer & Gaizauskas, 2000a) (Setzer & Gaizauskas, 2000b) y TIMEMJL (Radev & Sundheim, 2002b), los cuales han surgido en los últimos años gracias a que algunas instituciones de investigación han comenzado a trabajar en diferentes aspectos de la información temporal. Sin embargo, estudiando las necesidades del sistema aquí presentado, los esquemas de anotación existentes hasta el momento son demasiado extensos y contienen gran cantidad de información que no se necesita para el objetivo deseado, por lo que se decidió establecer un conjunto de etiquetas XML<sup>1</sup> propias con el objetivo de anotar los resultados del reconocimiento y resolución de las expresiones temporales de forma sencilla e intuitiva (Saquete *et al.*, 2002b).

En este sentido, se ha pretendido adoptar la filosofía de estandarización de un sistema de anotación defendida en Martínez-Barco (2001) y Palomar y Martínez-Barco (2001) para permitir la integración de dicha anotación en cualquier proyecto de PLN, independientemente de su implementación. Así, el uso de XML para el etiquetado de textos permite que los usuarios definan su propio modelo de documento a través de un fichero llamado DTD (*Document Type Definition*). Este DTD, construido según una

<sup>1</sup> <http://www.xml.org>



sintaxis concreta, permite que desde el programa informático se pueda interpretar el tipo de documento y la organización de las diferentes partes del mismo que se encuentran relacionadas entre sí. El documento etiquetado en XML y el DTD que lo interpreta, estandarizan el etiquetado permitiendo que cualquier herramienta posterior analice fácilmente el documento sin crear dependencias entre el etiquetador y la herramienta. En nuestro esquema de anotación usaremos el lenguaje XML como herramienta de etiquetado aprovechando la flexibilidad que posee.

Las etiquetas que conforman el esquema de anotación siguen la siguiente estructura:

- Para expresiones temporales explícitas:

```
<DATE_TIME ID='valor'
  TYPE='valor'
  VALDATE1='valor'
  VALTIME1='valor'
  VALDATE2='valor'
  VALTIME2='valor'
  VALORDER='valor'>
  expresión</DATE_TIME>
```

- Para expresiones temporales implícitas:

```
<DATE_TIME_REF ID='valor'
  TYPE='valor'
  VALDATE1='valor'
  VALTIME1='valor'
  VALDATE2='valor'
  VALTIME2='valor'
  VALORDER='valor'>
  expresión</DATE_TIME_REF>
```

DATE\_TIME es el nombre de la etiqueta para ETs explícitas y DATE\_TIME\_REF es el nombre de la etiqueta para ETs implícitas. Cada expresión tiene un ID numérico para identificarla y VALDATE# y VALTIME# almacenarán el rango de fechas y horas respectivamente obtenidos por el sistema, donde VALDATE2 y VALTIME2 son atributos únicamente usados para expresiones



temporales de tipo rango. Además, VALTIME1 podría omitirse si la expresión no contiene información sobre la hora. Por tanto, VALDATE2, VALTIME1 y VALTIME2 son argumentos opcionales. VALORDER es el atributo que utiliza la unidad de ordenación de eventos para indicar el orden de un evento. Anotando el texto de esta forma obtenemos un documento estructurado del que podemos extraer la información temporal de una manera eficiente.

## 4.2 Descripción del sistema

Tanto el modelo temporal como el esquema de anotación definido se usan como base para el desarrollo del sistema TERSEO puesto que en ellos se indica qué expresiones van a ser reconocidas por el sistema (ontología), cómo deben ser anotadas (esquema de anotación) y cuales son las reglas de resolución que se necesitan aplicar para cada una de las expresiones (reglas de resolución formalizadas en base a lógica de primer orden).

El conjunto de todas las expresiones temporales susceptibles de ser reconocidas así como las reglas de resolución de cada tipo de expresión conforman el modelo temporal (Anexos A y B) que el sistema TERSEO utilizará como base para la resolución de dicha tarea. En la formalización del modelo temporal las reglas aparecen expresadas en base a lógica de primer orden, lo que favorece que dichas reglas puedan ser aplicadas por cualquier tipo de sistema al tratarse de un lenguaje universal.

El sistema TERSEO realiza las siguientes tareas:

- Reconocimiento de las expresiones temporales
- Resolución de las expresiones temporales
- Anotación de las expresiones temporales en el texto
- Ordenación de los eventos del texto

En la Figura 4.1 se muestra la representación gráfica del sistema. Los textos son anotados con información léxica y morfológica por medio de un POS TAGGER (Carmona *et al.*, 1998) y esta información es la entrada de un analizador temporal. Este analizador está implementado utilizando una técnica ascendente (analizador chart) y está basado en una gramática temporal propia

definida dentro del modelo temporal. Una vez que el analizador reconoce las expresiones temporales (ET) en el texto, éstas son introducidas en la unidad de resolución, la cual actualiza el valor de la misma con una fecha o periodo concreto de acuerdo a la fecha de referencia de dicha expresión y con esta información se genera una etiqueta XML con la cual se anota la expresión en el texto. Finalmente, usando estas etiquetas como la entrada de la unidad de ordenación de eventos, el texto ordenado será devuelto como salida del sistema.

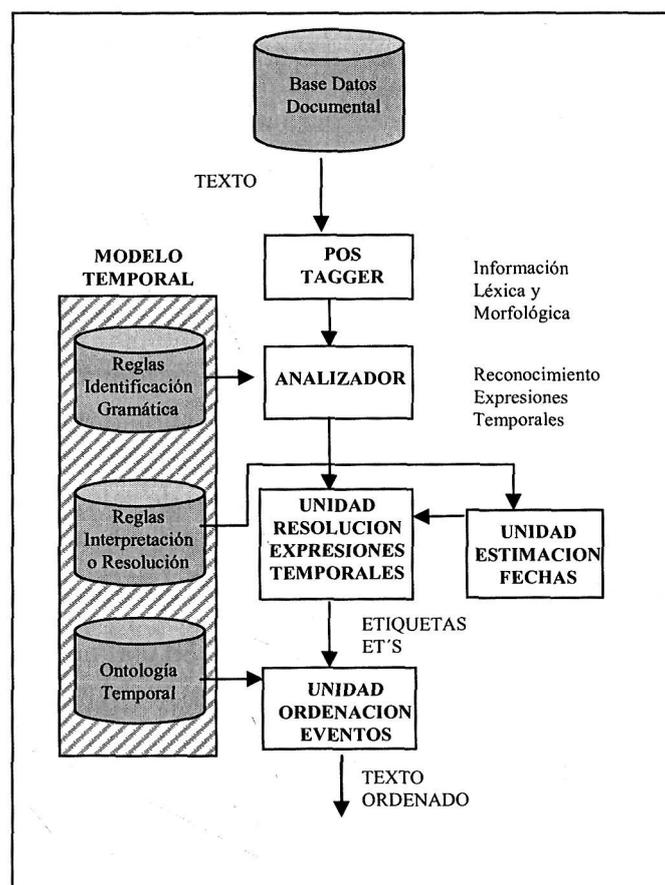


Figura 4.1. Representación gráfica de TERSEO



### 4.2.1 POS TAGGER

La herramienta que hemos utilizado como POS TAGGER para obtener la información léxica y morfológica del texto se denomina MACO (Carmona *et al.*, 1998). MACO es un analizador de la lengua española desarrollado por el Laboratori de Recerca en Lingüística Computacional de la Universitat de Barcelona y por el grupo de Lenguaje Natural de la Universitat Politècnica de Catalunya. El sistema está organizado internamente en un diccionario de raíces, un diccionario de sufijos y un conjunto de reglas de combinatoria de los mismos. Cada una de las raíces y sufijos tiene asociado un modelo de flexión. Las reglas indican que combinaciones de modelos son las correctas.

MACO puede utilizarse tanto para analizar textos como para generar formas a partir de un lema o de una raíz. El sistema trata diversos tipos de unidades:

- Formas simples: gato, solía, cantaré, ...
- Formas complejas: mientras que, en tanto que, verlo, ...
- Nombres propios
- Fechas

La cobertura actual del sistema es aproximadamente de un millón de formas, que se obtienen a partir de 100.000 raíces que corresponden a unos 93.000 lemas. El número de raíces es superior al de lemas ya que algunos lemas tienen más de una raíz, por ejemplo, el lema volver tiene como posibles raíces: volv- vuelv- vuelt-.

Los lemas corresponden a 12.000 verbos, 80.000 nombres y adjetivos y 1.000 preposiciones, adverbios, conjunciones y determinantes.

Un posible ejemplo de las etiquetas devueltas por MACO al analizar una expresión temporal sería:

```

el el DAOMSO
próximo próximo AQOMSO
mes mes NCMS000

```



### 4.2.2 Analizador Temporal

El analizador temporal se basa en las reglas de identificación del modelo temporal definido, utilizando una gramática temporal parcial para el análisis (Saquete *et al.*, 2002a). Se usa como elementos terminales de la gramática las etiquetas devueltas por MACO. Dicha gramática temporal únicamente tendrá reglas para reconocer expresiones temporales explícitas e implícitas que se encuentren dentro del marco del modelo temporal del capítulo 3 en el que se basa el sistema. La gramática define dos tipos diferentes de reglas para reconocer las expresiones temporales. Por un lado, se definen reglas para el reconocimiento de fechas y horas concretas como, por ejemplo, "12/06/1975". Estas expresiones se corresponden con las denominadas Expresiones Temporales Explícitas de nuestra primera taxonomía de expresiones y se denominan explícitas puesto que no hay nada que resolver para ellas. Por otro lado, se definen reglas para Expresiones Temporales Implícitas, es decir, expresiones temporales que hacen referencia a:

- La fecha del documento en el que aparecen, o,
- Una fecha nombrada anteriormente en el texto.

Algunos ejemplos de los tipos de expresiones que el sistema es capaz de reconocer se muestran en el Cuadro 4.1. En este Cuadro se indica también el tipo de valor temporal que devuelven las expresiones<sup>2</sup>.

Por tanto, utilizando esta gramática temporal como recurso y los textos anotados léxica y morfológicamente como entrada, el analizador temporal realiza un análisis sintáctico parcial ascendente que permite reconocer las posibles expresiones temporales en los textos.

El analizador desarrollado en este trabajo para la implementación de esta parte del sistema, al tratarse de un analizador parcial diseñado específicamente para el reconocimiento de expresiones temporales permite discriminar otro tipo de información en los

<sup>2</sup> La nomenclatura usada para las reglas arriba mencionadas es la misma que se usa en los diccionarios de datos de los sistemas de información. Si las palabras que aparecen entre paréntesis significa que son opcionales.



<b>Expresiones Explícitas</b>	fecha→ dd+ '/' +mm+ '/' +(yy)yy fecha→ dd+ 'de'+mm+ 'de'+(yy)yy fecha→ ('El')+díasemana+dd + 'de'+mes+ 'de'+(yy)yy
<b>Expresiones Implícitas Fecha del Documento Concretas</b>	referencia→ 'mañana' referencia→ 'ayer' referencia→ 'el próximo día'
<b>Expresiones Implícitas Fecha Previa Periodo</b>	referencia→ 'un mes después' referencia→ num+ 'años después'
<b>Exp. Implícitas Fecha Prev. Concretas</b>	referencia→ 'un día antes'
<b>Exp. Implícitas Fecha Previa Difusa</b>	referencia→ 'días después' referencia→ 'días antes'

**Cuadro 4.1.** Ejemplo de reglas de identificación de expresiones explícitas e implícitas

textos que no es necesaria para dicho propósito, quedándose únicamente con aquellas expresiones susceptibles de ser expresiones temporales. Este proceso es mucho más eficiente que si se hubiera utilizado un analizador parcial de propósito general, puesto que dicho analizador reconocería más elementos en los textos de los que realmente se necesitan.

#### 4.2.3 Unidad de Resolución de Expresiones Temporales

Una vez que las expresiones temporales son identificadas, éstas deben ser resueltas. Para la resolución de expresiones implícitas utilizamos el modelo temporal para interpretar cada referencia nombrada. Para ello, usa la ontología definida en el modelo temporal de la sección 3, que permite conocer sobre qué fecha de las nombradas anteriormente en el texto se va a aplicar la regla de resolución de la expresión. En algunos casos las referencias son estimadas según la fecha del documento en el que aparecen (FechaP) y en otros casos, se refieren a una fecha nombrada previamente en el texto que está siendo analizado (FechaA) como ya se comentaba en la sección anterior. La unidad de Resolución de Expresiones Temporales se basa en las dos reglas definidas a continuación, que sólo serán aplicables para aquellas expresiones que se refieren a una fecha anterior, puesto que las que se refieren a la

fecha del documento se resuelven directamente usando la fórmula adecuada.

1. Por defecto, la fecha del documento se usa como referente base si existe, si no, usaremos la fecha del sistema.
2. Si encontramos una ET explícita, ésta es almacenada como FechaA. Este valor será actualizado cada vez que encontremos una ET explícita en el texto.

En el Cuadro 4.2 se muestran algunas de las reglas de resolución del modelo temporal usadas por la unidad de Resolución de Expresiones Temporales. La unidad que se encarga de estimar las fechas accederá a la regla de resolución adecuada para cada expresión y aplicará dicha regla para obtener una fecha explícita en formato estandarizado. Por tanto, en este punto, la expresión implícita habrá sido resuelta y será anotada usando la etiqueta correspondiente del esquema de anotación definido.

REFERENCIA	ENTRADA DICCIONARIO
'ayer'	Dia(FechaP)-1/Mes(FechaP)/Año(FechaP)
'mañana'	Dia(FechaP)+1/Mes(FechaP)/Año(FechaP)
'anteayer'	Dia(FechaP)-2/Mes(FechaP)/Año(FechaP)
'el próximo día'	Dia(FechaP)+1/Mes(FechaP)/Año(FechaP)
'un día antes'	Dia(FechaA)-1/Mes(FechaA)/Año(FechaA)
'un mes después'	[DiaI/Mes(FechaA)+1/Año(FechaA) -- DiaF/Mes(FechaA)+1/Año(FechaA)]
num+'años después'	[01/01/Año(FechaA)+num -- 31/12/Año(FechaA)+num]
'días después'	>>>>FechaA
'días antes'	<<<<FechaA

Cuadro 4.2. Ejemplo de algunas reglas de resolución del modelo temporal

#### 4.2.4 Unidad de Ordenación de eventos

Finalmente, una vez que las expresiones han sido identificadas, resueltas y anotadas, la ordenación de los eventos en el texto puede llevarse a cabo, ubicando cada expresión en su lugar correspon-



diente de la ontología temporal. El sistema desarrollado considera como eventos todas aquellas frases del texto que contienen algún tipo de expresión temporal, ya sea explícita o implícita (Saquete *et al.*, 2003a) (Saquete *et al.*, 2004f). De esta forma, la identificación de eventos queda reducida a la identificación de ETs. La ordenación de eventos en textos escritos en lenguaje natural no es una tarea trivial. En el sistema propuesto la ordenación de eventos se determinará en función de las fechas estimadas para las expresiones temporales que son parte de dicho evento.

Una vez que el texto ha sido anotado usando el esquema de anotación definido en la sección 4.1, el módulo de ordenación de eventos construye una tabla con toda la información de las etiquetas de ese texto. Esta tabla incluye las columnas ID, VALDATE1, VALTIME1, VALDATE2, VALTIME2 y VALORDER (que está vacía en este momento). Una vez que la tabla ha sido construida, el módulo la ordena en orden ascendente de acuerdo a las siguientes reglas:

1. ET1 es previa a ET2 si el rango de VALDATE1, VALTIME1, VALDATE2 y VALTIME2 asociado con ET1 es previo y no solapa al rango asociado con ET2.
2. ET1 es concurrente a ET2 si el rango de VALDATE1, VALTIME1, VALDATE2 y VALTIME2 asociado con ET1 solapa al rango asociado con ET2.

Después de esto el sistema asigna un número de orden secuencial para cada ET en la tabla, teniendo el mismo número de orden aquellas expresiones concurrentes. Posteriormente, el texto es anotado nuevamente incluyendo el valor del orden en el atributo VALORDER para cada etiqueta del texto.

A continuación se muestra un ejemplo. Considerando como fecha del documento 30/12/2002, el texto quedaría anotado de la siguiente forma:

```
<DATE_TIME VALDATE1="13/11/2002" VALORDER = "2"> [El miércoles
13 de noviembre de 2002] </DATE_TIME>, el barco Prestige lanza un SOS
a 28 millas de Finisterre debido a una brecha. El barco había sido parchea-
do <DATE_TIME_REF VALDATE1 = "<<<<13/11/2002" VALORDER =
```



“1”> [unos meses antes] </DATE\_TIME\_REF> en un astillero chino. Al conocerse la noticia, la Agencia Marítima y de Guardacostas del Reino Unido ofrecen su ayuda al estado español para controlar la contaminación, pero no será aceptada hasta <DATE\_TIME VALDATE1 = “22/11/2002” VALORDER = “5”> [el 22 de noviembre] </DATE\_TIME>. <DATE\_TIME VALDATE1 = “15/11/2002” VALORDER=“3”> [El 15 de noviembre] </DATE\_TIME> comienzan a acercarse manchas y <DATE\_TIME VALDATE1 = “16/11/2002” VALORDER = “4”>[el 16 de noviembre] </DATE\_TIME> 190 kilómetros de costa se ven anegados por la marea negra. <DATE\_TIME\_REF VALDATE1 = “22/11/2002” VALORDER = “5”> [Seis días después] </DATE\_TIME\_REF> , Manuel Fraga visita por primera vez la zona afectada. <DATE\_TIME VALDATE1 = “01/12/2002” VALORDER=“7”> [El 1 de diciembre] </DATE\_TIME>, el bastiscafo francés Nautilus llega a la costa gallega. <DATE\_TIME\_REF VALDATE1 = “<<<<01/12/2002” VALORDER=“6”> [Previamente] </DATE\_TIME\_REF>, en el Prestige ya se habían producido nuevas grietas.

De esta forma sería posible colocar los eventos de este texto en la línea temporal. Definimos una variable para cada evento reconocido en el texto por el sistema:

- EV1: El barco Prestige lanza un SOS a 28 millas de Finisterre debido a una brecha
- EV2: El barco ha sido parcheado en un astillero chino
- EV3: La ayuda del Reino Unido es aceptada
- EV4: Comienzan a acercarse manchas
- EV5: 190 kilómetros de costa se ven anegados por la marea negra
- EV6: Manuel Fraga visita por primera vez la zona afectada
- EV7: El bastiscafo francés Nautilus llega a la costa gallega
- EV8: En el Prestige ya se habían producido nuevas grietas

La colocación de estos eventos ordenados en la línea temporal se muestra en la Figura 4.2.

### 4.3 Conclusiones del capítulo

Después de establecer una ontología de las expresiones temporales que aparecen en textos escritos y determinar un conjunto

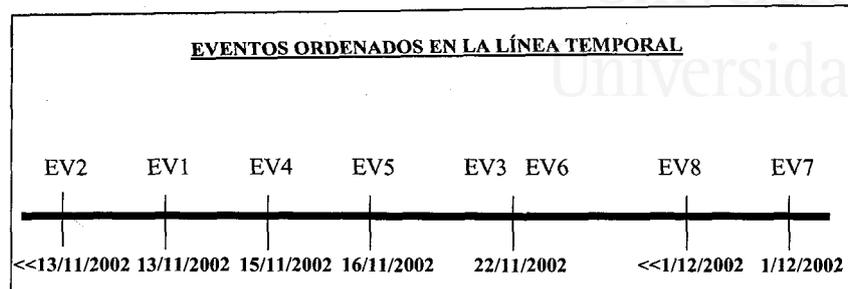


Figura 4.2. Ordenación de los eventos en la línea temporal

de etiquetas que conforman un esquema de anotación para dicha información, se establece cuáles serán las diferentes tareas que un sistema debería realizar para poder reconocer, resolver y ordenar esta información. Estas tareas son:

- Identificación de las expresiones temporales
- Resolución de las expresiones temporales
- Anotación de las expresiones temporales con sus resoluciones
- Ordenación de las expresiones temporales con sus eventos relacionados

El sistema TERSEO presentado en este capítulo es capaz de realizar todas estas tareas utilizando el modelo temporal definido y las reglas de identificación de las expresiones temporales para el castellano como recurso. Una vez completado todo el proceso, el resultado será un conjunto de eventos ordenados cronológicamente que pueden ser utilizados para diferentes aplicaciones, como por ejemplo, la realización de resúmenes. Además, aunque el sistema presentado en este capítulo es monolingüe, se extendió automáticamente, como se indica a continuación, a otra lengua, únicamente construyendo un conjunto de reglas de identificación para dicha nueva lengua. Además, este conjunto de reglas de identificación para la nueva lengua va a construirse de manera automática como se explica en detalle en el siguiente capítulo.



Universitat d'Alacant  
Universidad de Alicante

## 5. Extensión del sistema: Adquisición automática de conocimiento para otras lenguas

Tal y como se ha presentado en el capítulo anterior, TERSEO es un sistema de Reconocimiento y Resolución de expresiones temporales monolingüe, en concreto, en castellano. Sin embargo, la tendencia actual es la generación de sistemas que sean capaces de trabajar independientemente del idioma, evitando de esta forma la barrera lingüística en la utilización del sistema. En este sentido, surgen iniciativas como el Cross Language Evaluation Forum<sup>1</sup> (CLEF), que es una actividad de ámbito europeo nacida en el 2000 y celebrada anualmente, y cuyo objetivo principal es la especialización en diferentes tareas de recuperación de información multilingüe. Además autores como Mória (2001) trabajan en sistemas de reconocimiento y resolución de expresiones temporales en dos idiomas, en este caso portugués e inglés. Puesto que el trabajo aquí presentado está enmarcado en el proyecto de investigación del Ministerio de Ciencia y Tecnología “Recuperación de respuestas en documentos digitalizados”<sup>2</sup> y siendo uno de los objetivos del proyecto el desarrollo de sistemas de acceso a la información (Sistemas de Búsqueda de Respuestas y Recuperación de Documentos) a través del estudio por un lado, de los aspectos multilingües e interactivos y, por otro, de la robustez de los sistemas en documentos escritos y transcritos, se hace necesario que el sistema TERSEO sea un sistema multilingüe, es decir, que sea capaz de reconocer y resolver expresiones temporales en otras lenguas, además del castellano. Aun así, y debido que nuestro sistema está basado en Conocimiento, sería muy costoso adquirir conocimiento sobre cada lengua manualmente, puesto que además

<sup>1</sup> <http://www.clef-campaign.org/> Visitada el 03/01/2005

<sup>2</sup> <http://gplsi.dlsi.ua.es/r2d2/>

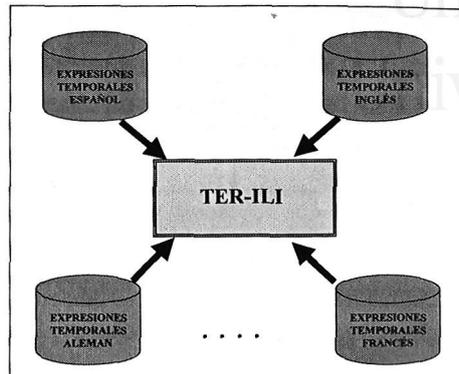


implicaría conocer la lengua y sus características para determinar qué es y qué no es una expresión temporal en la misma. Es por ello, que se desarrolla un mecanismo que nos permite extraer este conocimiento automáticamente para nuevas lenguas basándonos en la información que poseen las reglas de identificación de la lengua original, que es el castellano en este caso<sup>3</sup>. En las siguientes secciones podemos ver qué pasos son necesarios para adquirir este conocimiento de la nueva lengua automáticamente.

### 5.1 Multilingüidad automática

La adquisición automática de reglas permite crear un sistema multilingüe sin la necesidad de tener un conocimiento previo de dichas lenguas. Con este propósito, se ha planteado una arquitectura muy similar a la usada por EuroWordNet (Vossen, 2000), de forma que cada lengua tenga sus propias reglas de identificación pero todas ellas conectadas por una unidad que se va a denominar TER-ILI o *Temporal Expression Rules Interlingua Index*, tal y como se muestra en la Figura 5.1 (Saquete *et al.*, 2004a). Esta unidad contendrá tanto la ontología temporal como las reglas de resolución genéricas con las que estarán relacionadas las diferentes reglas de identificación que contemplan las expresiones en las diferentes lenguas. De esta forma el sistema TERSEO es capaz de resolver cualquier expresión temporal en cualquier lengua si se le indica las reglas de identificación que debe utilizar. Además, esta adquisición automática del conocimiento permite que el sistema TERSEO sea a su vez fácilmente extensible a otras lenguas. En principio, partimos de las reglas de identificación en castellano construidas manualmente. Estas reglas de identificación contemplan todas las expresiones temporales en castellano y cada una de estas expresiones está relacionada con una regla de resolución de la unidad TER-ILI, siendo el objetivo final tener todas las expresiones temporales en las diferentes lenguas (alemán, inglés, francés, ...) unidas a las reglas de resolución de la unidad TER-ILI.

<sup>3</sup> Consultar anexo C para más información



**Figura 5.1.** Representación gráfica de la unidad TER-ILI

Con este tipo de arquitectura se consigue que las reglas de resolución sean independientes de la lengua, por tanto, las diferentes expresiones temporales en las distintas lenguas estarán relacionadas con el mismo conjunto de reglas de resolución almacenadas en la unidad TER-ILI.

En la Figura 5.2 podemos ver la representación gráfica de la extensión del sistema TERSEO para la adquisición automática de reglas.

La extensión del sistema consta de seis unidades principales:

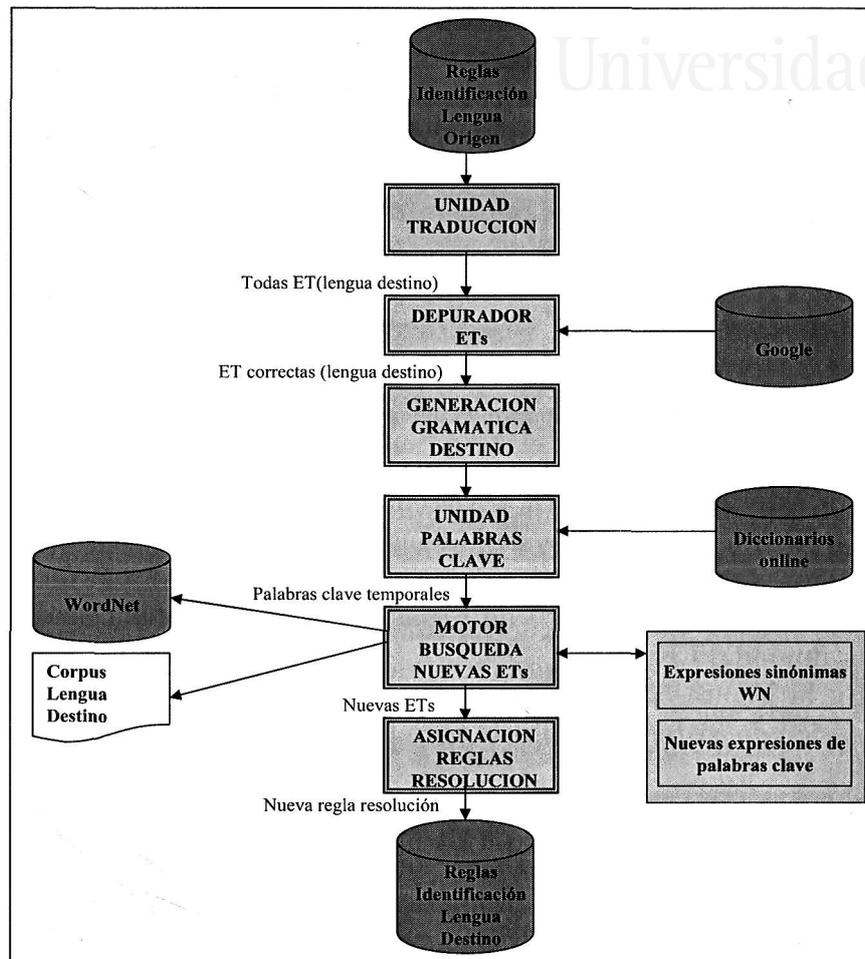
- Unidad de Traducción de Expresiones Temporales: Esta unidad, usando tres traductores automáticos (BabelFish<sup>4</sup>, FreeTranslator<sup>5</sup> y Power Translator<sup>6</sup>), realiza la traducción de todas las expresiones contempladas por el sistema para el castellano.
- Unidad de Depuración de Expresiones Temporales: Las expresiones traducidas serán la entrada de esta unidad, que utiliza Google<sup>7</sup> como recurso para determinar que expresiones son erróneas y eliminarlas.
- Unidad de Generación Automática de la Gramática en el idioma destino: Esta unidad, utilizando la información morfológica y sintáctica de las nuevas expresiones será capaz de generar las

<sup>4</sup> <http://world.altavista.com/>

<sup>5</sup> <http://www.free-translator.com/translator3.html>

<sup>6</sup> <http://www.outsource-sl.com/fabricantes/lh/pt8.htm>

<sup>7</sup> <http://www.google.com/>



**Figura 5.2.** Representación gráfica de la extensión automática multilingüe de TERSEO

reglas de la gramática que permita reconocer las expresiones temporales traducidas a la lengua destino.

- Unidad de Palabras Clave: Esta unidad obtiene un conjunto de palabras clave en la lengua destino.
- Motor de Búsqueda de Nuevas Expresiones Temporales: Esta unidad es capaz de aprender nuevas expresiones temporales automáticamente usando dos técnicas diferentes:
  - usando EuroWordNet para obtener expresiones sinónimas

- usando las palabras claves obtenidas en el paso previo y un corpus no anotado en la lengua destino.
- Unidad de Asignación de Reglas de Resolución: Esta unidad infiere la regla de resolución para cada nueva expresión obtenida en el paso anterior.

Una vez que el conjunto de expresiones temporales en la lengua destino se ha completado, el sistema TERSEO será capaz de trabajar con esta nueva lengua, identificando y resolviendo dichas expresiones temporales.

### 5.1.1 Unidad de Traducción

La unidad de Traducción se encarga de realizar una traducción directa de las expresiones que se contemplan en el sistema TERSEO (castellano) a la lengua destino a la que se desea extender el sistema (inglés, italiano, francés, catalán, etc.). Para esta traducción se han usado tres sistemas de traducción automática (BabelFish, FreeTranslator, PowerTranslator). El uso de tres traductores está basado en una conocida técnica de multilingüidad aplicada en una gran cantidad de sistemas multilingües (Llopis & Muñoz, 2004). Esta técnica implementa un sistema de votación entre las tres posibles traducciones y elige la expresión más votada, obteniéndose de esta forma la mejor traducción para cada expresión. Además, en esta unidad, cada expresión en la lengua destino se relaciona con la misma regla de resolución que tenía la expresión en la lengua origen.

Por ejemplo, la expresión en castellano “un día después” se relaciona con la Regla de Resolución Interlingua:

$DIA, FUTURO, CONCRETO^8 \equiv Day(F)+1/Month(F)/Year(F)^9$

para la lengua origen de TERSEO. Si esta expresión fuera la entrada de la unidad de traducción, los resultados de esta unidad serían los que se muestran en la Tabla 5.1.

Aplicando la técnica del voto, la expresión elegida sería “*a day later*”, y esta expresión será introducida en el conjunto de

<sup>8</sup> Todas las reglas de resolución se encuentran en el anexo B

<sup>9</sup> siendo F una fecha nombrada anteriormente en el texto



Translator	Result Expression
<i>BabelFish</i>	A day later
<i>FreeTranslator</i>	A day later
<i>PowerTranslator</i>	A day after

Cuadro 5.1. Ejemplo de traducciones para los tres traductores

expresiones de la lengua destino y relacionada con la misma regla de resolución que tiene “un día después” en la unidad TER-ILI.

### 5.1.2 Unidad de Depuración de Expresiones Temporales

Las traducciones obtenidas por la unidad de Traducción deberán ser depuradas para poder eliminar aquellas que sean erróneas. Esta tarea es la que realiza la unidad de Depuración y se lleva a cabo una vez que todas las nuevas expresiones en la lengua origen han sido traducidas.

La unidad de Depuración de Expresiones Temporales usa Google como recurso para eliminar aquellas expresiones que no se han traducido correctamente. Todas las expresiones traducidas son buscadas exactas en Google, de forma que la expresión será considerada incorrecta cuando Google no devuelve ningún resultado para dicha búsqueda. Los resultados, en caso de haberlos, no son analizados, únicamente se tiene en cuenta si ha habido alguna coincidencia en la búsqueda o no.

Por ejemplo, la expresión en castellano: “lo que va de mes” ha sido traducida por BabelFish como “*what goes of month*”. Sin embargo, la unidad de depuración de expresiones temporales ha buscado esta expresión usando Google y no ha encontrado ningún resultado, de forma que, esta expresión será considerada incorrecta y por tanto eliminada del conjunto de expresiones temporales de la lengua destino.

### 5.1.3 Unidad de Generación Automática de la Gramática Destino

Esta unidad, utilizando, por un lado el POS TAGGER MACO (Carmona *et al.*, 1998) que también puede ser usado para el



inglés únicamente cambiando los parámetros de entrada, y por otro, las expresiones temporales traducidas, es capaz de extraer la información morfológica y sintáctica de dichas expresiones en la lengua destino y generar un conjunto de reglas para establecer una gramática parcial de expresiones temporales en esta nueva lengua. Esta gramática parcial es utilizada por el sistema TERSEO para discriminar aquella información que no es necesaria para conseguir el objetivo final del sistema, quedándose únicamente con aquellas expresiones susceptibles de ser expresiones temporales.

#### 5.1.4 Unidad de Palabras Clave

El tercer paso es conseguir un conjunto de palabras clave en la lengua destino a partir un conjunto de palabras clave definidas previamente en la lengua origen con el objetivo de utilizar estas palabras clave para determinar nuevas expresiones temporales en un corpus de textos no anotados. Si para conseguir el conjunto de expresiones temporales en la lengua destino utilizáramos únicamente la traducción de dichas expresiones en la lengua origen, es muy posible que los resultados obtenidos por TERSEO en la tarea de reconocimiento y resolución de expresiones temporales no fueran muy altos, por el hecho de que determinadas expresiones no pueden obtenerse a partir de una traducción directa de otra lengua. Es por ello, que si se dispone de un conjunto de palabras clave temporales, se pueden realizar búsquedas en textos reales en la lengua destino, obteniendo de esta forma una mayor cantidad de expresiones que luego podrán ser reconocidas y resueltas.

Para llevar a cabo esta tarea se ha creado manualmente un repositorio de palabras clave en la lengua origen. Cada palabra clave tiene asociada un conjunto de características temporales que se corresponden con los elementos de la taxonomía definida en el capítulo 3. Estas características serán usadas posteriormente para conseguir determinar automáticamente la regla de resolución InterLingua de una nueva expresión temporal obtenida por el Motor de Búsqueda de Nuevas Expresiones. Estas palabras clave en castellano son traducidas a la lengua destino usando un diccio-



nario online<sup>10</sup> como recurso. Se ha usado un diccionario en este caso puesto que la traducción a realizar será únicamente de una palabra, por lo cual es más directo usar un diccionario que un sistema de traducción automática que requiere de un proceso más complejo para conseguir la traducción final.

Algunas de las palabras clave y sus características temporales son mostradas en la Tabla 5.2.

Lengua origen	Lengua destino	Características temporales
siguiente	following	Futuro Fecha Anterior
día	day	DIA Concreto

**Cuadro 5.2.** Ejemplo de algunas palabras clave temporales

Además, las palabras clave se han dividido en dos grandes grupos en función de su semántica temporal. Estos grupos son:

- Palabras clave de alta temporalidad: en este conjunto introduciremos las palabras clave que siempre tienen un sentido temporal, como por ejemplo: “día”, “enero”, “lunes”, etc.
- Palabras clave de baja temporalidad: introduciremos en este conjunto aquellas palabras que son susceptibles de formar parte de una expresión temporal pero que no siempre es así. Por ejemplo, estarían dentro de este grupo palabras como “próximo”, “anterior”, etc.

Esta división es necesaria puesto que, en el momento de realizar la búsqueda de nuevas expresiones temporales en corpus no anotados, se utilizarán las palabras clave de alta temporalidad, acotando de esta forma el espacio de búsqueda. Una vez encontrada una palabra clave de alta temporalidad se utilizan las palabras clave de baja temporalidad para obtener expresiones temporales completas en la lengua destino.

Todas las palabras clave que actualmente utiliza nuestro sistema, clasificadas en función de su grado de temporalidad, aparecen en el anexo C.

<sup>10</sup> <http://www.diccionarios.com>



### 5.1.5 Motor de búsqueda de Nuevas Expresiones Temporales

La búsqueda de nuevas expresiones se realiza de dos modos diferentes:

- Búsqueda de sinónimos de expresiones a través de EuroWordnet. Este submódulo obtiene nuevas expresiones temporales buscando expresiones sinónimas en el recurso léxico EuroWordNet. Por ejemplo, si el módulo busca la expresión: “mañana”, EuroWordnet devuelve las expresiones “*tomorrow*”, “*the day after today*”,... La expresión “*tomorrow*” ya está introducida en el conjunto de expresiones temporales, sin embargo “*the day after today*” no se había obtenido por medio de la traducción directa. Esta nueva expresión es, por tanto, introducida en dicho conjunto para la lengua destino y se une a la misma regla de resolución InterLingua que la expresión “mañana”.
- Búsqueda de nuevas expresiones a través de las palabras clave. Este submódulo usa las palabras clave obtenidas previamente para conseguir nuevas expresiones temporales. Para la búsqueda de nuevas expresiones se utiliza un corpus no anotado escrito en la lengua destino. Cuando en el texto se encuentra una palabra clave de alta temporalidad, la búsqueda se centra en esta palabra, usándola como referencia y el motor busca más palabras clave de alta o baja temporalidad hacia delante y hacia detrás de dicha palabra referencia, intentando así formar una nueva expresión temporal. Por ejemplo, para la frase: “Ellos tuvieron muchos regalos en Navidad y parecían felices”, el módulo encuentra la palabra clave de alta temporalidad “Navidad” y al buscar más palabras clave alrededor de ésta encuentra otra palabra de baja temporalidad “en”. Al no encontrar más palabras temporales, la búsqueda terminará devolviendo la expresión “en Navidad”. Esta expresión será introducida en el conjunto de expresiones para la lengua destino, pero no será relacionada con ninguna regla de resolución por el momento. Por supuesto, las nuevas expresiones obtenidas son también depuradas para eliminar posibles errores, utilizando Google de nuevo, de forma



que queden únicamente aquellas expresiones para las que Google devuelve alguna coincidencia.

### 5.1.6 Unidad de Asignación de Reglas de Resolución

La unidad de Asignación de Reglas de Resolución se encarga de asignar una regla de resolución para cada nueva expresión almacenada en el conjunto de expresiones temporales basándose en las características temporales de la palabra o palabras que forman la nueva expresión. Una vez que esta unidad encuentra la regla general de la unidad TER-ILI que permite resolver la expresión, la relaciona con ella. Una vez finalizado el proceso, toda expresión temporal en la lengua destino estará relacionada con alguna regla de resolución de la unidad TER-ILI.

En la práctica, si tenemos la expresión *“the following day”*, la unidad de Asignación de Reglas de Resolución buscará las características temporales de todas las palabras clave que forman la expresión. Estas características se corresponden con las propiedades temporales de las expresiones, que nos permitían clasificarlas en cinco diferentes taxonomías, formando así la ontología de todas las expresiones temporales (ver Capítulo 3). Por ejemplo, *“following”* se refiere a una expresión de futuro y siempre se encuentra en expresiones que hacen referencia a una fecha nombrada antes en el texto, y *“day”* se refiere a expresiones cuya unidad temporal es el día y su resolución, al ser singular, será concreta. Por tanto, la unidad de Asignación de Reglas de Resolución combinará todas estas características y buscará en la unidad TER-ILI qué regla de resolución se corresponde con dichas características. En este caso concreto, la regla que resuelve la expresión será una regla que expresa un día concreto en el futuro y referida a una fecha nombrada antes en el texto. La regla de Resolución Interlingua encontrada por la unidad de Asignación de Reglas será:

$$\text{DIA, FUTURO, CONCRETO}^{11} \equiv \text{Day}(F)+1/\text{Month}(F)/\text{Year}(F)^{12}$$

Una vez encontrada la regla, la unidad de Asignación de Reglas de Resolución relacionará la nueva expresión con esta regla de re-

<sup>11</sup> Más información en el anexo B

<sup>12</sup> siendo F una fecha nombrada antes en el texto



solución de la unidad TER-ILI. Finalmente, una vez terminado el proceso de la asignación de la regla, el sistema tendrá un conjunto de expresiones almacenadas para el lenguaje origen (castellano en este caso) y cada una de estas expresiones estará relacionada con una regla de resolución de la unidad TER-ILI. A su vez, el sistema tendrá el conjunto de expresiones equivalentes en la lengua destino y cada una de estas expresiones también relacionadas con sus correspondientes reglas de resolución de la unidad TER-ILI.

En el caso concreto de la regla de resolución DIA, FUTURO, CONCRETO que se especificaba previamente, el conjunto de expresiones temporales equivalentes para la lengua origen castellano y la lengua destino inglés serán:

- **Conjunto origen:** mañana, el próximo día, dentro de un día, el día próximo, al día siguiente
- **Conjunto destino:** tomorrow, the day after today, the next day, within a day, inside a day, the following day, on the following day, a day later

## 5.2 Conclusiones del capítulo

Debido a que el sistema inicial estaba basado en Conocimiento y era monolingüe (castellano), era necesaria la extensión del sistema para que este pudiera trabajar en un nivel multilingüe. Por ello, se desarrolla dicha extensión del sistema original que permite construir nuevos conjuntos de expresiones temporales para otras lenguas utilizando para ello el conocimiento previo de la lengua original del sistema. De esta forma, no será necesario conocer la lengua a la que queremos extender el sistema propuesto, pues las nuevas expresiones temporales y sus reglas de resolución serán obtenidas automáticamente. Para la obtención de dichas reglas se realizan dos procesos principales:

- Traducción automática de las expresiones temporales de la lengua origen
- Obtención de nuevas expresiones en la lengua destino usando un corpus no anotado en dicha lengua y un conjunto de palabras clave temporales



88 5. Extensión del sistema

Una vez obtenido el conjunto de expresiones temporales para la nueva lengua, las reglas de identificación y la relación entre estas y las reglas de resolución, el sistema TERSEO será capaz de reconocer y resolver expresiones temporales encontradas en textos escritos en dicha lengua.



## 6. Aplicaciones: Búsqueda de Respuestas Temporal

Los sistemas de Búsqueda de Respuestas son sistemas capaces de contestar a preguntas formuladas por usuarios, ya sean precisas o arbitrarias. Estos sistemas son especialmente útiles para obtener un determinado dato sin la necesidad de ir buscando manualmente por toda la información relacionada con el tema buscado.

Actualmente, la investigación en Búsqueda de Respuestas está centrada principalmente en el tratamiento de las preguntas factuales (Vicedo *et al.*, 2003), las cuales requieren como respuesta una información muy específica como, por ejemplo, fechas, nombres de entidades, cantidades, etc. Un ejemplo de este tipo de preguntas sería: “¿Cual es la capital de Brasil?”.

El análisis de algunas de las aproximaciones actuales relevantes (Moldovan *et al.*, 2002; Soubbotin & Soubbotin, 2002; Magnini *et al.*, 2002; Yang & Chua, 2002) permite identificar los componentes principales de un sistema de BR:

1. Análisis de la pregunta
2. Selección de documentos o pasajes
3. Extracción de respuestas

Estos componentes se relacionan entre sí procesando preguntas y documentos en diferentes niveles hasta obtener la respuesta.

En concreto en este trabajo se estudia la Búsqueda de Respuestas Temporal. La BR Temporal es una especialización de la Búsqueda de Respuestas dedicada exclusivamente a la resolución de preguntas temporales. Una pregunta será considerada temporal si cumple alguna de estas dos condiciones:

- La pregunta requiere de una fecha como respuesta.
- La pregunta contiene una fecha o expresión temporal en su formulación.



En este sentido, la Búsqueda de Respuestas Temporal no es un tarea trivial debido a la complejidad que las preguntas temporales pueden llegar a tener. En la actualidad, los sistemas de Búsqueda de Respuestas son capaces de resolver preguntas temporales muy simples, es decir, preguntas que no necesitan una interpretación previa de dichas preguntas o posteriormente de sus respuestas. Entre ellas nos encontramos preguntas simples que requieren únicamente una fecha como respuesta, por ejemplo: “¿Cuándo murió Bob Marley?” o preguntas simples que contienen una fecha o expresión temporal explícita, es decir, la expresión temporal no necesita ser resuelta sino que el sistema de Búsqueda de Respuestas la utiliza sin interpretarla para poder contestar a dicha pregunta, como por ejemplo: “¿Quién ganó el Open de USA en 1999?”.

El procesamiento de este tipo de preguntas simples suele llevarse a cabo identificando las expresiones temporales explícitas que contiene la pregunta y los documentos relevantes para poder detectar la expresión temporal que se necesita para contestarlas. Sería necesario destacar el sistema descrito en Breck *et al.* (2000) como el único que también utiliza el reconocimiento de expresiones temporales implícitas aplicándolo a la Búsqueda de Respuestas. Para ello utiliza el anotador temporal desarrollado por Mani y Wilson *et al.* (2000).

Esta metodología es útil para resolver las preguntas temporales simples, sin embargo, existen una serie de preguntas de mayor complejidad que actualmente no están siendo abordadas por los sistemas de Búsqueda de Respuestas:

- Preguntas muy complejas que necesitan obtener información de diferentes documentos y ordenar posteriormente esta información en el tiempo, como por ejemplo: ¿Quién fue el portavoz de la Embajada Soviética en Bagdad **durante** la invasión de Kuwait?”
- Preguntas que contienen expresiones temporales implícitas en su formulación, y por tanto, dichas expresiones deben ser interpretadas para que el sistema de Búsqueda de Respuestas sea

capaz de responder a la pregunta, por ejemplo: “¿Es Bill Clinton **actualmente** el presidente de los Estados Unidos?”

La aproximación que se presenta en este trabajo intenta imitar el comportamiento humano a la hora de resolver este tipo de preguntas complejas. Por ejemplo, si un humano necesita respuesta a una pregunta como: “¿Quién fue el portavoz de la Embajada Soviética en Bagdad durante la invasión de Kuwait?”, seguiría el siguiente proceso intuitivo:

1. Primero, descompondría la pregunta compleja en preguntas más sencillas y por tanto más fáciles de contestar: “¿Quién fue el portavoz de la Embajada Soviética en Bagdad?” y “¿Cuándo ocurrió la invasión de Kuwait?”.
2. El humano buscaría todas las posibles respuestas a la primera pregunta simple: “¿Quién fue el portavoz de la Embajada Soviética en Bagdad?”.
3. Después, buscaría todas las posibles respuestas de la segunda pregunta simple: “¿Cuándo ocurrió la invasión de Kuwait?”.
4. Finalmente, daría como respuesta final a la pregunta compleja una de las respuestas de la primera pregunta (si hay alguna) cuya fecha asociada cumpla las restricciones impuestas por las fechas devueltas como respuesta para la segunda pregunta simple. Se entiende por restricciones la ordenación implícita impuesta por el nexo de unión de las dos preguntas simples en la pregunta original. En este ejemplo concreto, el nexo de unión sería la señal temporal “durante”, lo cual indica que las posibles respuestas válidas son aquellas cuyas fechas se encuentren dentro de la fecha obtenida al responder la segunda pregunta simple. El resto serán descartadas por no cumplir las restricciones impuestas.

Por tanto, el tratamiento de las preguntas complejas está basado en la descomposición de estas preguntas complejas en preguntas más simples que pueden ser contestadas por sistemas de Búsqueda de Respuestas convencionales y finalmente, las respuestas a las preguntas simples serán combinadas para obtener la respuesta final a la pregunta original.



La idea de aplicar el sistema TERSEO para resolver preguntas complejas con características temporales en un sistema de Búsqueda de Respuestas ocasionó el desarrollo de una arquitectura multicapa que permitiera intercambiar módulos de tratamiento de preguntas y respuestas de un sistema general de Búsqueda de Respuestas. Para ello, en las dos siguientes secciones se presenta una taxonomía de preguntas temporales y la arquitectura multicapa propuesta.

## 6.1 Propuesta de una taxonomía de preguntas temporales

Antes de hacer una propuesta de resolución de preguntas temporales se impone la necesidad de determinar qué tipos de preguntas temporales podrían encontrarse en un proceso de búsqueda de respuestas. La clasificación de preguntas presentada a continuación está relacionada con la forma de resolver dichas preguntas. Para este propósito hemos usado dos tipos diferentes de corpus de preguntas temporales. El primero es el corpus de preguntas TimeML (versión 0.2) (Radev & Sundheim, 2002a) y el segundo son dos grupos de preguntas temporales de Excite y TREC8 (1999). Después de un profundo estudio de estos textos, se han distinguido varios tipos de preguntas diferentes. Por un lado existen preguntas simples, que pueden ser resueltas directamente, y por otro, preguntas complejas, que necesitan ser descompuestas en varias preguntas simples para poder ser resueltas. Un analizador de expresiones temporales procesa las expresiones de la pregunta, y un sistema de descomposición de preguntas divide la pregunta en preguntas más simples. Finalmente, las respuestas a las preguntas simples tienen que ser combinadas para dar una respuesta final a la pregunta compleja. Por tanto, la clasificación quedaría de la siguiente manera (Saquete *et al.*, 2004c):

- **Preguntas no temporales:** *Tipo 0*
- **Preguntas temporales simples**
  - *Tipo 1: Preguntas temporales de un único evento sin expresión temporal.* Este tipo de preguntas hacen referencia a un único



evento y serán resueltas por el sistema de Búsqueda de Respuestas directamente sin la necesidad de realizar un preproceso o postproceso de las mismas. No hay ninguna expresión temporal en la pregunta. Por ejemplo: “¿Cuándo cerró Jordania el puerto de Aqaba a Kuwait?”

- *Tipo 2: Preguntas temporales de un único evento con expresión temporal.* Estas preguntas hacen referencia a un único evento, pero hay una o más expresiones temporales formando parte de la pregunta. Estas expresiones temporales deben ser reconocidas, resueltas y anotadas. Toda esta información temporal podría ayudar en la búsqueda de una respuesta. Por ejemplo: “¿Quién ganó las elecciones en 2004?”. Expresión temporal: 2004
- **Preguntas temporales complejas**
  - *Tipo 3: Preguntas temporales de múltiples eventos con expresión temporal.* Se trata de preguntas que hacen referencia a dos o más eventos relacionados por una señal temporal. Esta señal establece el orden entre los eventos de la pregunta. Además, hay una o más expresiones temporales en la formulación de la pregunta. Estas expresiones temporales necesitan ser reconocidas, resueltas y anotadas y esta información establecerá unas restricciones temporales en las respuestas a dichas preguntas. Por ejemplo: “¿Qué hizo George Bush después de que las Naciones Unidas ordenarán un embargo global del mercado en Irak en Agosto del 90?”. En este ejemplo, la señal temporal es “después de”, las restricciones temporales serán “entre 1/8/1990 y 31/8/1990” y la pregunta será dividida en:
    - Q1: “¿Qué hizo George Bush?”
    - Q2: “¿Cuándo ordenaron las Naciones Unidas un embargo global del mercado en Irak en Agosto del 90?”
  - *Tipo 4: Preguntas temporales de múltiples eventos sin expresión temporal.* Se trata de preguntas que hacen referencia a dos o más eventos relacionados por una señal temporal. Esta señal establece el orden entre dichos eventos. Por ejemplo: “¿Qué ocurrió con el precio del petróleo después de la anexión iraquí de Kuwait?”. En este ejemplo, la señal temporal



será “después de” y la pregunta será dividida de la siguiente manera:

- o Q1: “¿Qué ocurrió con el precio del petróleo?”
- o Q2: “¿Cuándo fue la anexión iraquí de Kuwait?”

Cada tipo y su procesamiento será explicado en detalle en las siguientes secciones.

## 6.2 Arquitectura multicapa de un sistema de Búsqueda de Respuestas

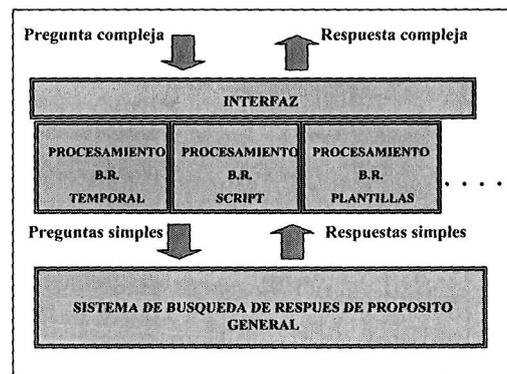
Para poder tratar preguntas complejas, se propone una arquitectura multicapa que será capaz de descomponer estas preguntas complejas en preguntas simples, de forma que las preguntas simples puedan ser tratadas por los sistemas actuales de Búsqueda de Respuestas. Esta arquitectura del sistema incrementa la funcionalidad de los actuales sistemas de Búsqueda de Respuestas, permitiendo resolver cualquier pregunta de tipo temporal. Además, este sistema puede ser fácilmente extendido con nuevas capas para tratar otro tipo de preguntas complejas no relacionadas con la temporalidad.

Algunos ejemplos de preguntas complejas son:

- Preguntas temporales como “¿Dónde estudió Michael Milken antes de ir a la Universidad de Pennsylvania?”. Este tipo de preguntas necesitan usar la información temporal y la ordenación de eventos para obtener la respuesta correcta.
- Preguntas de script como: “¿Cómo construyo una bicicleta?”. En estas preguntas, la respuesta final es un conjunto de respuestas ordenadas.
- Preguntas basadas en plantillas como “¿Cuáles son los datos biográficos más importantes de Nelson Mandela?”. Este tipo de preguntas deberían dividirse en un conjunto de subpreguntas necesarias para rellenar una plantilla relacionada con la pregunta.

Todos estos tipos de preguntas tienen en común la necesidad de un procesamiento adicional de la pregunta para poder ser con-

testada, y en algunos casos también de la respuesta. Nuestra propuesta es capaz de manejar todos estos tipos de preguntas complejas superponiendo una capa adicional de procesamiento, una para cada tipo, a un sistema Búsqueda de Respuestas de Propósito General como se muestra en la figura 6.1.



**Figura 6.1.** Representación gráfica de un sistema de Búsqueda de Respuestas multicapa

Las principales ventajas de esta arquitectura multicapa son:

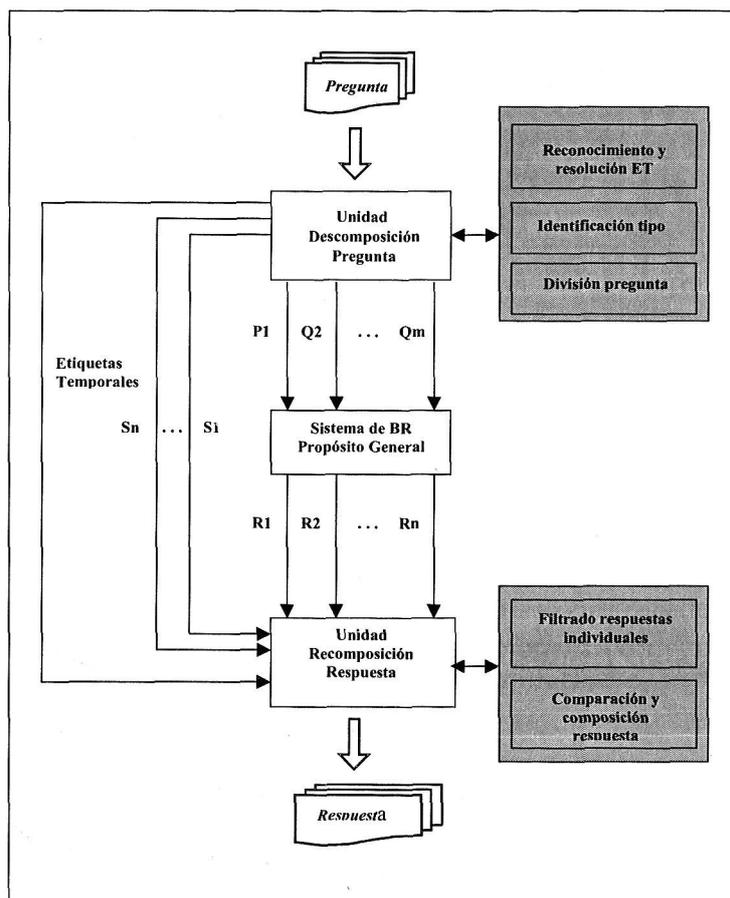
- Permite usar cualquier sistema de Búsqueda de Respuestas existente, con el único esfuerzo de adaptar la salida de la capa de procesamiento al tipo de entrada que dicho sistema requiera.
- Debido al hecho de que el procesamiento de la pregunta compleja se realiza en una capa superior, no es necesario modificar el sistema de Búsqueda de Respuestas cuando éste necesita resolver otro tipo de preguntas complejas.
- Cada capa de procesamiento adicional es independiente de todas las demás, con lo cual, si queremos añadir nuevas capas para tratar otro tipo de preguntas complejas, las ya existentes no se ven afectadas.

En concreto, en esta Tesis se presenta el desarrollo de la capa de procesamiento que es capaz de tratar preguntas temporales de acuerdo a la taxonomía definida en la sección 6.1.



### 6.3 Arquitectura de una sistema de Búsqueda de Respuestas Temporal

En la figura 6.2 se muestran las diferentes partes de la arquitectura del sistema de Búsqueda de Respuestas Temporal y la interacción entre sus unidades.



**Figura 6.2.** Representación gráfica del sistema de Búsqueda de Respuestas temporal

Los principales componentes del sistema de Búsqueda de Respuestas Temporal son:

- Unidad de Descomposición de la Pregunta



- Sistema de Búsqueda de Respuestas de Propósito General
- Unidad de Reconstrucción de la Respuesta

Todos estos componentes trabajan juntos para conseguir la respuesta final de la pregunta compleja. La unidad de Descomposición de Preguntas y la unidad de Reconstrucción de Respuestas forman la capa de procesamiento en la Búsqueda de Respuestas Temporal, la cual se encarga de procesar las preguntas temporales antes y después de que éstas usen el sistema de Búsqueda de Respuestas de Propósito General. Debido a que la división y la reconstrucción de la pregunta se va a hacer de acuerdo a características sintácticas, el sistema de procesamiento de preguntas temporales es dependiente del idioma. Sin embargo, puesto que el corpus de preguntas temporales que estamos utilizando está en inglés, la primera aproximación del sistema fue específica para el inglés, intentando, en la medida de lo posible, la mayor flexibilidad para hacerlo fácilmente extensible a otras lenguas como el castellano, lengua para la cual el sistema TERSEO se encuentra también en funcionamiento. A continuación se exponen las características generales de cada unidad.

- *Unidad de Descomposición de Preguntas:* Es una unidad de pre-procesamiento que tiene tres tareas principales.
  1. De acuerdo a la taxonomía de preguntas definida previamente, será necesario determinar el tipo de la pregunta que vamos a tratar puesto que el procesamiento de cada pregunta será diferente dependiendo del tipo del que se trate.
  2. Debido a que estamos tratando con preguntas relacionadas con temporalidad, será necesario reconocer y resolver las expresiones temporales que dicha pregunta pueda contener. La pregunta podría no contener este tipo de expresiones, en cuyo caso esta tarea no sería necesaria.
  3. Las preguntas de tipo complejo, es decir, las preguntas de Tipo 3 y Tipo 4 serán divididas en preguntas simples. Estas preguntas simples van a ser la entrada del sistema de Búsqueda de Respuestas de Propósito General.
- *Sistema de Búsqueda de Respuestas de Propósito General:* Cualquier sistema de Búsqueda de Respuestas genérico podría ser



usado aquí. La única condición es saber cual es el tipo de entrada y de salida que usa dicho sistema para poder adaptar la capa de procesamiento al mismo. Los principales componentes de un sistema general de Búsqueda de Respuestas son:

- Análisis de la pregunta
- Selección de pasajes o documentos
- Extracción de la respuesta
- *Unidad de Recomposición de Respuestas*: Éste sería el último paso del proceso. Esta unidad llevará a cabo la recomposición de las diferentes respuestas, usando la información temporal obtenida de la pregunta, como por ejemplo, las señales temporales o las expresiones temporales, y será capaz de devolver la respuesta correcta a la pregunta original.

En las siguientes secciones se explicará con detalle tanto la unidad de Descomposición como la unidad de Recomposición.

## 6.4 Unidad de Descomposición de Preguntas

La unidad de Descomposición de Preguntas (Saquete *et al.*, 2004h) está dividida en tres tareas principales:

1. Identificación de Tipo (de acuerdo a la taxonomía propuesta en la sección 6.1)
2. Reconocimiento y Resolución de Expresiones Temporales
3. División de la Pregunta

Cada uno de estos módulos será explicado con detalle a continuación. Una vez que la división de la pregunta ha sido realizada, la salida final de esta unidad estará formada por:

- Un conjunto de subpreguntas que serán la entrada de un sistema de Búsqueda de Respuestas de Propósito General.
- Un conjunto de etiquetas temporales que contienen las fechas concretas devueltas por el sistema TERSEO, y que serán usadas como entrada por la unidad de Recomposición de Respuestas, permitiéndole a esta unidad filtrar aquellas respuestas que no cumplan con las restricciones temporales.



- Un conjunto de señales temporales que serán también entrada para la Unidad de Recomposición de Respuestas y que servirán para realizar la composición de la respuesta final.

Una vez realizada la descomposición, el sistema de Búsqueda de Respuestas de Propósito General se usa para resolver las preguntas simples. Las respuestas a dichas preguntas simples serán, junto con las etiquetas temporales y las señales temporales, entrada para la unidad de Recomposición de Respuestas.

#### 6.4.1 Identificación de Tipo

La unidad de Identificación de Tipo clasifica la pregunta en uno de los cinco tipos de acuerdo a la taxonomía propuesta anteriormente. Esta identificación es necesaria porque cada tipo de pregunta provoca un comportamiento diferente en el sistema. Aquellas preguntas que no son temporales (Tipo 0) no serán tratadas por nuestra capa de procesamiento. Las preguntas de Tipo 1 y Tipo 2 son clasificadas como simples y la respuesta puede ser obtenida sin necesidad de dividir la pregunta en preguntas más simples para poder ser resueltas. Sin embargo, las preguntas de Tipo 3 y Tipo 4 son preguntas complejas y que por tanto necesitan ser divididas en preguntas simples. Una vez que las preguntas han sido divididas, las subpreguntas siempre serán de Tipo 0, Tipo 1 o Tipo 2. La identificación del tipo de la pregunta será establecido de acuerdo a las reglas que se muestran en la figura 6.3.

Algunos ejemplos de los diferentes tipos de preguntas en inglés serán:

- **Tipo 1:** “*When did Jordan close the port of Aqaba to Kuwait?*” (“¿Cuándo cerró Jordania el puerto de Aqaba a Kuwait?”)
- **Tipo 2:** “*Who won the 1988 New Hampshire republican primary?*” (“¿Quién ganó las primarias republicanas de New Hampshire en 1988?”)
- **Tipo 3:** “*What did George Bush do after the U.N. ordered a global embargo on trade with Iraq in August 90?*” (“¿Qué hizo George Bush después de que la U.N. ordenará un embargo global del mercado en Irak en Agosto del 90?”)

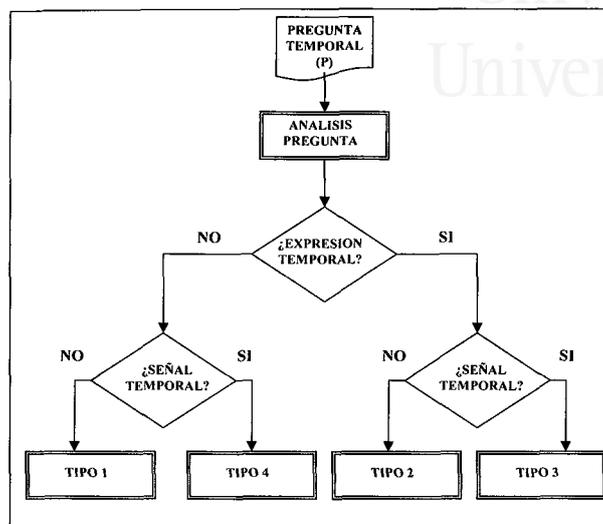


Figura 6.3. Árbol de decisión para Identificación de Tipo

- **Tipo 4:** “What happened to world oil prices *after* the Iraqi “*annexation*” of Kuwait?” (“¿Qué ocurrió con los precios mundiales del petróleo después de la anexión iraquí de Kuwait?”)

#### 6.4.2 Reconocimiento y Resolución de Expresiones Temporales

Este módulo utiliza el sistema TERSEO multilingüe, explicado en la capítulo 5, para reconocer, resolver y anotar las expresiones temporales de las preguntas. Como ya se explicó previamente, el sistema TERSEO se extendió automáticamente del castellano al inglés y el conjunto de reglas de identificación de expresiones temporales para el inglés obtenido automáticamente es el que actualmente se usa para procesar las expresiones temporales de las preguntas en esta unidad.

Las preguntas son tratadas por TERSEO, devolviendo como salida las etiquetas temporales asociadas a dichas preguntas. Esta información será usada en la unidad de Recomposición de Respuestas para poder filtrar las respuestas individuales obtenidas por el sistema de Búsqueda de Respuestas de Propósito General. Estas etiquetas, por tanto, están funcionando a modo de res-



tricciones temporales. Sólo aquellas respuestas que cumplan las restricciones serán posibles respuestas válidas a la pregunta final.

Mostramos con un ejemplo cómo funciona esta unidad con preguntas como: “*Which U.S. ship was attacked by Israeli forces during the war in the sixties?*” (“¿Qué barco de Estados Unidos fue atacado por fuerzas israelitas durante la guerra en los sesenta?”).

1. Primero, la unidad reconoce la expresión temporal en la pregunta, la resuelve y la etiqueta. La salida de la unidad sería la siguiente etiqueta:

```
<DATETIMEREf valdate1 = "01/01/1960"
valdate2 = "31/12/1969"> in
the sixties </DATETIMEREf>
```

2. Restricción temporal: las fechas asociadas a las respuestas deben estar dentro del rango de valores establecidos por los atributos de la etiqueta (valdate1 y valdate2).

### 6.4.3 División de la Pregunta

Esta tarea solamente será necesaria cuando el tipo de la pregunta, obtenido de la unidad de Identificación de Tipo, sea Tipo 3 o Tipo 4. Estas preguntas son consideradas complejas y necesitan ser divididas en preguntas más simples (Tipo 0, Tipo 1 o Tipo 2) para poder ser resueltas por un sistema de Búsqueda de Respuestas de Propósito General.

La descomposición de la pregunta compleja está basada en la identificación de las señales temporales, las cuales relacionan los diferentes eventos de la pregunta y establecen un orden entre las respuestas de las subpreguntas. Estas señales, junto con las preguntas divididas son las salidas de este módulo.

**Señales temporales.** Después de un profundo estudio del corpus de preguntas se determinaron un conjunto de señales temporales. Cada señal temporal denota una relación entre las fechas de los eventos que está relacionando. Asumiendo que F1 es la fecha relacionada con el primer evento de la pregunta y que F2 es la fecha relacionada con el segundo evento, la señal establecerá un cierto orden entre F1 y F2 y por tanto también entre el primer evento



y el segundo. A este orden lo hemos llamado “clave de ordenación”, y esta clave será usada en la unidad de recomposición de respuestas para componer la respuesta final del sistema. El conjunto de las señales temporales junto con sus claves de ordenación se muestra en el cuadro 6.1.

SEÑAL	CLAVE ORDENACIÓN
<i>After</i> (después de)	$F1 > F2$
<i>When</i> (cuando)	$F1 = F2$
<i>Before</i> (antes de)	$F1 < F2$
<i>During</i> (durante)	$F2i \leq F1 \leq F2f$
<i>From F2 to F3</i> (desde F2 hasta F3)	$F2 \leq F1 \leq F3$
<i>About F2 - F3</i> (aprox. F2 - F3)	$F2 \leq F1 \leq F3$
<i>On / in</i> (en)	$F1 = F2$
<i>While</i> (mientras)	$F2i \leq F1 \leq F2f$
<i>For</i> (durante)	$F2i \leq F1 \leq F2f$
<i>At the time of</i> (en el mismo momento de)	$F1 = F2$
<i>Since</i> (desde)	$F1 > F2$

Cuadro 6.1. Claves de ordenación para cada señal

**Implementación del módulo.** Para determinar el tipo de estructuras de preguntas se pueden encontrar y el tratamiento apropiado para una correcta división de las mismas se realizó un estudio del corpus de preguntas. En concreto, se detectó que la primera parte de la pregunta compleja no es un problema, porque no se necesita hacer ningún cambio en ella. La transformación debe realizarse en la parte de la pregunta que se encuentra detrás de la señal temporal. Esta segunda parte tiene que ser transformada a una pregunta de tipo “¿Cuándo?”. Sin embargo, la transformación es diferente según el tipo de estructura que tenga la pregunta compleja inicial. La implementación de este módulo se muestra en la figura 6.4.

Los tres posibles casos son:

- La pregunta que sigue a la señal temporal no tiene ningún verbo, como por ejemplo: “*What happened to the world oil prices after the Iraqí annexation of Kuwait?*” (“¿Qué ocurrió con los

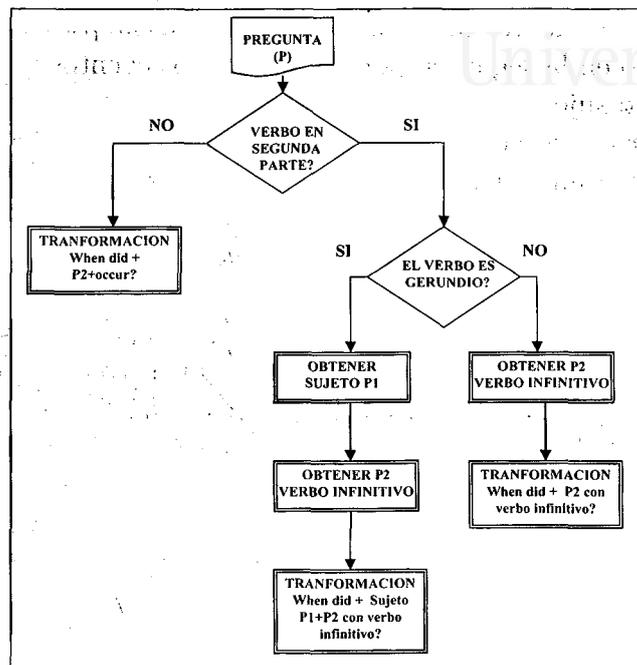


Figura 6.4. Árbol de decisión para la división de la pregunta

precios mundiales del petróleo después de la anexión iraquí de Kuwait?”). En este caso, el sistema devuelve la siguiente transformación de la subpregunta: “*When did the Iraqí annexation of Kuwait occur?*” (“¿Cuándo ocurrió la anexión iraquí de Kuwait?”). Como se puede observar, este caso es el más simple puesto que la única transformación necesaria implica añadir las palabras “*When did....occur?*” (“¿Cuándo ocurrió...?”) a la segunda parte de la pregunta compleja.

- La pregunta que sigue a la señal temporal contiene un verbo, pero este verbo es un gerundio, por ejemplo: “*Where did Bill Clinton study before going to Oxford University?*” (“¿Dónde estudió Bill Clinton antes de ir a la Universidad de Oxford?”). En este caso es necesario hacer dos pasos previos antes de transformar la pregunta:



1. Obtener el sujeto de la primera parte de la pregunta para utilizarlo en la segunda, puesto que ambos eventos se refieren al mismo sujeto.
2. Transformar el verbo de la segunda parte de la pregunta de gerundio a infinitivo.

La transformación final devuelta por el sistema será: “*When did Bill Clinton go to Oxford University?*” (“¿Cuándo fue Bill Clinton a la Universidad de Oxford?”)

- El último tipo de transformación se produce cuando la segunda parte de la pregunta compleja contiene un verbo conjugado y su propio sujeto, como por ejemplo: “*What did George Bush do after the U.N. ordered a global embargo on trade with Iraq?*” (“¿Qué hizo George Bush después de que la U.N. ordenará un embargo global del mercado en Irak?”). En este caso, se obtiene el lema y el tiempo verbal del verbo de la frase para transformarlo a un formato de pregunta adecuado, de forma que el resultado sería: “*When did the U.N. order a global embargo on trade with Iraq?*” (“¿Cuándo ordenó la U.N. un embargo global en el mercado con Irak?”)

#### 6.4.4 Ejemplo de funcionamiento de la unidad de Descomposición

En el siguiente ejemplo se muestra una parte del fichero de salida de la unidad de Descomposición de Preguntas.

1. **Where did Bill Clinton study before going to Oxford University?**  
 Tipo: 4  
 Señal Temporal: before  
 P1: Where did Bill Clinton study?  
 FechaP1:-  
 P2: When did Bill Clinton go to Oxford University?  
 FechaP2:-
2. **What did George Bush do after the U.N. Security Council ordered a global embargo on trade with Iraq in August 90?**  
 Tipo: 3  
 Señal temporal: after  
 Expresión temporal: in August 90  
 P1: What did George Bush do?  
 FechaP1:-  
 P2: When did the U.N. Security Council order a global embargo on trade with Iraq in August 90?  
 FechaP2:[01/08/1990-31/08/1990]



3. **When did Iraq invade Kuwait?**  
 Tipo: 1  
 Señal temporal: -  
 Expresión temporal: -  
 P1: When did Iraq invade Kuwait?  
 FechaP1:-  
 P2: -  
 FechaP2:-
4. **Who became governor of New Hampshire in 1949?**  
 Tipo: 2  
 Señal temporal: -  
 Expresión temporal: in 1949  
 P1: Who became governor of New Hampshire in 1949?  
 FechaP1:[01/01/1949-31/12/1949]  
 P2: -  
 FechaP2:-

## 6.5 Unidad de Reconstrucción de la Respuesta

La Reconstrucción de la Respuesta (Saquete *et al.*, 2004g) es el último paso de la capa de procesamiento propuesta. Como ya se especificó con detalle en la sección anterior, las preguntas complejas fueron divididas por la unidad de Descomposición en preguntas simples. Estas preguntas simples fueron procesadas por un sistema de Búsqueda de Respuestas genérico que devolvía un conjunto de respuestas y los pasajes donde dichas respuestas habían sido encontradas. El sistema general de Búsqueda de Respuestas utilizado para el inglés es IONAUT<sup>1</sup> (Singhal & Kaszkiel, 2001). Se trata de un sistema basado en web y de sencillo uso, que ante una pregunta devuelve un conjunto de posibles respuestas con los pasajes donde se encuentran dichas respuestas. Por tanto, para poder funcionar correctamente, la unidad de Reconstrucción de la Respuesta necesita un paso previo mediante el cual obtiene la fecha relacionada con esa respuesta, aplicando el sistema TERSEO sobre el pasaje devuelto. Esta fecha es necesaria para poder filtrar aquellas respuestas que no cumplen las restricciones temporales establecidas en la unidad de Descomposición. Por tanto, una vez obtenidas estas fechas, la unidad de Reconstrucción es capaz de

<sup>1</sup> <http://www.ionaut.com:8400/>



filtrar las respuestas simples individuales que no cumplen las restricciones, rechazándolas. Finalmente, usando la “clave de ordenación” impuesta por la señal temporal de la pregunta compleja, el sistema será capaz de obtener la respuesta final a la pregunta original.

Este proceso está dividido, por tanto, en tres módulos principales:

- Preprocesamiento de la salida del sistema general de Búsqueda de Respuestas
- Filtrado de las respuestas individuales
- Comparación y composición de la respuesta final

Esta unidad tendrá como entrada un conjunto de respuestas individuales, una serie de etiquetas temporales devueltas por el sistema TERSEO y la señal temporal (si existe) relacionada con la pregunta compleja. Con toda esta información, el sistema es capaz de devolver como salida la respuesta final.

### 6.5.1 Preprocesamiento de la salida del sistema general de Búsqueda de Respuestas

Debido a que el sistema aquí propuesto es independiente del sistema de Búsqueda de Respuestas que se esté utilizando para contestar a las preguntas simples, un módulo de Preprocesamiento de esta salida será necesario para convertir la salida del sistema en la entrada que la unidad de Recomposición está esperando. La entrada que la unidad está esperando es un fichero con las respuestas a las preguntas y las fechas asociadas a dichas respuestas. Por lo tanto, esta unidad de Preprocesamiento va a ser dependiente del sistema de Búsqueda de Respuestas utilizado. Normalmente, los sistemas de Búsqueda de Respuestas actuales devuelven la/s respuesta/s a las preguntas y los documentos o pasajes asociados a dichas respuestas. Nuestro módulo de Preprocesamiento utilizará el sistema TERSEO para buscar la fecha más cercana a la respuesta dada por el sistema y generará un fichero con dicha respuesta y la fecha encontrada. De esta forma, el resto de módulos de la unidad de Recomposición podrán trabajar sin ningún problema.

### 6.5.2 Filtrado de las respuestas individuales

Todas las posibles respuestas dadas por el sistema de Búsqueda de Respuestas general y sus fechas asociadas son la entrada del módulo de Filtrado. En principio, las respuestas individuales no están relacionadas unas con otras sino simplemente sujetas a la restricción temporal impuesta por el análisis de las expresiones temporales. Este módulo, por tanto, seleccionará solamente aquellas respuestas que satisfacen las restricciones temporales impuestas por las expresiones temporales obtenidas y anotadas por TERSEO en la unidad de Descomposición. Por tanto, la fecha asociada a la respuesta debería estar dentro del rango de valores de los atributos de la etiqueta temporal con la que estamos trabajando en ese momento. Como consecuencia final, serán rechazadas aquellas respuestas que no cumplan estas restricciones, reduciendo así el campo de búsqueda de la respuesta final.

**Ejemplo.** Continuando con el ejemplo usado para describir la sección anterior, pero añadiéndole además una expresión temporal, quedaría de la siguiente manera: “*Where did Bill Clinton study before going to Oxford University in the sixties?*” (“¿Dónde estudió Bill Clinton antes de ir a la Universidad de Oxford en los sesenta?”) y el proceso a seguir se detalla a continuación:

1. En el primer paso, la unidad de descomposición dividió la pregunta en P1: “*Where did Bill Clinton study?*” (“¿Dónde estudió Bill Clinton?”) y P2: “*When did Bill Clinton go to Oxford University in the sixties?*” (“¿Cuándo fue Bill Clinton a la Universidad de Oxford en los sesenta?”).
2. Se obtienen una serie de respuestas para cada subpregunta del sistema de Búsqueda de Respuestas de Propósito General, que después de ser tratadas por el módulo de Preprocesamiento tendría el siguiente aspecto:
  - Respuesta P1: “*Georgetown University*”(1964-1968)
  - Respuesta P1: “*Oxford University*” (1968-1970)
  - Respuesta P1: “*Yale Law School*” (1970-1973)
  - Respuesta P2: 1968 (1968)
3. Todas estas respuestas son filtradas por la unidad que nos ocupa, para determinar cuáles de ellas son válidas puesto que



cumplen las restricciones temporales. Las restricciones temporales fueron obtenidas en la unidad de Descomposición. En este ejemplo, la expresión temporal sería *“in the sixties”* (“en los sesenta”), esta expresión fue reconocida, resuelta y anotada estableciendo un rango de fechas entre **VALDATE1:** “01/01/1960” y **VALDATE2:** “31/12/1969”. Como se puede observar, las primeras dos respuestas se encuentran total o parcialmente dentro de este rango de fechas así como la respuesta de la P2, que es única en este caso. Sin embargo, la tercera respuesta de la P1 no cumple la restricción temporal y por tanto será rechazada y no se evaluará en el módulo de Composición de la respuesta final.

### 6.5.3 Comparación y composición de la respuesta final

Finalmente, una vez que las respuestas han sido filtradas, llega el momento de usar las señales temporales y la clave de ordenación impuesta por dicha señal para comparar los resultados de cada subrespuesta y devolver la respuesta final a la pregunta original.

En concreto, esta unidad comparará la fecha de cada respuesta de la primera pregunta con la fecha obtenida como respuesta para la segunda pregunta. Si el orden entre dichas fechas coincide con el orden indicado en la clave de ordenación querrá decir que la respuesta es válida, en otro caso será descartada. Al final, la respuesta válida será la salida de esta unidad y también de la capa de procesamiento de preguntas temporales, es decir, la respuesta a la pregunta compleja original.

**Ejemplo.** Siguiendo con el mismo ejemplo: *“Where did Bill Clinton study before going to Oxford University in the sixties?”* (“¿Dónde estudió Bill Clinton antes de ir a la Universidad de Oxford en los sesenta?”):

1. Las fechas de las respuestas se ordenan de acuerdo con la ordenación impuesta por la señal temporal *“before”* (antes). Esta señal nos está indicando que la fecha de la respuesta de la primera pregunta tiene que ser previa a la fecha de la respuesta de la segunda pregunta.



2. Después de esto, la única respuesta que cumple esta condición será: “*Georgetown University*”, puesto que la fecha asociada a la misma es la única previa a la fecha relacionada con la segunda pregunta cuyo intervalo de fechas va desde “01/01/1968” hasta “31/12/1968”.

## 6.6 Adaptación del sistema al castellano

Como ya se ha comentado previamente, la descomposición de preguntas complejas en preguntas más simples va a ser dependiente del idioma, es por ello que han sido necesarias dos tareas principales de adaptación para poder realizar la descomposición de preguntas complejas en castellano:

- Traducción al castellano e introducción en el sistema de todas las señales temporales que teníamos definidas para el inglés.
- Extensión del módulo de división de preguntas para que sea capaz de formar preguntas simples correctas también en castellano.

Por tanto, una vez traducidas las señales, la detección de las mismas es una tarea inmediata. La ampliación en la implementación del módulo de división de preguntas para el castellano se muestra en la figura 6.5.

Los tres posibles casos para el castellano son:

- La pregunta que sigue a la señal temporal no tiene ningún verbo, como por ejemplo: “¿Qué ocurrió con los precios mundiales del petróleo después de la anexión iraquí de Kuwait?”. En este caso, el sistema devuelve la siguiente transformación de la subpregunta: “¿Cuándo ocurrió la anexión iraquí de Kuwait?”. La única transformación que es necesaria aquí es añadir las palabras “¿Cuándo ocurrió...?” a la segunda parte de la pregunta compleja.
- La pregunta que sigue a la señal temporal contiene un verbo, pero este verbo es un infinitivo, por ejemplo: “¿Dónde estudió Bill Clinton antes de ir a la Universidad de Oxford?”. En este caso es necesario hacer dos pasos previos antes de transformar la pregunta:

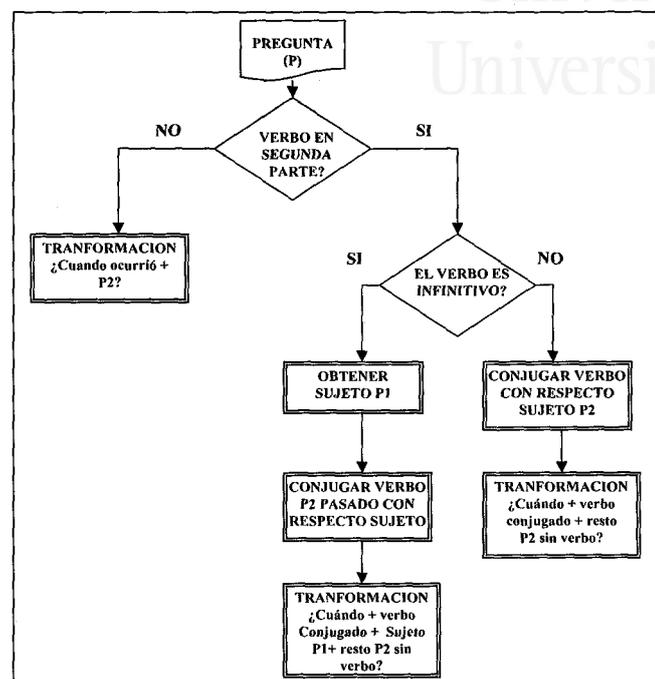


Figura 6.5. Árbol de decisión para la división de la pregunta en castellano

1. Obtener el sujeto de la primera pregunta para utilizarlo en ésta, puesto que ambos eventos se refieren al mismo sujeto.
2. Obtener la persona del sujeto y conjugar el verbo en pasado de acorde a este sujeto.

La ordenación correcta de la pregunta será: *¿Cuándo + verbo conjugado + Sujeto P1 + resto de P2 al quitarle el verbo infinitivo original?*. Por tanto, la transformación final devuelta por el sistema será: “¿Cuándo fue Bill Clinton a la Universidad de Oxford?”

- El último tipo de transformación es cuando la segunda parte de la pregunta compleja contiene un verbo conjugado y su propio sujeto, como por ejemplo: “¿Qué hizo George Bush después de que la U.N. ordenará un embargo global del mercado en Irak?”. En este caso se obtiene la persona del sujeto de la segunda pregunta, el lema del verbo que está conjugado y lo que hace el sistema es conjugar dicho verbo en pasado y de acuerdo



al sujeto al que acompaña. Finalmente, se combina toda la información para conseguir una sintaxis de pregunta correcta del tipo: *¿Cuándo + verbo conjugado + Sujeto P2 + resto de P2 al quitarle el verbo conjugado original y el sujeto?* La transformación final para nuestro ejemplo sería: “¿Cuándo ordenó la U.N. un embargo global en el mercado con Irak?”

### 6.6.1 Ejemplo de funcionamiento de la unidad de Descomposición para el castellano

A continuación se muestra un fichero con la salida que nuestro sistema a devuelto al procesar un conjunto de preguntas temporales en castellano.

1. **¿Qué hizo George Bush después de que la U.N. Security Council ordenara un embargo global con Irak en agosto de 1990?**  
 Tipo: 3  
 Señal temporal: después de que  
 Expresión temporal: en agosto de 1990  
 P1: ¿Qué hizo George Bush?  
 FechaP1:-  
 P2: ¿Cuándo ordenó la U.N. Security Council un embargo global con Irak en agosto de 1990?  
 FechaP2:[01/08/1990-31/08/1990]
2. **¿Qué le ocurrió a los precios de petróleo mundiales después de la unión de Irak a Kuwait?**  
 Tipo: 4  
 Señal temporal: después de  
 P1: ¿Qué le ocurrió a los precios de petróleo mundiales?  
 FechaP1:- P2: ¿Cuándo fue la unión de Irak a Kuwait?  
 FechaP2:-
3. **¿Quién ganó las primarias de New Hampshire en 1988?** Tipo: 2  
 Expresión temporal: en 1988  
 P1: ¿Quién ganó las primarias de New Hampshire en 1988?  
 FechaP1:[01/01/1988-31/12/1988]  
 P2: -  
 FechaP2:-
4. **¿Quién gobernaba en Francia en 1988?**  
 Tipo: 2  
 Expresión temporal: en 1988  
 P1: ¿Quién gobernaba en Francia en 1988?  
 FechaP1:[01/01/1988-31/12/1988]  
 P2: -  
 FechaP2:-
5. **¿Quién gobernaba en Argentina durante la guerra de las Malvinas?**  
 Tipo: 4  
 Señal temporal: durante  
 P1: ¿Quién gobernaba en Argentina?



- FechaP1:-  
P2: ¿Cuándo fue la guerra de las Malvinas?  
FechaP2:-
6. **¿Dónde estudió Michael Milken antes de ir a la Universidad de Pensilvania?**  
Tipo: 4  
Señal temporal: antes de  
P1: ¿Dónde estudió Michael Milken?  
FechaP1:-  
P2: ¿Cuándo fue Michael Milken a la Universidad de Pensilvania?  
FechaP2:-
7. **¿Qué le ocurrió al precio del crudo tras la invasión iraquí de Kuwait?**  
Tipo: 4  
Señal temporal: tras  
P1: ¿Qué le ocurrió al precio del crudo?  
FechaP1:-  
P2: ¿Cuándo fue la invasión iraquí de Kuwait?  
FechaP2:-

## 6.7 Conclusiones del capítulo

En este capítulo se ha presentado un sistema de Búsqueda de Respuestas Temporal capaz de contestar preguntas de carácter temporal usando una arquitectura multicapa para el tratamiento de dichas preguntas. El sistema utiliza el sistema TERSEO de reconocimiento y resolución de expresiones temporales, presentado en capítulos anteriores, para extraer la información temporal tanto de las preguntas como de los textos. El sistema propuesto en este trabajo consigue tratar preguntas complejas temporales, hoy en día fuera del ámbito de los sistemas de Búsqueda de Respuestas, dividiendo estas preguntas complejas en preguntas más simples que sí puedan ser contestadas por los sistemas actuales. Por tanto, los tres elementos clave del sistema de Búsqueda de Respuestas Temporal son:

- Descomposición de la pregunta compleja en preguntas simples.
- Estas preguntas simples pueden ser contestadas por cualquier sistema de Búsqueda de Respuestas de Propósito General.
- Las respuestas obtenidas son filtradas y recompuestas para dar la respuesta final a la pregunta compleja.



## 6.7 Conclusiones del capítulo 113

Puesto que la descomposición de la pregunta se hace según reglas sintácticas, el procesamiento de dicha división va a ser dependiente del idioma. En un principio se realizaron todos los pasos del procesamiento para el inglés y posteriormente se añadió una extensión para su funcionamiento correcto en castellano. Únicamente fue necesario adaptar el módulo de la división de la pregunta, el resto de módulos funcionan correctamente para todos los idiomas.



Universitat d'Alacant  
Universidad de Alicante



Universitat d'Alacant  
Universidad de Alicante

## 7. Evaluación del sistema

Una vez implementado el sistema TERSEO y su aplicación a la Búsqueda de Respuestas Temporal, se han realizado una serie de evaluaciones para medir la eficiencia del sistema tanto en la tarea de reconocimiento y resolución de expresiones temporales como en su aplicación a la ordenación de eventos y la Búsqueda de Respuestas.

El trabajo de evaluación realizado se ha desarrollado desde dos perspectivas. Por un lado, la evaluación del sistema de identificación y resolución de expresiones temporales (TERSEO) y por otro lado, la evaluación de las aplicaciones desarrolladas, que a su vez usan el sistema TERSEO.

La evaluación del sistema TERSEO se ha realizado en dos pasos. En primer lugar, se evaluó el sistema monolingüe en español y posteriormente, su extensión automática al inglés. El sistema en inglés fue evaluado también gracias a la participación del mismo en la conferencia TERN 2004<sup>1</sup> (*Temporal Expression Recognition and Normalization*), en la cual participaban diferentes organizaciones y universidades dedicadas al reconocimiento y resolución de información temporal en textos. Finalmente, se realizó la evaluación del sistema de Búsqueda de Respuestas Temporal mediante la utilización del corpus TERQAS<sup>2</sup> y gracias a la participación del sistema en la conferencia CLEF (2004).

<sup>1</sup> <http://timex2.mitre.org/tern.html>

<sup>2</sup> <http://www.cs.brandeis.edu/jamesp/arda/time/documentation.html>



## 7.1 Evaluación del sistema TERSEO monolingüe

Con la finalidad de realizar una evaluación del sistema TERSEO en castellano, se obtuvo de la web una colección de 100 textos digitales en dicha lengua y posteriormente se realizó una anotación manual de dichos textos. Esta anotación fue realizada por dos anotadores y su propósito era el de comparar dicha anotación manual con la anotación automática obtenida por el sistema TERSEO trabajando en castellano con el objetivo de determinar la eficiencia del sistema (Saquete *et al.*, 2004f).

Pero, es necesario confirmar que efectivamente la evaluación manual realizada por los dos anotadores es fiable y no altera los resultados del experimento, lo cual indicará que sirve como referencia para la evaluación del sistema TERSEO. De acuerdo con Carletta *et al.* (1997), para asegurar una buena anotación es necesario realizar una serie de mediciones directas que son:

- Estabilidad
- Reproducibilidad
- Precisión

Pero, además de estas medidas, para medir la fiabilidad del sistema se tiene también que tener en cuenta la cantidad de ruido existente en la información. Carletta *et al.* argumentan que, puesto que la cantidad de coincidencias por casualidad que se pueden esperar depende del número de frecuencias relativas de las categorías que se están evaluando, la fiabilidad para la clasificación de categorías tiene que ser medida usando el factor *kappa* definido en (Siegel & Castellan, 1988). El factor *kappa* ( $k$ ) mide la afinidad en el acuerdo entre un conjunto de anotadores cuando se realizan juicios de categorías.

En este caso, existe únicamente una clase de objetos que pueden ser de tres tipos:

- Objetos que se refieren a la fecha del artículo
- Objetos que se refieren a una fecha nombrada previamente
- Objetos que se refieren a otro tipo de fecha diferente a las dos anteriores



Después de llevar a cabo el cálculo del factor *kappa* teniendo en cuenta estas tres clases de objetos, el valor obtenido fue de  $k=0.953$ . De acuerdo con el trabajo de Carletta, para un valor de *kappa* de  $0.68 < k < 0.8$  las conclusiones eran favorables, y si  $k > 0.8$  significaba que existía fiabilidad absoluta en los resultados de ambos anotadores. Como se puede observar, el valor de *k* obtenido en esta anotación manual es mayor de 0.8, lo cual implica que se garantiza una fiabilidad total en la anotación realizada y los resultados de precisión y cobertura están garantizados. El factor *kappa* ha sido también estudiado para evaluaciones de anáfora por autores como Vieira (2002).

Todas las evaluaciones de reconocimiento y resolución de expresiones temporales presentadas en este capítulo se realizan utilizando las siguientes medidas:

- **CORR (correcto)**: Este valor se incrementa si dos elementos comparados se consideran idénticos.
- **INCO (incorrecto)**: Esta valor se incrementa si dos elementos comparados no son idénticos.
- **MISS (no detectado)**: Este valor se incrementa si una referencia existente no ha sido encontrada por el sistema.
- **SPUR (espúreo)**: Este valor se incrementa si el sistema devuelve una salida para un elemento pero ésta no es una referencia temporal real.
- **POS (posibles)**: Número de referencias que contribuyen al resultado final.

$$POS = CORR + INCO + MISS$$

- **ACT(tratadas)**: El número de resultados que el sistema devuelve.

$$ACT = CORR + INCO + SPUR$$

- **COB. (Cobertura)**: La cantidad de referencias temporales que se encuentran presenten en la salida del sistema.



$$COB = \frac{CORR}{POS}$$

- **PREC. (Precisión):** La cantidad de elementos en la salida del sistema que realmente son referencias temporales.

$$PREC = \frac{CORR}{ACT}$$

- **F-Medida:** Métrica usada para combinar los resultados de precisión y cobertura en un único valor.

$$F - Medida = \frac{(1+\beta^2)(P*C)}{(\beta^2*P+C)}$$

En la evaluación efectuada se ha considerado un valor de  $\beta = 1$  con el fin de asignar un mismo peso tanto a la precisión como a la cobertura.

En la figura 7.1 se muestra un gráfico explicativo en el que se relacionan las medidas de precisión y cobertura.

Para el evaluación del sistema TERSEO monolingüe se utilizaron dos colecciones de artículos de periódicos digitales en castellano. La primera colección se utilizó para entrenamiento del sistema y estaba formada por 50 artículos. Una vez realizados los ajustes oportunos, otro conjunto de 50 artículos fue utilizado para realizar la evaluación definitiva.

Los resultados del sistema pueden verse en el cuadro 7.1. En esta evaluación no se ha determinado el valor de elementos espúreos (SPUR), considerándolos como incorrectos.

Como se puede observar, el sistema alcanzó una precisión del 84% y una cobertura del 71% para el corpus de entrenamiento y una precisión del 88% y una cobertura del 69% para el corpus de test. Además, los resultados de precisión para el corpus de test fueron mejores que para el corpus de entrenamiento debido a que ambos conjuntos de documentos fueron elegidos al azar y se dio el caso de que el número de referencias temporales del corpus



## 7.1 Evaluación del sistema TERSEO monolingüe 119



$$\begin{aligned} \text{POS} &= \text{CORR} + \text{INCO} + \text{MISS} \\ \text{ACT} &= \text{CORR} + \text{INCO} + \text{SPUR} \\ \text{PREC} &= \text{CORR} / (\text{CORR} + \text{INCO} + \text{SPUR}) \\ \text{COB} &= \text{CORR} / (\text{CORR} + \text{INCO} + \text{MISS}) \end{aligned}$$

**COB** = están todos los que son

**PREC** = son todos los que están

Figura 7.1. Representación gráfica de las medidas de precisión y cobertura.

	ENTREN.	TEST
<b>DOCUMENTOS</b>	50	50
<b>POS</b>	238	199
<b>ACT</b>	201	156
<b>CORR</b>	170	138
<b>INCO</b>	31	18
<b>MISS</b>	37	43
<b>SPUR</b>	-	-
<b>PREC.</b>	84 %	88 %
<b>COB.</b>	71 %	69 %
<b>F-MEDIDA</b>	77 %	77 %

Cuadro 7.1. Evaluación del sistema TERSEO monolingüe

de test era menor que en el corpus de entrenamiento. Aunque los resultados obtenidos son altamente satisfactorios, después de un estudio en profundidad de los errores, se detectaron algunos



aspectos en los que el sistema podría mejorarse. Estos aspectos se detallan a continuación:

- Habitualmente, en los artículos de periódicos aparecen expresiones del tipo “el sábado habrá dos reuniones”. Para resolver este tipo de expresiones temporales es necesario tener en cuenta la información del contexto de la frase en la cual se encuentra la referencia. En este caso concreto, se necesita conocer el tiempo verbal del verbo de la frase en la que se encuentra la referencia. Si el tiempo verbal es pasado, la expresión “el sábado” se estaría refiriendo a “el sábado pasado”, mientras que si el tiempo verbal es futuro la expresión se estaría refiriendo a “el próximo sábado”. Puesto que el sistema TERSEO no trata este tipo de información, asumía siempre por defecto que este tipo de expresiones se refieren a un día pasado y no próximo, puesto que las noticias de los periódicos normalmente se refieren a hechos que han ocurrido previamente.
- El sistema evaluado no era capaz de resolver expresiones temporales que se refieren a un evento conocido, como por ejemplo: “dos días después de la guerra de Irak”. Para resolver este tipo de expresiones sería necesario un conocimiento extra del mundo y el sistema no dispone de este conocimiento contextual. Por ello, el sistema falla cuando intenta resolver este tipo de expresiones.
- El resto de fallos se debe a una incorrecta aplicación de las reglas asociadas a las expresiones o a la inexistencia de una regla capaz de resolver una determinada expresión.

## 7.2 Evaluación del sistema TERSEO multilíngüe

Una vez realizada la extensión automática del sistema TERSEO, éste fue evaluado para el inglés (Saquete *et al.*, 2004e). Tal y como se presentaba en el capítulo 5, la extensión automática se realizó traduciendo automáticamente las expresiones temporales contempladas por el sistema monolíngüe, en este caso, castellano,

utilizando como recursos traductores automáticos, palabras clave temporales y textos no anotados en la lengua destino, para llegar a conseguir un nuevo conjunto de expresiones temporales y reglas de identificación para la nueva lengua a la que queríamos extender el sistema. La extensión automática se aplicó para el inglés, puesto que los recursos de traducción automática disponibles eran más amplios que para otras lenguas.

El procedimiento de evaluación utilizando coincidía con el realizado para la evaluación monolingüe, es decir, se realizó la anotación previa manual de un conjunto de textos digitales y posteriormente la medición del factor kappa para dicha anotación. El resultado obtenido para el factor kappa fue de 0.9, lo cual, como ya se ha especificado en la sección anterior indica que la anotación es fiable y sirve como referencia para evaluar los resultados del sistema TERSEO multilingüe.

Para la evaluación del sistema TERSEO en inglés, la colección de artículos de entrenamiento no fue necesaria puesto que la finalidad de esta evaluación era determinar la eficiencia de la extensión automática del sistema sin realizar ningún tipo de ajuste manual en el mismo. Es por ello que, para la evaluación definitiva del sistema se utilizó una colección de 100 artículos de periódicos digitales en inglés obtenidos de la web. Al igual que en la evaluación del sistema monolingüe, se obtuvieron los resultados de precisión, cobertura y F-Medida (Cuadro 7.2).

	<b>INGLÉS</b>
	<b>TEST</b>
<b>DOCUMENTOS</b>	100
<b>POS</b>	634
<b>ACT</b>	511
<b>CORR</b>	393
<b>INCO</b>	118
<b>MISS</b>	123
<b>SPUR</b>	-
<b>PREC.</b>	77 %
<b>COB.</b>	62 %
<b>F-MEDIDA</b>	69 %

**Cuadro 7.2.** Evaluación del sistema TERSEO multilingüe



Como puede observarse en el cuadro se ha obtenido un 77% de precisión y un 62% de cobertura para 634 expresiones temporales contenidas en un total de 100 documentos. Comparando estos resultados con los obtenidos por el sistema en castellano y teniendo en cuenta que las reglas de identificación en castellano se obtuvieron de manera manual y las reglas de identificación en inglés se han obtenido de manera automática, los resultados son muy satisfactorios, puesto que únicamente hay una pérdida de precisión del 11% y una pérdida de cobertura del 7%. La pérdida de cobertura se debe a todas aquellas expresiones que no han sido traducidas correctamente y la pérdida de precisión se debe a aquellas expresiones que sí se han traducido correctamente, pero la regla de resolución que se le asignó automáticamente no era correcta.

Analizando los resultados más en profundidad, se han detectado las causas de estos errores:

- Errores debidos al analizador temporal: La gramática temporal que utiliza el analizador temporal en el sistema TERSEO ha sido obtenida también automáticamente a partir de las expresiones traducidas y de la información morfológico-léxica de dichas expresiones. Debido a que este proceso es automático, y a que algunas expresiones temporales pueden ser incorrectas o no obtenidas al realizar una traducción directa, las reglas generadas a partir de dichas expresiones también serían incorrectas o insuficientes. Por ejemplo, la expresión “24-hours” tiene un formato específico para el inglés, de forma que no se puede obtener una expresión de este tipo automáticamente partiendo del castellano como lengua origen. La regla gramatical para esta expresión sería  $RF - >Z Fg NNS$ . Sin embargo, esta regla no existe en la gramática puesto que dicha gramática se generó automáticamente a partir de las expresiones traducidas y por tanto, esta expresión nunca será detectada ni resuelta por el sistema TERSEO, a no ser que sea introducido manualmente en el conjunto de expresiones temporales contempladas.
- Errores debido a expresiones temporales que no han sido traducidas correctamente por los traductores automáticos, lo cual

provoca que se detecten como temporales expresiones que no lo son cuando TERSEO utiliza esta información o que no se detecten como temporales expresiones que sí lo son, por ejemplo, la expresión temporal en castellano “dentro de una año” fue traducida por el traductor automático on-line FreeTranslator como “*inside a year*”. Esta expresión no es una expresión temporal correcta y debería haber sido traducida como “*within a year*”.

- Errores debido a una mala resolución de las expresiones temporales provocada por una mala asignación automática de la regla de resolución a dicha expresión temporal o porque la regla de resolución no pudo ser encontrada de manera automática para alguna expresión. Por ejemplo, la expresión “*until the last day*” no tiene ninguna regla de resolución asociada en el sistema puesto que al intentar concatenar las características temporales de cada una de las palabras clave temporales que forman la expresión, algunas de estas características eran incompatibles y por ello la unidad de asignación de reglas no fue capaz de encontrar una regla que cumpliera con todas las características necesarias y por tanto, no devolvió ninguna.
- Errores debido a la ambigüedad que determinadas palabras tienen en inglés ya que poseen otros sentidos además del sentido temporal, como por ejemplo: “*may*” y “*May*” o “*march*” y “*March*”. Esto ha provocado que algunas expresiones se hayan reconocido como temporales cuando en realidad no lo eran.

Sería necesario realizar un refinamiento manual de las expresiones y las reglas de identificación obtenidas automáticamente para conseguir que muchos de los errores producidos por la automatización de la extensión del sistema fueran eliminados.

### 7.3 Evaluación TERN 2004

El sistema TERSEO multilingüe también participó en la evaluación TERN 2004<sup>3</sup> con el objetivo de realizar una comparativa de la eficiencia de dicho sistema frente a la de otros sistemas

<sup>3</sup> <http://timex2.mitre.org/tern.html>



que realizan tareas de reconocimiento y resolución de información temporal. Esta evaluación está basada en un trabajo que comenzó en 1999 con el desarrollo de una guía para la anotación de textos y el intercambio de información. Esta guía derivó en el esquema de anotación TIDES 2003, el cual define la etiqueta TIMEX2, que incluía a su vez un conjunto de atributos para almacenar la información obtenida de analizar las expresiones temporales marcadas. Este trabajo extiende al que ya comenzó en las conferencias “Message Understanding Conferences (MUC)”, explicadas en el capítulo 2, pioneras en la definición de una etiqueta para el marcado de entidades temporales. Dicha etiqueta se denominó TIMEX, y era utilizada para el etiquetado de una gran variedad de expresiones ofreciendo así un primer esquema de normalización.

En esta conferencia se podían acometer dos posibles tareas:

- Sólo reconocimiento: Únicamente se detecta y se anota la expresión temporal con una etiqueta sin atributos.
- Reconocimiento y normalización: La expresión se detecta, se resuelve y a la hora de anotarla, los valores de los atributos de dicha etiqueta contendrán la información resuelta para la misma.

Se podía participar para dos lenguas: inglés y chino. El objetivo era reconocer tanto expresiones explícitas, es decir, fechas concretas que no necesitan información de contexto para ser interpretadas, como expresiones implícitas. Este último tipo de expresiones hacen referencia a otra fecha nombrada previamente en el texto y necesitan de esta información para poder ser interpretadas, por ejemplo, “dos días después”. Todos los sistemas participantes debían generar su salida usando las etiquetas definidas en el esquema estándar de anotación TIDES 2003 para la anotación de expresiones temporales.

El etiquetado TIMEX2 únicamente es capaz de anotar expresiones temporales y los atributos que contiene esta etiqueta se muestran en el cuadro 7.3.

A continuación se muestran posibles ejemplos de anotaciones de diferentes tipos de expresiones. En algunos casos los resulta-



ATRIBUTO	FUNCIÓN	EJEMPLO
<i>VAL</i>	Contiene una fecha/hora	VAL="1964-10-16" VAL="P5Y"
<i>MOD</i>	Captura los modificadores temporales	MOD="APPROX"
<i>ANCHOR_VAL</i>	Contiene una forma normalizada de una fecha/hora de anclaje	ANCHOR_VAL="1964"
<i>ANCHOR_DIR</i>	Captura la dirección u orientación relativa entre VAL y ANCHOR_VAL	ANCHOR_DIR="ENDING"
<i>SET</i>	Identifica expresiones que denotan conjuntos de fechas	SET="YES"

Cuadro 7.3. Conjunto de atributos del etiquetado TIMEX2

dos son precisos y en otras impreciso por las características de la propia expresión:

- Fechas del calendario (años, meses, días) y horas (horas, minutos, segundos)

```
<TIMEX2 val="2000-10-03T20:04:56.18"> 10/03/2000
20:04:56.18</TIMEX2> <TIMEX2 val="T13:00">1 p.m.
</TIMEX2>
```

- Unidades de tiempo especiales (años fiscales, eras geológicas)

```
<TIMEX2 val="FY2000">fiscal year 2000</TIMEX2>
```

- Expresiones basadas en semanas

```
<TIMEX2 val="2000-W40">last week</TIMEX2>
```

- Duraciones

```
<TIMEX2 val="P30Y" anchor_val="1998"
anchor_dir="ENDING">30
years</TIMEX2>
```

- Expresiones que contienen modificadores

```
<TIMEX2 val="2000-W40" mod="END">late last
week</TIMEX2>
```



- Expresiones que se resuelven utilizando tokens definidos en el esquema de anotación para representar determinados periodos de tiempo.

```
<TIMEX2 val="2000-10-05TNI">Thursday night
</TIMEX2>
```

```
<TIMEX2 val="2000-Q3">third quarter</TIMEX2>
```

- Expresiones con valor indefinido

```
<TIMEX2 val="XXXX-XX-XX">day</TIMEX2>
```

```
<TIMEX2 val="PXY" anchor_val="2000"
anchor_dir="BEFORE">the
years</TIMEX2>
```

- Conjuntos

```
<TIMEX2 val="XXXX-XX-XX" set="YES">each day
</TIMEX2>
```

El corpus utilizado para la evaluación, tanto para el entrenamiento como para el test definitivo fue preparado por la corporación *MITRE*, bajo la supervisión de *SPAWAR Systems Center*<sup>4</sup>. Los documentos usados para la evaluación fueron seleccionados de aquellos utilizados para las tareas básicas de evaluación de la conferencia ACE 2004 y las fuentes utilizadas fueron tanto “*Broadcast news*” como “*Newswire news*”. El *Linguistic Data Consortium (LDC)* coordinó la distribución de todo el material a los diferentes participantes en la conferencia.

Los tres aspectos principales de la anotación con el esquema TIMEX2 que fueron medidos son:

- Detección (correcta/no detectada/espúrea): si una expresión ha sido detectada y se le ha asignado una etiqueta TIMEX2.
- Texto expresión (correcto/incorrecto): si el texto de la expresión detectado coincide exactamente con el texto que este esquema considera que debe ser marcado.

<sup>4</sup> <http://www.spawar.navy.mil/>



- Atributos (correcto/incorrecto/no detectado/espúreo): si los valores asignados para cada atributo dentro de la etiqueta TIMEX2 (VAL, ANCHOR\_DIR, ANCHOR\_VAL, SET, MOD) son correctos o no.

Con la finalidad de evaluar los resultados de esta evaluación se ha utilizado un programa escrito en Perl denominado `score_timex2.pl`<sup>5</sup>, que compara los ficheros producidos por cada sistema con los anotados manualmente, y analizando cada elemento etiqueta a etiqueta, produce un resumen de los métricas de precisión, cobertura y F-Medida.

En concreto, el sistema TERSEO participó en esta evaluación en la tarea de detección y normalización de expresiones temporales en inglés. Además, con el fin de conseguir el mismo conjunto de etiquetas definidas en el esquema TIMEX2 se realizaron una serie de transformaciones del etiquetado original devuelto por el sistema TERSEO. Dichas transformaciones se muestran en la siguiente subsección.

### 7.3.1 Transformación de las etiquetas TERSEO a las etiquetas TIMEX2

Debido a que el sistema TERSEO utiliza su propio esquema de anotación, específico para las aplicaciones en las que dicho sistema se emplea (Ordenación de eventos y Búsqueda de Respuestas Temporal), fue necesario introducir un módulo de transformación de las etiquetas generadas originalmente por dicho sistema a las etiquetas utilizadas en el esquema de anotación TIMEX2, con la finalidad de poder comparar los resultados del sistema TERSEO con otros sistemas.

#### Transformaciones para puntos en el tiempo

Cuando la expresión se refiere a una fecha concreta en el tiempo, la transformación es directa.

<sup>5</sup> [http://timex2.mitre.org/cgi-bin/download?file=score\\_timex2.pl](http://timex2.mitre.org/cgi-bin/download?file=score_timex2.pl)



Ejemplo: *I was sick yesterday.*

TERSEO:

```
I was sick <DATE_TIME_REF
type="C" valdate1="22/09/2004">
yesterday </DATE_TIME_REF>
```

TIMEX2:

```
I was sick <TIMEX2 VAL="2004-09-22">
yesterday</TIMEX2>
```

### Transformaciones para duraciones

En el esquema TIMEX2 se define una expresión de duración como aquella que indica un periodo de tiempo, informando de cuanto ha durado algo, por ejemplo, la expresión “*three hours long*”. Los formatos utilizados para expresar duraciones en el atributo VAL de la etiqueta TIMEX2 son los siguientes:

$$\begin{array}{c} PnYnMnDTnHnMnS \\ PnW \end{array}$$

El significado de cada elemento utilizado en el formato para expresar duraciones se muestra en el cuadro 7.4.

Además, los posibles valores del atributo ANCHOR\_DIR utilizados para expresar duraciones se muestran en el cuadro 7.5. El valor de este atributo va siempre relacionado con el valor del atributo ANCHOR\_VAL. Para realizar la transformación de los resultados obtenidos por TERSEO, el rango de fechas concretas resultante del sistema se transforma a una duración y los valores que tomarán cada uno de los atributos de la etiqueta TIMEX2 se obtienen usando las pautas que se muestran en el cuadro 7.6.

P	Elemento fijo que indica duración
n	Valor entero que indica el número de unidades existentes del elemento al que acompaña dependiendo de la duración
Y	Elemento usado si el valor de la duración se expresa en años
M	Elemento usado si el valor de la duración se expresa en meses
D	Elemento usado si el valor de la duración se expresa en días
W	Elemento usado si el valor de la duración se expresa en semanas
T	Elemento fijo que se añade al formato de la duración cuando la expresión también se refiere a una duración en horas
H	Elemento usado tras el elemento fijo T si el valor de la duración se expresa en horas
M	Elemento usado tras el elemento fijo T si el valor de la duración se expresa en minutos
S	Elemento usado tras el elemento fijo T si el valor de la duración se expresa en segundos

Cuadro 7.4. Significado de cada elemento del formato para duraciones

WITHIN	El contexto especifica que la fecha de referencia está dentro del periodo indicado por la duración
STARTING	El contexto especifica que el periodo indicado por la duración empieza en la fecha de referencia
ENDING	El contexto especifica que el periodo indicado por la duración termina en la fecha de referencia
AS_OF	Se utiliza cuando el periodo de duración de la expresión es presente pero indefinido
BEFORE	El contexto no indica cuando la duración empezó o terminó, pero el tiempo verbal (pasado) indica que fue antes de la fecha de referencia
AFTER	El contexto no indica cuando la duración empezó o terminó, pero el tiempo verbal (futuro) indica que fue después de la fecha de referencia

Cuadro 7.5. Posibles valores del atributo ANCHOR\_DIR

Ejemplo 1: *He will be in school the next three years.*

TERSEO:

```
<DATE_TIME_REF type="R"
valdate1="2004" valdate2="2007">
the next
three years </DATE_TIME_REF>
```

TIMEX2:

```
<TIMEX2 VAL="P3Y"
ANCHOR_DIR="STARTING"
```



VAL	PnYnMnDTnHnMnS (Se calcula la granularidad según el tipo de expresión y n será la diferencia entre valdate1 y valdate2)
ANCHOR_DIR	Si valdate1 y valdate2 son conocidos y (valdate1 >= Fecha Referencia >= valdate2) "WITHIN" Si sólo se conoce valdate1 y (valdate1 >= Fecha Referencia) "STARTING" Si sólo se conoce valdate2 y (valdate2 <= Fecha Referencia) "ENDING" En otro caso "BEFORE", TERSEO no utiliza información verbal y asume pasado.
ANCHOR_VAL	If (valdate1 >= Fecha Referencia) valdate1 If (valdate2 <= Fecha Referencia) valdate2

Cuadro 7.6. Transformación realizada para duraciones

```
ANCHOR_VAL="2004">
the next three years</TIMEX2>
```

Ejemplo 2: *He was in school the past four years.*

TERSEO:

```
<DATE_TIME_REF type="R"
valdate1="2000" valdate2="2004">
the past
four years</DATE_TIME_REF>
```

TIMEX2:

```
<TIMEX2 VAL="P4Y"
ANCHOR_DIR="ENDING"
ANCHOR_VAL="2004">the past
four years</TIMEX2>
```

### Transformaciones para expresiones difusas

En el esquema TIMEX2 se define una expresión como difusa cuando ésta se refiere a periodos de tiempo imprecisos, por tanto



los únicos valores posibles para estas expresiones indican si el atributo se refiere al presente, al pasado o al futuro.

Los elementos concretos utilizados por las etiquetas VAL y ANCHOR\_DIR para este tipo de expresiones difusas se muestran en el cuadro 7.7.

VAL	ANCHOR_DIR	EJEMPLO
PRESENT_REF	AS_OF	now, today, presently, nowadays,...
FUTURE_REF	AFTER	future, later, ahead, in a few years,...
PAST_REF	BEFORE	past, former, recently, long ago,...

**Cuadro 7.7.** Elementos utilizados para representar expresiones difusas

Las transformaciones realizadas a partir de los resultados obtenidos por el sistema TERSEO se muestran en el cuadro 7.8.

VAL	Si type="F" y (valdate1=Fecha Referencia=valdate2) "PRESENT_REF" Si type="F" y (valdate1<Fecha Referencia y valdate2=Fecha Referencia) "PAST_REF" Si type="F" y (valdate2>Fecha Referencia y valdate1=Fecha Referencia) "FUTURE_REF"
ANCHOR_DIR	Si type="F" y (valdate1=Fecha Referencia=valdate2) "AS_OF" Si type="F" y (valdate1<Fecha Referencia y valdate2=Fecha Referencia) "BEFORE" Si type="F" y (valdate2>Fecha Referencia y valdate1=Fecha Referencia) "AFTER"
ANCHOR_VAL	Fecha Referencia

**Cuadro 7.8.** Transformación para expresiones difusas

Ejemplo: *He was in school the last years.*



TERSEO:

```
<DATE_TIME_REF type="F" valdate1=">>>>1994"
valdate2="<<<<2004">
the last years</DATE_TIME_REF>
```

TIMEX2:

```
<TIMEX2 VAL="PAST_REF"
ANCHOR_DIR="BEFORE"
ANCHOR_VAL="2004">the
last years</TIMEX2>
```

### Transformaciones para expresiones modificadas

La semántica de una determinada expresión temporal puede verse modificada por adverbios o expresiones que la acompañan y para expresar esta modificación se añade el atributo MOD a la etiqueta TIMEX2. Los valores que puede tener este atributo se muestran en el cuadro 7.9.

	ELEMENTO	EJEMPLO
Puntos en el tiempo	BEFORE AFTER ON_OR_BEFORE ON_OR_AFTER	more than less than no less than no more than
Duraciones	LESS_THAN MORE_THAN EQUAL_OR_LESS EQUAL_OR_MORE	less than, nearly more than no more than at least
Puntos y duraciones	START MID END APPROX	start, beginning,... middle, mid-,... end, late,... about, around,...

Cuadro 7.9. Posibles valores atributo MOD

Para resolver este tipo de casos en el sistema TERSEO se definen un conjunto de palabras clave que denotan modificación en el valor que determina una expresión temporal como: more than, less than... y sus valores asociados para el atributo MOD. Si alguna de estas palabras clave son reconocidas acompañando a una



expresión, el atributo MOD tomará el valor adecuado dependiendo de la palabra clave que se trate.

### Transformaciones para expresiones que denotan conjuntos

Las expresiones que denotan conjuntos se refieren a cómo de frecuente ocurre o se realiza un determinado evento. Para marcar este tipo de expresiones con la etiqueta TIMEX2 se utiliza el atributo SET con valor "YES".

Para conseguir detectar las expresiones que denotan conjuntos, el sistema TERSEO analiza si en dicha expresión existe alguna palabra clave que implica periodicidad, como por ejemplo: cada, todos, mensualmente,... En ese caso el atributo SET se marca como verdadero.

#### 7.3.2 Resultados Evaluación TERN 2004

Previamente a la evaluación definitiva de la conferencia TERN 2004, el sistema fue probado para un conjunto de documentos de prueba facilitado por MITRE. En concreto, el sistema se probó con 53 documentos de la fuente *Broadcast News*. Los resultados obtenidos en la evaluación del corpus de entrenamiento se muestran en los cuadros 7.10 y 7.11 desglosados por atributos.

Etiqueta	POS	ACT	CORR	INCO	MISS	SPUR
TIMEX2	371	401	270	77	24	54
TIMEX2:ANCHOR_DIR	57	76	50	6	1	20
TIMEX2:ANCHOR_VAL	57	76	52	4	1	20
TIMEX2:MOD	2	6	0	0	2	6
TIMEX2:SET	6	5	4	0	2	1
TIMEX2:VAL	347	337	255	82	10	0

**Cuadro 7.10.** Resultados entrenamiento TERSEO desglosado por atributos (Broadcast News) (I)

Para este corpus los resultados de precisión, cobertura y F-Medida obtenidos son muy satisfactorios para todos los atributos



Etiqueta	PREC.	COB.	F-MEDIDA
TIMEX2	0.673	0.728	0.699
TIMEX2:ANCHOR_DIR	0.658	0.877	0.752
TIMEX2:ANCHOR_VAL	0.684	0.912	0.782
TIMEX2:MOD	0.000	0.000	0.000
TIMEX2:SET	0.800	0.667	0.727
TIMEX2:VAL	0.757	0.735	0.746

**Cuadro 7.11.** Resultados entrenamiento TERSEO desglosado por atributos (Broadcast News) (II)

excepto para el atributo MOD, principalmente porque esta característica no estaba contemplada inicialmente en el sistema TERSEO y la transformación realizada para contemplarla no siempre funciona correctamente puesto que muchas de las palabras claves necesarias para marcar los modificadores están fuera del ámbito de trabajo de TERSEO.

La colección de documentos sobre los que se ha realizado la evaluación definitiva procede de dos tipos de fuentes diferentes: *Broadcast News* y *Newswire News*. El número de documentos y de palabras para cada uno de estos dos conjuntos de textos se encuentran en la tabla 7.12. En concreto, el primer corpus contiene 986 expresiones temporales y el segundo 842, con lo que ambos corpus combinados contienen un total de 1828 expresiones temporales.

	Documentos	Palabras
<b>Broadcast News</b>	127	26.418
<b>Newswire</b>	65	28.196
<b>TOTAL</b>	192	54.614

**Cuadro 7.12.** Tamaño del corpus de test

La evaluación de los resultados obtenidos por los diferentes sistemas se ha realizado independientemente para cada atributo de la etiqueta. Los resultados obtenidos de las medidas presentadas en la sección 7.1 para cada atributo de la etiqueta TIMEX2 se muestran en los cuadros 7.13 y 7.14.

Etiqueta	POS	ACT	CORR	INCO	MISS	SPUR
TIMEX2	1828	1507	1437	0	391	70
TIMEX2:ANCHOR_DIR	302	209	171	19	112	19
TIMEX2:ANCHOR_VAL	302	209	147	43	112	19
TIMEX2:MOD	36	9	4	1	31	4
TIMEX2:SET	33	17	15	0	18	2
TIMEX2:TEXT	1828	1507	1036	401	391	70
TIMEX2:VAL	1390	1437	986	404	0	47

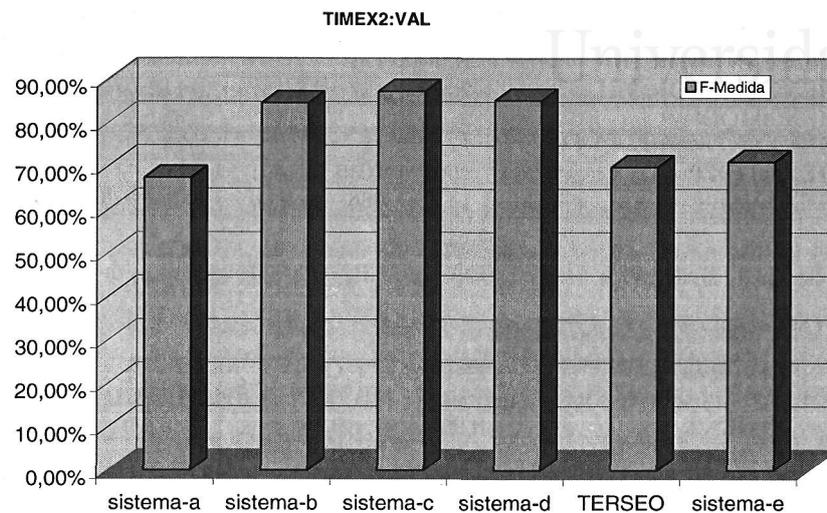
**Cuadro 7.13.** Resultados de TERSEO en TERN 2004 desglosado por atributos (I)

Etiqueta	PREC.	COB.	F-MEDIDA
TIMEX2	0.954	0.786	0.862
TIMEX2:ANCHOR_DIR	0.818	0.566	0.669
TIMEX2:ANCHOR_VAL	0.703	0.487	0.575
TIMEX2:MOD	0.444	0.111	0.178
TIMEX2:SET	0.882	0.455	0.600
TIMEX2:TEXT	0.687	0.567	0.621
TIMEX2:VAL	0.686	0.709	0.698

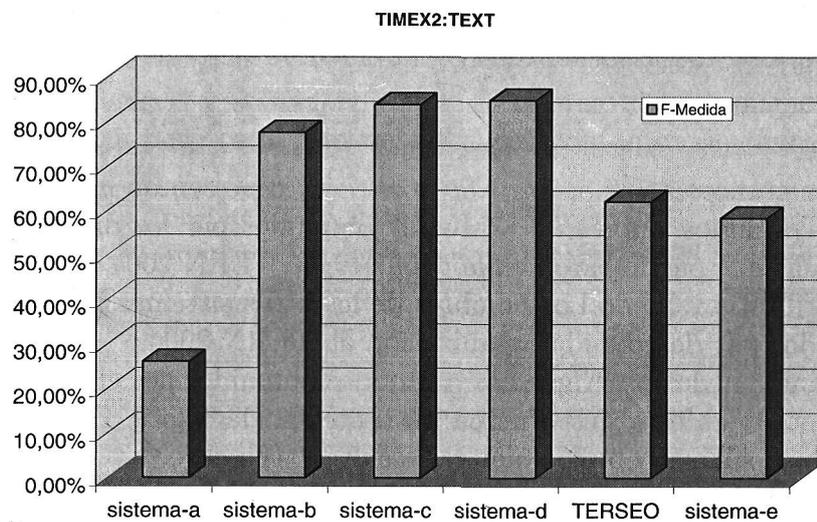
**Cuadro 7.14.** Resultados de TERSEO en TERN 2004 desglosado por atributos (II)

Los gráficos 7.2, 7.3, 7.4, 7.5, 7.6, 7.7 y 7.8 presentan una comparativa de los valores de F-Medida obtenidos por los diferentes sistemas que participaron en la competición TERN 2004 para cada atributo evaluado. Los nombres de los otros sistemas han sido omitidos por deseo de la organización de TERN 2004.

Como puede observarse, los resultados obtenidos por el sistema presentado en esta Tesis fueron bastante satisfactorios. Especialmente, el sistema obtiene mejores resultados en el reconocimiento de expresiones temporales y en la obtención del valor de una determinada expresión. Los resultados obtenidos en la evaluación definitiva están acordes a los obtenidos en el entrenamiento y no se han visto penalizados a pesar de que el número de expresiones aumenta considerablemente en el corpus definitivo. Es importante recordar que el sistema presentado fue el único cuyo conocimiento para el inglés había sido obtenido automáticamente.



**Figura 7.2.** Comparativa F-Medida de los diferentes sistemas para atributo VAL



**Figura 7.3.** Comparativa F-Medida de los diferentes sistemas para atributo TEXT

### 7.3.3 Análisis de errores

Se han analizado las posibles causas de los errores cometidos por parte del sistema TERSEO en la evaluación TERN 2004 para cada atributo.

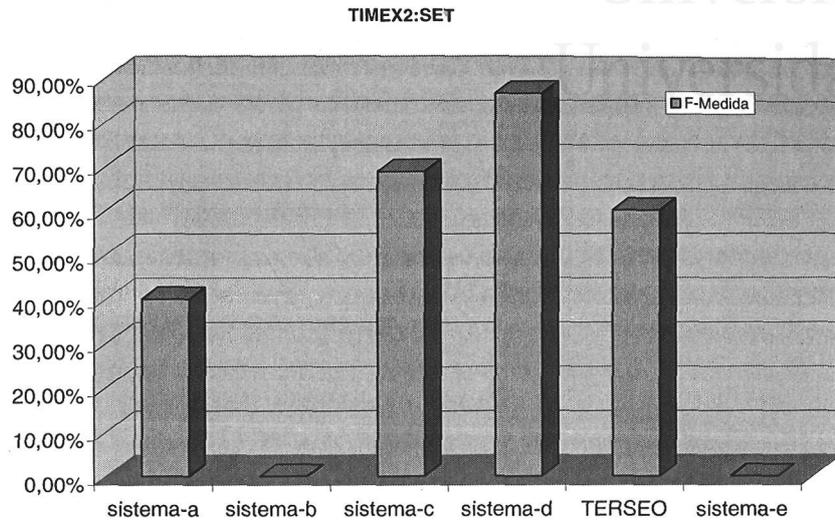


Figura 7.4. Comparativa F-Medida de los diferentes sistemas para atributo SET

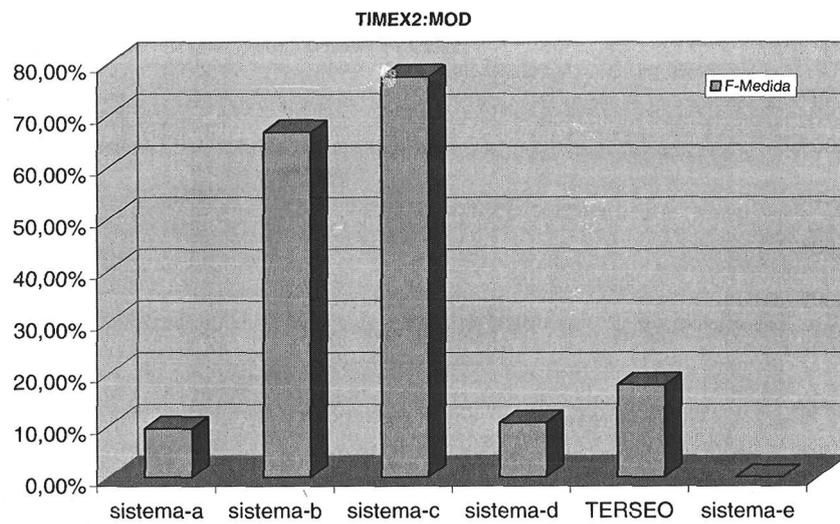
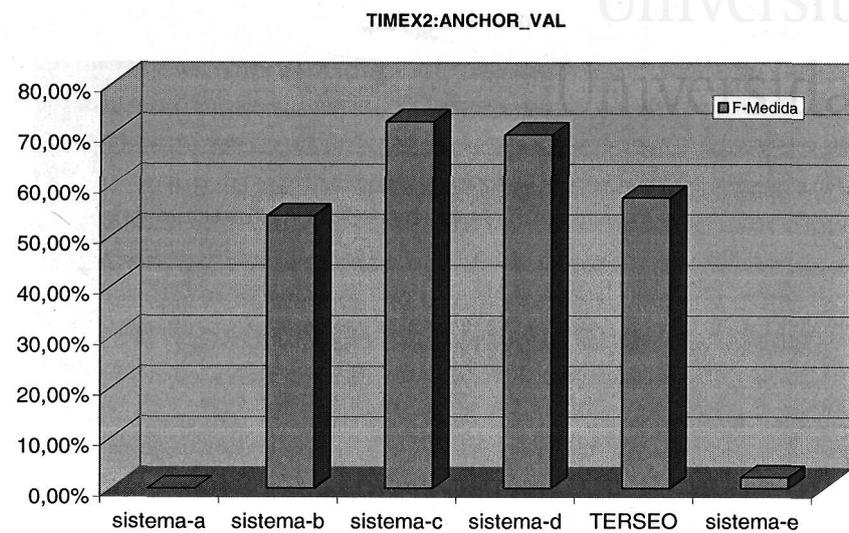
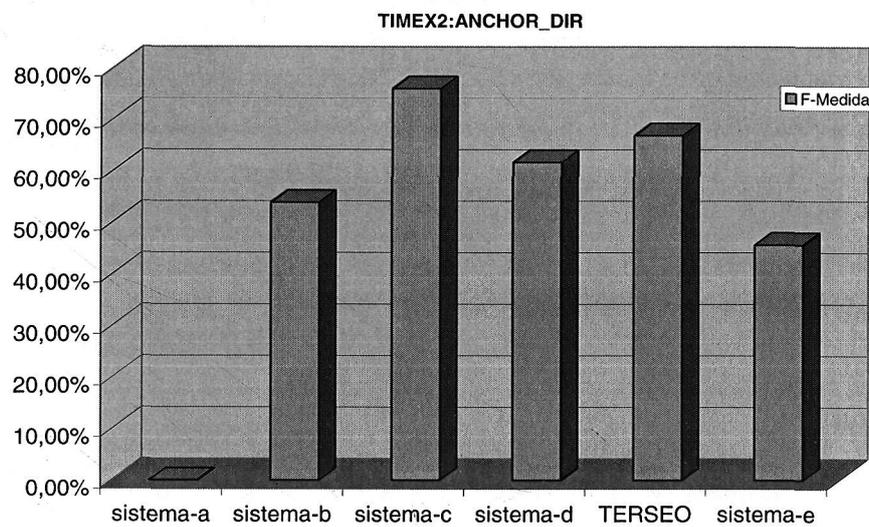


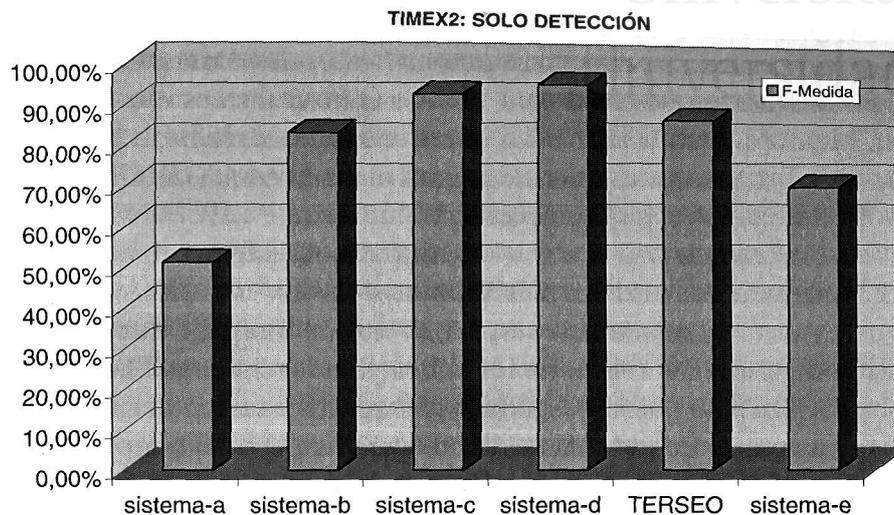
Figura 7.5. Comparativa F-Medida de los diferentes sistemas para atributo MOD



**Figura 7.6.** Comparativa F-Medida de los diferentes sistemas para atributo ANCHOR\_VAL



**Figura 7.7.** Comparativa F-Medida de los diferentes sistemas para atributo ANCHOR\_DIR



**Figura 7.8.** Comparativa F-Medida de los diferentes sistemas para sólo detección

### Errores en el atributo VAL

Los principales errores en este atributo se han producido al resolver expresiones difusas y también para determinadas expresiones no difusas, principalmente porque la interpretación que el sistema TERSEO hace del valor de dichas expresiones temporales es diferente a la que en el esquema TIDES está considerando como correcto. La solución para mejorar los resultados en este sentido es modificar las reglas de resolución de dichas expresiones coincidan con las que el esquema TIMEX2 está esperando.

### Errores en los atributos ANCHOR\_DIR y ANCHOR\_VAL

Los errores más destacados en el valor de estos dos atributos se deben a que el sistema TERSEO no está usando la información que le aporta el contexto en el cual se encuentra una determinada expresión temporal, como podría ser el tiempo verbal del evento con el que se encuentra relacionada dicha expresión. En concreto, cuando los posibles valores que puede tomar el atributo



ANCHOR\_DIR están entre BEFORE o AFTER, es necesario interpretar el tiempo verbal para poder llegar al valor correcto del atributo. Puesto que el sistema TERSEO no utiliza esta información, siempre asume BEFORE como valor de ANCHOR\_DIR y por tanto los resultados obtenidos para el atributo ANCHOR\_DIR serán en algunos casos incorrectos.

### Errores en el atributo MOD

Debido a que las reglas de identificación del sistema TERSEO en inglés han sido obtenidas automáticamente y teniendo en cuenta que las reglas de identificación en castellano no contemplan los modificadores como elementos independientes de las expresiones temporales, al realizar una traducción automática de las expresiones en castellano al inglés, modificadores como por ejemplo: “*up to*”, “*late*”,... no aparecen contemplados en las nuevas reglas puesto que tampoco aparecían para el castellano. Por tanto, estos elementos no pueden ser reconocidos actualmente por el sistema TERSEO. La solución a este problema radica en añadir estos términos a los contemplados actualmente por el sistema, mejorando de esta forma los resultados de precisión y cobertura para este atributo.

### Errores en el atributo TEXT

La mayor parte de los errores de reconocimiento de este atributo se deben a la extensión de la expresión temporal que se está reconociendo. El sistema TERSEO reconoce sintagmas preposicionales como posibles expresiones temporales, sin embargo, el esquema TIMEX2 no tiene en cuenta las preposiciones como parte de una expresión temporal. Por ejemplo, si en un texto aparece la expresión “*in two hours*”, para TERSEO la expresión completa será marcada como temporal mientras que la etiqueta TIMEX2 sólo marcará “*two hours*”. Resolver este problema únicamente implicaría cambiar la filosofía de TERSEO, de forma que las expresiones temporales contempladas no contengan preposiciones, aunque para resolver determinadas expresiones, las prepo-

siciones pueden proveer información muy útil, como la dirección de un determinado rango temporal en expresiones como “*before Christmas*”.

Otro problema causante de errores en este atributo es que TERSEO no está reconociendo expresiones como: “*24-hours*” o “*13-year-old*”. Es importante recordar que las expresiones temporales y las reglas de identificación que TERSEO está usando para el inglés han sido obtenidas automáticamente a partir de las expresiones temporales guardadas para el castellano. Es por ello, que, puesto que estas expresiones no existen en castellano con este formato tampoco se consiguieron en la extensión automática realizada para el inglés y por tanto, el sistema TERSEO, no es capaz de reconocerlas. Por ello, sería necesario realizar una revisión manual de dichas expresiones temporales y reglas de identificación para el inglés con el objetivo de añadir expresiones específicas para este idioma que la extensión de sistema no fue capaz de obtener automáticamente.

### **Errores en el atributo SET**

Los errores principales para este atributo se deben a que algunas palabras clave que denotan conjuntos de fechas no están siendo tratadas por TERSEO y por tanto, existen expresiones que no están siendo consideradas como conjuntos por nuestro sistema. La solución a este problema radica en añadir esta información.

### **Conclusión**

Los resultados de la evaluación son bastante altos teniendo en cuenta que las expresiones temporales y las reglas de identificación del sistema TERSEO en inglés han sido obtenidas automáticamente a partir de las expresiones temporales y las reglas de identificación originales del sistema, que era en castellano. Se observa, después de analizar los errores más destacados que sería necesario realizar un refinamiento de las expresiones temporales y las reglas de identificación para el inglés manualmente, añadiendo



esas peculiaridades del idioma que la extensión automática no ha tenido en cuenta.

### 7.3.4 Evaluación alternativa

Posteriormente a la conferencia TERN 2004 se decidió realizar una evaluación alternativa a la presentada en dicha conferencia. A grandes rasgos, la modificación consistía en no tener en cuenta las cabeceras y los pies de los documentos del corpus, en los cuales se almacena la fecha y hora del artículo explícitamente.

Los documentos del corpus TERN 2004 tienen el siguiente formato:

```
<DOC>
<DOCNO> ABC20010105.1830.1175 </DOCNO>
<DOCTYPE SOURCE="broadcast news"> NEWS STORY </DOCTYPE>
<DATE_TIME> 01/05/2001 18:49:35.55 </DATE_TIME>
<BODY>
<TEXT> ..... </TEXT>
</BODY>
<END_TIME> 01/05/2001 18:52:24.46 </END_TIME>
</DOC>
```

Debido a que las etiquetas `<DATE_TIME>` y `<END_TIME>` contiene la fecha y hora del documento explícitamente, los sistemas aumentan los resultados correctos gracias a estas expresiones. El motivo de la nueva evaluación es obtener los resultados que ofrecen los sistemas si únicamente se tiene en cuenta el contenido de las etiquetas `<TEXT>`.

Los resultados obtenidos por nuestro sistema al evaluarlo de esta forma se presentan en los cuadros 7.15 y 7.16.

Como puede observarse en los resultados de esta nueva evaluación el número de referencias tratadas y existentes es menor, puesto que ya no se tienen en cuenta las fechas explícitas de las

Etiqueta	POS	ACT	CORR	INCO	MISS	SPUR
TIMEX2	1427	1179	1111	0	316	68
TIMEX2:ANCHOR_DIR	302	209	171	19	112	19
TIMEX2:ANCHOR_VAL	302	209	147	43	112	19
TIMEX2:MOD	36	9	4	1	31	4
TIMEX2:SET	33	16	15	0	18	1
TIMEX2:TEXT	1427	1179	717	394	316	68
TIMEX2:VAL	1064	1111	699	365	0	47

**Cuadro 7.15.** Resultados de TERSEO en evaluación alternativa TERN 2004 desglosado por atributos (I)

Etiqueta	PREC.	COB.	F-MEDIDA
TIMEX2	0.942	0.779	0.853
TIMEX2:ANCHOR_DIR	0.818	0.566	0.669
TIMEX2:ANCHOR_VAL	0.703	0.487	0.575
TIMEX2:MOD	0.444	0.111	0.178
TIMEX2:SET	0.938	0.455	0.612
TIMEX2:TEXT	0.608	0.502	0.550
TIMEX2:VAL	0.629	0.657	0.643

**Cuadro 7.16.** Resultados de TERSEO en evaluación alternativa TERN 2004 desglosado por atributos (II)

cabeceras y los pies de los documentos. Estas fechas eran referencias que por el hecho de ser explícitas siempre se reconocían y resolvían correctamente. Además, para una fecha explícita, la etiqueta TIMEX2 asociada únicamente tiene valores en los atributos VAL y TEXT. Es por ello que, ahora, al no tener en cuenta estas expresiones temporales explícitas en la evaluación, los valores de los atributos TEXT y VAL se ven decrementados puesto que se han reducido el número de valores correctos para estos dos atributos. Por tanto, como resultado final, se ve reducida la F-Medida del reconocimiento de las expresiones y de los atributos TEXT y VAL. Aún así, la reducción es bastante pequeña, lo que indica que el sistema no está tratando esta información explícita de manera especial, sino que la trata como una fecha absoluta cualquiera y además está resolviendo eficientemente muchas expresiones implícitas, obteniendo buenos resultados para las mismas. Como conclusión, los resultados del sistema no se ven muy alterados por esta nueva evaluación, cosa que no ocurre igual en sistemas que



sí tratan de manera especial estas fechas absolutas, de forma que al eliminarlas de la evaluación esta ya no es tan buena. La comparación en este caso con otros sistemas no ha sido posible puesto que esta evaluación surgió posteriormente a la conferencia TERN 2004 y los resultados no han sido hechos públicos para el resto de sistemas.

## 7.4 Evaluación del sistema de Búsqueda de Respuestas Temporal en inglés

La evaluación del sistema de Búsqueda de Respuestas Temporal en inglés se ha realizado en dos pasos:

- Evaluación de la descomposición de la pregunta compleja en preguntas simples (Saquete *et al.*, 2004h).
- Comparativa de resultados entre un sistema general de Búsqueda de Respuestas y el sistema de Búsqueda de Respuestas Temporal implementado (Saquete *et al.*, 2004g) (Saquete *et al.*, 2004b).

A continuación se detallan cada una de estas evaluaciones más en profundidad.

### 7.4.1 Evaluación de la descomposición de la pregunta

Actualmente, una de las mejores formas de evaluar los sistemas de Búsqueda de Respuestas existentes es mediante la participación en foros internacionales como TREC (2002) o CLEF (2004). Sin embargo, después de la realización de un estudio de los corpus para el inglés que se utilizan en estos foros, se observa que las preguntas utilizadas en los mismos no tienen las características complejas que se necesitan para evaluar el sistema presentado en este trabajo. Por ello, para la evaluación del sistema presentado se ha utilizado un corpus de preguntas temporales en inglés usado en la conferencia TERQAS (Pustejovsky, 2002). Este corpus está formado por un conjunto de 124 preguntas relacionadas con la temporalidad, aunque 12 de estas preguntas se encuentran fuera



del ámbito de este trabajo. Por ello, la evaluación de la eficiencia del sistema en la descomposición de preguntas se ha realizado sobre las 112 preguntas restantes.

Esta evaluación ha sido realizada manualmente por tres anotadores y se han considerado importantes cuatro aspectos:

- **Reconocimiento y resolución de las expresiones temporales:** En este corpus hay 62 expresiones temporales y el sistema es capaz de reconocer 52 de las que 47 fueron correctamente resueltas por la unidad de reconocimiento y resolución de expresiones temporales.
- **Identificación de tipo:** Las 112 preguntas del corpus fueron tratadas por el sistema y de ellas 104 fueron correctamente identificadas según la taxonomía de preguntas propuesta en el capítulo 6.
- **Detección de la señal:** La detección de la señal sólo será realizada cuando la pregunta se considera de tipo complejo (Tipo 3 y Tipo 4). En este corpus en concreto, hay 17 preguntas que se consideran complejas y por tanto, contienen una señal temporal. El sistema propuesto fue capaz de tratar y reconocer correctamente las señales de 14 de estas preguntas.
- **División de la pregunta:** De este conjunto de 17 preguntas complejas, el sistema fue capaz de reconocer 14 y dividir correctamente 12 de ellas.

Los resultados obtenidos en cuanto a precisión, cobertura y fiabilidad teniendo en cuenta estos cuatro aspectos se muestran en el cuadro 7.17.

Observando los resultados, 19 preguntas fueron preprocesadas de manera incorrecta por el sistema. Los errores ocurridos que derivaron en este preproceso incorrecto se exponen a continuación:

### **Errores en Identificación de Tipo Pregunta**

Se producen 8 errores en la identificación del tipo de la pregunta y estos errores se deben a:

- Expresiones temporales no tratadas o reconocidas de manera incorrecta: 6 preguntas.



	POS	ACT	CORR	PREC.	COB.	F-MEDIDA
<b>Reconocimiento y Resolución ET</b>	62	52	47	90 %	75 %	86 %
<b>Identificación Tipo</b>	112	112	104	92 %	100 %	93 %
<b>Detección Señal</b>	17	14	14	100 %	82 %	95 %
<b>División Pre-gunta</b>	17	14	12	85 %	71 %	81 %
<b>UNIDAD DESCOMPOSICIÓN</b>	112	112	93	83 %	83 %	83 %

Cuadro 7.17. Evaluación de la descomposición de la pregunta en inglés

- Señales temporales detectadas de manera incorrecta: 2 preguntas.

Puesto que el sistema TERSEO es el que se encarga de reconocer las expresiones temporales y las señales temporales, será necesario introducir en el sistema estas expresiones y señales temporales no tratadas para solucionar los problemas que se producen actualmente en la identificación del tipo de la pregunta.

### Errores en la División de la Pregunta

Hay 5 errores en el módulo de la división de la pregunta debidos a:

- Señales temporales detectadas de manera incorrecta: 3 preguntas.
- Problemas en el análisis sintáctico: 2 preguntas.

A lo hora de dividir la pregunta es necesario, por un lado, reconocer correctamente la señal temporal, puesto que esta señal va a ser la que determine por donde se va a dividir la pregunta. Es por ello, que sería necesario contemplar el mayor número de señales temporales posible. Actualmente, el sistema TERSEO reconoce un conjunto reducido de señales temporales que aparecen



más comúnmente en los corpus, pero este conjunto sería fácilmente ampliable para poder reconocer una mayor cantidad de ellas. Por otro lado, también es necesario contemplar todas las posibilidades sintácticas en cuanto a la transformación de la pregunta compleja original en preguntas simples. Actualmente se contemplan los casos más generales de transformación, y sería necesario estudiar y añadir al sistema casos menos frecuentes.

### Otros errores

Hay 15 errores que no afectan a la división de la pregunta ni al tratamiento de dichas preguntas simples por parte del sistema general de Búsqueda de Respuestas, pero que sí que van a afectar a la recomposición de las respuestas de dichas preguntas. Estos errores se deben a:

- Expresiones temporales no tratadas o reconocidas de manera incorrecta: 6 preguntas.
- Expresiones temporales resueltas de manera incorrecta: 9 preguntas.

Estos errores están relacionados con la resolución que el sistema TERSEO realiza para las expresiones temporales de las preguntas. Es por ello que sería necesario refinar las reglas de resolución del sistema para aquellos casos en los que el sistema está funcionando de manera errónea actualmente.

En algunas de las preguntas se haya presente más de uno de estos problemas, por ejemplo si el tipo se identifica de manera incorrecta implica que la división de la pregunta también será incorrecta.

#### 7.4.2 Comparativa del sistema de BR Temporal con un sistema general de BR

Para realizar la evaluación del sistema completo de Búsqueda de Respuestas Temporal se ha utilizado la descomposición de preguntas realizada anteriormente. Se incluye también para la evaluación completa del sistema las 12 preguntas que están fuera del



ámbito del sistema TERSEO, para obtener una evaluación del corpus completo. Estas preguntas están dentro del grupo de preguntas de Tipo 1, para las cuales el sistema no realiza ningún tipo de tratamiento.

Los resultados obtenidos después de procesar las preguntas por el sistema de Búsqueda de Respuestas Temporal y un sistema de Búsqueda de Respuestas general se clasifican en tres grandes grupos:

1. Los resultados son los mismos para ambos sistemas.
2. Los resultados del sistema de Búsqueda de Respuestas Temporal son mejores que los obtenidos por el sistema general.
3. El sistema de Búsqueda de Respuestas Temporal obtiene peores resultados que el sistema general.

A continuación se detallan los resultados concretos para cada uno de estos grupos en base a la evaluación realizada.

1. Los resultados obtenidos son los mismos para ambos sistemas. Esto se debe a dos posibles razones:
  - El sistema de BR general no devuelve ninguna respuesta para las preguntas simples y por tanto, el sistema de BR Temporal, al basarse en los resultados del primero, tampoco devuelve nada. Existen 47 preguntas en las que se cumple esta situación. Las preguntas más afectadas en este caso son preguntas de Tipo 1 y 2.
  - El sistema de Búsqueda de Respuestas Temporal devuelve las mismas respuestas que el sistema de Búsqueda de Respuestas general. Esto ocurre siempre con las preguntas de Tipo 1 puesto que nuestro sistema no realiza en este caso ningún tipo de preproceso. Hay 34 preguntas de este tipo en el corpus.
2. El sistema de Búsqueda de Respuestas Temporal obtiene mejores resultado que el sistema general. En este caso se pueden dar cuatro posibles situaciones:
  - El sistema de BR Temporal no devuelve ninguna respuesta a pesar de que el sistema general si devolvió un conjunto de respuestas. Pero, sin embargo, ninguna de esas respuestas



## 7.4 Evaluación del sistema de Búsqueda de Respuestas Temporal en inglés 149

cumplía las restricciones temporales impuestas por la pregunta y por tanto fueron descartadas por el sistema de BR Temporal. En esta situación nuestro sistema actúa satisfactoriamente puesto que ninguna de las respuestas obtenidas por el sistema general eran correctas. Existen 12 preguntas de este tipo, de las cuales 11 se corresponden con preguntas de Tipo 2.

- El sistema de BR general no devuelve ninguna respuesta a una pregunta compleja, y sin embargo, gracias a la división realizada por el sistema de BR Temporal y la posterior recomposición de las respuestas a las subpreguntas, se consigue obtener respuesta para la pregunta compleja. En este corpus en concreto sólo tenemos una pregunta que cumpla esta condición.
  - El sistema de BR general sí que devuelve respuestas pero estas son incorrectas, sin embargo, el sistema de BR Temporal, gracias a la extracción y resolución de la información temporal de las preguntas, es capaz de obtener un conjunto de respuestas correctas para dichas preguntas. En este corpus de ejemplo, existe únicamente una pregunta con estas características.
  - El sistema de BR Temporal refina el conjunto de respuestas obtenidas por el sistema de BR genérico, eliminando aquellas que no cumplen las restricciones temporales. En este caso se considera por tanto que el sistema de BR Temporal está contestando mejor que el sistema general, puesto que aquellas respuestas que no cumplan las restricciones temporales no son respuestas válidas. En el corpus hay 28 preguntas de este tipo y se corresponden con preguntas de Tipo 2 y de Tipo 4, puesto que estos dos tipos son preguntas que contienen alguna expresión temporal (implícita o explícita) en su formulación.
3. El sistema de Búsqueda de Respuestas Temporal obtiene peores resultados que el sistema general.
- El sistema de BR general es capaz de responder y sin embargo el sistema de BR Temporal no. Este caso se debe a que cuando se está trabajando con preguntas complejas, al



dividir la pregunta en dos preguntas simples, pueden existir ciertas palabras clave en la segunda pregunta que ya no serán utilizadas para contestar la primera. Sin embargo, en algunos casos, el uso de estas palabras clave también en la primera pregunta podría ayudar a contestar dicha pregunta. Por ejemplo, en la pregunta: “¿Quién fue el secretario de Defensa durante la guerra del Golfo?”, un sistema de Búsqueda de Respuestas general que funcionara con palabras clave podría buscar “Secretario de Defensa” y “Guerra del Golfo” para intentar contestar a la pregunta. Sin embargo, al usar el sistema de BR Temporal, esta pregunta sería dividida en las dos preguntas simples siguientes: “¿Quién fue el secretario de Defensa?” y “¿Cuándo fue la guerra del Golfo?”. Cuando se intenta contestar a la primera pregunta simple con un sistema de BR general no se obtienen las respuestas adecuadas porque no puede utilizar información que necesita, que en este caso serían las palabras clave “Guerra del Golfo”.

Los resultados de este estudio se muestran en el Cuadro 7.18.

Preg	TOTAL	MEJOR	IGUAL	PEOR	MEJOR (%)	PEOR (%)
Tipo 1	47	-	47	-	-	-
Tipo 2	59	36	23	-	61 %	-
Tipo 3	3	-	3	-	-	-
Tipo 4	15	6	8	1	40 %	6 %
Todas	124	42	81	1	34 %	0.8 %

Cuadro 7.18. Evaluación del sistema de BR Temporal completo

Como se puede observar en el cuadro de evaluación, las preguntas de Tipo 2 y de Tipo 4 son las que se ven claramente mejoradas utilizando el sistema de Búsqueda de Respuestas Temporal frente a un sistema de Búsqueda de Respuesta general. El 61 % de las preguntas de Tipo 2 obtienen mejores respuestas usando el sistema aquí propuesto, y lo mismo ocurre con el 40 % de las preguntas de Tipo 4. Un pequeño porcentaje de las preguntas de Tipo



4 (6 %) ha obtenido peores respuestas como resultado del procesamiento realizado por el sistema temporal. Como conclusión es importante destacar que el sistema propuesto en este trabajo, teniendo en cuenta todas las preguntas, independientemente de su tipo, mejora un sistema de BR general en un 34 % de las preguntas y funciona peor tan sólo en un 1 %.

## 7.5 Evaluación CLEF2004

El sistema de Búsqueda de Respuestas Temporal en castellano ha sido evaluado en la tarea piloto en castellano de la conferencia CLEF 2004 (Saquete *et al.*, 2004d) (Herrera *et al.*, 2004). La tarea consistía en contestar 100 preguntas distribuidas de manera equitativa entre los siguientes tipos:

### ■ Preguntas factuales con restricciones temporales:

- Restricción de fecha: Una fecha concreta es la referencia temporal para contextualizar el objeto de la pregunta. La pregunta podrá circunscribirse al momento previo a dicha fecha, al mismo momento o al inmediatamente posterior. Por ejemplo: “¿Quién gobernaba en Francia en 1988?” o “¿Qué país visitó Berlusconi antes de junio de 1994?”. Este tipo de preguntas han sido clasificadas como preguntas de Tipo 2 por el sistema de BR Temporal.
- Restricción de periodo: En este caso es un período el que marca la referencia temporal. Por ejemplo: “¿Quién gobernaba en Irak entre 1985 y 1987?” o “¿Cuántos coches se vendieron en España en la década de los ochenta?”. Este tipo de preguntas también son consideradas de Tipo 2, aunque el sistema TERSEO que resuelve la expresión devuelve un rango de fechas para ella.
- Restricción de evento (pregunta anidada): Se trata de una restricción temporal en la que el delimitador temporal viene determinado por el momento en que ocurrió un segundo evento. Por ejemplo: “¿Quién gobernaba en Argentina durante la guerra de las Malvinas?” o “¿Qué le ocurrió al precio del crudo tras la invasión iraquí de Kuwait?”. Estas preguntas son



reconocidas por el sistema presentado aquí como preguntas complejas Tipo 3 y 4.

- Listas o enumeraciones:** Serán preguntas cuya respuesta es una lista de entidades, personas, organizaciones, etc. La respuesta será correcta de acuerdo con el grado de precisión y cobertura de la lista ofrecida con respecto a la lista esperada. No se limitará el número de respuestas. Por ejemplo: “Enumere los países que pertenecen a la UE.” o “¿Quién ha presidido el gobierno español desde 1992?”. Debido a que el sistema presentado en esta Tesis está enfocado a la resolución de expresiones temporales, muchas preguntas tipo lista no podrán ser resueltas. Sin embargo, teniendo en cuenta que la arquitectura en la que se basa el sistema es una arquitectura multicapa, es posible introducir una nueva capa capaz de tratar las preguntas de tipo lista para poder responderlas.

Los resultados obtenidos por el sistema presentado en esta Tesis en la conferencia CLEF 2004 se muestran en el Cuadro 7.19.

tipo pregunta	# preguntas	# respuestas dadas	# preguntas con respuesta correcta
Definiciones	2	2	0 (0%)
Factuales	18	42	4 (22.22%)
Listas	20	55	4 (20%)
Restricción Fecha	20	30	2 (10%)
Restricción Evento	20	42	2 (10%)
Restricción Periodo	20	29	3 (15%)
<b>TOTAL</b>	<b>100</b>	<b>200</b>	<b>15 (15%)</b>

**Cuadro 7.19.** Resultados Tarea Piloto CLEF 2004

Como puede observarse en el cuadro de evaluación, dado un conjunto de 100 preguntas temporales, se ha obtenido un total del 15% de preguntas con respuesta correcta utilizando el sistema de Búsqueda de Respuestas Temporal en castellano propuesto en este trabajo. Los resultados no son demasiado altos debido a que la arquitectura propuesta aquí requiere de un sistema de Búsqueda de Respuestas general para resolver las preguntas simples y por tanto los resultados globales del sistema son muy dependientes de



las características del sistema general de Búsqueda de Respuestas que se utilice.

## 7.6 Enriquecimiento del sistema TERSEO

Uno de los problemas que el sistema TERSEO evidencia en las evaluaciones anteriores fue el hecho de no utilizar la información del tiempo verbal para resolver aquellas expresiones temporales que resultaban ambiguas y sólo podían ser resueltas si se hacía uso de esta propiedad.

Con el fin de ver la mejora que aportaría a los resultados la utilización del tiempo verbal de la oración en la resolución de estas expresiones temporales, se ha introducido esta información como un factor más a tener en cuenta por el sistema para la resolución.

Puesto que se realiza un análisis morfológico inicial del texto, tanto el verbo como el tiempo verbal serán reconocidos en este paso del procesamiento y almacenados para su posterior uso. Una vez que la expresión temporal ha sido reconocida y previamente a proceder a su resolución, se comprobará si esta expresión es ambigua y dependiente del tiempo verbal, en cuyo caso será analizado, determinando así la correcta resolución de la expresión.

La evaluación de esta mejora del sistema se ha realizado sobre el mismo corpus de test que se utilizó para evaluar el sistema TERSEO en castellano y los resultados se muestran en el cuadro 7.20.

Después de realizar el enriquecimiento del sistema añadiendo la información del tiempo verbal, en concreto para este corpus, hay un total de 22 expresiones temporales que dependen del tiempo del verbo, de las que únicamente 2 deberán ser interpretadas como expresiones referidas al futuro. Las otras 20 se refieren a pasado y por tanto, puesto que el sistema estaba suponiendo el tiempo verbal como pasado por defecto, al realizar la mejora en la aplicación estas expresiones se aprecia únicamente un incremento del 2% en precisión.



	TEST ENRIQ.	TEST SIN EN-RIQ.
<b>DOCUMENTOS</b>	50	50
<b>POS</b>	199	199
<b>ACT</b>	156	156
<b>CORR</b>	140	138
<b>INCO</b>	16	18
<b>MISS</b>	43	43
<b>SPUR</b>	-	-
<b>PREC.</b>	90 %	88 %
<b>COB.</b>	70 %	69 %
<b>F-MEDIDA</b>	78 %	77 %

Cuadro 7.20. Evaluación del sistema TERSEO monolingüe enriquecido

## 7.7 Conclusiones

En este trabajo se presentan un conjunto de evaluaciones para medir la eficiencia de los diferentes aspectos del sistema TERSEO y sus aplicaciones.

Una primera evaluación se realiza sobre el sistema TERSEO de reconocimiento y resolución de expresiones temporales trabajando en castellano y obteniendo unos resultados de F-Medida del 77 %. Posteriormente, una vez que el sistema TERSEO fue extendido automáticamente al inglés se realizó una evaluación del sistema obteniendo una F-Medida del 69 % (Cuadro 7.21).

	MONOLINGÜE	MULTILINGÜE
<b>DOCUMENTOS</b>	50	100
<b>POS</b>	199	634
<b>ACT</b>	156	511
<b>CORR</b>	138	393
<b>INCO</b>	18	118
<b>MISS</b>	43	123
<b>PREC.</b>	88 %	77 %
<b>COB.</b>	69 %	62 %
<b>F-MEDIDA</b>	77 %	69 %

Cuadro 7.21. Comparación del sistema TERSEO monolingüe y multilingüe

A su vez, el sistema TERSEO en inglés también participó en la evaluación TERN 2004, dedicada a evaluar sistemas que tra-



tan información temporal en textos, obteniendo unos resultados muy competitivos con respecto al resto de sistemas presentados en dicha conferencia, especialmente en el reconocimiento de expresiones donde se obtuvo una precisión del 95 %, una cobertura del 78 % y una F-Medida del 86 %.

Además, el sistema TERSEO sirvió de base para el desarrollo de un sistema de Búsqueda de Respuestas Temporal Multicapa, que fue evaluado para el inglés en comparación con un sistema general de Búsqueda de Respuestas. El resultado fue que el sistema implementado en este trabajo mejoraba a un sistema de Búsqueda de Respuestas general en un 34 % y únicamente se comportaba peor en un 1 % de los casos. La evaluación del sistema de Búsqueda de Respuestas Temporal en castellano se realizó mediante la participación en CLEF2004. El sistema fue capaz de resolver satisfactoriamente un 15 % de las preguntas relacionadas con temporalidad del corpus utilizado en esta conferencia. Debido a que esta propuesta fue la única que participó en la conferencia para esta tarea, no puede llevarse a cabo una comparación del sistema con otros. Los resultados más altos fueron para las preguntas factuales, obteniendo un 22.2 % de preguntas con respuestas correctas. Sin embargo, en la edición de CLEF 2003, los resultados obtenidos por el sistema general de Búsqueda de Respuestas fueron algo mejores ( 40 % preguntas con respuesta correcta). Este resultado indica claramente que las preguntas propuestas para la tarea piloto de CLEF 2004 eran mucho más complejas que las de la edición anterior.

Como trabajo futuro, sería necesario refinar el sistema para que contemple aquellas expresiones y señales temporales que actualmente están fuera de su ámbito para mejorar los resultados de precisión y cobertura tanto para el castellano como para el inglés. Finalmente, sería conveniente la generación de un corpus de preguntas complejas temporales tanto para el inglés como para el castellano con el fin de evaluar en profundidad el sistema de Búsqueda de Respuestas Temporal con arquitectura multicapa propuesto en este trabajo.



Universitat d'Alacant  
Universidad de Alicante



Universitat d'Alacant  
Universidad de Alicante

## 8. Herramientas

En este capítulo se presentan las diferentes herramientas desarrolladas como consecuencia del trabajo presentado en esta Tesis. Las herramientas se han desarrollado sobre una plataforma web para un fácil acceso de los usuarios independientemente de su localización física. En concreto, se ha usado el lenguaje de programación *php* para la implementación de dichas herramientas. La primera herramienta desarrollada fue un analizador temporal capaz de realizar análisis sintácticos parciales indicándole en cada caso la gramática que se desea utilizar. Posteriormente, se desarrolló el sistema TERSEO, que permite reconocer y resolver las expresiones temporales de un texto digital, asignando a cada una de estas expresiones una etiqueta que indica que la expresión es una expresión temporal y almacenando en los atributos de las etiquetas la fecha o fechas concretas a las que se refiere la expresión, así como la información del tipo de expresión temporal de la que se trata. El sistema TERSEO utiliza a su vez el analizador morfológico sintáctico usando una gramática temporal definida especialmente para este sistema. Finalmente, el último paso del procesamiento del sistema TERSEO es la ordenación de los eventos y ésta se realiza utilizando la información de las fechas concretas a las que se refiere la expresión.

Por último, se presenta el sistema de Búsqueda de Respuestas Temporal realizado, que se divide en dos partes:

- Descomposición de las preguntas complejas en preguntas simples
- Recomposición de las respuestas de las preguntas simples para conseguir la respuesta final de la pregunta compleja



Cada una de estas herramientas se detalla en profundidad en las secciones posteriores.

## 8.1 Analizador temporal

La primera herramienta desarrollada fue el analizador sintáctico de expresiones temporales<sup>1</sup>. Se decidió realizar un analizador propio puesto que únicamente se necesitaba realizar un análisis parcial de los textos que nos permitiera obtener las expresiones temporales, descartando así el resto de información. Además se deseaba poder indicar la gramática temporal a utilizar para cada análisis. Estas gramáticas se encuentran almacenadas en una base de datos, de forma que el acceso por parte de la aplicación a las mismas es muy rápido. El analizador es un analizador sintáctico ascendente (chart) programado en php que utilizando la gramática como recurso es capaz de detectar las unidades consideradas como referencias temporales.

En la Figura 8.1 podemos ver la pantalla principal de la aplicación.

En la aplicación hay que especificar la siguiente información:

- Idioma del texto: Es necesario especificar el idioma del texto que se desea analizar.
- Tipo de Gramática: El analizador permite elegir la gramática que se desea utilizar para el análisis. Esta gramática fue definida previamente y almacenada en la base de datos de la aplicación. Las gramáticas definidas pueden verse en el Anexo B para el castellano y en el Anexo C para el inglés.
- Opción de procesamiento: El análisis que puede realizar la aplicación puede ser sólo morfológico, sólo sintáctico o ambos.
- Por último, se indica el texto que se desea analizar o el fichero en el cual se encuentra ese texto.

Después de realizar el procesamiento de la frase indicada en la figura: “Ayer fue martes”, los resultados obtenidos por la aplicación serían los que se muestran en el cuadro 8.1.

<sup>1</sup> <http://gplsi.dlsi.ua.es/~stela/analizador/>

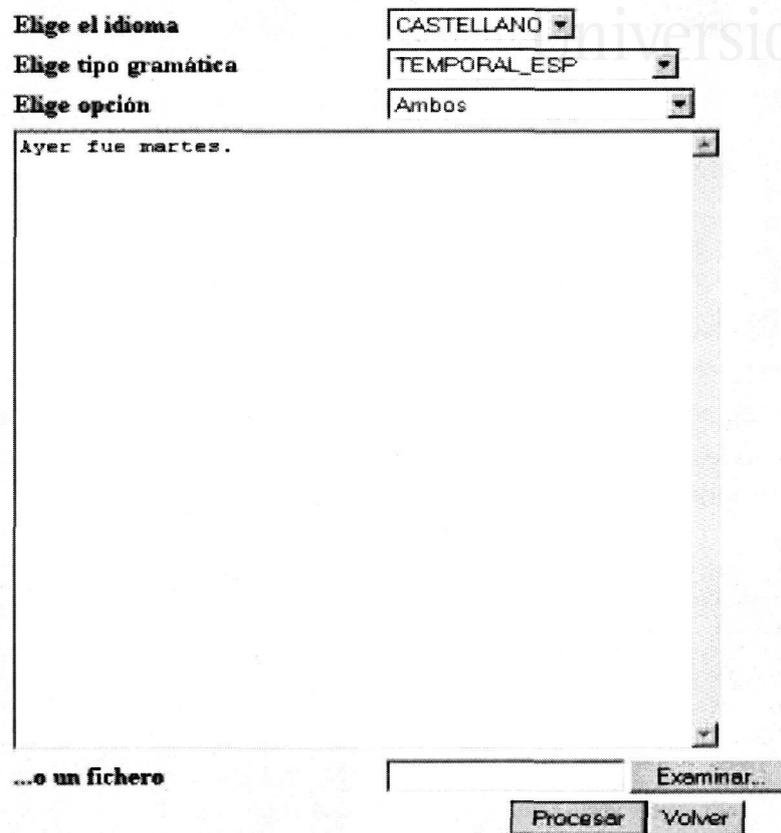


Figura 8.1. Analizador temporal

## 8.2 Sistema TERSEO

La aplicación TERSEO multilingüe<sup>2</sup> fue desarrollada con el objetivo de reconocer y resolver las expresiones temporales de textos digitales tanto en castellano como en inglés, pero además, esta aplicación es capaz de realizar la ordenación de los eventos contenidos en dicho texto. Se entiende por eventos todas aquellas oraciones del texto que contienen algún tipo de expresión temporal en la misma, de forma que, una vez que la expresión temporal

<sup>2</sup> <http://gplsi.dlsi.ua.es/~stela/sistemaMultilingue/>

**ANÁLISIS MORFOLÓGICO:**

Ayer ayer RG  
 fue ser VSIS3S0  
 martes [martes:??/??/?:?:?.??] W  
 . .Fp

**ANÁLISIS SINTÁCTICO:**

0: Ayer RF(0,0)  
 1: fue VSIS3S0(,) )  
 2: martes RF(2,2)  
 (S ( RF( RG Ayer) )( VSIS3S0 fue )( RF( W martes) ) )

**Cuadro 8.1.** Resultados obtenidos por el analizador

ha sido transformada a una fecha concreta, estas fechas pueden ser ordenadas, y dicha ordenación será devuelta por el sistema.

Se propone como ejemplo el siguiente texto en castellano:

La policía tailandesa ordenó ayer a los familiares, amigos y voluntarios que buscan a los miles de extranjeros desaparecidos por el maremoto mantenerse fuera de las áreas afectadas, incluidos los tanatorios instalados en templos.

“Con el fin de prevenir la manipulación de evidencias y la obstrucción en la recogida de ADN para la identificación de cuerpos, y prevenir las infecciones entre extranjeros, amigos, familiares y voluntarios, deben abstenerse de visitar las zonas afectadas por el tsunami, templos, mezquitas y todos los centros operativos y de autopsias”, indica la policía en un comunicado.

La orden, que cita “todas las zonas afectadas”, contrasta con la presencia de decenas de miles de tailandeses y turistas extranjeros que continúan con sus vacaciones en la playa de Patong, en la que perdieron la vida cientos de personas, la mayor parte extranjeros.

Cientos de occidentales, tailandeses y ciudadanos de otras naciones asiáticas emprendieron esta semana una desesperada búsqueda de familiares y amigos por los hospitales, centros de emergencia y tanatorios de la región.

El Ministerio del Interior tailandés señaló ayer que hay 5.046 muertos confirmados, incluidos 2.461 turistas, aunque redujo hasta 3.810 la cifra de desaparecidos, entre ellos al menos 1.600 extranjeros, la mayoría europeos. En los establecimientos hoteleros de las provincias de Phuket, Phang Nga y Krabi había al menos 100.000 turistas extranjeros, sobre todo europeos, cuando se produjo el maremoto, indicó el Ministerio del Interior.

En las Figuras 8.2, 8.4, 8.3 y 8.5 se muestran las diferentes fases del sistema para dicho ejemplo. En la pantalla principal de la aplicación (Figura 8.2) se debe indicar la fecha en la que el documento fue escrito. Al tratarse de artículos de periódicos dicha fecha coincide con la fecha del periódico. Además, se debe indicar el idioma del texto para que la aplicación sepa que expresiones

temporales y reglas de identificación debe utilizar para reconocer y resolver las expresiones temporales y por último, debe indicarse el fichero en el cual se encuentra el texto. Además, como puede verse en la figura se puede realizar el reconocimiento, resolución y ordenación del texto o únicamente reconocimiento y marcado de las expresiones temporales, sin indicar la resolución de las mismas.

Logo: L#
 Departamento de Lenguajes y Sistemas Informáticos

---

**TERSEO MULTILINGUE**

**Fecha Periódico:**  /  /

**Idioma:**

**Fichero noticia**

**Figura 8.2.** Pantalla principal sistema TERSEO

La Figura 8.3 muestra los resultados del sistema si únicamente se realiza el reconocimiento de las expresiones temporales del texto.

La Figura 8.4 muestra los resultados de reconocimiento y resolución de las expresiones temporales del texto. En la parte superior de la pantalla se marcan las expresiones temporales dentro del texto y en la parte inferior se indica para cada expresión la resolución obtenida.

Finalmente, en la Figura 8.5 se muestra la ordenación de los eventos realizada por el sistema, asignando a cada evento un número de orden.

La policía tailandesa ordenó ayer a los familiares , amigos y voluntarios que buscan a los miles de extranjeros desaparecidos por el maremoto mantenerse fuera de las áreas afectadas , incluidos los tanatorios instalados en templos. " Con el fin de prevenir la manipulación de evidencias y la obstrucción en la recogida de ADN para la identificación de cuerpos , y prevenir las infecciones entre extranjeros , amigos , familiares y voluntarios , deben abstenerse de visitar las zonas afectadas por el tsunami , templos , mezquitas y todos los centros operativos y de autopsias " , indica la policía en un comunicado. La orden , que cita " todas las zonas afectadas " , contrasta con la presencia de decenas de miles de tailandeses y turistas extranjeros que continúan con sus vacaciones en la playa de Patong , en la que perdieron la vida cientos de personas , la mayor parte extranjeros. Cientos de occidentales , tailandeses y ciudadanos de otras naciones asiáticas emprendieron esta semana una desesperada búsqueda de familiares y amigos por los hospitales , centros de emergencia y tanatorios de la región. El Ministerio del Interior tailandés señaló ayer que hay 5.046 muertos confirmados , incluidos 2.461 turistas , aunque redujo hasta 3.810 la cifra de desaparecidos , entre ellos al menos 1.600 extranjeros , la mayoría europeos. En los establecimientos hoteleros de las provincias de Phuket , Phang Nga y Krabi había al menos 100.000 turistas extranjeros , sobre todo europeos , cuando se produjo el maremoto , indicó el Ministerio del Interior.

[Pincha AQUÍ con el botón derecho para descargar archivo](#)

FIN DE LA APLICACIÓN

Figura 8.3. Pantalla resultado sólo resolución sistema TERSEO

### 8.3 Sistema Búsqueda de Respuestas Temporal

Como ya se explicó en la introducción de este capítulo, el sistema de Búsqueda de Respuestas Temporal esta dividido en dos partes. La primera parte realiza la descomposición de las preguntas complejas en preguntas simples y la segunda parte, dado un fichero con las respuestas a las preguntas simples y teniendo en cuenta las restricciones temporales impuestas por cada pregunta es capaz de recomponer estas respuestas para conseguir la respuesta final de la pregunta compleja. En las subsecciones posteriores se detalla cada uno de los pasos seguidos por estos dos subsistemas.

#### 8.3.1 Descomposición de la pregunta

Para mostrar un ejemplo de cómo el sistema de Búsqueda de Respuestas Temporal realiza la descomposición de las pregun-



La policía tailandesa ordenó **ayer** a los familiares, amigos y voluntarios que buscan a los miles de extranjeros desaparecidos por el maremoto mantenerse fuera de las áreas afectadas, incluidos los tanatorios instalados en templos. " Con el fin de prevenir la manipulación de evidencias y la obstrucción en la recogida de ADN para la identificación de cuerpos, y prevenir las infecciones entre extranjeros, amigos, familiares y voluntarios, deben abstenerse de visitar las zonas afectadas por el tsunami, templos, mezquitas y todos los centros operativos y de autopsias ", indica la policía en un comunicado. La orden, que cita " todas las zonas afectadas ", contrasta con la presencia de decenas de miles de tailandeses y turistas extranjeros que continúan con sus vacaciones en la playa de Patong, en la que perdieron la vida cientos de personas, la mayor parte extranjeros. Cientos de occidentales, tailandeses y ciudadanos de otras naciones asiáticas emprendieron **esta semana** una desesperada búsqueda de familiares y amigos por los hospitales, centros de emergencia y tanatorios de la región. El Ministerio del Interior tailandés señaló **ayer** que hay 5.046 muertos confirmados, incluidos 2.461 turistas, aunque redujo hasta 3.810 la cifra de desaparecidos, entre ellos al menos 1.600 extranjeros, la mayoría europeos. En los establecimientos hoteleros de las provincias de Phuket, Phang Nga y Krabi había al menos 100.000 turistas extranjeros, sobre todo europeos, cuando se produjo el maremoto, indicó el Ministerio del Interior.

Fecha				ORDENACIÓN DE EVENTOS			
Periódico:03/01/2005							
ID	POSI	POSE	EXPRESION	VALDATE1	VALTIME1	VALDATE2	VALTIME2
5060	4	4	ayer	02/01/2005	-	-	-
5061	164	165	esta semana	03/01/2005	-	09/01/2005	-
5062	189	189	ayer	02/01/2005	-	-	-

Figura 8.4. Pantalla resultado Reconocimiento y Resolución sistema TERSEO

tas temporales complejas en preguntas simples<sup>3</sup> vamos a utilizar una pregunta compleja de ejemplo. La pregunta sería la siguiente: "Where did Bill Clinton study after going to Oxford University?"

En la Figura 8.6 se muestra la pantalla principal de la aplicación de descomposición.

Como se puede apreciar en la figura, al igual que pasaba con el sistema TERSEO es necesario indicar una fecha de referencia, el idioma de las preguntas y el fichero dónde se encuentran dichas preguntas. Una vez especificada esta información, procesaremos el fichero con la pregunta ejemplo y el resultado obtenido se muestra a continuación:

<sup>3</sup> <http://gplsi.dlsi.ua.es/~stela/TQA/>



ORDEN	EVENTO
1	El Ministerio del Interior tailandés señaló ayer que hay 5.046 muertos confirmados , incluidos 2.461 turistas , aunque redujo hasta 3.810 la cifra de desaparecidos , entre ellos al menos 1.600 extranjeros , la mayoría europeos
1	La policía tailandesa ordenó ayer a los familiares , amigos y voluntarios que buscan a los miles de extranjeros desaparecidos por el maremoto mantenerse fuera de las áreas afectadas , incluidos los tanatorios instalados en templos
2	Cientos de occidentales , tailandeses y ciudadanos de otras naciones asiáticas emprendieron esta semana una desesperada búsqueda de familiares y amigos por los hospitales , centros de emergencia y tanatorios de la región

Figura 8.5. Pantalla resultado Ordenación Eventos sistema TERSEO


 Departamento de  
Lenguajes y Sistemas Informáticos 

---

**TEMPORAL QUESTION ANSWERING**

**Fecha Referencia:** [03] / [01] / [2005]  
**Idioma:** [INGLES]   
**Fichero preguntas complejas**

Figura 8.6. Pantalla principal sistema BR Temporal(Descomposición)

- Where did Bill Clinton study after going to Oxford University?  
 Type: 4  
 Temporal Signal: after  
 Temporal Expression:  
 Q1: Where did Bill Clinton study? FechaQ1:-  
 Q2: When did Bill Clinton go to Oxford University?  
 FechaQ2:-

### 8.3.2 Recomposición de la pregunta

El último paso del sistema de Búsqueda de Respuestas Temporal es la recomposición de las preguntas simples obtenidas por

el sistema de Búsqueda de Respuestas General para devolver las respuestas finales a las preguntas complejas<sup>4</sup>.

Para comprobar los resultados con el ejemplo propuesto, procesamos la pregunta compleja con el sistema de Búsqueda de Respuestas general, obteniendo los siguientes resultados:

1. Georgetown University
2. Sociology New York
3. Oxford
4. Yale University
5. Park End St.
6. York
7. MI
8. UK
9. Virginia
10. Ann Arbor
11. Montserrat

Ahora realizamos el proceso usando la unidad de recomposición. En la Figura 8.7 se muestra la pantalla principal de esta aplicación, y como se puede observar la única diferencia en cuanto a la pantalla de descomposición es que en lugar de las preguntas complejas, lo que necesita esta unidad es un fichero con las respuestas simples obtenidas para las subpreguntas de cada una de las preguntas complejas.

AS# Departamento de Lenguajes y Sistemas Informáticos

**TEMPORAL QUESTION ANSWERING (RECOMPOSITION)**

Fecha Referencia: 20 / 01 / 2004

Idioma: INGLES

Fichero respuestas simples  Examinar...

**QUESTION RECOMPOSITION UNIT**

Figura 8.7. Pantalla principal sistema BR Temporal (Recomposición)

<sup>4</sup> <http://gplsi.dlsi.ua.es/~stela/TQA/index2.php>



El fichero de respuestas simples para cada subpregunta después de procesar estas preguntas con un sistema general de Búsqueda de Respuestas online sería el siguiente:

1. Q1: Georgetown University (1964-1968)
2. Q1: Oxford University (1968-1970)
3. Q1: Yale University (1973)
4. Q2: 1968

Una vez introducida la información necesaria, el sistema genera las respuestas finales para la pregunta compleja. El resultado se muestra a continuación:

1. Where did Bill Clinton study after going to Oxford University?  
Tipo:4  
Respuesta: Yale University

El sistema de Búsqueda de Respuestas general devuelve muchas respuestas pero ninguna acertada mientras que nuestro sistema es capaz de filtrar las respuestas obtenidas para conseguir la respuesta correcta.

## 8.4 Conclusiones

En este capítulo se presentan las tres herramientas concretas desarrolladas en este trabajo de Tesis, sirviendo la ejecución de las mismas para obtener los diferentes resultados de evaluación presentados en el capítulo 7. Las tres herramientas presentadas son:

- Analizador temporal de expresiones temporales: Permite realizar un análisis parcial de los textos para extraer únicamente las expresiones temporales en los mismos.
- Sistema TERSEO: Aplicación que realiza el reconocimiento y resolución de las expresiones temporales de los textos, así como la ordenación de los eventos en dichos textos.
- Sistema de Búsqueda de Respuestas Temporal: Se ha desarrollado, por un lado, la capa de preprocesamiento de las preguntas complejas (Unidad de Descomposición), dividiéndolas en preguntas simples que pueden ser procesadas por un sistema de



#### 8.4 Conclusiones 167

Búsqueda de Respuestas general. Por otro lado, se ha implementado la capa de postprocesamiento de las respuestas simples obtenidas para conseguir las respuestas finales de las preguntas complejas (Unidad de Recomposición).



Universitat d'Alacant  
Universidad de Alicante



Universitat d'Alacant  
Universidad de Alicante

## 9. Conclusiones

En esta Tesis se ha desarrollado un sistema para el reconocimiento y resolución de expresiones temporales en textos escritos, convirtiendo dichas expresiones en fechas o periodos de fechas absolutas según un modelo temporal formal. Esta transformación permite la ordenación de los diferentes eventos expuestos en el texto, así como la obtención de información concreta referida a temporalidad a través de su aplicación en la Búsqueda de Respuestas.

El sistema desarrollado se denomina TERSEO (*Temporal Expression Resolution System applied to Event Ordering*) y aunque en sus orígenes se trataba de un sistema monolingüe (castellano), posteriormente se modificó para que fuera capaz de trabajar en un nivel multilingüe, gracias a la adaptación automática del sistema a otras lenguas.

Las principales aportaciones de esta Tesis han sido:

### Estudio del estado del arte

Realización de un profundo estudio de aquellos autores que han tratado algún aspecto de la información temporal en textos. El primer paso para este tratamiento consiste en establecer una ontología temporal que define cuáles son las entidades susceptibles de ser reconocidas. Estas ontologías sirven como base para poder hacer frente a las principales tareas relacionadas con el tratamiento de la información temporal. Estas tareas son:

- Anotación temporal de textos usando esquemas.
- Ubicación de eventos con fechas concretas asociadas en la línea temporal.



- Establecimiento de relaciones de orden entre los eventos que no pueden ser directamente relacionados con una fecha concreta.

A la hora de afrontar estas tareas, algunos autores desarrollan sistemas basados en Conocimiento, mientras que otros desarrollan sistemas basados en Corpus.

### **Definición de un Modelo Temporal**

Definición de un modelo temporal sobre el que se basa el sistema y que consta de dos partes principales:

- Un ontología temporal, en la que se indican los conceptos temporales relevantes en el dominio en el que nos encontramos (artículos de periódicos) así como las relaciones existentes entre dichos conceptos.
- Un conjunto de reglas de dos tipos: reglas de identificación (dependientes de la lengua) y reglas de resolución (independientes de la lengua).

La ontología temporal permite establecer el ámbito de anotación y resolución de las expresiones temporales que el sistema propuesto será capaz de tratar. La ontología temporal propuesta contempla todos los conceptos relevantes a tener en cuenta en el dominio en el que nos encontramos y las relaciones entre dichos conceptos. Por ello fue necesario construir una taxonomía temporal que contemple todos los posibles aspectos temporales de una determinada expresión. La taxonomía definida se basa en cinco clasificaciones posibles para las expresiones temporales atendiendo a las siguientes propiedades:

- El tipo de resolución
- La fecha de referencia
- La unidad temporal
- El carácter temporal
- El valor temporal

Existen dos relaciones entre los conceptos temporales definidos previamente:



- Relación **es un**: es la relación que definen las diferentes taxonomías, es decir, “ayer” es un **DIA** o “hace unos días” es una **Expresión Difusa**.
- Relación de **precedencia**: es la relación que define un orden temporal entre los diferentes conceptos de la ontología y que posteriormente se utilizará para realizar la ordenación de eventos, por ejemplo: “ayer” precede a “hoy” y a su vez “hoy” precede a “mañana”.

Como conclusión, cada expresión contemplada tiene asignada cinco aspectos temporales que la ubican en la ontología propuesta. Una vez que la expresión ha sido ubicada en la ontología, la regla de resolución de dicha expresión es directa puesto que depende de la combinación de estos aspectos temporales.

Además, el modelo temporal se ha completado con la formalización de las reglas de resolución definidas en el mismo. Para la definición de la formalización se utilizó como punto de referencia la formalización de la ontología temporal definida en (Hobbs, 2002) y que se basa en lógica de primer orden, siendo el modelo temporal definido en este trabajo una extensión de éste. El sistema de reconocimiento y resolución de expresiones temporales propuesto en este trabajo se basa en este modelo temporal.

### Definición de un Esquema de Anotación Temporal

Definición de un esquema de anotación temporal propio puesto que, después de estudiar las necesidades del sistema aquí presentado y las características de los esquemas existentes, estos esquemas eran demasiado extensos y complejos. La decisión final fue la creación de un esquema de anotación temporal nuevo capaz de anotar los resultados del reconocimiento y resolución de las expresiones temporales de una forma sencilla e intuitiva. Además, el esquema propuesto cumple las características básicas que todo esquema de anotación debe tener:

- **Completitud**: es capaz de captar todas las propiedades que serán necesarias.
- **Simplicidad**: es sencillo de entender.



- **Minimalidad:** no expresa conceptos en función de otros.
- **Formalidad:** cada concepto tiene una interpretación única, precisa y definida en lenguaje XML.

El esquema de anotación definido incluye un conjunto de etiquetas para marcar las expresiones temporales explícitas (fechas, horas, etc.) y las expresiones temporales implícitas (ayer, al día siguiente, etc.), así como sus atributos para almacenar los valores obtenidos por el sistema TERSEO al procesar dichas expresiones.

### **Desarrollo de un analizador temporal**

Desarrollo de un analizador temporal que se basa en una gramática temporal y realiza un análisis parcial ascendente (analizador chart). La gramática temporal únicamente tendrá reglas para reconocer expresiones temporales y desechará el resto de la información, realizando un reconocimiento de las expresiones temporales mucho más rápido que si se realizará un análisis completo de los textos. El analizador temporal permite obtener la información sintáctica del modelo temporal.

### **Desarrollo de un sistema de Reconocimiento y Resolución de Expresiones Temporales monolingüe**

Desarrollo de un sistema basado en Conocimiento, cuyo nombre es TERSEO y que es capaz de:

- Identificar las expresiones temporales en los textos
- Resolver las expresiones temporales en los textos
- Anotar las expresiones temporales junto con sus resoluciones
- Ordenar las expresiones temporales con sus eventos relacionados.

Para realizar estas tareas, el sistema utiliza el analizador temporal desarrollado y un modelo temporal en castellano como recurso. Este modelo temporal contiene todas las expresiones que el sistema es capaz de tratar y las reglas de resolución asociadas a cada una de ellas.

El resultado final del sistema será un texto en el cual, las expresiones temporales aparecen anotadas y resueltas con valores de fechas concretos.

### **Definición de un método de Ordenación de eventos**

Una vez que las expresiones han sido identificadas, resueltas y anotadas, la ordenación de los eventos en el texto puede llevarse a cabo. Esta ordenación se determinará en función de las fechas estimadas para las expresiones temporales que acompañan a los eventos. El resultado final del sistema será un conjunto de eventos ordenados cronológicamente que pueden ser utilizados para diferentes aplicaciones como, por ejemplo, la realización de resúmenes.

### **Extensión automática del sistema**

Desarrollo de la extensión del sistema TERSEO para poder obtener expresiones temporales y reglas de identificación para nuevas lenguas automáticamente. De esta forma, no será necesario conocer la lengua a la que queremos extender el sistema, puesto que las nuevas expresiones temporales y la asociación con sus reglas de resolución serán obtenidas automáticamente. Para la obtención de dichas reglas se realizan dos procesos principales:

- Traducción automática de las expresiones temporales de la lengua origen.
- Obtención de nuevas expresiones en la lengua destino usando un corpus no anotado en dicha lengua y un conjunto de palabras clave temporales.

Una vez obtenida las reglas de identificación para la nueva lengua, ésta será utilizada por el sistema TERSEO para poder reconocer y resolver expresiones temporales en dicha lengua.

### **Aplicación del sistema a la Búsqueda de Respuestas**

La aplicación del sistema TERSEO a una tarea concreta del Procesamiento de Lenguaje Natural como es la Búsqueda de Res-



puestas, en concreto, respuestas a preguntas temporales. Este tipo de preguntas temporales pueden ser preguntas simples con alguna fecha en su formulación o preguntas complejas que necesitan ser divididas en preguntas más simples y cuyas respuestas deben de ser reordenadas basándose en la información temporal para llegar a responder a la pregunta original.

Para poder tratar preguntas temporales complejas se propone una arquitectura multicapa que será capaz de descomponer estas preguntas complejas en preguntas simples, de forma que las preguntas simples puedan ser tratadas por los sistemas actuales de Búsqueda de Respuestas.

Las principales ventajas de utilizar una arquitectura multicapa son:

- Permite usar cualquier sistema de Búsqueda de Respuestas existente adaptando la salida de la capa con la entrada del sistema.
- El sistema de Búsqueda de Respuesta no necesita modificarse puesto que el procesamiento se realiza en una capa independiente.
- El sistema puede ser fácilmente extendido añadiendo capas para tratar otros tipos de preguntas complejas que actualmente no pueden ser tratadas por los sistemas de Búsqueda de Respuestas existentes.

En concreto, en este trabajo, se ha desarrollado la capa de procesamiento de preguntas temporales y dicha capa consta de las siguientes unidades:

- Unidad de Descomposición de la Pregunta: En esta unidad la información temporal es extraída por el sistema TERSEO y las preguntas complejas son divididas en preguntas simples.
- Sistema de Búsqueda de Respuestas de Propósito General: Cualquier sistema de Búsqueda de Respuestas puede ser utilizado.
- Unidad de Recomposición de la Respuesta: Utilizando la información temporal obtenida en la primera capa y las respuestas a las preguntas simples es capaz de recomponer la respuesta final para la pregunta compleja.

Puesto que la descomposición de la pregunta se hace según reglas sintácticas, el procesamiento de dicha división va a ser de-

pendiente del idioma. En un principio se realizaron todos los pasos del procesamiento para el inglés y posteriormente se añadió una extensión para su funcionamiento correcto en castellano. Únicamente fue necesario adaptar el módulo de la división de la pregunta, el resto de módulos funcionan correctamente para todos los idiomas.

### **Evaluación del sistema propuesto**

La evaluación del sistema en todos los aspectos desarrollados utilizando artículos digitales de periódicos como corpus. Una primera evaluación se realiza sobre el sistema TERSEO de reconocimiento y resolución de expresiones temporales trabajando en castellano y obteniendo unos resultados de F-Medida del 77%. Posteriormente, una vez que el sistema TERSEO fue extendido automáticamente al inglés se realizó una evaluación del sistema obteniendo una F-Medida del 69%.

A su vez, el sistema TERSEO en inglés también participó en la evaluación TERN 2004, dedicada a evaluar sistemas que tratan información temporal en textos, obteniendo unos resultados muy competitivos con respecto al resto de sistemas presentados en dicha conferencia. Puesto que la evaluación se realizó por partes, se sabe que los mejores resultados se obtuvieron en el reconocimiento de las expresiones, donde se obtuvo una precisión del 95%, una cobertura del 78% y una F-Medida del 86%.

Además, el sistema de Búsqueda de Respuestas Temporal Multicapa fue evaluado para el inglés en comparación con un sistema de Búsqueda de Respuestas general. El resultado obtenido demostró que el sistema propuesto en este trabajo mejoraba al sistema general en un 34% de los casos y se comportaba peor en el 1% de los casos. La evaluación del sistema de Búsqueda de Respuestas Temporal en castellano fue evaluado mediante la participación del sistema en la tarea piloto de CLEF 2004. El sistema fue capaz de resolver satisfactoriamente un 15% de las preguntas relacionadas con temporalidad del corpus utilizado en esta conferencia.



### **Herramientas desarrolladas**

Las tres herramientas desarrolladas para este trabajo de Tesis son:

- **Analizador sintáctico temporal:** Permite realizar un análisis parcial para extraer únicamente la información temporal de los textos utilizando una gramática temporal.
- **Sistema TERSEO:** Aplicación que realiza el reconocimiento y resolución de las expresiones temporales en los textos, así como la ordenación de los eventos en dichos textos.
- **Sistema de Búsqueda de Respuestas Temporal:** Se ha desarrollado, por un lado, la capa de preprocesamiento de las preguntas complejas (Unidad de Descomposición), dividiéndolas en preguntas simples que pueden ser procesadas por un sistema de Búsqueda de Respuestas general. Por otro lado, se ha implementado la capa de postprocesamiento de las respuestas simples obtenidas para conseguir las respuestas finales de las preguntas complejas (Unidad de Recomposición).

### **9.1 Trabajos en progreso**

Como continuación de esta Tesis existen una serie de trabajos que se están llevando a cabo o que se prevén interesantes a abordar, entre los que nos encontramos:

- **Introducción de información de contexto** para que pueda ser usada por el sistema TERSEO a la hora de resolver determinadas expresiones temporales que necesitan dicha información para poder ser correctamente interpretadas. Por información de contexto se entiende fechas de eventos conocidos que pueden aparecer en los textos, por ejemplo: “la Guerra de Irak”. Si en un texto tenemos una expresión temporal que se refiere a este evento, esta expresión únicamente podrá ser resuelta si la fecha de dicho evento se encuentra almacenada en el sistema. La información de los eventos conocidos deberá actualizarse constantemente, a medida que nuevos eventos vayan teniendo



lugar. Almacenar este tipo de información supone una mejora importante respecto a un sistema que no la posee, puesto que en artículos de periódicos se hace referencia múltiples veces a eventos de este tipo.

- Utilización de corpus paralelos en dos idiomas para la ampliación de las reglas de identificación del sistema TERSEO a otras lenguas. Los corpus paralelos nos permiten obtener la traducción exacta y correcta de una determinada expresión, y su uso, junto con la traducción automática que ya se realiza actualmente, puede conseguir unas reglas de identificación que permitan obtener resultados altamente satisfactorios para la nueva lengua introducida.
- Refinamiento manual de las expresiones temporales y las reglas de identificación del sistema TERSEO en inglés, puesto que han sido generadas automáticamente y existen expresiones o formatos de expresiones que, al ser particulares para cada lengua, no han sido obtenidos al realizar la extensión automática. Es necesario, por tanto, realizar un estudio de aquellas expresiones que se han reconocido o resuelto de manera incorrecta por parte del sistema TERSEO para posteriormente introducir esta información de manera correcta en el sistema.
- Mejora del sistema TERSEO a la hora de etiquetar un texto usando el esquema TIMEX2. Como se pudo ver en la evaluación de la conferencia TERN 2004, el sistema obtenía resultados satisfactorios a la hora de reconocer y resolver las expresiones temporales, pero los resultados eran demasiado bajos al tratar el atributo MOD de dicha etiqueta. El atributo MOD captura los modificadores temporales que pueden aparecer en una expresión, como por ejemplo: “*aproximadamente* dos días después”. El sistema TERSEO presentado aquí no posee prácticamente ninguno de estos modificadores y por tanto no es capaz de asignar un valor correcto a dicho atributo. Por ello, será necesario realizar un estudio de todos los posibles modificadores que aparecen en las expresiones temporales así como su interpretación y esta información deberá ser introducida también en el sistema.
- Realización de un corpus de preguntas temporales complejas, tanto para el castellano como para el inglés que permita evaluar



en profundidad el sistema de Búsqueda de Respuestas Temporal desarrollado en este trabajo. Actualmente, el corpus de preguntas temporales complejas es escaso puesto que este tipo de preguntas no suelen utilizarse para competiciones internacionales como CLEF o TREC.

- Extensión de la metodología usada para el tratamiento de preguntas temporales complejas, usando una arquitectura multicapa, para tratar otro tipo de preguntas complejas que actualmente se encuentran fuera del ámbito de los sistemas de Búsqueda de Respuestas generales. Entre estas preguntas se podrían destacar las preguntas de script, como por ejemplo: “¿Cómo construyo un ordenador?”. La respuesta final a esta pregunta es un conjunto de respuestas ordenadas. Por tanto, la capa de procesamiento de preguntas de script debería recoger todas las posibles respuestas relacionadas con la construcción de un ordenador y ser capaz de realizar una ordenación de las mismas. Otro tipo de preguntas complejas que podrían ser tratadas en una nueva capa del sistema son las preguntas basadas en plantillas, como por ejemplo: “¿Cuáles son los datos biográficos de Nelson Mandela?”. La estructura de esta capa sería similar a la de la capa de preguntas temporales. La unidad de descomposición convertiría esta pregunta en un conjunto de subpreguntas relacionadas con datos biográficos, por ejemplo: “¿Cuándo nació Nelson Mandela?”, “¿En qué año murió?”, ... Estas preguntas simples se procesarían a través del sistema de Búsqueda de Respuestas general y finalmente, la capa de procesamiento rellenaría una plantilla con todos los datos biográficos obtenidos.

## 9.2 Trayectoria investigadora

Mi actividad investigadora se inicia en el año 1998 con la realización de los cursos del programa de doctorado “**INGENIERÍA LINGÜÍSTICA, APRENDIZAJE AUTOMÁTICO Y RECONOCIMIENTO DE FORMAS**” perteneciente al Departamento de Lenguajes y Sistemas Informáticos (DLSI) de la Universidad de Alicante (UA) y posteriormente mi incorporación al



grupo de investigación “Procesamiento de Lenguajes y Sistemas de Información (GPLSI)” como Ayudante.

En concreto, mi trayectoria investigadora se divide en tres etapas:

- Estudio inicial
- Consolidación y especialización
- Expansión y aplicación

### 9.2.1 Estudio inicial (1998-2000)

Esta etapa inicial se dedicó al estudio de las diferentes tareas del Procesamiento del Lenguaje Natural mediante las diferentes asignaturas del programa de doctorado, con la posterior especialización en la información temporal, en concreto en las expresiones temporales mediante el estudio de artículos presentados por otros autores en congresos y revistas. Este estudio y su aplicación derivó en las siguientes aportaciones científicas:

- Especificación de una gramática para el reconocimiento de expresiones temporales publicada bajo el título de *Grammar specifications for the recognition of temporal expressions*. E. Saquete, P. Martínez-Barco. Proceedings of Machine Translation and Multilingual applications in the new millenium, MT2000. Exeter (UK). Enero 2000 (Saquete & Martínez-Barco, 2000).
- Trabajo de suficiencia investigadora con título *Especificación de una gramática para el reconocimiento de referencias temporales* bajo la dirección del Dr. Patricio Martínez Barco y presentado en el año 2000.

### 9.2.2 Consolidación y especialización (2000-2003)

Una vez conseguida la suficiencia investigadora comienza la consolidación de mi etapa investigadora en la que profundicé en el estudio del reconocimiento y resolución de expresiones temporales en castellano, aplicándolo a la ordenación de los eventos en los textos en función de la información temporal. Esto llevó a la publicación de las siguientes aportaciones científicas:



- Desarrollo de un sistema basado en una gramática para la resolución de expresiones temporales en castellano que se presentó bajo el título *A grammar-based system to solve temporal expressions in Spanish Texts*. E. Saquete, P. Martínez-Barco, R. Muñoz. Lecture Notes in Computer Science. PorTal 2002. Faro, PORTUGAL. 2002. pp 53-62 (Saquete *et al.*, 2002a).
- Posteriormente, se presentó no sólo el reconocimiento sino también la posterior anotación de los resultados obtenidos utilizando un esquema de anotación propio: *Recognising and Tagging Temporal Expressions in Spanish*. E. Saquete, P. Martínez-Barco, R. Muñoz. 3rd International Conference on Language Resources and Evaluation (LREC'2002). Las Palmas, SPAIN. 2002 (Saquete *et al.*, 2002b).
- La aplicación de ordenación de eventos usando la resolución de la información temporal fue presentada en el artículo *TERSEO: Temporal Expression Resolution System Applied to Event Ordering*. E. Saquete, R. Muñoz, and P. Martínez-Barco. Lecture Notes in Computer Science. International Conference on Text Speech and Dialogue (TSD). Ceske Budejovice. Czech Republic. September 2003. pp. 220-228 (Saquete *et al.*, 2003b) y en *Event Ordering through Temporal Expression Resolution*. E. Saquete, R. Muñoz and P. Martínez-Barco. Recent Advances in Natural Language Processing (RANLP). Borovets, BULGARIA. September 2003 (Saquete *et al.*, 2003a).

### 9.2.3 Expansión y aplicación (2004)

Una vez desarrollado el sistema, la última etapa se divide en dos partes. Una etapa de expansión del sistema desarrollado previamente para que pueda ser utilizado de forma multilingüe y de ahí surgen las siguientes contribuciones:

- Desarrollo de un sistema de extensión automática del sistema original para que sea capaz de trabajar a nivel multilingüe. Este desarrollo se presentó en *Automatic multilinguality for Time Expression Resolution*. E. Saquete, P. Martínez-Barco, R. Muñoz. Lecture Notes in Computer Science. Mexican International Con-

ference on Artificial Intelligence (MICAI). Mexico City, MEXICO. April 2004. pp. 458-467 (Saquete *et al.*, 2004a).

- Evaluación del sistema TERSEO trabajando a nivel multilingüe una vez extendido automáticamente, en concreto para el inglés. Los resultados de esta evaluación se presentaron en *Evaluation of the Automatic Multilinguality for Time Expression Resolution*. E. Saquete, P. Martínez-Barco, R. Muñoz. 4th International Workshop on Natural Language and Information Systems (NLIS). Zaragoza, SPAIN. August 2004 (Saquete *et al.*, 2004e).

Posteriormente, se realizaron determinadas aportaciones relacionadas con las aplicaciones para las que el sistema TERSEO ha sido utilizado, como por ejemplo, ordenación de eventos, desambiguación de palabras, pero más en profundidad en la Búsqueda de Respuestas Temporal, aplicación para la cual se ha realizado un sistema multicapa basado en TERSEO y capaz de trabajar en varios idiomas. Las aportaciones científicas de esta etapa fueron:

- Evaluación del sistema de ordenación de eventos de TERSEO multilingüe, presentado bajo el título *Event Ordering using TERSEO system*. E. Saquete, R. Muñoz, P. Martínez-Barco. Lecture Notes in Computer Science. International Conference on Applications of Natural Language to Information Systems (NLDB) Manchester. UK. June 2004 (Saquete *et al.*, 2004f).
- Este artículo es a su vez ampliado y publicado en la revista SCI Data and Knowledge Engineering Journal (indizada en Science Citation Index). Actualmente se encuentra aceptado y pendiente de publicación (Saquete *et al.*, 2005)
- Estudio del posible beneficio de la resolución de expresiones temporales en la desambiguación del sentido de las palabras presentado en *The Role of Temporal Expressions in Word Sense Disambiguation*. S. Vázquez, E. Saquete, A. Montoyo, P. Martínez-Barco, R. Muñoz. Lecture Notes in Computer Science. Fifth International Conference on Intelligent Text Processing and Computational Linguistics (CICLING). Seoul. Korea. February 2004. pp. 209-212 (Vázquez *et al.*, 2004).
- Definición de una arquitectura multicapa para el procesamiento de preguntas complejas temporales para complementar un



sistema de Búsqueda de Respuestas de propósito general, realizando una división de las preguntas complejas en preguntas simples. El método de división fue presentado en *Splitting Complex Temporal Questions for Question Answering systems*. E. Saquete, P. Martínez-Barco, R. Muñoz, J.L. Vicedo. Association for Computational Linguistics (ACL) Barcelona, SPAIN. July 2004 (Saquete *et al.*, 2004h).

- La posterior recomposición de las respuestas simples obtenidas para la reconstrucción de la respuesta final en este sistema multicapa se presentó bajo el título de *Multilayered Question Answering system applied to Temporality evaluation*. E. Saquete, P. Martínez-Barco, R. Muñoz, J.L. Vicedo. XX Congreso de la SEPLN. Barcelona, SPAIN. July 2004 (Saquete *et al.*, 2004g).
- El sistema de Búsqueda de Respuestas Temporal para el inglés fue evaluado utilizando un corpus de 124 preguntas temporales, y se propuso una mejora a dicho sistema que también fue evaluada y comparada con la inicial. Los resultados se presentaron en *Comparison and Evaluation of two approaches of a Multilayered QA system applied to Temporality*. E. Saquete, R. Muñoz, P. Martínez-Barco, J.L. Vicedo. 4th International Conference. ESTAL 2004. Lecture Notes in Artificial Intelligence. Alicante, SPAIN. October 2004 (Saquete *et al.*, 2004b).
- El sistema de Búsqueda de Respuestas Temporal en castellano participó en la tarea piloto del CLEF 2004 para responder a preguntas temporales complejas y los resultados fueron publicados en *Evaluation of Complex Temporal Question in CLEF-QA*. E. Saquete, J.L. Vicedo, P. Martínez-Barco, R. Muñoz Cross-Lingual Evaluation Forum-2004, Lecture Notes in Computer Science (Saquete *et al.*, 2004d).



## A. Anexo Formalización de las reglas de resolución del modelo temporal

### A.1 Reglas de resolución desde la perspectiva de la lógica de 1er orden

- Funciones primitivas definidas por Hobbs y usadas en nuestro modelo temporal:
  - $interval(x)$  : Indica que  $x$  es un intervalo.
  - $day(x)$  :  $x$  es un día del calendario.
  - $start - of(x)$  : Indica el día de inicio del intervalo  $x$ .
  - $end - of(x)$  : Indica el día de fin del intervalo  $x$ .
  - $Hath(S, n, u, x)$  : Un conjunto  $S$  de  $n$  intervalos de calendario de tipo  $u$  (\*Day\*, \*Month\*, ...) forman el intervalo de calendario  $x$ .
  - $interval - between(z, x, y)$ :  $z$  es un intervalo que comienza en el instante  $x$  y acaba en el instante  $y$ .
  - $in - interval(y, x)$ : El instante  $y$  pertenece al intervalo  $x$ .
  - $duration(z, u) = n$ : El intervalo  $z$  está formado por  $n$  unidades temporales de tipo  $u$ .
  - $before(x, y)$ : La entidad temporal  $x$  ocurre antes que la entidad temporal  $y$ .
  - $after(x, y)$ : La entidad temporal  $x$  ocurre después de la entidad temporal  $y$ .
  - $Monday(x, y), Tuesday(x, y), \dots, Sunday(x, y)$ : Indica que  $x$  es el día de la semana (Lunes, martes, ...) de la semana  $y$ .
- Funciones primitivas nuevas: La resolución de todas las expresiones está basada o en la fecha del documento o en una fecha nombrada previamente. Por tanto tendremos las funciones:
  - $documentDate(T)$
  - $previousDate(T)$
 Además necesitamos una función que nos devuelva en número de semana dentro del año para una fecha concreta que se pasa como argumento:
  - $weekOfYear(T)$
- Reglas de resolución relacionadas con la unidad temporal DIA:
  - Reglas de resolución para expresiones de DIA, PASADO y CONCRETO:
    - $yesterday(x) \equiv \exists y, z [documentDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = 1 \wedge day(x)]$
    - $daysAgo(x) \equiv \exists y, z [documentDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = n \wedge day(x)]$
    - $dayBefore(x) \equiv \exists y, z [previousDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = 1 \wedge day(x)]$
    - $daysBefore(x) \equiv \exists y, z [previousDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = n \wedge day(x)]$



- Reglas de resolución para expresiones de DIA,PASADO y RANGO:
  - $daysLast(x, n) \equiv \exists y, z, k, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Day*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Day*) = n \wedge interval - between(z, end - of(x), y) \wedge duration(k, *Day*) = 1]$
  - $daysPrevious(x, n) \equiv \exists y, z, k, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Day*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Day*) = n \wedge interval - between(z, end - of(x), y) \wedge duration(k, *Day*) = 1]$
- Reglas de resolución para expresiones de DIA,PASADO y DIFUSO:
  - $daysLastF(x, n) \equiv \exists y, z, k, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Day*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Day*) = 7 \wedge interval - between(z, end - of(x), y) \wedge duration(k, *Day*) = 1]$
  - $daysPreviousF(x, n) \equiv \exists y, z, k, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Day*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Day*) = 7 \wedge interval - between(z, end - of(x), y) \wedge duration(k, *Day*) = 1]$
- Reglas de resolución para expresiones de DIA,MISMO y CONCRETO:
  - $today(x) \equiv documentDate(x)$
- Reglas de resolución para expresiones de DIA,PASADO PRESENTE y RANGO:
  - $daysSince(x, n) \equiv \exists y, z, S[documentDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, n, *Day*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Day*) = n \wedge end - of(x) = y]$
  - $daysSinceThosc(x, n) \equiv \exists y, z, S[previousDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, n, *Day*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Day*) = n \wedge end - of(x) = y]$
- Reglas de resolución para expresiones de DIA,FUTURO y CONCRETO:
  - $tomorrow(x) \equiv \exists y, z[documentDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = 1 \wedge day(x)]$
  - $daysWithin(x) \equiv \exists y, z[documentDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = n \wedge day(x)]$
  - $dayLater(x) \equiv \exists y, z[previousDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = 1 \wedge day(x)]$
  - $daysLater(x) \equiv \exists y, z[previousDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Day*) = n \wedge day(x)]$
- Reglas de resolución para expresiones de DIA, FUTURO y RANGO:
  - $daysNext(x, n) \equiv \exists y, z, k, S[documentDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Day*, x) \wedge interval \neg between(z, start - of(x), y) \wedge duration(z, *Day*) = 1 \wedge interval - between(z, end - of(x), y) \wedge duration(k, *Day*) = n]$
  - $daysFollowing(x, n) \equiv \exists y, z, k, S[previousDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Day*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Day*) = 1 \wedge interval - between(z, end - of(x), y) \wedge duration(k, *Day*) = n]$
- Reglas de resolución para expresiones de DIA,FUTURO y DIFUSO:
  - $daysNextF(x, n) \equiv \exists y, z, k, S[documentDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Day*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Day*) = 1 \wedge interval - between(z, end - of(x), y) \wedge duration(k, *Day*) = 7]$
  - $daysFollowingF(x, n) \equiv \exists y, z, k, S[previousDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Day*, x) \wedge interval - between(z, start - of(x), y)$

$$\begin{aligned} \wedge duration(z, *Day*) &= 1 \wedge interval - between(z, end - of(x), y) \\ \wedge duration(k, *Day*) &= 7] \end{aligned}$$

■ Reglas de resolución relacionadas con la universal MES:

- Reglas de resolución para expresiones de MES,PASADO y CONCRETO:
  - $monthAgo(x) \equiv \exists y, z[documentDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = 1 \wedge day(x)]$
  - $monthsAgo(x) \equiv \exists y, z[documentDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n \wedge day(x)]$
  - $monthBefore(x) \equiv \exists y, z[previousDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = 1 \wedge day(x)]$
  - $monthsBefore(x) \equiv \exists y, z[previousDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n \wedge day(x)]$
- Reglas de resolución para expresiones de MES,PASADO y RANGO:
  - $monthLast(x) \equiv \exists y, z, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = 1]$
  - $monthsLast(x, n) \equiv \exists y, z, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n]$
  - $monthPrevious(x) \equiv \exists y, z, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = 1]$
  - $monthsPrevious(x, n) \equiv \exists y, z, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n]$
- Reglas de resolución para expresiones de MES,PASADO y FUZZY CONCRETO:
  - $monthLastF(x) \equiv \exists y, z, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = 1]$
  - $monthLastF(x, n) \equiv \exists y, z, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n]$
  - $monthPreviousF(x) \equiv \exists y, z, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = 1]$
  - $monthsPreviousF(x, n) \equiv \exists y, z, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n]$
- Reglas de resolución para expresiones de MES,PASADO y FUZZY:
  - $monthsAgoF(x) \equiv \exists y, z, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Year*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Year*) = 1 \wedge end - of(x) = y]$
  - $monthsBeforeF(x) \equiv \exists y, z, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Year*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Year*) = 1 \wedge end - of(x) = y]$
- Reglas de resolución para expresiones de MES,MISMO y RANGO:
  - $monthThis(x) \equiv \exists y, S[documentDate(y) \wedge in - interval(y, x) \wedge interval(x) \wedge Hath(S, 1, *Month*, x)]$
  - $monthThat(x) \equiv \exists y, S[previousDate(y) \wedge in - interval(y, x) \wedge interval(x) \wedge Hath(S, 1, *Month*, x)]$



- Reglas de resolución para expresiones de MES,PASADO PRESENTE y RANGO:
  - $monthSince(x) \equiv \exists y, z, S[documentDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Month*) = 1 \wedge end - of(x) = y]$
  - $monthsSince(x, n) \equiv \exists y, z, S[documentDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Month*) = n \wedge end - of(x) = y]$
  - $monthSinceThat(x) \equiv \exists y, z, S[previousDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Month*) = 1 \wedge end - of(x) = y]$
  - $monthsSinceThose(x, n) \equiv \exists y, z, S[previousDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Month*) = n \wedge end - of(x) = y]$
- Reglas de resolución para expresiones de MES,FUTURO y CONCRETO:
  - $monthWithin(x) \equiv \exists y, z[documentDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = 1 \wedge day(x)]$
  - $monthsWithin(x) \equiv \exists y, z[documentDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n \wedge day(x)]$
  - $monthLater(x) \equiv \exists y, z[previousDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = 1 \wedge day(x)]$
  - $monthsLater(x) \equiv \exists y, z[previousDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n \wedge day(x)]$
- Reglas de resolución para expresiones de MES,FUTURO y RANGO:
  - $monthsNext(x, n) \equiv \exists y, z, S[documentDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n]$
  - $monthsFollowing(x, n) \equiv \exists y, z, S[previousDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n]$
- Reglas de resolución para expresiones de MES,FUTURO y FUZZY CONCRETO:
  - $monthsNextFC(x, n) \equiv \exists y, z, S[documentDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n]$
  - $monthsFollowingFC(x, n) \equiv \exists y, z, S[previousDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Month*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Month*) = n]$
- Reglas de resolución para expresiones de MES,FUTURO y FUZZY:
  - $monthsNextF(x) \equiv \exists y, z, S[documentDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, 1, *Year*, x) \wedge interval - between(z, end - of(x), y) \wedge duration(z, *Year*) = 1 \wedge start - of(x) = y]$
  - $monthsFollowingF(x) \equiv \exists y, z, S[previousDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, 1, *Year*, x) \wedge interval - between(z, end - of(x), y) \wedge duration(z, *Year*) = 1 \wedge start - of(x) = y]$
- Reglas de resolución relacionadas con la unidad temporal AÑO:
  - Reglas de resolución para expresiones de AÑO,PASADO y CONCRETO:



## A.1 Reglas de resolución desde la perspectiva de la lógica de 1er orden 187

- $yearAgo(x) \equiv \exists y, z [documentDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = 1 \wedge day(x)]$
- $yearsAgo(x) \equiv \exists y, z [documentDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = n \wedge day(x)]$
- $yearBefore(x) \equiv \exists y, z [previousDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = 1 \wedge day(x)]$
- $yearsBefore(x) \equiv \exists y, z [previousDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = n \wedge day(x)]$
- Reglas de resolución para expresiones de AÑO, PASADO y RANGO:
  - $yearLast(x) \equiv \exists y, z, S [documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Year*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = 1]$
  - $yearsLast(x, n) \equiv \exists y, z, S [documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Year*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = n]$
  - $yearPrevious(x) \equiv \exists y, z, S [previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Year*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = 1]$
  - $yearsPrevious(x, n) \equiv \exists y, z, S [previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Year*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = n]$
- Reglas de resolución para expresiones de AÑO, PASADO y FUZZY CONCRETO:
  - $yearLastF(x) \equiv \exists y, z, S [documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Year*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = 1]$
  - $yearsLastF(x, n) \equiv \exists y, z, S [documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Year*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = n]$
  - $yearPreviousF(x) \equiv \exists y, z, S [previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Year*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = 1]$
  - $yearsPreviousF(x, n) \equiv \exists y, z, S [previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Year*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Year*) = n]$
- Reglas de resolución para expresiones de AÑO, PASADO y FUZZY:
  - $yearsAgoF(x) \equiv \exists y, z, S [documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 10, *Year*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Year*) = 10 \wedge end - of(x) = y]$
  - $yearsBeforeF(x) \equiv \exists y, z, S [previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 10, *Year*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Year*) = 10 \wedge end - of(x) = y]$
- Reglas de resolución para expresiones de AÑO, MISMO y RANGO:
  - $yearThis(x) \equiv \exists y, S [documentDate(y) \wedge in - interval(y, x) \wedge interval(x) \wedge Hath(S, 1, *Year*, x)]$
  - $yearThat(x) \equiv \exists y, S [previousDate(y) \wedge in - interval(y, x) \wedge interval(x) \wedge Hath(S, 1, *Year*, x)]$
- Reglas de resolución para expresiones de AÑO, PASADO PRESENTE y RANGO:
  - $yearSince(x) \equiv \exists y, z, S [documentDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, 1, *Year*, x)]$



- $\wedge interval - between(z, start - of(x), y) \wedge duration(z, *Year*) = 1$   
 $\wedge end - of(x) = y]$
- $yearsSince(x, n) \equiv \exists y, z, S[documentDate(y) \wedge before(start - of(x), y)$   
 $\wedge interval(x) \wedge Hath(S, 1, *Year*, x)$   
 $\wedge interval - between(z, start - of(x), y) \wedge duration(z, *Year*) = n$   
 $\wedge end - of(x) = y]$
  - $yearSinceThat(x) \equiv \exists y, z, S[previousDate(y) \wedge before(start - of(x), y)$   
 $\wedge interval(x) \wedge Hath(S, 1, *Year*, x)$   
 $\wedge interval - between(z, start - of(x), y) \wedge duration(z, *Year*) = 1$   
 $\wedge end - of(x) = y]$
  - $yearsSinceThose(x, n) \equiv \exists y, z, S[previousDate(y) \wedge before(start - of(x), y)$   
 $\wedge interval(x) \wedge Hath(S, 1, *Year*, x)$   
 $\wedge interval - between(z, start - of(x), y) \wedge duration(z, *Year*) = n$   
 $\wedge end - of(x) = y]$
  - Reglas de resolución para expresiones de AÑO, FUTURO y CONCRETO:
    - $yearWithin(x) \equiv \exists y, z[documentDate(y) \wedge after(x, y) \wedge interval -$   
 $between(z, x, y) \wedge duration(z, *Year*) = 1 \wedge day(x)]$
    - $yearsWithin(x) \equiv \exists y, z[documentDate(y) \wedge after(x, y) \wedge interval -$   
 $between(z, x, y) \wedge duration(z, *Year*) = n \wedge day(x)]$
    - $yearLater(x) \equiv \exists y, z[previousDate(y) \wedge after(x, y) \wedge interval -$   
 $between(z, x, y) \wedge duration(z, *Year*) = 1 \wedge day(x)]$
    - $yearsLater(x) \equiv \exists y, z[previousDate(y) \wedge after(x, y) \wedge interval -$   
 $between(z, x, y) \wedge duration(z, *Year*) = n \wedge day(x)]$
  - Reglas de resolución para expresiones de AÑO, FUTURO y RANGO:
    - $yearsNext(x, n) \equiv \exists y, z, S[documentDate(y) \wedge after(x, y) \wedge$   
 $interval(x) \wedge Hath(S, n, *Year*, x) \wedge interval - between(z, x, y)$   
 $\wedge duration(z, *Year*) = n]$
    - $yearsFollowing(x, n) \equiv \exists y, z, S[previousDate(y) \wedge after(x, y) \wedge$   
 $interval(x) \wedge Hath(S, n, *Year*, x) \wedge interval - between(z, x, y)$   
 $\wedge duration(z, *Year*) = n]$
  - Reglas de resolución para expresiones de AÑO, FUTURO y FUZZY CONCRETO:
    - $yearsNextFC(x, n) \equiv \exists y, z, S[documentDate(y) \wedge after(x, y) \wedge$   
 $interval(x) \wedge Hath(S, n, *Year*, x) \wedge interval - between(z, x, y)$   
 $\wedge duration(z, *Year*) = n]$
    - $yearsFollowingFC(x, n) \equiv \exists y, z, S[previousDate(y) \wedge after(x, y) \wedge$   
 $interval(x) \wedge Hath(S, n, *Year*, x) \wedge interval - between(z, x, y)$   
 $\wedge duration(z, *Year*) = n]$
  - Reglas de resolución para expresiones de AÑO, FUTURO y FUZZY:
    - $yearsNextF(x) \equiv \exists y, z, S[documentDate(y) \wedge after(x, y) \wedge$   
 $interval(x) \wedge Hath(S, 10, *Year*, x) \wedge interval - between(z, end - of(x), y)$   
 $\wedge duration(z, *Year*) = 10 \wedge start - of(x) = y]$
    - $yearsFollowingF(x) \equiv \exists y, z, S[previousDate(y) \wedge after(x, y) \wedge$   
 $interval(x) \wedge Hath(S, 10, *Year*, x) \wedge interval - between(z, end - of(x), y)$   
 $\wedge duration(z, *Year*) = 10 \wedge start - of(x) = y]$
  - Reglas de resolución relacionadas con la unidad temporal **DIA-SEMANA**:
    - Reglas de resolución para expresiones de DIASEMANA, PASADO y CONCRETO:
      - $mondayLast(x) \equiv \exists y[documentDate(y) \wedge$   
 $Monday(x, (weekOfYear(y) - 1)) \wedge day(x)]$



## A.1 Reglas de resolución desde la perspectiva de la lógica de 1er orden 189

- $tuesdayLast(x) \equiv \exists y[documentDate(y) \wedge Tuesday(x, (weekOfYear(y) - 1)) \wedge day(x)]$
- $wednesdayLast(x) \equiv \exists y[documentDate(y) \wedge Wednesday(x, (weekOfYear(y) - 1)) \wedge day(x)]$
- $thursdayLast(x) \equiv \exists y[documentDate(y) \wedge Thursday(x, (weekOfYear(y) - 1)) \wedge day(x)]$
- $fridayLast(x) \equiv \exists y[documentDate(y) \wedge Friday(x, (weekOfYear(y) - 1)) \wedge day(x)]$
- $saturdayLast(x) \equiv \exists y[documentDate(y) \wedge Saturday(x, (weekOfYear(y) - 1)) \wedge day(x)]$
- $sundayLast(x) \equiv \exists y[documentDate(y) \wedge Sunday(x, (weekOfYear(y) - 1)) \wedge day(x)]$
- Reglas de resolución para expresiones de DIASEMANA, MISMO y CONCRETO:
  - $mondayThis(x) \equiv \exists y[documentDate(y) \wedge Monday(x, weekOfYear(y)) \wedge day(x)]$
  - $tuesdayThis(x) \equiv \exists y[documentDate(y) \wedge Tuesday(x, weekOfYear(y)) \wedge day(x)]$
  - $wednesdayThis(x) \equiv \exists y[documentDate(y) \wedge Wednesday(x, weekOfYear(y)) \wedge day(x)]$
  - $thursdayThis(x) \equiv \exists y[documentDate(y) \wedge Thursday(x, weekOfYear(y)) \wedge day(x)]$
  - $fridayThis(x) \equiv \exists y[documentDate(y) \wedge Friday(x, weekOfYear(y)) \wedge day(x)]$
  - $saturdayThis(x) \equiv \exists y[documentDate(y) \wedge Saturday(x, weekOfYear(y)) \wedge day(x)]$
  - $sundayThis(x) \equiv \exists y[documentDate(y) \wedge Sunday(x, weekOfYear(y)) \wedge day(x)]$
- Reglas de resolución para expresiones de DIASEMANA, FUTURO y CONCRETO:
  - $mondayNext(x) \equiv \exists y[documentDate(y) \wedge Monday(x, (weekOfYear(y) + 1)) \wedge day(x)]$
  - $tuesdayNext(x) \equiv \exists y[documentDate(y) \wedge Tuesday(x, (weekOfYear(y) + 1)) \wedge day(x)]$
  - $wednesdayNext(x) \equiv \exists y[documentDate(y) \wedge Wednesday(x, (weekOfYear(y) + 1)) \wedge day(x)]$
  - $thursdayNext(x) \equiv \exists y[documentDate(y) \wedge Thursday(x, (weekOfYear(y) + 1)) \wedge day(x)]$
  - $fridayNext(x) \equiv \exists y[documentDate(y) \wedge Friday(x, (weekOfYear(y) + 1)) \wedge day(x)]$
  - $saturdayNext(x) \equiv \exists y[documentDate(y) \wedge Saturday(x, (weekOfYear(y) + 1)) \wedge day(x)]$
  - $sundayNext(x) \equiv \exists y[documentDate(y) \wedge Sunday(x, (weekOfYear(y) + 1)) \wedge day(x)]$
- Reglas de resolución relacionadas con la unidad temporal **FIN DE SEMANA**:
  - Reglas de resolución para expresiones de FIN DE SEMANA, PASADO y RANGO:
    - $weekendLast(x) \equiv \exists y, S[documentDate(y) \wedge interval(x) \wedge Hath(S, 2, *Day*, x) \wedge Saturday(start - of(x), (weekOfYear(y) - 1)) \wedge Sunday(end - of(x), (weekOfYear(y) - 1))]$



- Reglas de resolución para expresiones de FIN DE SEMANA, MISMO y RANGO:
  - $weekendThis(x) \equiv \exists y, S[documentDate(y) \wedge interval(x) \wedge Hath(S, 2, *Day*, x) \wedge Saturday(start - of(x), weekOfYear(y)) \wedge Sunday(end - of(x), weekOfYear(y))]$
- Reglas de resolución para expresiones de FIN DE SEMANA, FUTURO y RANGO:
  - $weekendNext(x) \equiv \exists y, S[documentDate(y) \wedge interval(x) \wedge Hath(S, 2, *Day*, x) \wedge Saturday(start - of(x), (weekOfYear(y) + 1)) \wedge Sunday(end - of(x), (weekOfYear(y) + 1))]$
- Reglas de resolución relacionadas con la unidad temporal SEMANA:
  - Reglas de resolución para expresiones de SEMANA, PASADO y CONCRETO:
    - $weekAgo(x) \equiv \exists y, z[documentDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = 1 \wedge day(x)]$
    - $weeksAgo(x) \equiv \exists y, z[documentDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n \wedge day(x)]$
    - $weekBefore(x) \equiv \exists y, z[previousDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = 1 \wedge day(x)]$
    - $weeksBefore(x) \equiv \exists y, z[previousDate(y) \wedge before(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n \wedge day(x)]$
  - Reglas de resolución para expresiones de SEMANA, PASADO y RANGO:
    - $weekLast(x) \equiv \exists y, z, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = 1]$
    - $weeksLast(x, n) \equiv \exists y, z, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n]$
    - $weekPrevious(x) \equiv \exists y, z, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = 1]$
    - $weeksPrevious(x, n) \equiv \exists y, z, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n]$
  - Reglas de resolución para expresiones de SEMANA, PASADO y FUZZY CONCRETO:
    - $weekLastF(x) \equiv \exists y, z, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = 1]$
    - $weeksLastF(x, n) \equiv \exists y, z, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n]$
    - $weekPreviousF(x) \equiv \exists y, z, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = 1]$
    - $weeksPreviousF(x, n) \equiv \exists y, z, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, n, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n]$
  - Reglas de resolución para expresiones de SEMANA, PASADO y FUZZY:



## A.1 Reglas de resolución desde la perspectiva de la lógica de 1er orden 191

- $weeksAgoF(x) \equiv \exists y, z, S[documentDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Month*) = 1 \wedge end - of(x) = y]$
- $weeksBeforeF(x) \equiv \exists y, z, S[previousDate(y) \wedge before(x, y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Month*) = 1 \wedge end - of(x) = y]$
- Reglas de resolución para expresiones de SEMANA, MISMO y RANGO:
  - $weekThis(x) \equiv \exists y, S[documentDate(y) \wedge in - interval(y, x) \wedge interval(x) \wedge Hath(S, 1, *Week*, x)]$
  - $monthThat(x) \equiv \exists y, S[previousDate(y) \wedge in - interval(y, x) \wedge interval(x) \wedge Hath(S, 1, *Week*, x)]$
- Reglas de resolución para expresiones de SEMANA, PASADO PRESENTE y RANGO:
  - $weekSince(x) \equiv \exists y, z, S[documentDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, 1, *Week*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Week*) = 1 \wedge end - of(x) = y]$
  - $weeksSince(x, n) \equiv \exists y, z, S[documentDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, 1, *Week*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Week*) = n \wedge end - of(x) = y]$
  - $weekSinceThat(x) \equiv \exists y, z, S[previousDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, 1, *Week*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Week*) = 1 \wedge end - of(x) = y]$
  - $weeksSinceThose(x, n) \equiv \exists y, z, S[previousDate(y) \wedge before(start - of(x), y) \wedge interval(x) \wedge Hath(S, 1, *Week*, x) \wedge interval - between(z, start - of(x), y) \wedge duration(z, *Week*) = n \wedge end - of(x) = y]$
- Reglas de resolución para expresiones de SEMANA, FUTURO y CONCRETO:
  - $weekWithin(x) \equiv \exists y, z[documentDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = 1 \wedge day(x)]$
  - $weeksWithin(x) \equiv \exists y, z[documentDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n \wedge day(x)]$
  - $weekLater(x) \equiv \exists y, z[previousDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = 1 \wedge day(x)]$
  - $weeksLater(x) \equiv \exists y, z[previousDate(y) \wedge after(x, y) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n \wedge day(x)]$
- Reglas de resolución para expresiones de SEMANA, FUTURO y RANGO:
  - $weeksNext(x, n) \equiv \exists y, z, S[documentDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n]$
  - $weeksFollowing(x, n) \equiv \exists y, z, S[previousDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n]$
- Reglas de resolución para expresiones de SEMANA, FUTURO y FUZZY CONCRETO:
  - $weeksNextFC(x, n) \equiv \exists y, z, S[documentDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n]$
  - $weeksFollowingFC(x, n) \equiv \exists y, z, S[previousDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, n, *Week*, x) \wedge interval - between(z, x, y) \wedge duration(z, *Week*) = n]$
- Reglas de resolución para expresiones de SEMANA, FUTURO y FUZZY:
  - $weeksNextF(x) \equiv \exists y, z, S[documentDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval - between(z, end - of(x), y) \wedge duration(z, *Month*) = 1 \wedge start - of(x) = y]$



192 A. Anexo Formalización de las reglas de resolución

- $weeksFollowingF(x) \equiv \exists y, z, S [previousDate(y) \wedge after(x, y) \wedge interval(x) \wedge Hath(S, 1, *Month*, x) \wedge interval-between(z, end-of(x), y) \wedge duration(z, *Month*) = 1 \wedge start-of(x) = y]$



Universitat d'Alacant  
Universidad de Alicante

## B. Anexo Modelo Temporal castellano

### B.1 Gramática en castellano

#### B.1.1 Reglas para la identificación de Expresiones Temporales Explícitas

Lo primero que vamos a especificar es un cuadro indicando el significado de la simbología que posteriormente será usada en las reglas (Cuadro B.1):

dd	→	valor numérico de día entre 1 y 31
mm	→	valor numérico de mes entre 1 y 12
MES	→	enero, febrero, marzo, ...
DIASEMANA	→	lunes, martes, ...
NUM	→	valor en letra o dígitos representando un número cualquiera
NumDia	→	valor en letra o dígitos representando un día
NumMes	→	valor en letra o dígitos representando un mes
NumAño	→	valor numérico representando un año
EVENTO	→	Navidad, Semana Santa, ...

Cuadro B.1. Simbología utilizada en el anexo

Una vez definida la simbología, en el Cuadro B.2 se muestran todas las reglas para la identificación de expresiones de fechas explícitas.



fecha→ dd+ '/' +mm+ '/' +(yy)yy	(12/06/1975)
fecha→ dd+ '-' +mm+ '-' +(yy)yy	(12-06-1975)
fecha→ dd+ '-' +MES+ '-' +(yy)yy	(12-junio-1975)
fecha→ dd+ 'de'+MES+ 'de'+(yy)yy	(12 de junio de 1975)
fecha→ dd+ 'de'+MES+ 'del'+(yy)yy	(12 de junio del 75)
fecha→ dd+ 'de'+mes+ ', '+ (yy)yy	(12 de junio, 1975)
fecha→ ('El')+DIASEMANA+dd+ 'de'+MES+ 'de'+(yy)yy	(El domingo 12 de junio de 1975)
hora→ hh+ ':' +mm+ ( ':' +ss)	(tiempo)

Cuadro B.2. Reglas para la identificación de expresiones explícitas en castellano

### B.1.2 Reglas para la identificación de Expresiones Temporales Implícitas

Para la gramática (Cuadro B.3) hemos utilizado la notación del analizador morfológico MACO, y hemos detectado las siguientes reglas que implican que una expresión puede ser una expresión temporal:

Además, una vez que una referencia es detectada, se comprueba que contenga alguna que implica temporalidad. A este conjunto de palabras las hemos denominado como “Palabras clave temporales” y son las que se muestran en el Cuadro B.4:

## B.2 Expresiones temporales en castellano clasificadas según la ontología temporal

En los Cuadros B.5, B.6, B.7, B.8, B.9, B.10, B.11 y B.12 se presentan las expresiones temporales en castellano que el sistema es capaz de reconocer y resolver actualmente. Las expresiones se dividen en aquellas que se resuelven utilizando la fecha del documento (FechaP) y las que se resuelven utilizando una fecha nombrada anteriormente (FechaA).



## B.2 Expresiones temporales en castellano clasificadas según la ontología temporal 195

RF->DI NC SP DA NC PR VMI	RF->DI NC SP VMI DN NC
RF->DA NC SP NC PR VMI	RF->SP DA W CC DA W
RF->DI NC SP DA NC AQ	RF->SP DA NC SP W
RF->VMI DI NC SP NC	RF->VMI DN NC SP NC
RF->SP W SP AQ NC	RF->SP DA AQ SP NC
RF->DI NC SP DD NC	RF->DA AQ NC SP NC
RF->DA PR VMI SP NC	RF->RG SP Z NC
RF->RG SP DI NC	RF->RG SP DN NC
RF->SP W CC W	RF->SP W SP W
RF->SP DD AO NC	RF->SP DD Z NC
RF->SP DD DN NC	RF->SP VMI Z NC
RF->SP VMI DN NC	RF->SP VMI DI NC
RF->SP DA Z NC	RF->SP DA AO NC
RF->SP DA DI NC	RF->DI NC SP W
RF->DI NC RG RG	RF->DN NC RG RG
RF->SP DA W NC	RF->DA Z AQ NC
RF->DA Z NC AQ	RF->DA Z AO NC
RF->DA AO Z NC	RF->DA AQ Z NC
RF->DA Z NC AQ	RF->DA NC SP NC
RF->DA DN AQ NC	RF->DA DN AO NC
RF->DA DN NC AQ	RF->DA AO DN NC
RF->DA AQ DN NC	RF->DA DN NC AQ
RF->SP DA Z	RF->SP DN NC
RF->SP Z NC	RF->SP DI NC
RF->SP NC NC	RF->DA W NC
RF->DA Z NC	RF->RG SP NC
RF->DD DI NC	RF->SP DA NC
RF->SP DD NC	RF->SP NC AQ
RF->SP VMI NC	RF->DA AQ NC
RF->DA NC AQ	RF->DA AO NC
RF->DI NC RG	RF->VMI DI NC
RF->VMI DN NC	RF->VMI Z NC
RF->DN NC RG	RF->Z NC RG
RF->Z NC AQ	RF->Z Fd Z
RF->SP DA W	RF->SPSOO W
RF->DA W	RF->DD NC
RF->SP NC	RF->SP W
RF->NC RG	RF->NC AQ
RF->VMI NC	RF->Z NP
RF->RG	RF->NC
RF->W	RF->Z

Cuadro B.3. Reglas de la gramática en castellano con nomenclatura MACO



1999	abril
agosto	año
años	ayer
día	días
diciembre	domingo
domingos	enero
febrero	fin de semana
hoy	jueves
julio	junio
lunes	mañana
martes	marzo
mayo	mes
meses	miércoles
navidad	navidades
noviembre	octubre
pasado mañana	sábado
sábados	semana
semanas	septiembre
viernes	

**Cuadro B.4.** Listado de palabras clave temporales en castellano

<p>1.DIA,PASADO,CONCRETA SIMPLE  <i>FECHAP</i>  ayer, el pasado día, el día pasado, el último día, hace un día  hace NUM días, anteayer, anoche, de ayer, el pasado día NumDia  durante el día de ayer, durante todo el día de ayer,  <i>FECHAA</i>  un día antes, el día anterior, NUM días antes, durante la noche,</p> <p>2.DIA, PASADO, CONCRETA PERIODO  <i>FECHAP</i>  los NUM días pasados, los pasados NUM días, los últimos NUM días  los NUM últimos días, después de NUM días  <i>FECHAA</i>  los NUM días anteriores, durante NUM días,</p> <p>3.DIA, PASADO, DIFUSA PERIODO  <i>FECHAP</i>  hace unos días, hace días, durante días, los días pasados  los últimos días, desde hace días, estos últimos días  durante estos últimos días, los pasados días,  <i>FECHAA</i>  unos días antes, varios días antes, días antes, días anteriores  los días anteriores, en esos días, durante unos días,</p> <p>4.DIA, PRESENTE, CONCRETA SIMPLE  <i>FECHAP</i>  hoy, esta madrugada, en las últimas horas  esta mañana, esta noche, de la madrugada, las últimas horas  la pasada madrugada, las próximas horas, hoy DIASEMANA  la jornada de hoy,  <i>FECHAA</i>  horas después, a lo largo del día, aquella noche, durante el día  durante varias horas, horas antes, pocas horas después,</p> <p>5.DIA, PASADO PRESENTE, CONCRETA PERIODO  <i>FECHAP</i>  desde hace NUM días, en estos NUM días, durante estos NUM días,  <i>FECHAA</i>  en esos NUM días,</p> <p>6.DIA, FUTURO, CONCRETA SIMPLE  <i>FECHAP</i>  mañana, pasado mañana, el próximo día  dentro de un día, el día próximo, dentro de NUM días  el próximo día NumDia,  <i>FECHAA</i>  al día siguiente, NUM días después, un día después,  el día siguiente,</p>
---

Cuadro B.5. Expresiones temporales en castellano (I)



7.DIA, FUTURO, CONCRETA PERIODO  
*FECHAP*  
 los NUM próximos días,  
*FECHAA*  
 los NUM días siguientes,  
 8.DIA, FUTURO, DIFUSA PERIODO  
*FECHAP*  
 dentro de unos días, dentro de días, los próximos días,  
*FECHAA*  
 los días siguientes, los siguientes días, días después,  
 9.SEMANA, PASADO, CONCRETA SIMPLE  
*FECHAP*  
 hace una semana, hace NUM semanas,  
*FECHAA*  
 una semana antes, NUM semanas antes,  
 10.SEMANA, PASADO, CONCRETA PERIODO  
*FECHAP*  
 la semana pasada, la pasada semana, la última semana,  
 las NUM semanas pasadas,  
*FECHAA*  
 las NUM semanas anteriores, durante una semana,  
 durante NUM semanas,  
 11.SEMANA, PASADO, DIFUSA SIMPLE  
*FECHAP*  
 un día de la semana pasada, un día de hace NUM semanas,  
*FECHAA*  
 un día de la semana anterior,  
 12.SEMANA, PASADO, DIFUSA PERIODO  
*FECHAP*  
 hace semanas, desde hace semanas, las últimas semanas,  
 hace algunas semanas,  
*FECHAA*  
 varias semanas antes, las anteriores semanas, las semanas anteriores  
 durante semanas, durante varias semanas, durante algunas semanas  
 durante unas semanas,  
 13.SEMANA, PRESENTE, CONCRETA PERIODO  
*FECHAP*  
 durante la semana, durante esta semana, esta semana,  
*FECHAA*  
 durante esa semana,  
 14.SEMANA, PASADO PRESENTE, CONCRETA PERIODO  
*FECHAP*  
 desde hace una semana, desde hace NUM semanas, en estas NUM semanas,  
*FECHAA*  
 en esas NUM semanas,

Cuadro B.6. Expresiones temporales en castellano (II)



15.SEMANA, FUTURO, CONCRETA PERÍODO  
*FECHAP*  
 la próxima semana, la semana próxima,  
*FECHAA*  
 la semana siguiente, NUM semanas después, una semana después,  
 16.SEMANA, FUTURO, DIFUSA SIMPLE  
*FECHAP*  
 un día de la semana próxima, un día de la semana que viene,  
*FECHAA*  
 un día de la semana siguiente, un día de la siguiente semana,  
 17.SEMANA, FUTURO, DIFUSA PERÍODO  
*FECHAP*  
 las próximas semanas, las semanas próximas,  
*FECHAA*  
 varias semanas después,  
 18.MES, PASADO, CONCRETA SIMPLE  
*FECHAP*  
 hace un mes, hace NUM meses, el pasado NumDia de MES  
 el pasado NumDia de MES de NumAño, el NumDia del mes pasado  
 el pasado día NumDia de MES, hasta el NumDia de MES  
 desde el NumDia de MES del NumAño,  
*FECHAA*  
 un mes antes, NUM meses antes,  
 19.MES, PASADO, CONCRETA PERÍODO  
*FECHAP*  
 el pasado mes, el mes pasado, los NUM meses pasados  
 los pasados NUM meses, el último mes, los últimos NUM meses  
 el pasado MES, en MES pasado, el MES pasado  
 hasta MES pasado, el pasado mes de MES, el MES  
*FECHAA*  
 el mes anterior, los NUM meses anteriores, durante NUM meses,  
 20.MES, PASADO, DIFUSA SIMPLE  
*FECHAP*  
 un día del mes pasado, un día de MES pasado  
 un día de hace NUM meses, una semana del mes pasado  
*FECHAA*  
 un día del mes anterior, una semana del mes anterior,  
 21.MES, PASADO, DIFUSA PERÍODO  
*FECHAP*  
 hace meses, hace unos meses, hace unos cuantos meses,  
 los últimos meses, hacía meses,  
*FECHAA*  
 varios meses antes, los anteriores meses, los meses anteriores,  
 durante meses, durante varios meses,  
 en estos meses, tras meses,

Cuadro B.7. Expresiones temporales en castellano (III)



<p>22.MES, PRESENTE, CONCRETA PERIODO  <i>FECHAP</i>  este mes, este mismo mes, en MES, en MES de NumAño,  entre MES de NumAño, de MES,  MES de NumAño, del mismo mes de NumAño,  el mismo mes de NumAño, el mes de MES,  del mes de MES, a MES de NumAño, a MES del NumAño,  en MES del NumAño, de MES de NumAño,  de MES del NumAño,  <i>FECHAA</i>  ese mes, ese mismo mes, aquel mes,  en el mismo mes, mismo mes de NumAño,</p> <p>23.MES, PASADO PRESENTE, CONCRETA PERIODO  <i>FECHAP</i>  desde hace un mes, desde hace NUM meses, en estos NUM meses,  desde MES, desde MES de NumAño,  desde el NumDia de MES de NumAño, desde MES del NumAño,  <i>FECHAA</i></p> <p>24.MES, FUTURO, CONCRETA SIMPLE  <i>FECHAP</i>  dentro de un mes, dentro de NUM meses,  el próximo NumDia de MES, el próximo NumDia de MES de NumAño,  el NumDia de MES próximo,  <i>FECHAA</i>  NUM meses después, un mes después, al mes siguiente,  el mes siguiente, un mes más tarde,</p> <p>25.MES, FUTURO, CONCRETA PERIODO  <i>FECHAP</i>  el próximo mes, el mes próximo, los próximos NUM meses,  los NUM meses próximos, el mes que viene, el próximo MES,  el MES próximo, el próximo MES de NumAño,  en MES próximo,  <i>FECHAA</i>  los NUM meses siguientes, en MES siguiente,</p> <p>26.MES, FUTURO, DIFUSA SIMPLE  <i>FECHAP</i>  un día del mes próximo, un día del mes que viene,  <i>FECHAA</i>  un día del mes siguiente, un día del siguiente mes,</p> <p>27.MES, FUTURO, DIFUSA PERIODO  <i>FECHAP</i>  los próximos meses, los meses próximos,  <i>FECHAA</i>  varios meses después, meses después,</p>
--

Cuadro B.8. Expresiones temporales en castellano (IV)



## B.2 Expresiones temporales en castellano clasificadas según la ontología temporal 201

## 28. AÑO, PASADO, CONCRETA SIMPLE

*FECHAP*

hace un año, hace NUM años, el NumDia de MES del año pasado,

*FECHAA*

un año antes, NUM años antes,

## 29. AÑO, PASADO, CONCRETA PERIODO

*FECHAP*

el pasado año, el año pasado, el último año,  
los NUM años pasados, los pasados NUM años, los últimos NUM años,  
en MES del año pasado, en MES del pasado año,

*FECHAA*

el año anterior, los NUM años anteriores, durante NUM años,  
durante un año,

## 30. AÑO, PASADO, DIFUSA SIMPLE

*FECHAP*

un día del año pasado, un día de NumAño,

un día de MES del año pasado,

*FECHAA*

un día del año anterior, una semana del año anterior,  
un mes del año anterior

## 31. AÑO, PASADO, DIFUSA PERIODO

*FECHAP*

hace años, hace unos años, hace unos cuantos años,  
los últimos años, desde hace años,

*FECHAA*

varios años antes, años antes, durante muchos años,  
en años anteriores, durante años,

## 32. AÑO, PRESENTE, CONCRETA PERIODO

*FECHAP*

este mismo año, en NumAño, de NumAño, el año NumAño,  
del presente año, del NumAño, el NumAño,  
a NumAño, entre NumAño, NumAño, a principios de NumAño,  
a principios del NumAño, de NumAño a NumAño, desde NumAño  
a NumAño, desde NumAño hasta NumAño,  
entre el pasado NumDia de MES y el NumDia de MES,  
entre MES y MES, entre MES de NumAño y MES de NumAño,  
entre el NumDia de MES de NumAño y el NumDia de MES de NumAño,  
entre el NumDia de MES y el NumDia de MES,  
de MES a MES, de MES hasta MES, desde MES hasta MES,  
desde MES a MES, del NumDia de MES al NumDia de MES,  
del NumDia de MES de NumAño al NumDia de MES de NumAño,  
desde el NumDia de MES de NumAño hasta el NumDia de MES de NumAño,  
desde el NumDia de MES hasta el NumDia de MES,  
de MES de NumAño a MES de NumAño, del año,  
durante NumAño, durante NumAño, entre NumAño y NumAño

*FECHAA*

ese mismo año, en el año,

Cuadro B.9. Expresiones temporales en castellano (V)



33.AÑO, PASADO PRESENTE, CONCRETA PERIODO  
*FECHAP*  
 desde hace un año, desde hace NUM años, en estos NUM años,  
*FECHAA*  
 tras estos NUM años,  
 34.AÑO, FUTURO, CONCRETA SIMPLE  
*FECHAP*  
 dentro de un año, dentro de NUM años, el NumDia del MES  
 del próximo año,  
*FECHAA*  
 NUM años después, un año después, al año siguiente,  
 el año siguiente, NUM años más tarde,  
 35.AÑO, FUTURO, CONCRETA PERIODO  
*FECHAP*  
 el próximo año, los próximos NUM años,  
 los NUM años próximos, el año que viene, el MES del próximo año,  
 el MES del año que viene, en menos de NUM años, en NUM años,  
*FECHAA*  
 los NUM años siguientes,  
 36.AÑO, FUTURO, DIFUSA SIMPLE  
*FECHAP*  
 un día del año que viene, un día del próximo año,  
*FECHAA*  
 un día del año siguiente, una semana del año siguiente,  
 un mes del año siguiente,  
 37.AÑO, FUTURO, DIFUSA PERIODO  
*FECHAP*  
 los años próximos, los próximos años,  
*FECHAA*  
 varios años después, años después,  
 38.DIASEMANA, FUTURO, CONCRETA SIMPLE  
*FECHAP*  
 el próximo DIASEMANA, el DIASEMANA próximo,  
 dentro de NUM DIASEMANA,  
*FECHAA*  
 39.DIASEMANA,PASADO, CONCRETA SIMPLE  
*FECHAP*  
 el pasado DIASEMANA, este pasado DIASEMANA, el DIASEMANA pasado,  
 hace NUM DIASEMANA, pasado DIASEMANA,  
*FECHAA*

Cuadro B.10. Expresiones temporales en castellano (VI)



## B.2 Expresiones temporales en castellano clasificadas según la ontología temporal 203

<p>40. DIASEMANA, PRESENTE, CONCRETA SIMPLE  <i>FECHAP</i>  el DIASEMANA, DIASEMANA, del DIASEMANA, al DIASEMANA,  este DIASEMANA,  <i>FECHAA</i></p> <p>41. EVENTO, PASADO, CONCRETA PERIODO  <i>FECHAP</i>  el pasado EVENTO, la pasada EVENTO, hace NUM EVENTO,  <i>FECHAA</i></p> <p>42. EVENTO, PASADO, DIFUSA SIMPLE  <i>FECHAP</i>  un día del pasado EVENTO, un día de la pasada EVENTO,  <i>FECHAA</i></p> <p>43. EVENTO, PRESENTE, CONCRETA PERIODO  <i>FECHAP</i>  esta EVENTO, en EVENTO, este EVENTO, durante el verano de NumAño,  <i>FECHAA</i>  ese EVENTO,</p> <p>44. EVENTO, PRESENTE, DIFUSA SIMPLE  <i>FECHAP</i>  un día de este EVENTO, un día de esta EVENTO,  <i>FECHAA</i>  un día de ese EVENTO, un día de esa EVENTO,</p> <p>45. EVENTO, FUTURO, CONCRETA PERIODO  <i>FECHAP</i>  el próximo EVENTO, la próxima EVENTO, dentro de NUM EVENTO,  <i>FECHAA</i></p> <p>46. EVENTO, FUTURO, DIFUSA SIMPLE  <i>FECHAP</i>  un día del próximo EVENTO, un día de la próxima EVENTO,  <i>FECHAA</i></p> <p>47. AÑO, PASADO PRESENTE, CONCRETA PERIODO  <i>FECHAP</i>  desde NumAño, lo que va de año, desde el NumAño,  <i>FECHAA</i></p> <p>48. AÑO, PRESENTE FUTURO, CONCRETA PERIODO  <i>FECHAP</i>  lo que queda de año, hasta el NumAño, hasta NumAño,  antes de NumAño,  <i>FECHAA</i></p>
--

Cuadro B.11. Expresiones temporales en castellano (VII)



B. Anexo Modelo Temporal castellano

<p>49.MES, PRESENTE FUTURO, CONCRETA PERIODO <i>FECHAP</i> lo que queda de mes, <i>FECHAA</i></p> <p>50.MES, PASADO PRESENTE, CONCRETA PERIODO <i>FECHAP</i> lo que va de mes, <i>FECHAA</i></p> <p>51.AÑO, PRESENTE, CONCRETA SIMPLE <i>FECHAP</i> el NumDia de MES, el NumDia del MES, el día NumDia de MES, el día NumDia del MES, de NumDia de MES, el NumDia de Mes de NumAño, el NumDia de MES del NumAño, el día NumDia de MES de NumAño, el día NumDia de MES del NumAño, DIASEMANA NumDia de MES, <i>FECHAA</i></p> <p>52.FINSEMANA, PASADO, CONCRETA PERIODO <i>FECHAP</i> el pasado fin de semana, hace un fin de semana, este fin de semana, hace NUM fines de semana, el fin de semana, del fin de semana, <i>FECHAA</i></p> <p>53.FINSEMANA, FUTURO, CONCRETA PERIODO <i>FECHAP</i> el próximo fin de semana, dentro de un fin de semana, dentro de NUM fines de semana, el fin de semana que viene, <i>FECHAA</i></p> <p>54.MES, PRESENTE, DIFUSA PERIODO <i>FECHAP</i> antes de MES de NumAño, <i>FECHAA</i></p>
--

Cuadro B.12. Expresiones temporales en castellano (VIII)

Universitat d'Alacant  
Universidad de Alicante

## C. Anexo Modelo Temporal inglés

### C.1 Gramática en inglés

#### C.1.1 Reglas para la identificación de Expresiones Temporales Explícitas

Lo primero que vamos a especificar es un cuadro indicando el significado de la simbología que posteriormente será usada en las reglas (Cuadro C.1):

dd	→	valor numérico de día entre 1 y 31
mm	→	valor numérico de mes entre 1 y 12
MES	→	January, February, March, ...
DIASEMANA	→	Monday, Tuesday, ...
NUM	→	valor en letra o dígitos representando un número cualquiera
NumDia	→	valor en letra o dígitos representando un día
NumMes	→	valor en letra o dígitos representando un mes
NumAño	→	valor numérico representando un año
EVENTO	→	Christmas, Easter, ...

Cuadro C.1. Simbología utilizada en el anexo

Una vez definida la simbología, en el Cuadro C.2 se muestran todas las reglas para la identificación de expresiones de fecha explícitas.



fecha	→	mm+ '/' +dd+ '/' +(yy)yy	(10/06/1975)
fecha	→	mm+ '-' +dd+ '-' +(yy)yy	(10-06-1975)
fecha	→	MES+ dd+ ', '+(yy)yy	(June 1,1999)
fecha	→	DIASEMANA+ ', '+MES+ dd+ ', '+(yy)yy	(Friday, October 1, 1999)
hora	→	hh+ ':' +mm+ (' : ' +ss)	(tiempo)

Cuadro C.2. Reglas para la identificación de expresiones explícitas en inglés

### C.1.2 Reglas para la identificación de Expresiones Temporales Implícitas

Para la gramática (Cuadros C.3, C.4 y C.5) hemos utilizado la notación del analizador morfológico MACO, y hemos detectado las siguientes reglas que implican que una expresión puede ser una expresión temporal:

Además, una vez que una referencia es detectada, se comprueba que contenga alguna que implica temporalidad. A este conjunto de palabras las hemos denominado como “Palabras clave temporales” y son las que se muestran en el Cuadro C.6:

## C.2 Expresiones temporales en inglés clasificadas según la ontología temporal

En los Cuadros C.7, C.8, C.9, C.10, C.11, C.12, C.13, C.14, C.15 y C.16 se presentan las expresiones temporales en inglés que el sistema es capaz de reconocer y resolver actualmente. Las expresiones se dividen en aquellas que se resuelven utilizando la fecha del documento (FechaP) y las que se resuelven utilizando una fecha nombrada anteriormente (FechaA).

RF->IN WP VBZ JJ IN DT NN	RF->IN DT NN IN Z W
RF->DT NPO0000 IN DT JJR NN	RF->IN DT JJ NN IN NPO0000
RF->DT NPO0000 NN IN DT NPO0000	RF->DT NPO0000 IN DT JJ NN
RF->DT NN IN DT JJ NN	RF->DT NN IN DT JJR NN
RF->DT NN IN DT VBG NN	RF->IN NPO0000 IN DT JJ NN
RF->NN NN IN DT JJR NN	RF->IN DT NN IN DT NPO0000
RF->IN DT DT NN IN NN	RF->IN DT NN IN W
RF->IN NPO0000 IN NPO0000 W	RF->DT NN IN Z NNS
RF->DT NN IN Z NN	RF->DT NN IN DI NN
RF->DT NN IN DT NN	RF->DT JJ IN DT NN
RF->DT NN IN JJ NN	RF->IN DT JJR NPO0000 NNS
RF->DT NN IN JJR NN	RF->DT NPO0000 IN JJ NN
RF->IN NPO0000 IN JJ NN	RF->DT JJ NN NPO0000 NN
RF->DT NPO0000 IN JJR NN	RF->IN NPO0000 Z Fc W
RF->IN DT JJR JJR NNS	RF->IN DT JJ NPO0000
RF->IN DT JJ JJ NN	RF->IN DT JJ Z NN
RF->DT NN IN NPO0000 Z	RF->DT NPO0000 IN JJR NPO0000
RF->DT NN IN DT NPO0000	RF->DT JJ NN IN NPO0000
RF->DT NPO0000 NN IN NPO0000	RF->IN DT NN IN NPO0000
RF->NN VBZ DT JJ NNS	RF->IN DT NPO0000 NNS IN
RF->IN NPO0000 NNS IN NPO0000	RF->IN CC IN NPO0000 NNS
RF->DT VBN NN IN NPO0000	RF->IN DT JJ NPO0000 NNS
RF->IN DT NN IN NN	RF->IN DT JJ JJ NNS
RF->IN DT W W	RF->IN DT W Z
RF->IN DT Z W	RF->IN DT JJ W
RF->IN DT NN Z	RF->IN NPO0000 Z Z
RF->IN NPO0000 Z W	RF->IN W IN W
RF->IN DT NN W	RF->IN W CC W
RF->DT JJ Z NNS	RF->DT JJ NN NNS
RF->NN IN Z NN	RF->NN IN DI NN
RF->NN IN DT NN	RF->JJ IN DT NN
RF->IN DT NPO0000 W	RF->DT JJR NPO0000 NNS
RF->DT JJ NPO0000 NNS	RF->IN DT NPO0000 VBG
RF->IN DT NN VBG	RF->IN DT JJR NPO0000
RF->IN DT NPO0000 NNS	RF->IN DT VBG NNS
RF->NN IN NPO0000 Z	RF->NN NN IN W
RF->NNS NN IN W	RF->IN DT NPO0000 Z
RF->DT JJR NPO0000 NN	RF->IN DT NNS VBG
RF->JJ NN IN NPO0000	RF->DT NN IN NN
RF->IN DT VBG NN	RF->NPO0000 NNS DT NN
RF->NPO0000 NNS JJR NN	RF->DT JJ JJ NN
RF->NN NN IN NNP	RF->NN NNS IN NNP
RF->DT NPO0000 JJR NNS	RF->DT NN IN VBZ
RF->DT NN IN NPO0000	RF->IN DT JJ NN
RF->DT NPO0000 JJ NNS	RF->NN VBZ NPO0000 NNS

Cuadro C.3: Reglas de la gramática en inglés con nomenclatura MACO(I)



RF->DT NPO0000 NNS VBG	RF->IN DT JJR NN
RF->IN DT JJ NNS	RF->DT NPO0000 VBG NNS
RF->DT VBG NPO0000 NNS	RF->VBG DT JJR NN
RF->DT NNS NPO0000 NNS	RF->IN NN NN IN
RF->DT NPO0000 JJ NN	RF->DT JJ NNS IN
RF->IN NPO0000 NN VBZ	RF->IN NPO0000 CC NPO0000
RF->IN NPO0000 IN NPO0000	RF->DT JJR JJR NNS
RF->DT NPO0000 VBN NNS	RF->DT VBN NPO0000 NNS
RF->IN DT NN IN	RF->VBZ DT JJ NNS
RF->DT NN JJR NN	RF->DT JJR NN NPO0000
RF->IN DT JJR NNS	RF->NN IN Z NNS
RF->IN NPO0000 NNS IN	RF->IN JJR NPO0000
RF->DT JJ NN NPO0000	RF->IN DT NNS IN
RF->VBD VBG NNS	RF->DT NPO0000 JJ
RF->IN DT VBN	RF->IN DT W
RF->IN NPO0000 W	RF->JJ NN JJ
RF->NN NNS IN	RF->Z NPO0000 W
RF->IN NPO0000 Z	RF->DT NPO0000 Z
RF->JJ NNS JJ	RF->Z NN IN
RF->IN DT NN	RF->DT JJR NPO0000
RF->DT NN JJ	RF->VBZ DT NN
RF->IN DT NPO0000	RF->NN IN NPO0000
RF->IN Z NNS	RF->DT JJ NN
RF->DT NPO0000 NN	RF->IN NPO0000 NN
RF->NPO0000 NNS IN	RF->DT JJR NN
RF->DT NN IN	RF->IN JJ NN
RF->DT NPO0000 IN	RF->IN JJ NPO0000
RF->DT JJ NPO0000	RF->JJR NNS JJ
RF->DT VBG NN	RF->NN NN IN
RF->IN NN NN	RF->DT VBG NNS
RF->IN NPO0000 NNS	RF->IN DT NNS
RF->DT VBN NPO0000	RF->DT NPO0000 VBN
RF->DT NPO0000 JJR	RF->DT NN NPO0000
RF->JJ NNS IN	RF->IN VBZ NNS
RF->IN NN VBG	RF->DT VBN NNS
RF->DT NNS IN	RF->DT VBN NN
RF->VBZ NPO0000 NNS	RF->IN VBG NPO0000
RF->DT NNP NN	RF->NN NN JJ
RF->IN JJ NNS	RF->IN JJR NNS
RF->VBZ NPO0000 NN	RF->DT NN RB
RF->DT NN NNS	RF->NPO0000 NNS RB
RF->IN VBN NPO0000	RF->DT NN NN
RF->NPO0000 NNS JJ	RF->NNP IN NN

Cuadro C.4. Reglas de la gramática en inglés con nomenclatura MACO (II)



## C.2 Expresiones temporales en inglés clasificadas según la ontología temporal 209

RF->VBZ DT NNS	RF->JJ NPO0000 IN
RF->DT NPO0000 NNS	RF->IN JJR NN
RF->IN NNP NPO0000	RF->DT JJ NNS
RF->JJR NNS IN	RF->DT JJR NNS
RF->JJ NN IN	RF->Z NNS IN
RF->IN NNS VBG	RF->NNP Z
RF->Z NNS	RF->Z NN
RF->NNP NNP	RF->NPO0000 W
RF->NNP W	RF->NN W
RF->VBZ NNS	RF->VBN NN
RF->DT VBN	RF->JJ NPO0000
RF->VBN NPO0000	RF->NNP NPO0000
RF->DT NNS	RF->JJ NNS
RF->NNS IN	RF->JJR NNS
RF->JJ NN	RF->DT NPO0000
RF->IN NPO0000	RF->IN NN
RF->JJR NN	RF->NNS JJ
RF->DT NN	RF->IN NNS
RF->NN Z	RF->NPO0000 Z
RF->IN W	RF->Z
RF->W	RF->NNP
RF->JJR	RF->NPO0000
RF->NN	

Cuadro C.5. Reglas de la gramática en inglés con nomenclatura MACO (III)

1999	April
August	year
years	yesterday
day	days
December	Sunday
January	February
weekend	today
Thursday	July
June	Monday
tomorrow	Tuesday
March	May
month	months
Wednesday	Christmas
November	October
Saturday	week
weeks	September
Friday	

Cuadro C.6. Listado de palabras clave temporales en inglés



<p>1.DIA,PASADO,CONCRETA SIMPLE  <i>FECHAP</i>  yesterday,the past day,the last day,for a day,it does NUM days  the day before yesterday,last night,of yesterday  during the day of yesterday,the passed day,last day  a day ago,does NUM days,day before yesterday  during all the day of yesterday,the day past,makes a day  makes NUM days <i>FECHAA</i>  a day before,the previous day,NUM days before  during the night,NUM days ago,the day before  one day before,that NUM days before,in the day before</p> <p>2.DIA, PASADO, CONCRETA PERIODO  <i>FECHAP</i>  the NUM last days,the last NUM days,after NUM days  the NUM passed days,the passed NUM days,the NUM past days  the pasts NUM days,in the last NUM days  <i>FECHAA</i>  the NUM previous days,during NUM days,in the past NUM days  the past NUM days,during the past NUM days  the previous NUM days,in the NUM days before</p> <p>3.DIA, PASADO, DIFUSA PERIODO  <i>FECHAP</i>  for days,during days,the last days,these last days  during these last days,some days ago,days ago  the passed days,for days,makes some days,makes days  the past days,from makes days,during these past few days  the day pasts,these days,during the last days  in the last days,in the final days,over the past few days</p> <p><i>FECHAA</i>  days before,several days before,previous days  the previous days,in those days,some days before  during some days,in the days before,several days ago  in the previous days</p> <p>4.DIA, MISMO, CONCRETA SIMPLE  <i>FECHAP</i>  today,this dawn,in the last hours,this morning,tonight  of the dawn,the last hours,the last dawn,the next hours  today <i>DIASEMANA</i>,the today day,this early morning,this  tomorrow,this night,of the early morning,the day of today  this evening,this awoken,of the awoken,the session of today  <i>FECHAA</i>  hours later,throughout the day,that night,during the day  during several hours,hours before,few hours later  along the day,hours after,to what is long of the day  few hours after</p>
--

Cuadro C.7. Expresiones temporales en inglés (I)



<p>5.DIA, PASADO PRESENTE, CONCRETA PERIODO  <i>FECHAP</i>  for NUM days,in these NUM days,during these NUM days  in these days</p> <p><i>FECHAA</i>  in those NUM days,these NUM days</p> <p>6.DIA, FUTURO, CONCRETA SIMPLE  <i>FECHAP</i>  tomorrow,the day after tomorrow,the next day,within a day  within NUM days,the next NumDia day,past tomorrow  inside a day,inside NUM days,the next day NumDia  passed tomorrow,the day after,within the next day  within the day,that next day,within the next NUM days</p> <p>on the following day,NUM days later,a day later  the following day,one day after,to the following day  NUM days after,a day after,following the present day  within NUM days after</p> <p>7.DIA, FUTURO, CONCRETA PERIODO  <i>FECHAP</i>  the NUM next days,the next NUM days,NUM days</p> <p><i>FECHAA</i>  the NUM following days,in the next NUM days  in the NUM days after,the NUM days following  that NUM days after,the following NUM days</p> <p>8.DIA, FUTURO, DIFUSA PERIODO  <i>FECHAP</i>  within days,the next days,inside some days,inside days  within some days,in the days after,within the next NumDia days  in the next day,in the next days</p> <p><i>FECHAA</i>  the following days,days later,days after,after several days  in the days following,within several days</p> <p>9.SEMANA, PASADO, CONCRETA SIMPLE  <i>FECHAP</i>  for one week,it does NUM weeks,a week ago,does NUM weeks  one week ago,makes NUM weeks</p> <p><i>FECHAA</i>  one week before,NUM weeks before,a week before  the week before,NUM weeks ago,before NUM weeks</p> <p>10.SEMANA, PASADO, CONCRETA PERIODO  <i>FECHAP</i>  the last week,the NUM last weeks,last week,the passed week  the week past</p>
--

Cuadro C.8. Expresiones temporales en inglés (II)



<p><i>FECHAA</i> the NUM previous weeks,during one week,during NUM weeks in the last NUM weeks,during a week,the past NUM weeks during the last NUM weeks,in the previous NUM weeks in the past NUM weeks,in the NUM weeks before,in NUM weeks before,the previous NUM weeks,during the past NUM weeks</p> <p>11.SEMANA, PASADO, DIFUSA SIMPLE <i>FECHAP</i> a day of the last week,a day of last week <i>FECHAA</i> a day of the previous week</p> <p>12.SEMANA, PASADO, DIFUSA PERIODO <i>FECHAP</i> for weeks,the last weeks,for some weeks,weeks ago,does some weeks,makes weeks,makes some weeks,in the last week,in last weeks,in the past week,last weeks,during the past week,during last weeks,in last week,in past weeks,that last weeks,during last week,in that last week,the past weeks,for several weeks,over the past few weeks,in recent weeks <i>FECHAA</i> several weeks before,the previous weeks,during weeks,during several weeks,during some weeks,several weeks ago,in the previous week,in the weeks before,in previous weeks,in the week before,during previous weeks</p> <p>13.SEMANA, MISMO, CONCRETA PERIODO <i>FECHAP</i> during the week,during this week,this week,until this week <i>FECHAA</i> during that week</p> <p>14.SEMANA, PASADO PRESENTE, CONCRETA PERIODO <i>FECHAP</i> for NUM weeks,inthese NUM weeks,since a week ago,from one week ago,in the final NUM weeks <i>FECHAA</i> in those NUM weeks</p> <p>15.SEMANA, FUTURO, CONCRETA PERIODO <i>FECHAP</i> the next week,within the next NUM to NUM weeks,next week,one week from tonight,one week from today,one week</p>
--

Cuadro C.9. Expresiones temporales en inglés (III)

<p><i>FECHAA</i> the following week, NUM weeks later, one week later, a week later, NUM weeks after, a week after, after NUM weeks, the NUM weeks following, in the next NUM weeks</p>
<p>16. SEMANA, FUTURO, DIFUSA SIMPLE <i>FECHAP</i> a day of the next week, a day of next week <i>FECHAA</i> a day of the following week</p>
<p>17. SEMANA, FUTURO, DIFUSA PERIODO <i>FECHAP</i> the next weeks, in the weeks after, after weeks, in next weeks, in the next week, that next weeks, after NumDia weeks, in the week after, after next weeks, within the next NumDia weeks, within the next several weeks, several weeks, within weeks <i>FECHAA</i> several weeks later, the next several weeks, several weeks after, in the week following, within weeks following, the following weeks, in the following weeks</p>
<p>18. MES, PASADO, CONCRETA SIMPLE <i>FECHAP</i> for a month, it does NUM months, the past NumDia of MES, the NumDia of the last month, until the NumDia of MES, a month ago, does NUM months, the NumDia of last month, to the NumDia of MES, makes NUM months, NUM months ago, before the MES NumDia <i>FECHAA</i> a month before, NUM months before, the month before, before MES, the last one MES, before the MES, one month before, before that MES, before last MES, last MES before, before MES NumAño, one month ago, before the MES NumAño, the MES before</p>
<p>19. MES, PASADO, CONCRETA PERIODO <i>FECHAP</i> the past month, the last month, the NUM last months, the last NUM months, in last MES, until last MES, the past month of MES, the MES, the passed month, last month, the passed NUM months, in passed MES, to passed MES, the passed month of MES, the MES, the pasts NUM months, in the last NUM months, last MES <i>FECHAA</i> the previous month, the NUM previous months, during NUM months the month previous</p>
<p>20. MES, PASADO, DIFUSA SIMPLE <i>FECHAP</i> a week of last month, a LAST MES day, a week of the last month <i>FECHAA</i> a day of the previous month</p>

Cuadro C.10. Expresiones temporales en inglés (IV)



## 21. MES, PASADO, DIFUSA PERIODO

*FECHAP*

for months, it does a few months, the last months, months ago, some months ago, does a few months, makes months, makes some months, makes a few months, was making months, in the last month, in the past month, in last MES, during the past month, during last month

*FECHAA*

several months before, the previous months, during months, during several months, in these months, after months, various months before, during various months, in the previous month, in the previous MES, during the previous month, in the month before

## 22. MES, MISMO, CONCRETA PERIODO

*FECHAP*

this month, this same month, in MES, in MES of NumAño, of MES, MES of NumAño, of the same month of NumAño, the same month of NumAño, between MES of NumAño, the month of MES, of the month of MES, to MES of NumAño, to MES of the NumAño, in MES of the NumAño, of MES of NumAño, of MES of the NumAño, from MES to MES, from MES until MES, until MES NumDia, until MES, to MES, from MES NumDia to MES, to the MES, from this MES to MES, on MES, until this month, until MES NumDia NumAño, in MES, in the MES, in MES NumAño, by MES, MES NumAño

*FECHAA*

that month, that same month, in the same month, same month of NumAño

## 23. MES, PASADO PRESENTE, CONCRETA PERIODO

*FECHAP*

for NUM months, in these NUM months, from MES, from MES of NumAño, from MES of the NumAño, since a month ago, from a month ago, since MES, these MES, from NUM months

*FECHAA*

## 24. MES, FUTURO, CONCRETA SIMPLE

*FECHAP*

within a month, within NUM months, the next NumDia of MES, the NumDia of next MES, inside a month, inside NUM months, within the next month, within the month

*FECHAA*

NUM months later, a month later, the following month, monthly following, NUM months after, a month after, to the following month a month afterwards, one month after, within one month

Cuadro C.11. Expresiones temporales en inglés (V)



<p>25.MES, FUTURO, CONCRETA PERIODO  <i>FECHAP</i>            the next month,the next NUM months,the NUM next months,the month that comes,in next MES,next month,after MES NumDia NumAño,one month from today  <i>FECHAA</i>            the NUM following months,in following MES</p>
<p>26.MES, FUTURO, DIFUSA SIMPLE  <i>FECHAP</i>            a day of the next month  <i>FECHAA</i>            a day of the following month</p>
<p>27.MES, FUTURO, DIFUSA PERIODO  <i>FECHAP</i>            the next months,in the next month  <i>FECHAA</i>            several months later,months later,various months later,several months after,months after,in the MES following</p>
<p>28.AÑO, PASADO, CONCRETA SIMPLE  <i>FECHAP</i>            for a year,it does NUM years,a year ago,the NumDia of MES of last year,makes NUM years,NUM years ago  <i>FECHAA</i>            a year before,NUM years before</p>
<p>29.AÑO, PASADO, CONCRETA PERIODO  <i>FECHAP</i>            the past year,the last year,the NUM last years,the last NUM years,in MES of the last year,the passed year,last year,the passed NUM years,in MES of last year,the NUM last year,the pasts NUM years,in MES of the past year,nearly NUM years,the past NUM years  <i>FECHAA</i>            the previous year,the NUM previous years,during NUM years during a year,the year before</p>
<p>30.AÑO, PASADO, DIFUSA SIMPLE  <i>FECHAP</i>            a day of NumAño,a day of last year,a last year MES day  <i>FECHAA</i>            a day of the previous year,one week of the previous year a month of the previous year,a week of the previous year</p>

Cuadro C.12. Expresiones temporales en inglés (VI)

<p>31. AÑO, PASADO, DIFUSA PERIODO  <i>FECHAP</i>            for years, it does a few years, the last years, years ago            some years ago, makes years, a few years ago, in recent years  <i>FECHAA</i>            several years before, years before, during many years            in previous years, during years, various years before</p>
<p>32. AÑO, MISMO, CONCRETA PERIODO  <i>FECHAP</i>            this same year, in NumAño, of NumAño, the NumAño year, of the present            year, of the NumAño, the NumAño, to NumAño, between NumAño, NumAño, of            NumAño to NumAño, from NumAño to NumAño, between NumAño and            NumAño, between MES and MES, of MES to MES, of MES until            MES, from MES to MES, of the year, the year NumAño, among            NumAño, from NumAño until NumAño, from MES until MES, this            year, in the year NumAño, on the NumAño, in the NumAño, during the            NumAño, during the period of NumAño NumAño, from the NumAño NumAño  <i>FECHAA</i>            that same year, in the year</p>
<p>33. AÑO, PASADO PRESENTE, CONCRETA PERIODO  <i>FECHAP</i>            for NUM years, in these NUM years, since a year ago, from a year ago  <i>FECHAA</i>            after these NUM years</p>
<p>34. AÑO, FUTURO, CONCRETA SIMPLE  <i>FECHAP</i>            within a year, within NUM years, inside a year, inside NUM years            NUM years from now  <i>FECHAA</i>            NUM years later, a year later, the following year,            NUM years after, a year after, per year following,            to the following year, NUM years afterwards</p>
<p>35. AÑO, FUTURO, CONCRETA PERIODO  <i>FECHAP</i>            the next year, the next NUM years, the NUM next years, next year, the            MES of the next year, in NUM years, the MES of next year, in less            than NUM years, the year that comes, one year from today, one year  <i>FECHAA</i>            the NUM following years, for NUM more years</p>

Cuadro C.13. Expresiones temporales en inglés (VII)



## C.2 Expresiones temporales en inglés clasificadas según la ontología temporal 217

<p>36. AÑO, FUTURO, DIFUSA SIMPLE  <i>FECHAP</i>            a day of next year  <i>FECHAA</i>            a month of the following year</p>
<p>37. AÑO, FUTURO, DIFUSA PERIODO  <i>FECHAP</i>            the next years, in a few years  <i>FECHAA</i>            several years later, years later, various years later            several years after, years after</p>
<p>38. DIASEMANA, FUTURO, CONCRETA SIMPLE  <i>FECHAP</i>            the next DIASEMANA, the next one DIASEMANA, the DIASEMANA next            next DIASEMANA  <i>FECHAA</i></p>
<p>39. DIASEMANA, PASADO, CONCRETA SIMPLE  <i>FECHAP</i>            the past DIASEMANA, this last DIASEMANA, the last            DIASEMANA, DIASEMANA, passed DIASEMANA, the DIASEMANA passed            this past DIASEMANA, past DIASEMANA  <i>FECHAA</i>            before DIASEMANA, before last DIASEMANA, the DIASEMANA before</p>
<p>40. DIASEMANA, MISMO, CONCRETA SIMPLE  <i>FECHAP</i>            the DIASEMANA, of the DIASEMANA, to the DIASEMANA, this DIASEMANA            on DIASEMANA, until DIASEMANA  <i>FECHAA</i></p>
<p>41. EVENTO, PASADO, CONCRETA PERIODO  <i>FECHAP</i>            the past EVENT, the last EVENT, the passed EVENT, the EVENT past  <i>FECHAA</i></p>
<p>42. EVENTO, PASADO, DIFUSA SIMPLE</p>
<p>43. EVENTO, MISMO, CONCRETA PERIODO  <i>FECHAP</i>            in EVENT, this EVENT, to EVENT, to the EVENT, until EVENT            NumAño, until EVENTO  <i>FECHAA</i>            that EVENT</p>

Cuadro C.14. Expresiones temporales en inglés (VIII)



<p>44.EVENTO, MISMO, DIFUSA SIMPLE <i>FECHAP</i> a day of this EVENT <i>FECHAA</i> a day of that EVENT</p> <p>45.EVENTO, FUTURO, CONCRETA PERIODO <i>FECHAP</i> the next EVENT,within NUM EVENT,inside NUM EVENT <i>FECHAA</i></p> <p>46.EVENTO, FUTURO, DIFUSA SIMPLE <i>FECHAP</i> un día del próximo EVENTO, un día de la próxima EVENTO, <i>FECHAA</i></p> <p>47.AÑO, PASADO PRESENTE, CONCRETA PERIODO <i>FECHAP</i> from NumAño,from the NumAño,since the NumAño,since NumAño <i>FECHAA</i></p> <p>48.AÑO, PRESENTE FUTURO, CONCRETA PERIODO <i>FECHAP</i> until the NumAño,until NumAño,to the NumAño <i>FECHAA</i></p> <p>49.MES, PRESENTE FUTURO, CONCRETA PERIODO <i>FECHAP</i> lo que queda de mes, <i>FECHAA</i></p> <p>50.MES, PASADO PRESENTE, CONCRETA PERIODO <i>FECHAP</i> lo que va de mes, <i>FECHAA</i></p>
---

Cuadro C.15. Expresiones temporales en inglés (IX)



<p>51. AÑO, MISMO, CONCRETA SIMPLE</p> <p><i>FECHAP</i></p> <p>the NumDia of MES, the NumDia of the MES, the NumDia day of MES, the NumDia day of the MES, of NumDia of MES, DIASEMANA NumDia of MES, the day NumDia of MES, on the MES NumDia, on MES NumDia, MES NumDia, on MES NumDia NumAño, for MES NumDia NumAño, NumDia MES NumAño</p> <p><i>FECHAA</i></p>
<p>52. FINSEMANA, PASADO, CONCRETA PERIODO</p> <p><i>FECHAP</i></p> <p>the past weekend, for a weekend, this weekend, the weekend, of the weekend, the passed weekend, does a weekend, the weekend past, makes a weekend, makes NUM weekend</p> <p><i>FECHAA</i></p>
<p>53. FINSEMANA, FUTURO, CONCRETA PERIODO</p> <p><i>FECHAP</i></p> <p>the next weekend, within a weekend, within NUM week ends, the weekend that comes, inside a weekend, within NUM weekend the weekend that comes</p> <p><i>FECHAA</i></p>
<p>54. MES, MISMO, DIFUSA PERIODO</p> <p><i>FECHAP</i></p> <p><i>FECHAA</i></p>

Cuadro C.16. Expresiones temporales en inglés (X)



Universitat d'Alacant  
Universidad de Alicante



## Referencias

- Aerotext. 2004. *Lockheed Martin*. <http://www.lockheedmartin.com>.
- Agirre, E., & Martínez, D. 2004. Smoothing and Word Sense Disambiguation. *Pages 360–371 of: González, José Luis Vicedo, Martínez-Barco, Patricio, Muñoz, Rafael, & Saiz-Noeda, Maximiliano (eds), EsTAL. Lecture Notes in Artificial Intelligence*, vol. 3230. Springer.
- Allan, J., Connell, M.E., Croft, W.B., Feng, F., Fisher, D., & Li, X. 2000. IN-QUERY and TREC-9. *Pages 551–562 of: Ninth Text REtrieval Conference*. NIST Special Publication, vol. 500-249. National Institute of Standards and Technology.
- Allen, J. 1995. *Natural Language Understanding*. The Benjamin/Cummings Publishing Company, Inc.
- Allen, J., Byron, D., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. 2000. An Architecture for a Generic Dialogue Shell. *NLENG: Natural Language Engineering, Cambridge University Press*, 6.
- Allen, J.F. 1983. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, 26(11), 832–843.
- Amsterdam. 2004. <http://lit.science.uva.nl/Research/>.
- Arnold, D., Balkan, L., Meijer, S., Humphreys, R., & Sadler, L. 1994. *Machine Translation: An Introductory Guide*. NCC Blackwell, Oxford.
- Atserias, J., Carmona, J., Castellón, I., Cervell, S., Civit, M., Màrquez, L., Martí, M.A., Padró, L., Placer, R., Rodríguez, H., Taulé, M., & Turmo, J. 1998. Morphosyntactic Analysis and Parsing of Unrestricted Spanish Text. *Pages 1267–1272 of: Proceedings of First International Conference on Language Resources and Evaluation (LREC'98)*.
- Baldwin, B., & Ross, A. 2001. Baldwin Language Technology's DUC Summarization System. *In: Proceedings of the 1st Document Understanding Conference*.
- Bernsen, N.O., Dybkjaer, H., & Dybkjaer, L. 1998. *Designing interactive Speech Systems: From first Ideas to User Testing*. Berlin/New York: Springer.
- Boguraev, B., & Neff, M.S. 2000. The effects of analysing cohesion on document summarisation. *Pages 76–82 of: COLING 2000, 18th International Conference on Computational Linguistics*. Universität des Saarlandes, Saarbrücken, Germany: Morgan Kaufmann.
- Breck, E., Burger, J., Ferro, L., Greiff, W., Light, M., Mani, I., & Rennie, J. 2000. Another Sys Called Quanda. *Pages 369–378 of: Ninth Text REtrieval Conference*. NIST Special Publication, vol. 500-249. National Institute of Standards and Technology.
- Brill, E. 1995. Unsupervised Learning of Disambiguation Rules for Part of Speech Tagging. *Pages 1–13 of: Yarovsky, David, & Church, Kenneth (eds), Proceedings of the Third Workshop on Very Large Corpora*. Somerset, New Jersey: Association for Computational Linguistics.



## 222 Referencias

- Brown, P.F., Cocke, J., Pietra, S.A.D., Pietra, V.J.D., Jelinek, F., Lafferty, J.D., Mercer, R.L., & Roossin, P.S. 1990. A statistical approach to machine translation. *Computational Linguistics*, **16**(2), 79–85.
- Cardie, C., & Pierce, D. 1998. Error-Driven Pruning of TreeBank Grammars for Base Noun Phrase Identification. *Pages 218–224 of: ACL (ed), Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98)*.
- Carl, M., Iomdin, L.L., Pease, C., & Streiter, O. 2001. Towards a dynamic linkage of example-based and rule-based machine translation. *Machine Translation*, **15**(3), 223–257.
- Carletta, J., Isard, A., Isard, S., Kowtko, J.C., Doherty-Sneddon, G., & Anderson, A.H. 1997. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics*, **23**(1), 13–32.
- Carmona, J., Cervell, S., Márquez, L., Martí, M.A., Padró, L., Placer, R., Rodríguez, H., Taulé, M., & Turmó, J. 1998. Morphosyntactic Analysis and Parsing of Unrestricted Spanish Text. *In: LREC (ed), Proceedings of First International Conference on Language Resources and Evaluation, LREC 1998*.
- Carpenter, B. 2004. Phrasal Queries with LingPipe and Lucene. *In: 13th Text REtrieval Conference*. NIST Special Publication. National Institute of Standards and Technology.
- Castellón, I., Civit, M., & Atserias, J. 1998. Syntactic Parsing of Unrestricted Spanish Text. *In: Proceedings of First International Conference on Language Resources and Evaluation (LREC'98)*.
- Chomsky, N. 1980. *Rules and Representation*. New York: Columbia University Press.
- CLAC. 2004. *Universidad Concordia*. <http://www.cs.concordia.ca/CLAC/>.
- CLEF (ed). 2004 (August). *Cross-Lingual Evaluation Forum-2004, Lecture Notes in Artificial Intelligence (to be published)*.
- CLSR. 2004. *Universidad de Colorado*. <http://cslr.colorado.edu/>.
- Cormack, G.V., Clarke, C.L.A., Kisman, D.I.E., & Palmer, C.R. 1999. Fast Automatic Passage Ranking (MultiText Experiments for TREC-8). *Pages 735–742 of: Eight Text REtrieval Conference*. NIST Special Publication, vol. 500-248. National Institute of Standards and Technology.
- Cymfony. 2004. <http://www.cymfony.com/>.
- Dowty, D. 1982. Tenses, time-adverbs, and compositional semantic theory. *Linguistics and Philosophy*, 23–55.
- Dowty, D. 1986. The effects of aspectual class on the temporal structure of discourse: semantics or pragmatics? *Linguistics and Philosophy*, 37–62.
- Fernández, M.G. 2000. *Un modelo para la especificación lingüística y la gestión computacional en diálogos hombre-máquina mediante instrucciones expresadas en lenguaje natural*. Ph.D. thesis, Universidad de Sevilla, Departamento de Filología Inglesa. Facultad de Filología, Sevilla.
- Ferrández, A. 1998. *Aproximación computacional al tratamiento de la anáfora pronominal y de tipo adjetivo mediante gramáticas de unificación de huecos*. Ph.D. thesis, Universidad de Alicante, Departamento de Lenguajes y Sistemas Informáticos.
- Filatova, E., & Hovy, E. 2001. Assigning Time-Stamps to Event-Clauses. *Pages 88–95 of: ACL (ed), Proceedings of the 2001 ACL Workshop on Temporal and Spatial Information Processing*.
- Forcada, M.L., & Pérez, J.A. 2004. *Informàtica Aplicada a la Traducció: notes de classe amb exercicis i problemes resolts*. Departamento de Lenguajes y Sistemas Informáticos: Universidad de Alicante.

- Frege, G. 1892. Sobre sentido y referencia. *Pages 24-45 of: Villanueva, Luis Ml. Valdés (ed), La búsqueda del significado: Lecturas de filosofía del lenguaje.* Tecnos.
- Fuller, M., Kaszkiel, M., Zobel, J., Wilkinson, R., & Wu, M. 1999. The RMIT/CSIRO Ad Hoc, QA, Web, Interactive, and Speech Experiments at TREC-8. *Pages 549-564 of: Eight Text REtrieval Conference.* NIST Special Publication, vol. 500-248. National Institute of Standards and Technology.
- Gaizauskas, R., & Wilks, Y. 1998. Information Extraction: Beyond Document Retrieval. *Journal of Documentation*, **54**(1), 70-105.
- Hausser, R. 1999. *Foundations of Computational Linguistics.* Berlin Heidelberg New York: Springer-Verlag.
- Herrera, J., Peñas, A., & Verdejo, F. 2004 (August). Question Answering Pilot Task at CLEF 2004. *Pages 445-452 of: CLEF (ed), Cross-Lingual Evaluation Forum-2004, Lecture Notes in Computer Science (to be published).*
- Hinrichs, E. 1986. Temporal Anaphora in Discourses of English. *Linguistics and Philosophy*, 63-82.
- Hirst, G. 1981. *Anaphora in Natural Language Understanding.* Berlin: Springer-Verlag.
- Hobbs, J. R. 2002. *DAML Time homepage.* <http://www.cs.rochester.edu/ferguson/daml/>.
- Hutchins, J. 2001. *Machine translation over fifty years.* Histoire Epistémologie Langage.
- Hutchins, W., & Somers, H. 1992. *An introduction to machine translation.* Academic Press.
- ITC-irst. 2004. <http://tcc.itc.it>.
- Ittycheriah, A., Lita, L.V., Kambhatla, N., Nicolov, N., Roukos, S., & Stys, M. 2003. Identifying and Tracking Entity Mentions in a Maximum Entropy Framework. *In: ACL (ed), Proceedings of the North American Chapter Association for Computational Linguistic (NAACL) Workshop WordNet and Other Lexical Resources: Applications, Extensions and Customizations.*
- Katz, G., & Arosio, F. 2001. The Annotation of Temporal Information In Natural Language Sentences. *Pages 104-111 of: ACL (ed), Proceedings of the 2001 ACL Workshop on Temporal and Spatial Information Processing.*
- KSL, Ontología. 1991. <http://www-ksl-svc.stanford.edu:5915/>.
- Lancaster, F. 1968. *Information Retrieval System. Characteristics, Testing and Evaluation.* New York: Wiley.
- Llopis, F., & Muñoz, R. 2004. Cross-language experiments with IR-n system. *Pages 122-132 of: Peters, Carol, Gonzalo, Julio, Braschler, Martin, & Kluck, Michael (eds), CLEF. Lecture Notes in Computer Science, vol. 3237.* Springer.
- Llopis, F., Muñoz, R., Suárez, A., & Montoyo, A. 1998. EXIT: Propuesta de un sistema de extracción de información de textos notariales. *Novática*, **133**, 26-30.
- Llopis, F., Ferrández, A., & Vicedo, J.L. 2002. Utilización de pasajes de tamaño variable para mejorar el proceso de recuperación de información. *Procesamiento del Lenguaje Natural*, **28**, 89-98.
- Magnini, B., & Strapparava, C. 2000. Experiments in Word Domain Disambiguation for Parallel Texts. *In: Proceedings of the ACL Workshop on Word Senses and Multilinguality.*
- Magnini, B., Negri, M., Prevete, R., & Tanev, H. 2002. Mining Knowledge from Repeated Co-Occurrences: DIOGENE at TREC 2002. *In: Eleventh Text REtrieval Conference.* NIST Special Publication, vol. 500-251. National Institute of Standards and Technology.

- Magnini, B., Romagnoli, S., Vallin, A., Herrera, J., Peñas, A., Peinado, V., Verdejo, F., & de Rijke, M. 2003. The Multiple Language Question Answering Track at CLEF 2003. *Pages 471–486 of: Peters, Carol, Gonzalo, Julio, Braschler, Martin, & Kluck, Michael (eds), CLEF. Lecture Notes in Computer Science, vol. 3237. Springer.*
- Mani, I., & Maybury, M.T. 2001. Automatic Summarization. *Page 5 of: ACL (ed), Proceedings of the 2001 ACL Workshop on Temporal and Spatial Information Processing.*
- Martínez-Barco, P. 2001. *Resolución computacional de la anáfora en diálogos: estructura del discurso y conocimiento lingüístico*. Ph.D. thesis, Departamento de Lenguajes y Sistemas Informáticos. Universidad de Alicante.
- Martínez-Barco, P., & Palomar, M. 2000. Dialogue Structure Influence Over Anaphora Resolution. *Pages 515–525 of: Cairó, Osvaldo, Sucar, Luis Enrique, & Cantu, Francisco J. (eds), MICAI. Lecture Notes in Computer Science, vol. 1793. Springer.*
- Martínez-Barco, P., Peral, J., Ferrández, A., Moreno, L., & Palomar, M. 1998. Analizador Parcial SUPP. *Pages 329–341 of: Proceedings of VI biennial Iberoamerican Conference on Artificial Intelligence, IBERAMIA '98. Lecture Notes in Computer Science, vol. 1484. Lisboa, Portugal: Springer.*
- Martínez-Santiago, F., García-Cumbreras, M.A., & Ureña-López, L.A. 2004. The Merging Problem in Distributed Information Retrieval and the 2-Step RSV Merging Algorithm. *Pages 442–453 of: González, José Luis Vicedo, Martínez-Barco, Patricio, Muñoz, Rafael, & Saiz-Noeda, Maximiliano (eds), EsTAL. Lecture Notes in Artificial Intelligence, vol. 3230. Springer.*
- METACARTA. 2004. <http://www.kornai.com/pub.html>.
- Mihalcea, R., & Moldovan, D.I. 2001. A Highly Accurate Bootstrapping Algorithm for Word Sense Disambiguation. *International Journal on Artificial Intelligence Tools, 10(1-2), 5–21.*
- Mitkov, R. 1995. An uncertainty reasoning approach to anaphora resolution. *Pages 149–154 of: Proceedings of the Natural Language Pacific Rim Symposium.*
- Mitkov, R. 1997. Factors in anaphora resolution: they are not the only things the matter. A case study based on two different approaches. *Pages 14–21 of: ACL (ed), Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL/EACL).*
- Mitkov, R. 1998. Robust pronoun resolution with limited knowledge. *Pages 869–875 of: ACL (ed), Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL '98).*
- Moens, M., & Steedman, M. 1988. Temporal Ontology and Temporal Reference. *Computational Linguistics, 14(2), 15–28.*
- Moia, T. 2001. Telling apart temporal locating adverbials and time-denoting expressions. *In: ACL (ed), Proceedings of the 2001 ACL Workshop on Temporal and Spatial Information Processing.*
- Moldovan, D.I., Harabagiu, S.M., Girju, R., Morarescu, P., Lacatusu, V.F., Novischi, A., Badulescu, A., & Bolohan, O. 2002. LCC Tools for Question Answering. *In: Eleventh Text REtrieval Conference. NIST Special Publication, vol. 500-251. National Institute of Standards and Technology.*
- Molina, A., Pla, F., & Segarra, E. 2002. A Hidden Markov Model Approach to Word Sense Disambiguation. *Pages 655–663 of: IBERAMIA.*
- Montoyo, A., & Palomar, M. 2000a. Word Sense Disambiguation with Specification Marks in Unrestricted Texts. *Pages 103–107 of: Proceedings of 11th International Workshop on Database and Expert Systems Applications (DEXA 2000). Greenwich, London, UK: IEEE Computer Society.*

- Montoyo, A., & Palomar, M. 2000b. WSD Algorithm Applied to a NLP System . *Pages 54–65 of: Bouzeghoub, Mokrane, Kedad, Zoubida, & Métais, Elisabeth (eds), Proceedings of 5th International conference on Applications of Natural Language to Information Systems (NLDB-2000). Natural Language Processing and Information Systems. Lecture Notes in Computer Science. Versailles, France: Springer-Verlag.*
- Montoyo, A., & Palomar, M. 2001. Specification Marks for Word Sense Disambiguation: New Development. *Pages 182–191 of: Gelbukh, A. (ed), Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science, vol. 2004. Mexico City: Springer-Verlag.*
- Moreno, L., Palomar, M., Molina, A., & Ferrández, A. 1999. *Introducción al Procesamiento del Lenguaje Natural*. Alicante: Servicio de Publicaciones de la Universidad de Alicante.
- Muñoz, R., & Ferrández, A. 2000. Definite Descriptions Resolution in Spanish. *Pages 140–145 of: Proceeding of the International Conference on Artificial and Computational Intelligence for Decision, Control and Automation in Engineering and Industrial Applications (ACIDCA'2000): Corpora and Natural Language Processing.*
- Muñoz, R., Montoyo, A., Llopis, F., & Suárez, A. 1998. Reconocimiento de entidades en el sistema EXIT. *Procesamiento del Lenguaje Natural*, **23**, 47–53.
- Muñoz, R., Palomar, M., & Ferrández, A. 2000. Processing of Spanish Definite Descriptions. *Pages 526–537 of: Cairó, Osvaldo, Sucar, Luis Enrique, & Cantu, Francisco J. (eds), MICAI. Lecture Notes in Computer Science, vol. 1793. Springer.*
- Orasan, C., Mitkov, R., & Hasler, L. 2003. CAST: A computer-aided summarisation tool. *Pages 135–138 of: Association for Computational Linguistics (EACL).*
- Palomar, M., & Martínez-Barco, P. 2001. Computational approach to anaphora resolution in Spanish dialogues. *Journal of Research in Artificial Intelligence*, **15**, 263–287.
- Partee, B. 1973. Some Structural Analogies between Tenses and Pronouns in English. *Journal of Philosophy*, 601–609.
- Peral, J., & Ferrández, A. 2002. Pronominal Anaphora Generation in an English-Spanish MT Approach. *Pages 187–196 of: Gelbukh, Alexander F. (ed), CILing. Lecture Notes in Computer Science, vol. 2276. Springer.*
- Poesio, M., & Vieira, R. 1998. A Corpus-based Investigation of Definite Description Use. *Computational Linguistics*, **24**(2), 183–216.
- Pustejovsky, J. 2002. *TERQAS: Time and Event Recognition for Question Answering Systems*. <http://time2002.org/>.
- Pustejovsky, J., Wiebe, J., & Maybury, M. 2002. Multiperspective and Temporal Question Answering. *In: LREC (ed), Proceedings of the LREC Workshop on Temporal Annotation Standards, 2002.*
- Radev, D., & Sundheim, B. 2002a. *Using TimeML in Question Answering*. Tech. rept. ARDA, Brandeis University.
- Radev, D., & Sundheim, B. 2002b. *Using TimeML in Question Answering*. <http://www.cs.brandeis.edu/~jamesp/arda/time/documentation/TimeML-use-in-qa-v1.0.pdf>.
- Reichenbach, H. 1947. *Elements of Symbolic Logic*. MacMillan, London.
- Rigau, G., Atserias, J., & Agirre, E. 1997. Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation. *Pages 48–55 of: ACL (ed), Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and European Chapter Association for Computational Linguistics (ACL/EACL).*

- Rigau, G., Rodríguez, H., & Agirre, E. 1998. Building Accurate Semantic Taxonomies from Monolingual MRDs. *Pages 1103-1109 of: ACL (ed), Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98).*
- Russell, B. 1919. Descripciones. *Pages 46-56 of: Villanueva, Luis Ml. Valdés (ed), La búsqueda del significado: Lecturas de filosofía del lenguaje.* Tecnos.
- Sag, I., & Hankamer, J. 1984. Toward a Theory of Anaphoric Processing. *Linguistics and Philosophy*, June, 325-345.
- Saiz-Noeda, M., Palomar, M., & Moreno, L. 2001. Pronoun Resolution in Spanish from Full Parsing. *Pages 84-91 of: Matousek, Václav, Mautner, Pavel, Moucek, Roman, & Tauser, Karel (eds), TSD. Lecture Notes in Computer Science*, vol. 2166. Springer.
- Saquete, E., & Martínez-Barco, P. 2000. Grammar specifications for the recognition of temporal expressions. *Pages 21.1-21.7 of: MT (ed), Proceedings of Machine Translation and Multilingual applications in the new millenium, MT2000.*
- Saquete, E., Martínez-Barco, P., & Muñoz, R. 2002a. A grammar-based system to solve temporal expressions in Spanish Texts. *Pages 53-62 of: Springer-Verlag (ed), Proceedings of the Portugal for Natural Language Processing PORTAL-2002. Lecture Notes in Artificial Intelligence.*
- Saquete, E., Martínez-Barco, P., & Muñoz, R. 2002b. Recognising and Tagging Temporal Expressions in Spanish. *Pages 44-51 of: LREC (ed), Proceedings of the LREC Workshop on Temporal Annotation Standards, 2002.*
- Saquete, E., Muñoz, R., & Martínez-Barco, P. 2003a (September). Event Ordering through Temporal Expression Resolution. *Pages 417-423 of: RANLP (ed), Recent Advances in Natural Language Processing.*
- Saquete, E., Muñoz, R., & Martínez-Barco, P. 2003b. TERSEO: Temporal Expression Resolution System Applied to Event Ordering. *Pages 220-228 of: Matousek, Václav, & Mautner, Pavel (eds), TSD. Lecture Notes in Artificial Intelligence*, vol. 2807. Springer.
- Saquete, E., Martínez-Barco, P., & Muñoz, R. 2004a. Automatic multilinguality for Time Expression Resolution. *Pages 458-467 of: Monroy, Raul, Arroyo-Figueroa, Gustavo, Sucar, Luis Enrique, & Azuela, Juan Humberto Sossa (eds), MICAI. Lecture Notes in Artificial Intelligence*, vol. 2972. Springer.
- Saquete, E., Muñoz, R., Martínez-Barco, P., & Vicedo, J.L. 2004b. Comparison and Evaluation of two approaches of a Multilayered QA system applied to Temporality. *Pages 91-102 of: González, José Luis Vicedo, Martínez-Barco, Patricio, Muñoz, Rafael, & Saiz-Noeda, Maximiliano (eds), EsTAL. Lecture Notes in Artificial Intelligence*, vol. 3230. Springer.
- Saquete, E., Martínez-Barco, P., Muñoz, R., & Vicedo, J.L. 2004c. *Decomposition of Complex Temporal Questions for Question Answering Systems.* Tech. rept. DLSI.
- Saquete, E., Vicedo, J.L., Martínez-Barco, P., & Muñoz, R. 2004d (August). Evaluation of Complex Temporal Question in CLEF-QA. *In: CLEF (ed), Cross-Lingual Evaluation Forum-2004, Lecture Notes in Artificial Intelligence (to be published).*
- Saquete, E., Martínez-Barco, P., & Muñoz, R. 2004e. Evaluation of the Automatic Multilinguality for Time Expression Resolution. *Pages 25-30 of: DEXA Workshops. IEEE Computer Society.*
- Saquete, E., Muñoz, R., & Martínez-Barco, P. 2004f. Event Ordering using TERSEO system. *Pages 39-50 of: Meziane, Farid, & Métais, Elisabeth (eds), NLDB. Lecture Notes in Computer Science*, vol. 3136. Springer.

- Saquete, E., Martínez-Barco, P., Muñoz, R., & Vicedo, J.L. 2004g. Multilayered Question Answering system applied to Temporality evaluation. *Procesamiento del Lenguaje Natural*, **33**, 25–32.
- Saquete, E., Martínez-Barco, P., Muñoz, R., & Vicedo, J.L. 2004h (July). Splitting Complex Temporal Questions for Question Answering systems. *Pages 566–573 of: ACL (ed), 42nd Annual Meeting of the Association for Computational Linguistics*.
- Saquete, E., Muñoz, R., & Martínez-Barco, P. 2005. Event Ordering using TERSEO system. *Data and Knowledge Engineering Journal*, (To be published).
- Savoy, J. 2004. Combining Multiple Strategies for Effective Monolingual and Cross-Language Retrieval. *Inf. Retr.*, **7**(1-2), 121–148.
- Schilder, F., & Habel, C. 2001. From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages. *Pages 65–72 of: ACL (ed), Proceedings of the 2001 ACL Workshop on Temporal and Spatial Information Processing*.
- Setzer, A., & Gaizauskas, R. 2000a. Annotating events and temporal information in newswire texts. *In: LREC (ed), Proceedings of the Information Extraction Meets Corpus Linguistic, LREC 2000*.
- Setzer, A., & Gaizauskas, R. 2000b. Building a temporally annotated corpus for information extraction. *In: LREC (ed), Proceedings of the Information Extraction Meets Corpus Linguistic, LREC 2000*.
- Setzer, A., & Gaizauskas, R. 2001. A Pilot Study On Annotating Temporal Relations In Text. *In: ACL (ed), Proceedings of the 2001 ACL Workshop on Temporal and Spatial Information Processing*.
- Setzer, A., & Gaizauskas, R. 2002. On the Importance of Annotating Event-Event Temporal Relations in Text. *Pages 52–60 of: LREC (ed), Proceedings of the LREC Workshop on Temporal Annotation Standards, 2002*.
- Siegel, S., & Castellan, J. 1988. *Nonparametric Statistics for the Behavioral Sciences*. 2nd edn. McGraw-Hill.
- Singhal, A., & Kaszkiel, M. 2001. A Case Study in Web Search using TREC Algorithm. *In: Proceedings of the Tenth International World Wide Web Conference*.
- Smeaton, A.F. 1997. Information Retrieval: Still butting heads with natural language processing. *SCIE*, 115–138.
- Soubbotin, M.M., & Soubbotin, S.M. 2002. Use of Patterns for Detection of Likely Answer Strings: A Systematic Approach. *In: Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. National Institute of Standards and Technology.
- Steedman, M. 1997. *Temporality*. Handbook of Logic and Linguistics ( J. Van Benthem, A. ter Meulen, eds. ).
- Suarez, A. 2004. *Resolución de la ambigüedad semántica de las palabras mediante modelos de probabilidad de máxima entropía*. Ph.D. thesis, Universidad de Alicante, Departamento de Lenguajes y Sistemas Informáticos.
- Suárez, A., & Palomar, M. 2002. Desambiguación del sentido y del dominio de las palabras con modelos de probabilidad de Máxima Entropía. *Procesamiento Lenguaje Natural*, **28**(1), 45–54.
- TREC-11. 2002. *Eleventh Text REtrieval Conference*. NIST Special Publication, vol. 500-251. National Institute of Standards and Technology.
- TREC-13. 2004. *13th Text REtrieval Conference*. NIST Special Publication. National Institute of Standards and Technology.
- TREC-8. 1999. *Eight Text REtrieval Conference*. NIST Special Publication, vol. 500-248. National Institute of Standards and Technology.
- Turmo, J. 2003. Information Extraction, Multilinguality and Portability. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, **22**, 57–78.

- Vicedo, J.L., Rodríguez, H., Peñas, A., & Massot, M. 2003. Los Sistemas de Búsqueda de Respuestas desde una Perspectiva Actual. *Procesamiento del Lenguaje Natural*, 31(31), 351–367.
- Vieira, R. 2002. *How to evaluate systems against human judgment on the presence of disagreement?* [http://acdc.linguatca.pt/aval\\_conjunta/Faro2002/Renata\\_Vieira.pdf](http://acdc.linguatca.pt/aval_conjunta/Faro2002/Renata_Vieira.pdf).
- Vossen, P. 2000. EuroWordNet: Building a Multilingual Database with WordNets in 8 European Languages. *The ELRA Newsletter*, 5(1), 9–10.
- Voutilainen, A., & Järvinen, T. 1995. Specifying a Shallow Grammatical Representation for Parsing Purposes. In: *Proceedings of the 7th European Chapter of the Association for Computational Linguistics (EACL)*.
- Vázquez, S., Saquete, E., Montoyo, A., Martínez-Barco, P., & Muñoz, R. 2004. The Role of Temporal Expressions in Word Sense Disambiguation. *Pages 208–212 of: Gelbukh, Alexander F. (ed), CICLing. Lecture Notes in Computer Science*, vol. 2945. Springer.
- Webber, B. 1988. Tense as Discourse Anaphor. *Computational Linguistics*, 14(2), 61–73.
- Wilks, Y. 1997. Information Extraction as a Core Language Technology. *Pages 1–9 of: Pazienza, Maria Teresa (ed), SCIE. Lecture Notes in Computer Science*, vol. 1299. Springer.
- Wilson, G., Mani, I., Sundheim, B., & Ferro, L. 2000 (October). Robust Temporal Processing of News. *Pages 69–76 of: ACL (ed), Proceedings of the 38th Meeting of the Association of Computational Linguistics (ACL 2000)*.
- Wilson, G., Mani, I., Sundheim, B., & Ferro, L. 2001. A Multilingual Approach To Annotating And Extracting Temporal Information. *Pages 81–87 of: ACL (ed), Proceedings of the 2001 ACL Workshop on Temporal and Special Information Processing*.
- Yang, H., & Chua, T. 2002. The Integration of Lexical Knowledge and External Resources for Question Answering. In: *Eleventh Text REtrieval Conference. NIST Special Publication*, vol. 500-251. National Institute of Standards and Technology.
- Yarowsky, D. 1995. Unsupervised word sense disambiguation rivaling supervised methods. *Pages 189–196 of: Proceedings of the 33th Annual Meeting of the Association for Computational Linguistic*.