

## COMPRESIÓN DE TEXTOS: INFORMACIÓN TEMPORAL EN DOMINIOS DEPORTIVOS

*Núria Castell \* / Lowdes Díaz †*

\*Departamento LSI, Facultad de Informática  
Pau Gargallo, 5, 08028 Barcelona

†Departamento de Filología española, Facultad de Filología – UB  
Gran vía, 585, 08007 Barcelona

### Resumen

El presente trabajo es una ampliación de las técnicas de análisis, interpretación y representación de la información temporal contenida en textos, desarrolladas previamente en el sistema COTEM. Se diseñó un prototipo de sistema de comprensión de textos aplicado a un dominio específico: noticias de la sección de economía del diario El País. Este primer prototipo trataba la representación de la información temporal verbal y de las expresiones temporales como puntos e intervalos, dejando fuera las cadenas temporales. En el presente trabajo se ha considerado como dominio de aplicación las noticias de deportes con objeto de incluir el tratamiento de las cadenas en el mecanismo de interpretación y representación de textos. Este nuevo dominio ofrece además la posibilidad de considerar textos más ricos en cuanto a su estructura por lo que permiten abordar con más detalle las cuestiones relativas a la estructura del discurso.

### 1. INTRODUCCIÓN

El presente trabajo forma parte de un proyecto de comprensión automática de textos emprendido con anterioridad y que condujo al diseño del sistema COTEM [Castell, 1989]. La finalidad de dicho proyecto fue la comprensión de estructuras textuales (incluidas las temporalizadas), tomando como dominio de aplicación un corpus formado por noticias de la sección de economía del diario El País. De los textos analizados se extraen los hechos relevantes, desde el punto de vista del "mundo económico", que son almacenados para posterior consulta.

En el dominio de aplicación se establecieron una serie de criterios, basados tanto en las relaciones temporales como en el contenido semántico de algunos verbos, que permitían eliminar información no relevante para los fines del sistema: comentarios, opiniones, etc.

Para llevar a cabo el proceso de comprensión se creó una base de conocimiento que incluye información lingüística, de dominio y de carácter general. A esta base de conocimiento se le incorporó un sistema temporal para manejar la información temporal asociada a los elementos del dominio, así como las relaciones temporales entre los mismos. La base de conocimiento se usa en la fase de comprensión del texto y crece a medida que se incorporan textos analizados.

Los verbos y las expresiones temporales constituyen las principales fuentes de información temporal, tanto para la concreción de momentos temporales como para la definición de relaciones temporales. Las expresiones temporales fueron clasificadas según criterios sintácticos y semánticos. Se detectaron relaciones entre expresiones temporales y tiempos verbales y entre expresiones temporales y tipos de verbos.

La tipificación de la información temporal permite definir el esquema de interacción entre la base de conocimiento y el sistema temporal. Asimismo, el estudio de este tipo de información evidenció la relevancia de la misma tanto en el plano de la frase como en el de discurso.

En el dominio de las noticias económicas se consideraban los hechos como un todo. La descripción de un hecho consistía en la descripción de los participantes, su localización temporal, y otros datos relevantes. No se trataban hechos estructurados, es decir, aquellos que permiten un descomposición en hechos parciales, con entidad propia. Por esta razón sólo

fueron considerados dos tipos de elementos temporales: puntos (días) e intervalos (secuencias de días).

El interés por incorporar al sistema temporal un nuevo tipo de elemento más complejo, las cadenas (secuencias de intervalos no necesariamente consecutivos), provocó la selección de un dominio en el que los hechos fueran estructurables. Este nuevo dominio son las noticias deportivas.

En el siguiente apartado se presentan los nuevos objetivos planteados. En el tercer apartado describimos el marco teórico en que se sitúa nuestro trabajo. En el apartado 4 se expone brevemente el modelo de representación temporal utilizado. En el apartado 5 se plantean cuestiones generales relativas a un dominio de aplicación. La especificación de dichas cuestiones para nuestro dominio se presenta en el apartado 6. El estudio realizado hasta el momento se describe en el apartado 7 y las líneas futuras se plantean en el apartado 8.

## 2. OBJETIVOS

En la misma línea del trabajo realizado anteriormente, los objetivos planteados en esta segunda fase son los siguientes:

- La creación de una base de conocimiento en función de los hechos relevantes en el nuevo dominio seleccionado.
- El diseño de los mecanismos adecuados para la extracción de la información y su representación en esta nueva base de conocimiento.
- La descripción del funcionamiento general de las relaciones temporales entre los hechos del dominio y su relación en el texto, teniendo en cuenta consideraciones intrínsecas a los hechos (información aspectual de los verbos seleccionados) y extrínsecas (organización de los hechos en la estructura textual de acuerdo con los requisitos del propio texto – composición, estilo... – o interacción entre lo narrado y la propia narración).

En este último apartado hay que destacar el papel desempeñado por las expresiones adverbiales temporales, cuya interacción con los verbos modifica, en ocasiones, la lectura temporal–aspectual simple. También hay que considerar los cambios de acepción constantes en el léxico propio del dominio.

Queda al margen, de momento, un grupo de expresiones de carácter metafórico que, si bien tienen cabida en el dominio, no siempre presentan la misma relevancia. Dado que algunas de ellas son de alta frecuencia, su tratamiento deberá ser abordado en algún momento por el sistema.

## 3. MARCO TEÓRICO

Este trabajo se inserta parcialmente en dos líneas de análisis de la información temporal.

- Por un lado, los principios teóricos que subyacen en la Teoría de Representación del Discurso de Kamp [Kamp, 1979; Kamp, 1981; Kamp y Rohrer, 1983]. Éstos han sido aplicados al francés, con vistas a la implementación, por el grupo de investigación del IRIT de Toulouse [Borillo, Borillo y Bras, 1989]. De un modo más descriptivo otros autores han abordado el análisis del funcionamiento de los tiempos verbales del francés en el ámbito del discurso [Engel, 1990; Vet, 1980; Vet, 1985; Vet y Molendijk, 1986]. En el caso del español, ambos trabajos están por hacer y la nuestra es una primera aproximación. La versión lingüística puramente teórica de Kamp parte de la información temporal aportada por el verbo y los modificadores temporales. Éstos son también los elementos que consideramos en nuestra aplicación.

- Por otro lado, las propuestas de modelización computacional de la información temporal formulada por diversos autores [Allen, 1983; Bruce, 1972; Kahn y Gorry, 1977] también han sido contempladas en la aplicación desarrollada.

Los trabajos más destacados hasta la fecha se han ocupado de la información temporal aportada por los verbos como operadores temporales [Reichenbach, 1947], así como de la incidencia de los tiempos verbales en el conjunto del significado del texto [Kamp, Kamp y Rohrer, y Dowty, 1986]. Respecto a este último punto, la conclusión más importante es la existencia de un "sistema de tiempos relevantes", en nuestro dominio básicamente es el pretérito indefinido, que proporciona las referencias y la instrucción de avance en el eje de

localización frente a otros cuya función es la creación de situaciones (p.e. el imperfecto) sin implicar el avance de las referencias en la línea temporal. Estas referencias permiten localizar e integrar los hechos de una frase en una estructura de hechos textual o discursiva más amplia que se va construyendo progresivamente y necesariamente de acuerdo con las referencias contenidas en el discurso previo.

Inicialmente, las interpretaciones concebían los momentos temporales como puntos de un sistema [Reichenbach, 1947; Hornstein, 1977] y únicamente se contemplaba el verbo como aportador de información temporal. Sólo en trabajos posteriores se ha atribuido también una cierta relevancia a los adverbios y expresiones temporales [Hornstein, 1990; Partee, 1984] junto al tiempo y aspecto verbal. Este ensanchamiento de perspectiva se ha reflejado en sistemas de representación que tratan los momentos temporales no ya como puntos, sino como intervalos [Eurotra, 1985]; siendo la principal referencia el modelo computacional de Allen [Allen, 1983].

Debido a esa tendencia anterior, tanto los adverbios como las expresiones temporales han tenido un papel secundario, de modificadores, en estos análisis, mientras que en el terreno de los verbos se ha profundizado mucho más: hay análisis de la estructura interna (aspecto y modo de acción) [Eberle y Kasper, 1989], así como de su contribución a la representación. También son numerosos los trabajos sobre las funciones informativas específicas de algunos tiempos, especialmente para el francés (Passé Simple, Passé composé como señaladores de hechos discursivos frente al Imparfait como señalador de situaciones discursivas en la narración [Engel, 1990; Molendijk, 1985; Vet, 1980; Vet, 1985; Vet y Molendijk, 1986]).

Con todo, el panorama actual es poco sistemático en el análisis y la representación de las aportaciones de las expresiones temporales en general. En nuestro trabajo nos ocupamos de ellas y proponemos una modelización, ciñéndonos a un dominio específico de textos. Presentamos a continuación el estudio realizado. Sólo mostramos algunas cuestiones relativas a las expresiones temporales (en adelante, ETs) y a los verbos y sintagmas nominales que consideramos relevantes para el dominio, es decir, los que queremos representar, aunque obviaremos abordarlos detalladamente.

#### 4. MODELIZACIÓN TEMPORAL

El modelo de representación temporal utilizado es una ampliación del definido para COTEM [Castell, 1989], basado a su vez en los trabajos de Allen [Allen, 1983].

La modelización temporal consiste en una colección de elementos temporales de tres tipos (punto, intervalo y cadena) y las relaciones temporales entre ellos.

La definición de los mencionados elementos temporales varía en función del grado de refinamiento del modelo (nivel de granularidad). Los tres tipos considerados son:

1. Punto: es la unidad temporal. En un determinado dominio puede ser un día (p.e. en las noticias de economía); mientras en otros puede ser una hora, o un minuto (p.e. en nuestro dominio es el minuto, ya que es donde se suelen localizar los hechos "gol", "llegada de X a la meta"...). El punto es el índice de granularidad del sistema. Punto es, por extensión, toda expresión lingüística (sintagmas nominales, ETs) localizable en un momento temporal, puntual, del eje.

2. Intervalo: es una secuencia consecutiva de puntos. Por extensión, toda expresión lingüística localizable en un conjunto de puntos (pero que no es un punto en sí misma).

3. Cadena: es una secuencia de intervalos, no necesariamente consecutivos; pero también podría ser una secuencia de puntos no consecutivos y, en un caso extremo, una secuencia combinada de puntos e intervalos (cadena heterogénea). En general, aquellas expresiones lingüísticas que denotan iteración son cadenas. Y en sentido restringido, aquellas expresiones, por lo general sintagmas nominales, cuyo contenido conceptual dentro del dominio implican iteración en el tiempo.

En el dominio seleccionado, las noticias deportivas aparecidas semanalmente en el suplemento del periódico "El País", los hechos propios del dominio imponen un grado de refinamiento particular. Se han considerado noticias de fútbol, baloncesto, golf, tenis, motociclismo, fórmula uno, rallies, esquí y ciclismo. En esta presentación mostraremos

ejemplos obtenidos exclusivamente del área futbolística, por ajustarse a toda la escala de unidades descritas anteriormente.

A modo de ejemplo aplicamos a continuación las definiciones anteriores al deporte "fútbol":

1.- *punto*: un minuto. Un hecho que tiene asociado un punto en este deporte es, por ejemplo, el marcar un gol.

- Ej. 1-0. Minuto 16 Hugo transforma una falta....  
 2-0. Minuto 34 Feli marca por bajo.  
 3-0. Minuto 44 Fernando Hierro saca un pelotazo... que entra por la escuadra (...) *El País* 26-3-90

2.- *intervalo*: secuencia de minutos. Ej. "la primera parte del partido", "la segunda mitad del encuentro" o toda una "jornada" futbolística.

3.- *cadena*: secuencia no necesariamente consecutiva de intervalos o puntos. Ej. "Campeonato", "Liga" o "la presente temporada".

En cada caso, se trata de hechos y expresiones temporales que tienen asociados un elemento también temporal perteneciente a uno de los tres tipos propuestos: punto, intervalo o cadena.

## 5. NECESIDAD DE OPERAR EN UN DOMINIO

Debido al carácter heterogéneo de las ETs tanto en el terreno sintáctico como léxico, se plantea la necesidad del análisis en un ámbito temático reducido. Tratar un dominio específico reduce el nivel de aplicación, pero es un requisito necesario para conseguir un mecanismo eficaz de interpretación automática de textos, dado que, hasta la fecha, sólo en ámbitos reducidos se obtienen resultados satisfactorios. No obstante, la metodología seguida tiene un carácter general aplicable a cualquier dominio, siendo posible obtener algunas tipologías generales.

Proponemos una modelización de los dominios desde dos puntos de vista:

- a) Semántico: asociado al significado de los hechos y, por tanto, no verbal ("partido", "carrera"). Cada elemento léxico tendrá una representación propia en el dominio que puede diferir de su significado en un ámbito general.
- b) Temporal: Conceptualmente, para cada hecho, se define el tipo de elemento temporal asociado. El procedimiento que se emplea es la clasificación de los hechos del dominio en dos grupos:
  - Los no descomponibles en el tiempo: tienen asociados elementos temporales básicos (puntos o intervalos).
  - Los descomponibles en el tiempo: tienen asociados elementos temporales estructurados (cadenas).

El criterio de descomposición se refiere a la posibilidad de desarrollo del hecho en etapas en el curso del tiempo, con una cierta linealidad.

La modelización completa de un dominio (las noticias deportivas, en nuestro caso) consiste en:

- i) tipificar los hechos y agentes del dominio, estableciendo todas las relaciones conceptuales sin carácter temporal.
- ii) establecer correspondencias entre hechos, ETs y sus correspondientes elementos temporales.

El modelo así definido permitirá la interpretación del texto con menor grado de ambigüedad.

## 6. VISIÓN DEL DOMINIO SELECCIONADO

En el dominio de noticias de economía, un hecho como "inversión" no permitía descomposición, sino que se insertaba directamente en un calendario de operaciones

financieras. La única información que se extraía era la fecha del hecho y la descripción de los participantes (agentes), o la relación con otros hechos que formaban parte de la historia de algún agente.

En el caso del nuevo dominio seleccionado, los deportes y específicamente el fútbol, algunos hechos son susceptibles de descomposición. El resultado es que podemos incorporar no sólo el hecho, sino las partes que lo componen a una estructura de calendario organizada ("campeonatos", "jornadas", "temporadas"...).

Los hechos descomponibles tienen asociada una estructura temporal que sirve de referencia para la ubicación de sus componentes en la base de conocimiento, lo que resulta útil en el caso de que aparezcan desordenados cronológicamente en el texto.

Por lo que se refiere al tratamiento de los agentes, la representación de hechos simples, el tratamiento de fechas, etc., la problemática es similar en ambos dominios.

A continuación, mostramos de forma ejemplificada lo expuesto en el apartado 5. Por conocimiento de dominio, sabemos que una ET como "la primera parte de un partido" define un intervalo o un subintervalo. Pero necesitamos, además de concretar el tipo de elemento temporal de cada hecho o ET, situarlo en relación con los demás. Es decir, relacionar "un partido", o "la primera parte", con la "Liga" o con otros partidos. Incluso, si se puede concretar más, en relación con un calendario.

Las conexiones temporales podrán ser directas o indirectas, según se desprendan explícitamente del texto o bien se deduzcan por razonamiento temporal accediendo al sistema.

Para conseguir esta interpretación del texto establecemos cuáles son los elementos que aportan información temporal :

- los propios hechos (sean sintagmas nominales del tipo "partido" o verbos como "marcar"): definidos a partir del estudio que hemos realizado del dominio, basándonos en un vaciado de elementos léxicos que toman una significación temporal específica y aportan, por tanto, información sobre el tipo de elemento temporal asociado.

P.e. Un partido de fútbol dura 90 minutos repartidos en dos partes iguales de 45; un partido de básquet dura 40 en total. En cambio, en las carreras, el tiempo sólo cuenta en la llegada y funciona como índice para las puntuaciones...

- los verbos y las expresiones temporales (ETs) en general: su aportación es la concreción del momento temporal y las relaciones temporales.

Respecto a este último punto, por el momento se ha seleccionado un grupo particular de verbos cuyo significado es "marcar un gol", verbos relevantes en el dominio. También se ha hecho un vaciado de lo que constituyen las ETs en el dominio, así como sus correspondientes elementos temporales. La combinación de verbos y expresiones temporales puede realizarse siguiendo la metodología ya aplicada en COTEM [Castell, 1989; Castell y Díaz, 1991].

El análisis de las ETs permite diferenciar cuándo una ET actúa como referencia general ( la ET "ayer" en la expresión "el gol de ayer") o como el momento localizador asociado al hecho ("gol marcado por Husillos en el minuto 26 de la primera parte").

Ejemplo de combinación de la información: "gol de ayer"

- \* El hecho "gol" sabemos, por conocimiento de dominio, que tiene asociado un elemento temporal del tipo punto.

- \* "de ayer" es una ET que define un intervalo en nuestro sistema donde el nivel de granularidad se sitúa en el minuto. Por tanto, esta ET define una referencia marco en la que situar el gol.

El sistema es capaz de asignar la lectura adecuada en cada caso a partir de las descripciones dadas.

## 7. DESCRIPCIÓN DEL TRABAJO EFECTUADO

Hasta el momento se ha procedido a la selección de una muestra de carácter general que recoge información relativa a los deportes antes mencionados. Sobre esa muestra se han realizado las siguientes tareas:

1. Clasificación de elementos léxicos del dominio que tienen información temporal (sintagmas nominales): "partido", "liga", "carrera", "temporada", "edición", "jornada", "prueba", "Copa", "Premio de X", "Campeonato del Mundo de X", "ronda", "semifinales", "finales", "fase previa", "cuartos de final", "Mundial femenino de X"...

La tarea consiste en la detección de estos elementos léxicos y la asignación del tipo de elemento temporal adecuado: gol-punto, partido-intervalo, liga-cadena.

Hay que notar que la descripción varía en cada dominio y que, aunque el mismo tipo de elemento temporal pueda aparecer en otro dominio, hay palabras que adquieren una connotación temporal especial de la que carecen fuera del ámbito deportivo, ya sea ganando precisión o modificando su significado ("temporada", "jornada").

Otros términos sólo cobran significación temporal en dominios deportivos ("Copa X", "partido", "encuentro"). Esto ya se constató en el análisis de noticias de economía, donde "ejercicio" tomaba una significación específica: año fiscal que se cierra normalmente en octubre. En el terreno deportivo, en particular en el fútbol, ocurre lo mismo con "temporada" (que tiene un calendario específico que excluye el verano y va de septiembre a junio); con "jornada" (que abarca un fin de semana o un día entre semana); etc.

Entre los elementos léxicos que adquieren una connotación temporal especial en el dominio podemos retomar "partido", cuyo intervalo temporal asociado varía en función del deporte (desde 40 minutos hasta 90 en los casos del básquet y el fútbol, por ejemplo); "el Gran Premio X", que se refiere a una secuencia de carreras mundiales o puntuables con un calendario específico, tanto para motociclismo como para fórmula uno. Lo mismo ocurre con las copas de fútbol, golf, tenis...

## 2. Estudio de las expresiones temporales propiamente dichas

Siguiendo la misma metodología aplicada en COTEM, clasificamos las ETs (adverbiales o no) en dos niveles: las específicas del dominio y las no específicas. Con ejemplos del dominio seleccionado, se ilustraría así:

– independientes del dominio: "la tarde de ayer"  
"la pasada semana"  
"el próximo miércoles"

– dependientes o propias del dominio: "a los 35 minutos de partido"  
"a cinco minutos del final"  
"correspondiente a la 5a. jornada..."

Dentro de este conjunto de expresiones habría que incluir aquellas que tienen una estructura semejante pero que se basan en parámetros espaciales y no temporales:

"en la última vuelta"  
"a cuatro vueltas del final"  
"en los últimos metros"

Estas expresiones se dan en deportes como la fórmula uno, motociclismo, atletismo y parecen asociadas a deportes de carácter más individual, donde el espacio está fijado y sirve de coordenada, siendo el objetivo recorrerlo sin una limitación de tiempo. La duración se asocia entonces a la puntuación ("vuelta rápida", "primero en cruzar la línea de salida", "último en cruzar la línea de meta", "corredores con vueltas perdidas").

Por el contrario, los deportes de equipo que tienen delimitado un espacio de juego y la duración ("partidos" y no "pruebas", siendo el objetivo marcar "tantos" (goles, canastas); estos hechos se ubican en un momento determinado de la duración establecida, que es fija y sirve de coordenada ("minuto 32 de la primera parte").

Por otro lado, las ETs también se estudian en función de:

- qué elemento temporal definen (punto, intervalo, cadena)
- qué relaciones temporales pueden establecer (enlaza, acaba, durante, solapa, empieza, etc.) [Allen, 1983] entre hechos.

Ejemplo:

"última media hora de juego": define un intervalo de una duración determinada y que tiene una relación *acaba* con el intervalo asociado al hecho "partido".

## 3. Estudio de los verbos

De manera análoga a como se ha procedido en el ámbito de las ETs, se ha realizado una clasificación de los verbos para nuestro dominio. De acuerdo con el criterio de relevancia se han seleccionado los verbos cuyo significado es:

- marcar un tanto (sujeto: jugador / equipo)

- efectuar un cambio (sujeto: entrenador /equipo)
- fallar un tanto (sujeto: jugador / equipo)
- encajar un gol (sujeto: portero / equipo)
- señalar una falta (sujeto: árbitro)
- enseñar una tarjeta (sujeto: árbitro)

En la mayoría de los casos se trata de hechos localizables específicamente en la línea temporal y en el intervalo "partido". Asimismo forman parte del hecho "partido" que es su referencia-paraguas y el hecho situable en una estructura temporal más general.

Los verbos, en tanto que descripción de acciones, son dotados de una estructura de casos que es utilizada para la representación de hechos relevantes en la base de conocimiento.

#### 4. Consideraciones sobre la estructura textual

Respecto a la estructura textual, los hechos relevantes pueden presentarse de dos formas, una opcional y la otra obligatoria. Los hechos relevantes pueden aparecer en la cabecera de la noticia acompañando al sujeto de la acción y la especificación temporal. Sin embargo, su aparición es obligatoria en el cuerpo de la noticia y el tiempo verbal más frecuente es el pretérito indefinido (más raramente, el imperfecto).

La variación en el tiempo verbal se ha sistematizado en función de su posición en la estructura del texto periodístico, de las ETs que le preceden o acompañan y del tipo aspectual al que pertenece el verbo específicamente.

#### 5. Representación del conocimiento del dominio

El análisis de la muestra nos permite definir los hechos y agentes que configuran el dominio, tal y como se esbozó en el apartado 6.

Estos conceptos los representamos mediante estructuras de tipo "frame". Cada concepto se define mediante una serie de descriptores que lo caracterizan y las relaciones que lo sitúan en la red conceptual de descripción del dominio.

El valor de alguno de estos descriptores será un elemento temporal. A través de este tipo de descriptores, el sistema podrá establecer relaciones temporales entre hechos, que complementan las relaciones conceptuales.

Ejemplo: Seguimos en el ámbito del fútbol.

##### Descripción de "Campeonato"

NOMBRE: Liga, Recopa, Copa del Rey, C. de la UEFA...  
 NIVEL: Nacional, continental, internacional, mundial...  
 CALENDARIO: secuencia de jornadas  
 MOMENTO TEMPORAL: cadena  
 CLASIFICACION FINAL: lista ordenada de equipos  
 ORGANIZADOR: Entidad

##### Descripción de "Jornada"

NOMBRE: Nombre de campeonato  
 No. DE JORNADA: número  
 PARTIDOS: lista de partidos  
 MOMENTO TEMPORAL: intervalo (normalmente un día o dos)

##### Descripción de "Partido"

JORNADA/CAMPEONATO: Identificador de un objeto "jornada"  
 EQUIPO-LOCAL: nombre  
 ALINEACION LOCAL: lista personas  
 EQUIPO-VISITANTE: nombre  
 ALINEACION VISITANTE: lista personas  
 LUGAR: ciudad  
 CAMPO: nombre campo  
 ARBITRO: persona

GOLES: (número, lista de "goles")  
 MOMENTO TEMPORAL: intervalo  
 EXPULSIONES: lista de "expulsiones"  
 TARJETAS: lista de "tarjetas"

#### Descripción de "Gol"

PARTIDO/JORNADA/CAMPEONATO: identificador de un objeto "partido"  
 EQUIPO: local, visitante  
 JUGADOR: persona  
 TIEMPO: punto (minuto)

## 8. CONCLUSIONES

El presente trabajo es la continuación del estudio y la implementación realizados anteriormente y que dieron lugar a COTEM, prototipo de sistema de comprensión automática de textos.

Básicamente se han abordado dos nuevas cuestiones: la ampliación de la tipología de la información temporal, añadiendo el tipo cadena, y la consideración de textos más ricos en cuanto a estructura y que permiten la aplicación de teorías del discurso.

Se ha elegido como dominio de aplicación las noticias deportivas y, en particular, las noticias referentes al fútbol. En este nuevo dominio, siguiendo la metodología aplicada en el desarrollo de COTEM, se han realizado diversas tareas:

- Definición de los hechos relevantes del dominio y su formalización para construir la base de conocimiento adecuada.
- Estudio lingüístico del corpus para obtener y clasificar los elementos léxicos que contienen información temporal, las expresiones temporales y los verbos.
- Análisis de las estructuras textuales y del papel de la información temporal en el desarrollo del discurso. Y en consecuencia, el papel de la información temporal en el proceso de comprensión del texto.

Queda pendiente una importante tarea de formalización, tanto de las expresiones temporales (tal y como se hizo durante el desarrollo de COTEM) como de la estructuras textuales.

En una fase posterior, se trasladará el trabajo realizado a otros dominio deportivos. El objetivo final será conseguir un sistema de comprensión de noticias sobre cualquier deporte.

## 9. REFERENCIAS

ALLEN, J.F. 1983. "Maintaining knowledge about temporal intervals", *Communications of the ACM*, vol. 26, no. 11, pp. 832-843.

BORILLO, A.-BORILLO, M.-BRAS, M. 1989. "Sémantique du temps dans la phrase et dans le discours: de la théorie aux contraintes liées à l'implémentation", Comunicación presentada en el Primer Workshop *Sémantique Logique du temps, de l'espace et du Movement dans le Langage Naturel*, Toulouse, mayo, 1989.

BRUCE, B.C. 1972. "A Model for temporal references and its application in a Question answering program", *Artificial Intelligence*, vol. 3, no. 1, pp. 1-25.

CASTELL, N. 1989. "Un model pel tractament de la informació temporal en un sistema de comprensió automàtica de notícies", Tesis doctoral, Barcelona, U.P.C., 1989.

CASTELL, N.-DÍAZ, L. 1991. "Temporal Information and Narrative Understanding", Comunicación presentada en el Tercer Workshop *Semantics of Time, Space, and Movement*, Toulouse, setiembre, 1991.

DOWTY, D. 1986. "Tense and aspect in discourse", *Linguistics and Philosophy*, 9(1), Special Issue.

- EBERLE, K.-KASPER, W. 1989. "Tenses as anaphora", European Chapter of the Association of Computational Linguistics, Manchester, 1989 (ACL), pp. 43-50.
- ENGEL, D.M. 1990. "Tense and text", Room Helm Series, London: Routledge.
- EUROTRA 1985. "ELS-3 Eurotra Linguistic Specification", Report Interno de la CEE.
- HORNSTEIN, N. 1977. "Towards a theory of Tense", Linguistic Inquiry, vol. 8, no. 3, pp. 521-557.
- HORNSTEIN, N. 1990. "As Time goes long". London: Routledge.
- KAHN, K.-GORRY, G.A. 1977. "Mechanizing Temporal Knowledge", Artificial Intelligence, vol.9, no.2, pp. 39-64.
- KAMP, H. 1979. "Events, Instants and temporal Reference", en Bäuerle/Egli/von Stechow (Eds.) *Semantics from different points of view*, Berlin: De Gruyter.
- KAMP, H. 1981. "Événements, représentations discursives et référence temporelle", Langages, 64, pp. 39-64.
- KAMP, H.-ROHRER, Ch. 1983. "Tense in texts", en Bäuerle et alt. (Eds.) *Meaning, use and interpretation of language*, Berlin: De Gruyter.
- MOLENDIJK, A. 1985. "Point référentiel et imparfait", Langue Française 67, septembre, pp.78-94.
- PARTEE, B. 1984. "Nominal and Temporal Anaphora", Linguistics and Philosophy.
- REICHENBACH, H. 1947. "Elements of Symbolic Logic", New York: Academic Press.
- VET, C. 1980. "Temps, aspects et adverbess de temps en français contemporain", Genève : Droz.
- VET, C. 1985. "Universe de Discours et Universe d'enonciation", Langue Française, sept.
- VET, C.-MOLENDIKJ, A. 1986. "The Discourse Functions of the past tenses in french", en Lo Cascio, V and C. Vet (eds.) *Temporal Structure in Sentence and Discourse*, Dordrecht: Foris.