

EL MÓDULO LÉXICO EN EL SISTEMA ATLAS

Joseba Abaitua, Elisabet Cayuelas, Lluís Hernández, Xavier Lloré y Jorge Vivaldi¹
Grupo de Tratamiento del Lenguaje Natural
Fujitsu, I+D Barcelona

1. Presentación

Exponemos aquí las directrices básicas que el grupo de traducción automática de Fujitsu está aplicando en la construcción del diccionario para el sistema ATLAS. Uno de nuestros principales objetivos es que tanto la configuración como el contenido del diccionario sean independientes de la aplicación. Con ello se facilitará la reutilización del material para otras aplicaciones distintas de ATLAS y, en definitiva, servirá para que el diccionario constituya un producto en sí mismo. El diccionario tiene la arquitectura de una base de datos léxica, lo que permite una gestión eficaz de la información introducida. Los datos léxicos se actualizan siguiendo las instrucciones de un sistema de menús predefinidos que, además de guiar sobre el tipo de información requerida, realiza comprobaciones sobre los datos introducidos y añade información por defecto. De esta manera se consigue que las nuevas incorporaciones y las modificaciones sean coherentes con el resto de las entradas. El diccionario contiene en la actualidad algo más de 15.000 entradas y está previsto que alcance las 30.000 en el próximo año.

1.1. ATLAS, un Sistema de Traducción Automática

ATLAS es un sistema diseñado en los laboratorios de Fujitsu en Kawasaki, Japón (cf. Uchida et al 1985). La versión moderna del sistema data de finales de los setenta, cuando se decidió el empleo de la técnica *interlingual* en lugar de la traducción directa ensayada hasta entonces. Esta elección estuvo muy influida por los experimentos en inteligencia artificial y representación del conocimiento de la época (Charniak & Wilks 1976, Shank 1975, etc.), casi todos ellos basados en las gramáticas de casos de Fillmore.

Favorecido por la estrategia de la interlingua, ATLAS está concebido como un sistema de traducción multilingüe. En la actualidad existen versiones que traducen entre japonés, inglés y alemán, y existen prototipos para el español, francés y otras lenguas asiáticas (chino, malayo y coreano). Todavía no existe una versión operativa que integre todos los módulos formando un solo sistema. De la experiencia del trabajo entre lenguas tan dispares ha surgido una nueva definición de las primitivas de la interlingua, a la que se han comenzado a adecuar la totalidad de los módulos y prototipos. El módulo de español ha sido diseñado desde su comienzo según esta nueva versión.

1.2. La Interlingua y las Redes Semánticas

ATLAS utiliza el formalismo de las redes semánticas para representar el análisis de una oración de la lengua origen (véase una exposición más completa en Abaitua, Soler y Vivaldi, 1989a y 1989b). La red semántica ordena los conceptos jerárquicamente. El concepto nuclear de la oración figura como nodo del que cuelgan el resto de los conceptos, en una secuencia muy similar al que establecen las gramáticas de dependencias (Tesnière 1959, Schubert 1987 o Mel'čuk 1988). Una red semántica

¹ Los apartados 3, 4 y 5 han sido desarrollados por Lluís Hernández, el sintagma nominal; Xavier Lloré, subcategorización verbal; y Elisabet Cayuelas, modificación adverbial. La base de datos léxica ha sido programada en dBase IV sobre DOS para AT compatibles.

Una primera versión de este estudio fue presentada en uno de los Seminarios sobre Inteligencia Artificial que organiza el Departament de Llenguatges i Sistemes Informàtiques de la Universitat Politècnica de Catalunya, Marzo 1991.

constituye la *instanciación* de un subconjunto de los conceptos de la interlingua, aquellos que aparecen en el texto a traducir.

La suma de todos los conceptos posibles o existentes en las lenguas consideradas forman el dominio conceptual de la interlingua. Este dominio es en la práctica vastísimo y por ello está organizado mediante diccionarios: Un diccionario de conceptos y un diccionario descriptivo de tales conceptos. En las versiones operativas actuales, los diccionarios conceptuales se limitan a dominios restringidos, pero Fujitsu tiene previsto que en el futuro se utilicen los diccionarios del proyecto EDR (cf. EDR 1988).

1.3. Elementos de la Interlingua

La interlingua de ATLAS se compone de dos tipos de unidades bien definidas: un inventario finito de relaciones conceptuales (relaciones y conceptos especiales) y el diccionario de conceptos.

En el apartado 4. se muestra el repertorio de relaciones de Caso, que son utilizadas para el análisis de los constituyentes de la oración. Estas relaciones están complementadas por cuatro grupos diferentes de relaciones que tratan fenómenos como la coordinación, la subordinación y otras dependencias sintácticas. Existe, además, un grupo de relaciones que pertenecen al diccionario descriptivo de conceptos (*part_of, element_of, similar, equivalent, super_sub*) y que rara vez son instanciadas en una red semántica.

El inventario de relaciones se complementa con una colección de conceptos subsidiarios que modifican al núcleo de una red semántica indicando valores de modalidad, tiempo y aspecto.

1.4. El diccionario de conceptos y el EDR japonés

El diccionario de conceptos está siendo desarrollado por Fujitsu en Japón, dentro de un consorcio de empresas y entes públicos que forman parte del proyecto japonés de diccionario electrónico (EDR).

El diccionario de conceptos tiene dos componentes: el diccionario de clasificación de conceptos y el diccionario de descripción de conceptos. El primero organiza los conceptos de forma jerárquica, emulando la estructura de un 'thesaurus'. Los conceptos son clasificados mediante las relaciones *super-sub, equivalent* o *synonymous*. Por otro lado, el diccionario de descripción de conceptos establece relaciones de coocurrencia entre conceptos, p.e. propiedades selectivas, restricciones semánticas, etc. (cf. EDR 1988).

Los diccionarios de conceptos son independientes de los diccionarios de palabras. La conexión se establece mediante índices conceptuales.

2. El diccionario español para ATLAS

Independientemente de su índice conceptual, la descripción de una palabra debe contener información adicional de tipo morfológico, sintáctico y semántico. Esta información se especifica en el diccionario de palabras, que es independiente de los diccionarios conceptuales. Existe un diccionario para cada lengua y en él se describe la información sobre el comportamiento superficial de las palabras. Una palabra en los diccionarios de palabras puede corresponderse con varios conceptos del diccionario de conceptos, y viceversa.

2.1. Diccionario de palabras

El diccionario de palabras está especializado en el comportamiento idiosincrático de las palabras en cada lengua. Esta especialización reporta el beneficio de la modularidad. Por este motivo, el grupo de Fujitsu en Barcelona ha buscado un tratamiento de la información que pueda satisfacer otras aplicaciones distintas a las de ATLAS.

2.2. Formato, definición de los campos

Cada palabra en español está asociada a una o más fichas que contienen:

- un índice conceptual
- la forma sin flexión o raíz ('headword')
- la forma de citación: el infinitivo para los verbos, el masculino singular para los adjetivos ('reading')
- categoría sintáctica: N, A, V, etc.
- información gramatical
- códigos morfológicos

El campo más problemático es el de la información gramatical. Si distinguimos entre categorías abiertas y cerradas, el énfasis de nuestro estudio recae sobre la información gramatical necesaria para describir las categorías abiertas: categorías nominales, sustantivos y adjetivos, categorías verbales y modificadores adverbiales.

2.3. Automatización del proceso de entrada de palabras

Ilustramos la metodología seguida para la actualización de la base de datos léxica con un ejemplo del tratamiento de la morfología. El proceso de generación de una palabra realiza dos operaciones:

- la selección de la raíz apropiada;
- la selección del sufijo correspondiente.

Para el tratamiento de los cambios de raíz, el diccionario asigna a cada palabra las raíces con los cambios ya materializados. Es decir, no existen reglas alomórficas ni en la gramática ni en el diccionario. La correspondencia entre raíces y sufijos se indica mediante códigos de adyacencia. Estos códigos definen de manera exhaustiva los distintos paradigmas flexivos nominales y verbales para el español. La figura 1. ilustra un fragmento del paradigma de flexión nominal del español.

Código Morfológico	Masculino singular	Masculino Plural	Femenino Singular	Femenino Plural
186	φ	φ	-	es
187	o	os	a	as
188	e	os	a	as

Figura 1. Fragmento de morfología nominal del español

Para el cambio de raíz en los nombres y adjetivos, el sistema utiliza una tabla, similar a la que se presenta en la figura 2, que permite seleccionar la raíz apropiada dado un determinado tipo de irregularidad.

Tipo de Irreg	Masculino singular	Masculino Plural	Femenino Singular	Femenino Plural
T0	R1	R1	R1	R1
T1	R1	R2	R1	R2
T2	R1	R2	R2	R2

Figura 2. Tabla de tipos de cambio en las raíces nominales

La morfología verbal tiene un tratamiento paralelo a la nominal, como se muestran las tablas de las figuras 3 y 4.

Código morfológico	101	102
Ejemplo	AMAR	ANDAR
PRI	o as a amos áis an	idem
PSI	é aste ó amos asteis aron	e iste o imo isteis ieron
IMP	a en	a en
INF	ar	ar
GRUND	ando	ando
PART	ado	ado

Figura 3. Paradigma morfología verbal

Tipo de irreg.	Significado	Ejemplo
T13	R1: todas las formas excepto R2: 1SPRI, PRS, 3PIMP R3: 3S+PPSI, IMS, GRUND	caer
T14	R1: todas las formas excepto R2: 3S+PPSI, IMS, GRUND	leer
T15	R1: todas las formas excepto R2: 1SPRI, PRS, 3PIMP R3: PSI, IMS R4: GRUND	traer

Figura 4. Tabla de cambios de raíz en formas verbales

El programa de actualización y entrada de palabras guía al usuario sobre los paradigmas morfológicos y tipo de irregularidad, de forma que no se necesita otra información que unos conocimientos gramaticales básicos de español para producir entradas a un gran rendimiento. Las figuras 5, 6 y 7 muestran los menús presentados al usuario para que éste elija las formas que le corresponden a la palabra que se está actualizando.

Las figuras 5 y 6 muestran los menús para la entrada de un nombre con cambio de raíz. En la figura 5 se ofrecen cambios en la raíz según cinco paradigmas (singular/plural, singular masculino/todos los demás, etc.). Para seleccionar la flexión apropiada, se permiten tres métodos: a través de ejemplos, por preguntas sucesivas o simplemente introduciendo los códigos directamente. Para la detección del paradigma verbal, el programa permite además introducir un código de referencia (que se corresponde

== CATEGORIA ==

DEFINICION DEL TIPO DE RAIZ MEDIANTE EJEMPLOS

Se comporta como : S/N ?

T1 camión camiones - -	N
T2 león leon/es /a /as	N
T3 rey /es reina /s	S
T4 actor /es actriz actrices	
T5 barón barones baronesa /s	

seleccione la categoría apropiada y pulse <ENTER>

Figura 5.

16/02/91 ** I N T R O D U C C I O N D E P A L A B R A S ** 16:43:37
 Memoria disponible :29KB Entrada en proceso : rey (N)

== FORMAS DE ENTRAR LA MORFOLOGIA ==

== CATEGORIA ==

a través de ejemplos
por preguntas sucesivas
los códigos ya son conocidos

DEFINICION DE LA MORFOLOGIA MEDIANTE PREGUNTAS SUCESIVAS

Con '.' se indica que no se agregan letras
 Con '_' se indica que esta combinación no es posible

Para formar el :	Agregar :
masculino singular	el rey.
masculino plural	los reyes
femenino singular	la reina.
femenino plural	las reinas

Elija la manera de entrar la morfología y pulse <RETURN>

Figura 6.

16/02/91 ** I N T R O D U C C I O N D E P A L A B R A S ** 16:46:19
 Memoria disponible :29KB

== CATEGORIA ==

Número de referencia : 18

Raiz 1 : preven	Ejemplo : VENIR
Raiz 2 : previen	Raiz 1 : ven
Raiz 3 : preveng	Raiz 2 : vien
Raiz 4 : previn	Raiz 3 : veng
Raiz 5 : prevendr	Raiz 4 : vin
Participio : prevenido	Raiz 5 : vendr

Son correctas estas raices (S/N) ? S

seleccione la categoría apropiada y pulse <ENTER>

Figura 7.

a los códigos del *Larrouse de la conjugación*) de manera que el sistema puede proponer un modelo de flexión al usuario, facilitando su introducción (figura 7).

El programa permite señalar cambios de categoría gramatical (p.e. nombre/adjetivo, adjetivo/adverbio, etc.) y el tratamiento de palabras compuestas, construcciones locutivas, y colocaciones. Siguiendo un método análogo al indicado para la morfología se introducen también los atributos de carácter sintáctico y semántico. En los apartados siguientes se da cuenta de las soluciones adoptadas para las formas nominales, verbales y adverbiales.

3. Tratamiento de las entradas léxicas para el *sintagma nominal*

Hemos establecido una primera división en dos grupos de lexemas, según que las clases a las que se adscriban sean abiertas (lexemas propiamente dichos) o cerradas (morfemas o morfemas lexicales), que podemos definir por extensión, es decir, a partir de la enumeración de todos sus elementos. Las clases abiertas o léxicas son: Nombre, Nombre Propio, Adjetivo, Participio e Infinitivo.

3.1 Los grupos léxicos

El cuadro siguiente contiene una lista de información a la que la gramática debe tener acceso para cada uno de los miembros de las clases marcadas y que deberá ir contenida en el diccionario.

	Nom- bre	Nombre propio	Adj.	Parti- cipio	Inf.
Rasgos de formación fonético-gráfica (RFFG)	x	x	x	x	x
Especificación de n. propio		x			
Puede formar un adverbio con el sufijo <i>-mente</i>			x	x	
Puede formar grado morfo.			x		
No puede formar grado ni morfológica ni sintácticamente.			x		
Forma superl. o comparativa			x		
Información sobre género	x	x			
Puede llevar artículo u otros determinantes		x			
Posición respecto al núcleo			x	x	
Puede funcionar como atributo de los verbos <i>ser</i> o <i>estar</i>			x	x	
Puede acompañar a un verbo atributivo de los que permutan el atributo con <i>lo</i> (1)			x	x	
Puede acompañar a un verbo atributivo de la otra categ. (2)			x	x	

(1). Por ejemplo: *ser, estar, parecer*

(2). Por ejemplo: *volverse*

3.2. Los grupos morfemáticos

Los grupos o clases morfemáticos presentan más divergencias. La relación completa de los grupos tratados es la siguiente:

(a) Pronombres

(a.1.) Pronombres (personales) que pueden formar SN en solitario.

(a.2.) Pronombres (personales) que llevan incorporada preposición.

- (b) Determinantes
 - (b.1.) Indefinidos que pueden formar SN en solitario.
 - (b.2.) Artículo.
 - (b.3.) Demostrativos.
 - (b.4.) Posesivos.
 - (b.4.1.) Posesivos átonos.
 - (b.4.2.) Posesivos tónicos que pueden llevar artículo.
 - (b.4.3.) Posesivos tónicos que pueden preceder al núcleo del SN.
 - (b.5.) Indefinidos que actúan como determinantes pre-nominales.
 - (b.5.1.) Y que pueden ir seguidos de *otro, -a, -os, -as*.
 - (b.5.2.) Y que no pueden ir seguidos de *otro, -a, -os, -as*.
 - (b.6.) Indefinidos que actúan como determinantes post-nominales.
 - (b.7.) [todo]
 - (b.8.) [mismo]
 - (b.9.) [otro]

La distinción entre (a) y (b) no tendrá función en el sistema de distribución, y se mantiene a efectos clarificadores y para poder hacer referencias conjuntas a alguno de los dos grupos que delimitan. Por problemas de espacio, nos limitaremos a dar como muestra la clasificación de los demostrativos.

Los rasgos para los demostrativos son los siguientes:

- * El género: masculino, femenino o neutro.
- * El número: singular, plural o neutro (no marcado).
- * El nivel de proximidad: (1, 2, 3).
- * La posibilidad de que el referente (si se utiliza como pronombre) o el nombre al que acompañan sean de persona, de cosa, o ambos.
- * La posibilidad de que, además de la utilización primaria, se pueda utilizar como adjetivo.

	Género	Nombre	Proxim.	Pers.	Adjetivo
este	m	s	1	- +	+
aquello			3	-	-
ello			(3)	-	-
tal		s	(3)	- +	+
tales		p	(3)	- +	+

3.3. Enlace con la información requerida por el verbo. Atributos semánticos

Hemos definido una partición de los nombres, que es la que figura en la columna inicial del cuadro siguiente, bajo el encabezamiento de **atributo nominal**; cada término que haya de ser introducido en el diccionario será clasificado de acuerdo a esta partición, y recibirá por defecto los valores que, para cada una de las columnas restantes, se especifican en la tabla (la *s* expresa un valor positivo, la *n* un valor negativo, y la presencia tanto de *s* como de *n* indica que según el contexto, el sentido de la oración, o el uso habitual de la lengua, las palabras de esa clase pueden recibir uno u otro valor). A su vez, los verbos incluirán en sus entradas de diccionario la información de sus restricciones expresada en términos de la fila inicial, de acuerdo con los valores semánticos que exijan de sus elementos categorizados. (Aquí, **persona humana** se refiere a personas o entes asimilables de la realidad o de la ficción; **animado** incluye cualquier ser dotado de vida; **movimiento interno** quiere expresar la idea quizás demasiado vaga de movimiento o cambio en relación a la propia unidad, y en este epígrafe por tanto, incluiríamos tanto el crecimiento de una planta como un parpadeo o un fruncir de labios; **desplazamiento**, en oposición al anterior, se refiere al movimiento en relación a objetos o referencias exteriores a la unidad que se mueve; **natural**, por fin, en oposición a artificial, tiene valor

positivo en aquellos nombres que se refieren a entes no creados por el hombre.)

Rasgos del sintagma verbal	H u m a n o	A n i m a d o	P r o c e s o	L u g a r	G o a l	T i m e	I m p l e m e n t	M a n n e r	S o u r c e	M e a s u r e
A Contables										
Persona física	+	+	-			-	-	-	-	
Animal	-	+	-			-		-	-	-
Planta - vegetal	-		-			-		-		-
Lugar; institución	α	α		α		-			+	-
Persona jurídica	-	α		α		-			+	-
Parte del cuerpo	+	-		+		-		-		?
Automatismo	-	α								-
Auto - móviles	-									-
Otros concretos	-	-	-	?	?	?		?	-	?
Abstractos, etc.	-	-	?	?	?	?		?		?
B MEDIBLES										
Concretos De materia	-	-	-			?				-
Intangibles	-	-	+	-		?			-	?
Abstractos De acción	-	-	+	-		?			-	-
De estado	-	-	-	-		?			-	-

La gramática relacionará los valores solicitados por el verbo con los contenidos en esta tabla para los nombres, y cuidará de la correcta selección del término correspondiente al concepto buscado de acuerdo con las restricciones impuestas.²

²Significado del contenido del cuadro:

(+) indica que el rasgo del sintagma nominal correspondiente a la línea implica el rasgo del sintagma verbal correspondiente a la columna.

(-) indica que el rasgo del sintagma nominal correspondiente a la línea implica la negación del rasgo del sintagma verbal correspondiente a la columna.

(α) indica que no existe una relación necesaria entre los rasgos de la fila y de la columna, dado que el rasgo correspondiente al sintagma nominal puede ser interpretado como implicador o como del rasgo correspondiente al sintagma verbal o de su negación; pero hay que tener en cuenta que en un contexto determinado solamente será posible simultáneamente uno de los dos valores, de manera que cuando se ha producido una identificación, por ejemplo, de una persona jurídica con un elemento animado, no suele ser posible que en el mismo contexto inmediato aparezca como no animado.

(?) indica que en el momento de realizar la entrada léxica el programa deberá preguntar el valor que dicha entrada léxica tendrá con relación a ese atributo verbal. Hay que tener en cuenta, además, que no solamente deben ser posibles en ese caso la respuesta + y -, sino también la α, explicadas más arriba y la "casilla en blanco", que se explica a continuación.

() La casilla en blanco indica que no se puede establecer una relación general entre el rasgo del sintagma nominal (de la fila) y el del sintagma verbal (de la columna). Por tanto no se puede obtener información a partir de la entrada léxica nominal sobre el valor del atributo requerido por el sintagma verbal. Por ejemplo, un nombre con atributo de persona física puede funcionar

4. Subcategorización y clasificación verbal

Se exponen en este apartado los aspectos considerados relevantes para el tratamiento de la subcategorización verbal y la estructura del módulo de diccionario diseñado para recoger esta información. Los principios generales de los que se ha partido son los siguientes:

- Tratamos con tres niveles de descripción:
 - a) Temático-conceptual.
 - b) Funcional.
 - c) Sintagmático.
- Establecemos un ámbito de subcategorización para cada una de las piezas léxicas consideradas como verbos plenos en nuestra lengua. La estructura temática que ofrezca cada una de ellas en el nivel conceptual deberá ser objeto de un proceso de transferencia para asociarlas con las descripciones de Interlingua con que tratar en los procesos de análisis y generación.
- Discriminamos entre verbos (plenos) que entran en estructuras de control y auxiliares. Estos últimos no presentan marco de subcategorización.
- Establecemos para cada "verbo" un correlato conceptual, sin desarrollar ningún tipo de descomposición léxica en primitivas semánticas ni estructuras de información asociadas a taxonomías de base aspectual.
- Identificamos y caracterizamos las posiciones subcategorizadas por un verbo partiendo de los elementos que concurren con él en sus construcciones oracionales básicas.
- A través de procedimientos operativos, en particular, tests de "so do" y supresión, discriminamos entre esos elementos los que conforman el ámbito de subcategorización: complementos obligatorios y opcionales, y los que caen fuera de éste: adjuntos y extraperiféricos.
- Incluimos la posición correspondiente al sujeto dentro de las subcategorizadas.
- La gramática deberá tratar la aceptabilidad de los distintos tipos de adjuntos para los verbos a partir de la información sobre la matriz temática, y también de información relativa al modo de la acción, que no se recoge en este diseño.

4.1. Parámetros

La información considerada relevante para dar razón de los fenómenos propios de la subcategorización verbal se distribuye en los siguientes parámetros:

- Identificador conceptual.
- Funciones gramaticales.
- Correferencia.
- Selección de cláusula inserta.
- Marcadores semánticos.
- Relación léxica.
- Proyección funciones-estructuras sintácticas.
- Roles temáticos.
- Proyección roles-funciones.
- Selección de modo-tiempo.
- Pronominalidad.

como lugar o como goal, pero si tal funcionamiento es correcto o no no será deducible a partir de la información léxica correspondiente: Véanse los siguientes ejemplos:

"Venid y vamos todos con flores a María"
"Venid a mí"

4.2. Matriz temática.

Discriminamos entre posiciones subcategorizadas a las que corresponde una caracterización argumental definida: roles temáticos y aquellas otras sin correspondencia definida en un cuadro actancial: índices sobre posiciones argumentales.

ROLES TEMATICOS:

Agent	Object	Benefactive	Time	Place
Goal	Source	Manner	Measure	

INDICES SOBRE POSICIONES ARGUMENTALES:

- 0. Sobre posiciones argumentales sin caracterización actancial. Verbos ambientales, existenciales, atributivos.
- @. Sobre posiciones argumentales con rol no precisado. Principalmente en verbos ditransitivos:
- &. Marca posiciones para las que se exige algún tipo de complemento, no necesariamente realizable sintagmáticamente, sin rol determinado.

4.3. Patrón funcional.

FUNCIONES GRAMATICALES:

SUJETO	COMPLEMENTO TEMÁTICO DE PLACE
OBJETO DIRECTO	COMPLEMENTO TEMÁTICO DE GOAL
OBJETO INDIRECTO	COMPLEMENTO TEMÁTICO DE SOURCE
ATRIBUTO DE SUJETO	COMPLEMENTO TEMÁTICO DE TIME
ATRIBUTO DE OBJETO	COMPLEMENTO TEMÁTICO DE MANNER
COMPLEMENTO	OBLICUO RESTRINGIDO
COMPLEMENTO NO RESTRINGIDO	

4.4. Correspondencia entre niveles.

Postulamos la existencia de unos principios generales de proyección de roles temáticos a estructuras funcionales y de una correspondencia entre éstas y estructuras sintácticas.

4.4.1. Correspondencia entre roles y funciones

Exponemos algunas de las reglas consideradas:

- R.1.- A partir de cualquier patrón agentivo <AGT ... > se establece la correspondencia AGT - Su.
- R.2.- BEN siempre se proyecta a Oi.
- R.3.- OBJ si forma parte de una matriz sin AGT se proyecta a Su.
- R.4.- OBJ si forma parte de una matriz con AGT se podrá proyectar a Od o Oblr. La discriminación deberá efectuarse en el momento de efectuar las entradas.
- R.5.- MANNER, TIME, SOURCE, IMP y GOAL se proyectarán siempre al complemento temático correspondiente.
- R.6.- PLACE se proyectará a CtPlace siempre que concurra con AGT u OBJ, si no se da esta circunstancia, deberá determinarse para cada verbo en particular si se proyecta a Su o CtPlace.

4.4.2. Correspondencia entre funciones y estructuras sintácticas

Esta correspondencia se trata en relación a tres tipos de determinación:

DETERMINACION GENERAL

Son sujeto de este tipo de tratamiento las funciones que se realizan en determinados sintagmas con independencia del tipo de verbo y de los rasgos semánticos propios de las lexias particulares : CtManner, CtTime, CtSource, CtPlace y CtGoal.

DETERMINACION DEPENDIENTE INDIRECTA

Son sujeto de este tratamiento las funciones cuya realización sintáctica es función de marcadores semánticos o relaciones de correferencia :Su.

DETERMINACION DEPENDIENTE DIRECTA

Serán funciones directamente determinadas aquellas en las que la estructura sintáctica del complemento en cuestión venga establecida de modo singular en la entrada léxica de al menos uno de los verbos que las requieran: Oblr, ASu, AOb, Oi y Od.

4.5. Clasificación verbal

Entendemos una denominación de clase verbal como una macro con un conjunto de información asociada. Tal asociación deberá ser recuperable explícitamente del diccionario.

Las clases verbales establecidas responden a la distinta conducta de los verbos en relación a su matriz temática, patrón funcional y proyección entre ambos. En algún caso se han adoptado también fenómenos sintácticos como criterio de discriminación.

Exponemos parte de la tabla Matrices temáticas - Categoría - Patrón funcional. (Los signos '?' indican elemento opcional.)

La discriminación VA1 y VA2 responde a la capacidad de los primeros y no de los segundos para pronominalizar el Atributo de Sujeto.

4.6. Diseño del módulo de diccionario.

El Diccionario se concibe desarrollado en tres módulos principales: Gestión, Estructuras de Conocimiento y Base de Datos Léxica.

La Base de Datos Léxica corresponde a las fichas de los distintos verbos, con la información físicamente presente en ellos.

Como Estructuras de Conocimiento entendemos todas aquellas donde se recoge la información pertinente para la importación, exportación y procesamiento de Datos.

El Módulo de Gestión recoge los procedimientos que deben permitir estas tareas. Comprende la Interface con el usuario y los procesos de comunicación entre módulos.

4.6.1. Estructuras de Conocimientos

Algunas de las estructuras de conocimiento diseñadas son las siguientes:

- Tablas de correspondencia entre funciones y roles.
- Tablas de correspondencia entre funciones y estructuras sintácticas.
- Tablas de correspondencia entre preposiciones y complementos temáticos.
- Asociación entre categorías, códigos implicados y acciones de consulta.
- Reglas léxicas.
- Acciones de implicación para activación de funciones de asignación de valores.
- Ejemplos frasales y sintagmáticos correspondientes a las distintas categorías y funciones.
- Módulos de formateo para la exportación.

Matriz temática	Cal.	Patrón funcional
<0 TIME?>	VAMB	{CtTime? Co}
< >	VAMB	{ }
<OBJ>	VI	{Su}
<OBJ @?>	VI	{Su Oblr?}
<AGT OBJ?>	VTD	{Su Od?}
<AGT OBJ?>	VTOBL	{Su Oblr?}
<AGT OBJ @?>	VDT	{Su Od Oblr?}
<AGT OBJ BEN?>	VDT	{Su Od Oi?}
<AGT OBJ BEN?>	VDT	{Su Oi? Od}
<AGT OBJ TIME>	VDT	{Su Od CtTime}
<AGT OBJ GOAL? BEN? >	VDT	{Su Od CtGoal? Oi?>
<AGT>	VE	{Su}
<AGT @?>	VE	{Su Oblr?}
<AGT &>	VE	{Su Cnr}
<AGT TIME?>	VE	{Su CtTime?}
<AGT SOURCE? GOAL?>	VE	{Su CtSource? CtGoal?}
<OBJ BEN>	VD	{Oi Su}
<OBJ BEN>	VD	{Su Od}
< 0 0 >	VA1	{Su ASu}
< 0 0 >	VA2	{Su ASu}
< 0 MEASURE >	VA1	{Su Co}
<AGT 0 OBJ>	VTA	{Su AOb Od}
<0 PLACE?>	VEX	{Su Co?}
<0 PLACE?>	VEX	{Co CtPlace?>
<0 PLACE?>	VEX	{Su CtPlace?

Tabla de matrices temáticas / categorías / patrón funcional.

4.6.3. Módulo Gestión

Algunas de las funciones que desarrollan este módulo son:

MODO INTRODUCCIÓN

a) SOBRE BASE DE DATOS LÉXICA

- Gestión de menús para selección de categoría.
- Procedimientos de consulta activados a partir de la categoría.
- Procedimientos de corroboración.

b) SOBRE ESTRUCTURAS DE CONOCIMIENTO

- Desarrollo de módulos de exportación.

MODOS DE RECUPERACIÓN

a) SOBRE BASE DE DATOS LÉXICA.

- Activación de las tablas F-Cat-R, F-Marc-S y Categoría valores implicados.
- Selección y activación módulo exportación.

b) SOBRE ESTRUCTURAS DE CONOCIMIENTO

- Procedimientos de consulta de la información en ellas recogida.

4.7. Síntesis de los procesos de tratamiento del diccionario

La introducción de la información concerniente a cada verbo en el diccionario se desarrolla a partir de la identificación de la categoría verbal que le corresponde. Los menús para selección de categoría ofrecerán ejemplos frasales correspondientes a las distintas categorías. Una vez identificada la categoría se activará la tabla de asociación entre categorías, códigos implicados y acciones de consulta.

Categoría	Funciones	Corref.	Proy_f_s	Pronom
VAMB	RestfVamb*		Depproy*	
VI	{Su RestfVi*}		Depproy*	Selpro*
VTD	{Su RestfVtd*}	SelCorSt*	Depproy*	Selpro*
VTOBL	{Su RestfVtobl*}	SelCorSt*	Depproy*	Selpro*
VDT	{Su RestfVdt*}	SelCorSt*	Depproy*	Selpro*
VE	{Su RestfVe*}		Depproy*	Selpro*
VD	RestfVd*		Depproy*	
VA1	{Su ASu}	CSuAsu	Depproy*	
VA2	{Su RestfVA1*}	CSuASu	Depproy*	Selpro*
VTA	{Su AOb Od}	COdAsu	Depproy*	Selpro*
VEX	RestfVex*			

Cada una de las funciones (*), activadas a partir de la categoría, desarrolla los procesos de consulta a través de menús para seleccionar el valor pertinente. En particular, las llamadas de la columna "Funciones" permitirán precisar la estructura funcional dentro de las contempladas para la categoría en cuestión.

Toda la información introducida tras el procesamiento de la tabla activará la matriz de ejemplos frasales y sintagmáticos, que ofrecerá ejemplos que recojan las posibles configuraciones sintácticas correspondientes a la información introducida. Esto permitirá verificar su corrección y proceder a rectificaciones si fuera preciso.

La información que se recoja en la Base de Datos Léxica será la que se haya introducido a partir de la activación de las funciones (*). En los procesos de recuperación, a través de la tabla 18 y otras matrices de conocimiento asociadas a la categoría, en particular la tabla de Funciones-Categoría-Roles, podrá accederse a los valores que presente el verbo en cuestión para todos los parámetros contemplados.

La activación de los módulos de exportación permitirá precisar cuál es la información que se desea recuperar y en qué formato se quiere recoger.

5. Modificaciones *adverbiales*

Esta descripción se centra básicamente en el adverbio, entendiendo como tal al elemento que funciona como núcleo de un sintagma adverbial. Ahora bien, la descripción que se presenta es también aplicable al sintagma nominal y al sintagma preposicional en su función adverbial. Simplemente, habría que tener en cuenta aquellas funciones adverbiales desempeñadas obligatoriamente por un sintagma preposicional, como es el caso de las expresiones de duración temporal ('He trabajado hasta las nueve') y de los locativos que expresan dirección u origen.

También para el adverbio, como ya se ha indicado en el punto 2, el diccionario contendrá información morfológica, sintáctica y semántica.

5.1 Información morfológica

Morfológicamente el adverbio carece de flexión de género y número. Las únicas variaciones que presenta son la derivación cualitativa (aumentativos, diminutivos), que aquí no trataremos, y el grado. El grado positivo es el no marcado y el superlativo absoluto se forma mediante sufijación o bien mediante variaciones léxicas, formas heredadas del sistema flexivo latino. El grado comparativo se realiza mediante procedimientos sintácticos, excepto en algunos casos en los que también encontramos las variaciones léxicas antes mencionadas.

En el diccionario, los adverbios que no admiten grado se marcarán con el rasgo [-GR] ('hoy'). Los que sí lo aceptan no llevarán ninguna marca, lo cual querrá decir que permiten formar tanto el superlativo morfológico como el sintáctico. En este caso, como el superlativo morfológico tendrá una entrada distinta, esta se marcará con el símbolo [GSUP]. Por último, los que no aceptan el superlativo morfológico llevarán el símbolo [-GMOR], con lo cual se generará sólo el superlativo sintáctico.

Ejemplos:	hoy	[-GR]
	tarde	
	tardísimo	[GSUP]
	bien	[-GMOR]

El grado comparativo se forma en general sintácticamente. Sólo hay cuatro comparativos léxicos: más ("mucho"), menos ("poco"), mejor ("bueno") y peor ("malo"). El grado de estos cuatro adverbios se puede representar en el diccionario con el símbolo [GCOM].

5.2 Clasificación semántica de los adverbios

Hemos establecido ocho clases semánticas de adverbios, algunas de ellas con subdivisiones adicionales.

- 1 Adverbios performativos o de acto de habla.
'Francamente, considero que has fracasado.'
- 2 Adverbios de punto de vista.
'Económicamente este proyecto no es rentable.'
- 3 Adverbios actitudinales.
'Lamentablemente, Pepe no dijo nada.'
- 4 Adverbios modales.
'Probablemente habrá otra guerra en el Golfo.'
- 5 Adverbios de manera.
'Pedro contestó amablemente'
- 6 Adverbios locativos.
'María come aquí'

7 Adverbios temporales.

7.1 Adverbios de posición temporal

'Llegará mañana'

7.2 Adverbios de frecuencia

'Nos reunimos semanalmente'

8 Adverbios de grado.

8.1 Amplificadores: 'Una actitud profundamente irónica'

8.2 Enfatizadores: 'Es francamente simpático'

8.3 Reductores: 'Llegó ligeramente tarde'

Somos conscientes del considerable solapamiento que existe entre las distintas clases semánticas de adverbios que hemos establecido. Por ejemplo, entre adverbios de grado y adverbios temporales de frecuencia y duración:

Veo mucho/frecuentemente a Juan
Esta noche he dormido mucho/ diez horas

El criterio que hemos seguido para esta clasificación es la posibilidad o imposibilidad de coaparición de cada clase en una misma oración. Consideramos que dos adverbios pertenecen a la misma clase si no pueden aparecer en una misma oración. En caso contrario, pertenecen a clases distintas:

Juan trabaja a menudo cuatro horas los sábados

En el ejemplo anterior encontramos tres adverbiales temporales: el primero de frecuencia, el segundo de duración y el tercero de posición temporal. En cambio, no sería posible encontrar dos o más adverbios de frecuencia, de manera, actitudinales, etc.

5.3 Descripción sintáctica

Hemos establecido seis categorías sintácticas de adverbios:

- 1 Modificadores del adjetivo - MADJ
muy joven, sorprendentemente alto
- 2 Modificadores del adverbio - MAV
llegó muy / increíblemente tarde
- 3 Modificadores del pronombre y del sintagma nominal - MSN
solamente él / Juan
- 4 Modificadores del verbo - MV
María se comportó bien
- 5 Modificadores del predicado nuclear (V + argumentos) - MPN
Pepe come mucho / correctamente
- 6 Modificadores oracionales - MO
Degradadamente, María tuvo un accidente

Debido a la imposibilidad de establecer correspondencias entre clases semánticas y categorías sintácticas sin riesgos de ambigüedad, hemos optado por establecer una lista de posiciones preferentes en la oración.

Los adverbios de acto de habla, actitudinales, de punto de vista y modales son modificadores oracionales o extraperiféricos, pero su posición en la oración no es homogénea. Así, los dos primeros prefieren la posición inicial seguidos de coma (aunque también pueden aparecer en otras posiciones, pero siempre con marca de independencia de entonación), mientras que los dos segundos no exigen ir seguidos de pausa en posición inicial.

CÓDIGO	POSICIÓN	EJEMPLO
P1	inicial seguido de coma	Francamente, estoy decepcionado
P2	inicial sin coma	Ayer llegué tarde
P3	entre sujeto y verbo entre comas	Él, lamentablemente, se murió
P4	final precedido de coma	Pepe no ha venido, desgraciadamente
P5	final sin coma	Pepe come manzanas frecuentemente
P6	antes del adjetivo	muy simpático
P7	antes del adverbio	demasiado lejos
P8	antes del verbo	Pepe casi se muere
P9	después del verbo	Juan llegará tarde
P10	antes del nombre o del pronombre	Solamente Juan/él puede hacerlo
P11	después del pronombre	Yo personalmente haré el trabajo
P12	después del nombre	el día antes

En el diccionario, cada adverbio contendrá la siguiente información: categoría gramatical, grado, clase semántica, función sintáctica y posición o posiciones preferentes en la oración. Si un mismo adverbio pertenece a dos clases semánticas, tendrá dos entradas en el diccionario.

Veamos, por ejemplo, el tratamiento que se daría a un adverbio como francamente:

	FRANCAMENTE 1	FRANCAMENTE 2	FRANCAMENTE 3	FRANCAMENTE 4
CATEGORÍA	ADVERBIO	ADVERBIO	ADVERBIO	ADVERBIO
CLASE SEMÁNT.	ACTO DE HABLA	ENFATIZADOR	AMPLIFICADOR	MANERA
GRADO	[-GR]	[-GR]	[-GR]	[+GR]
FUNCIÓN SINT.	MODIFICADOR ORACIONAL	MODIFICADOR DEL ADJETIVO	MODIFICADOR DEL ADVERBIO	MOD. PREDICADO NUCLEAR
POSICIÓN	P1 - P4	P6	P7	P9

Ejemplos:

- 1 Francamente, no me lo esperaba.
- 2 Juan es francamente simpático.
- 3 Llegó francamente tarde.
- 4 Pepe me habló francamente.

Comentario final

Hemos presentado los aspectos más sobresalientes tratados hasta ahora para la elaboración de un diccionario de palabras para el español. Dos criterios han prevalecido sobre los demás. En primer lugar, la concepción del diccionario como una base de datos léxica; y en segundo lugar, la disposición de unos mecanismos de ayuda para que la entrada de datos atienda a un patrón definido. En los apartados 3, 4 y 5 se han mostrado los criterios que se han tenido en cuenta. Este trabajo supone una contribución importante en la descripción exhaustiva de las categorías nominal, verbal y adverbial para el español.

Referencias

- Abaitua, J., J. Soler & J. Vivaldi. 1989a. Problemática de la Interlingua como estrategia de Traducción Automática. *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural* 7:157-166.
- Abaitua, J., J. Soler & J. Vivaldi. 1989b. Problemas de interpretación de estructuras conceptuales: La pasiva. *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural* 8:29-44.
- Charniak, E. & Y.A. Wilks. 1976. *Computational Semantics: An introduction to Artificial Intelligence and Natural Language Comprehension*. North Holland.
- EDR. 1988. *Concept Dictionary (ver.1)*. EDR Technical Report. Japan Electronic Research Institute, LTD.
- Mel'čuk, I.A. 1988. *Dependency Syntax: Theory and Practice*. State University Press of New York.
- Schank, R. C. 1975. *Conceptual Information Processing*. North Holland.
- Schubert, Klaus. 1987. *Metataxis: Contrastive Dependency Syntax for Machine Translation*. Foris.
- Tesnière, Lucien. 1959. *Éléments de la Syntaxe Structurale*. Klincksieck.
- Uchida, H., Hayasi, T., Kushima, H. 1985. Atlas: Automatic Translation System. *Fujitsu Scientific and Technical Journal* 21:317-329.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100