

PEDECIBA Informática
Instituto de Computación – Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay

Reporte Técnico RT 06-02

**Generación de un diccionario de formas
verbales a partir de información
obtenida de la Web**

Mónica Martínez Amarante

2006

Generación de un diccionario de formas verbales a partir de información obtenida de la Web
Martínez Amarante, Mónica
ISSN 0797-6410
Reporte Técnico RT 06-02
PEDECIBA
Instituto de Computación – Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay, 2006

Generación de un diccionario de formas verbales a partir de información obtenida de la Web

Mónica Martínez
Instituto de Computación (InCo), Facultad de Ingeniería, Universidad de la Republica,
Montevideo, Uruguay

mmartine@fing.edu.uy

Resumen: En este trabajo se presenta la generación de un diccionario de formas verbales, formas no personales y las conjugaciones en los distintos modos y tiempos, para un conjunto de verbos. Para esto se realizó un programa Prolog que obtiene la información de las páginas públicas de la Real Academia Española y genera un archivo de salida con las formas verbales que permite su uso en distintas aplicaciones de procesamiento de lenguaje.

Palabras claves: Programación Lógica, Tratamiento de Lenguaje Natural, Diccionarios Electrónicos

Introducción

El objetivo de este trabajo consiste en la generación de un diccionario de formas verbales (formas no personales y las conjugaciones en los distintos modos y tiempos) marcado con etiquetas EAGLES¹ para de un conjunto de verbos.

La información correspondiente a las formas no personales y conjugaciones de los verbos se obtienen del diccionario online de la Real Academia Española².

Para esto se desarrolló un programa prolog que partiendo del infinitivo de los verbos accede a las páginas de la RAE, procesa la información allí contenida generando las etiquetas correspondientes las cuales se vuelcan en un archivo de salida.

El archivo de etiquetas se genera con el formato compatible a la entrada de un programa (ya disponible) que permite agrupar las distintas etiquetas para una misma palabra como es requerido para el uso del diccionario de Freeling.

Se generaron dos predicados: conjugacion y leoDicc.

Con el primero es posible obtener las etiquetas correspondientes para un verbo y con el segundo se obtienen las mismas para un conjunto de verbos que se encuentran en un archivo dado como entrada.

En las siguientes secciones se describe el procesamiento que se realiza para la obtención del resultado; se describe las condiciones de uso del programa desarrollado, conclusiones sobre el trabajo y finalmente en los anexos se detallan la estructura de la página de conjugaciones de verbos de la RAE y las reglas de construcción de etiquetas EAGLES para verbos.

Proceso de generación de etiquetas

Proceso general

A grandes rasgos el procesamiento de cada verbo se puede dividir en las siguientes etapas:

1. Captura de la página con las conjugaciones correspondientes al verbo
2. Procesamiento de la información capturada para preparar la información necesaria para la generación de las etiquetas
3. Generación de las etiquetas correspondientes

Para acceder a la información del diccionario de la RAE se utilizaron los predicados disponibles en las bibliotecas HTTP y SGML/XML de swi-prolog.

Se utilizó el predicado `http_open` que permite obtener el contenido de una página web en un stream y luego se utilizó el predicado `load_html_file` que carga el stream anterior en una estructura que permite su procesamiento con mayor simplicidad.

Parte importante de este trabajo consistió en determinar la estructura de la página de forma tal de poder obtener a partir de ella los valores correspondientes a cada forma no personal y a cada conjugación. La estructura identificada se detalla en el anexo A.

Para cada verbo de la página de la RAE se obtiene la siguiente información:

Las formas no personales del verbo: infinitivo, participio, gerundio

Las conjugaciones correspondientes al

Modo Indicativo en los tiempos: Presente, Futuro simple, Pretérito imperfecto o Copretérito, Condicional simple o Pospretérito, Pretérito perfecto simple o Pretérito

¹ <http://www.lsi.upc.es/~nlp/freeling/parole-es.html>

² <http://www.rae.es/>

Modo Subjuntivo en los tiempos: Presente, Pretérito imperfecto o Pretérito, Futuro simple o Futuro

Modo Imperativo

Esta información se procesa directamente sin ningún tipo de chequeo y/o generación. Para el caso particular del participio se vio que se requería de un tratamiento especial ya que la información brindada por el diccionario de la RAE es parcial. Este tratamiento especial se detalla más adelante.

Otro caso particular esta dado por los verbos defectivos, verbos para los cuales no están definidas (o lo están en forma incompleta) todos los elementos, por ejemplo el verbo abolir. Estos verbos también fueron procesados en forma particular y su tratamiento se detalla más adelante.

En el caso no excepcional una vez capturada la información se procesan en forma secuencial cada una de las formas no personales (infinitivo, participio, gerundio) y cada uno de los modos/tiempos verbales en principio sin ningún tipo de chequeo y/o generación.

Para esto se definieron los predicados proceso_FNP y proceso_Conj. Cada uno de ellos en primera instancia obtiene de la página capturada la información relevante (dar_FNP, dar_Conjugacion, dar_Imperativo), procesa esta información generando una lista de valores a partir de la cual se generan las etiquetas correspondientes.

La obtención de la información relevante en cada caso (formas no personales, conjugación correspondiente a un determinado modo tiempo) se basa en una búsqueda en la estructura de la página hasta obtener la información. Se obtiene una lista con los valores buscados. Por ejemplo se obtiene la lista [camino,caminas / caminás, camina, caminamos,camináis / caminan, caminan] para el presente del modo indicativo del verbo caminar.

El procesamiento de esta información se basa en la manipulación de lista de valores, de forma de agregarle a la información obtenida todos los datos necesarios para la generación de las etiquetas (modo, tiempo, persona, etc.)

Es en este punto que se determina y marca la lista obtenida con la información correspondiente a si existe la forma no personal o conjugación buscada y si la misma es completa. En el caso de no existir (lista vacía) se marca como no definida y se ignora en el momento de generar las etiquetas. En caso de no ser completa (menos de 6 elementos para conjugaciones del modo indicativo y subjuntivo y menos de 2 conjugaciones para el modo imperativo) se marca como incompleta, por lo que no se generarán etiquetas pero sí se registrará en el archivo correspondiente que se genera como salida junto al archivo de las etiquetas.³

Para las conjugaciones completas se agrega al valor de cada una de ellas la persona a la que corresponden (1 a 6, en el modo indicativo y subjuntivo; 3 y 6 en el modo imperativo).

Y tanto en las formas no personales como en las distintas conjugaciones es posible encontrar valores múltiples. Por ejemplo: el verbo freír tiene 2 participios masculinos singulares (freído y frito); el verbo caminar en el presente del indicativo tiene 2 formas posibles para la conjugación correspondiente a la segunda persona del singular (caminas / caminás), etc. En todos estos casos es necesario separar las formas múltiples sin perder referencia a qué elementos corresponden ni a qué persona corresponden (en el caso de las conjugaciones) de forma tal de obtener una etiqueta para cada una de ellas. Para los ejemplos antes mencionados se obtienen las siguientes listas de valores:

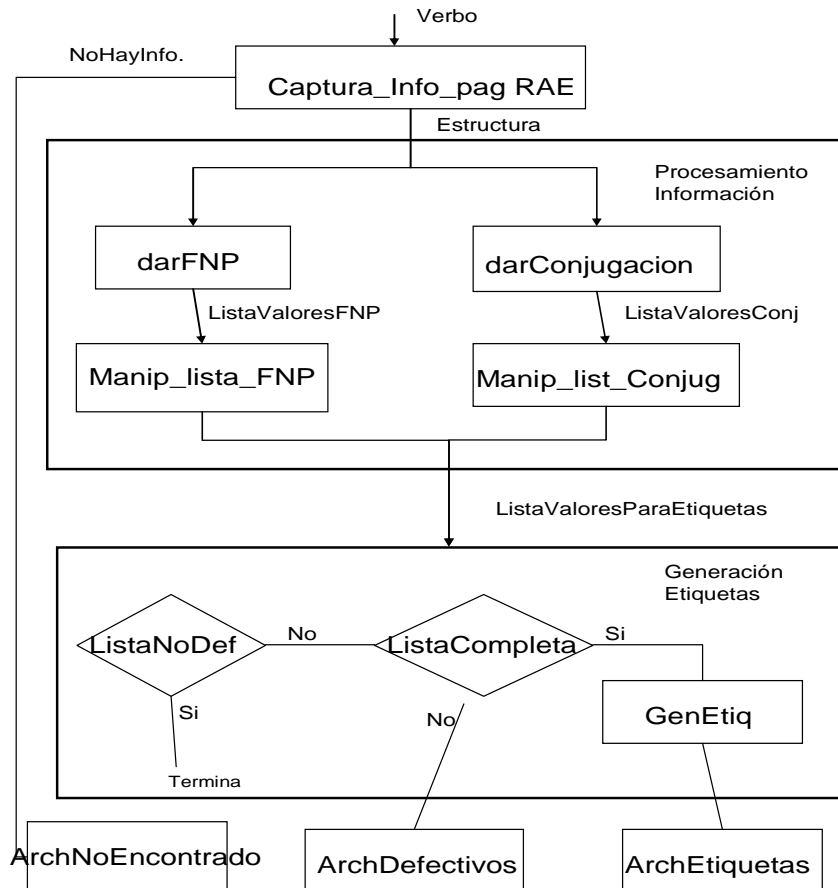
[participio, [completa, ['freído, frito']]]

³ Ver proceso particular para verbos defectivos

[indicativo, presente, [completa, [(1, camino), (2, caminas), (2, caminás), (3, camina), (4, caminamos), (5, camináis), (5, caminan), (6, caminan)]]]]

Una vez que se tiene esta lista de valores simplemente se recorre esta lista y para cada uno de sus elementos se genera una etiqueta aplicando las reglas definidas para su construcción.⁴

El proceso antes descrito se puede ver gráficamente en la siguiente figura:



Proceso particular para los participios

Al observar la información brindada por la RAE en la página correspondiente a la conjugación de un verbo se observó que con respecto a los participios solo tenían la forma correspondiente al masculino singular. Sin embargo para una gran cantidad de verbos están definidas 4 formas correspondientes a las combinaciones de género (masculino, femenino) y de número (singular, plural).

Se consideró de importancia no limitar las etiquetas generadas solo al caso masculino-singular, por lo cuál surgió la necesidad de generar las otras formas del participio a partir de la forma brindada por la RAE.

La primera dificultad en este punto fue determinar en qué casos, para qué verbos están definidas las 4 formas del participio.

⁴ Ver Anexo B : reglas de construcción de las etiquetas

Considerando que no existen reglas que permitan identificar estos casos (por lo menos con la información que se tiene: infinitivo del verbo), se tomó como regla que todos los verbos transitivos y/o pronominales tienen las 4 formas.

Por lo tanto, al momento de generar la lista con los valores correspondientes al participio es necesario determinar si está dentro de los verbos transitivos y/o pronominales.

Para esto es necesario consultar otra página correspondiente al diccionario de la RAE donde se encuentra la definición del verbo y se enuncian allí las características del tipo mencionadas.

Se carga la página correspondiente y se determina si tiene las referencias buscadas (verbo transitivo o verbo pronominal). En caso afirmativo se procede a generar las 3 formas restantes en caso negativo se devuelve como lista de valores para el participio únicamente el brindado por la RAE.

Para la formación del participio femenino se utilizó como regla de construcción: sustitución de la última letra del participio disponible (supuestamente una **o**) por una **a**.

Y para la generación de los respectivos plurales simplemente se agrega una **s** al final de cada uno de ellos.

Se consideró como forma de obtener una mayor garantía sobre las palabras que se estaban generando la verificación de la existencia de las mismas en el diccionario. Este mecanismo no resultó bueno ya que el diccionario no es completo, en el sentido que si un participio no corresponde a otra forma gramatical (sustantivo, adjetivo, etc.) no se encuentra en el diccionario sin que esto signifique que no es una palabra válida. (Ej. Abolida, cocinada).

Si bien no existen reglas específicas que garanticen la construcción utilizada, empíricamente se comprobó que se obtenían los resultados esperados.

Proceso particular para los Verbos defectivos

En el idioma español existen verbos, llamados defectivos, que no tienen todas sus formas personales y/o no tienen conjugaciones definidas para todos los modos/tiempos verbales. (ejemplo: usucapir, embaír, colorir).

Incluso puede suceder que tengan para algún modo/tiempo definidas las conjugaciones para algunas personas pero no en forma completa. (ejemplo: presente del modo indicativo del verbo abolir o embaír)

Desde el punto de vista de la generación de las etiquetas, el caso en que una forma personal o una conjugación no esté definida no son problemas. Simplemente no se generan etiquetas para esa forma personal o conjugación ya que en realidad no deben existir.

El problema se presenta en el caso de que una conjugación se encuentre en forma parcial, no está definida para todas las personas.

En este caso como no se cuenta con ningún tipo de información que permita identificar a qué personas corresponden las conjugaciones definidas (son una simple lista sin referencias ni lugares vacíos) se optó por no generar ninguna etiqueta correspondiente al modo/tiempo definido en forma parcial.

Para no perder totalmente esta información se genera junto al archivo de las etiquetas un segundo archivo con estos datos, los cuales podrían ser considerados en un trabajo futuro.

En este archivo se incluye: infinitivo del verbo, modo/tiempo para el cual la lista de conjugaciones es incompleta y la lista de las conjugaciones capturadas.

Forma de uso

En el programa desarrollado (principal.pl) se generaron dos predicados que pueden ser consultados:

- conjugacion(Modo,ArchEtiq,ArchDef,ArchNoEsta)
- leoDicc(Modo,ArchEntrada,ArchEtiq,ArchDef,ArchNoEsta)

Con el primero se generaran las etiquetas a partir del infinitivo de un verbo ingresado por pantalla.

En el segundo caso se generan a partir de un conjunto de infinitivos de verbos dados como entrada en un archivo (ArchEntrada), que contiene el infinitivo de un verbo seguido de un punto por línea.

Para cualquiera de ellos existen dos modos de trabajo, que dependen de si se utiliza un servidor proxy para navegar en Internet. Esto se indica en el parámetro Modo de los predicados y los posibles valores para el mismo son: noproxy y proxy.

En caso de utilizar servidor proxy se asume el correspondiente a la facultad de Ingeniería de la Universidad de la Republica (httpproxy, 3128)⁵

En todos los casos se generan tres archivos de salida: ArchEtiq, ArchDef y ArchNoEsta, los cuales se describen a continuación.

ArchEtiq

Contiene las etiquetas generadas para cada verbo, en el siguiente formato:.

Palabra lema etiqueta

Con una línea por etiqueta generada. Esto lo hace compatible con la entrada esperada de un programa ya existente (morf_gral.pl)⁶ que permite agrupar todas las etiquetas de una misma palabra en una línea de la forma:

Palabra lema1 etiqueta1 lema2 etiqueta2.....lemaN etiquetaN

Obteniendo el formato requerido por el diccionario de Freeling.

ArchDef

Contiene la información relativa a las conjugaciones incompletas de verbos defectivos para las cuales no se generaron etiquetas.

El formato del archivo es el siguiente:

Infinitivo_verbo modo tiempo [conjugaciones_encontradas]

Con una línea para cada modo/tiempo de cada verbo para el cual no se generaron las etiquetas.

Por ejemplo:

abolir indicativo presente [abolís, abolimos, abolís /]

donde se indica que para el verbo abolir en el modo indicativo en el tiempo presente no fueron generadas etiquetas, las únicas conjugaciones encontradas para este verbo, en este modo/tiempo fueron: [abolís, abolimos, abolís /]

⁵ De ser necesario una modificación en el servidor proxy utilizado se debe modificar la flia de predicados paginaproxy y paginaTipoF (en el archive paginas.pl).

⁶ Para utilizar este programa es necesario previamente ordenar alfabéticamente el archivo de etiquetas.

Este archivo deberá ser tomado en trabajos futuros como insumo y determinar como obtener las etiquetas que no fueron generadas en este trabajo para los verbos defectivos.

Es importante marcar que aunque el verbo de partida sea un verbo defectivo, todas las formas personales y conjugaciones para determinados modo/tiempo que estuvieran completas, las etiquetas correspondientes fueron generadas sin problema. Los únicos casos pendientes de resolución son los que figuran en este archivo. Por ejemplo para el verbo abolir solo resta definir las etiquetas correspondientes al presente del modo indicativo.

ArchNoEsta

Este archivo contiene la lista de verbos para los cuáles no se encontró página de conjugación en el diccionario online de la RAE, por lo tanto fue imposible la generación de etiquetas.

El que no se encontrara página de conjugación para un verbo determinado puede estar ocasionado por dos motivos: por no encontrarse la palabra en el diccionario o por que el diccionario referencia a otra palabra.

Por ejemplo: el verbo abalastrar no se encuentra en el diccionario y lo indica de esa forma, en cambio para el verbo abetunar no existe una página de conjugaciones pero se referencia al verbo embetunar.

Quedo fuera del alcance de este trabajo el tomar algún tipo acción con estos casos más allá de reportarlos.

Nota: Al momento de utilizar estos predicados, durante su ejecución se desplegarán en pantalla una serie de warnings que deberán ser ignorados. Estos son originados por problemas de formato de la página de la RAE (estructuras incompletas) pero no afectan la generación que se esta realizando.

Conclusiones

Los predicados desarrollados fueron aplicados con éxito a un conjunto de verbos brindados por los docentes.

El archivo de entrada constaba de 6376 verbos para los cuales se obtuvo como salida un archivo de 440579 etiquetas (332306 palabras luego de agruparlas), se encontraron 181 modo/tiempos correspondientes a verbos defectivos para los cuales no se generaron etiquetas y finalmente 368 verbos de la entrada no fueron encontrados en el diccionario.

Estos resultados permitirán ampliar en forma considerable el diccionario de Freeling, con los beneficios que eso implica a los usuarios del mismo.

Es importante marcar que el funcionamiento correcto de los predicados desarrollados es altamente dependiente de la estructura de la página de conjugaciones de la RAE.

En caso de modificaciones de esta página, si solo se realizan redistribución de los elementos dentro de la tabla donde se encuentra la información son manejables cambiando los hechos que representan la ubicación de los elementos, de lo contrario es necesario modificar los predicados que obtienen la información.

El uso del lenguaje Prolog para el desarrollo de esta herramienta resultó adecuado, particularmente las bibliotecas utilizadas para la conexión a internet y la carga de la página en una estructura resultaron de gran utilidad.

Por otro lado la velocidad de generación puede resultar en algunos momentos lenta pero esto es totalmente dependiente de la velocidad de conexión.

Quedaron fuera del alcance de este trabajo y deberán ser considerados en el futuro los siguientes puntos:

- La evaluación de los criterios usados para la generación de los participios en las formas singular-femenino y los dos plurales. Si bien empíricamente se observó que los resultados obtenidos eran los esperados no existe una demostración de su correctitud.
- Generación de etiquetas para las conjugaciones incompletas de los verbos defectivos, reportados en el archivo correspondiente.
- Determinar cómo tratar los verbos no encontrados en el diccionario online de la RAE.

Referencias

Diccionario de la Real Academia Española [en línea]

<http://www.rae.es>

(Consulta: 22 de diciembre de 2005).

García-Pelayo y Gross, Ramón; García-Pelayo y Gross, Fernando; Durand, Micheline. Larousse : Diccionario Práctico Conjugación. México: Larousse, 1982. 175 p. ISBN 970-607-018-4

Introducción a las etiquetas Eagles (v.2.0) [en línea]

www.lsi.upc.es/~nlp/freeling/parole-es.html#verbos

(Consulta: 22 de diciembre de 2005).

SWI-Prolog 5.6.0 Reference Manual [en línea]

<http://www.swi-prolog.org/>

(Consulta: 22 de diciembre de 2005).

SWI-Prolog HTTP support

<http://www.swi-prolog.org/packages/http.html> [en línea]

(Consulta: 22 de diciembre de 2005).

SWI-Prolog SGML/XML parserVersion 2.0.5, March 2005 [en línea]

<http://www.swi-prolog.org/packages/sgml2pl.html>

(Consulta: 22 de diciembre de 2005).

ANEXO A - Estructura de la página de conjugaciones de la RAE

La página de conjugaciones de la RAE tiene el siguiente aspecto:

FORMAS NO PERSONALES		
Infinitivo caminar	Participio caminado	Gerundio caminando
INDICATIVO		SUBJUNTIVO
Presente camino caminas / caminás camina caminamos camináis / caminan caminan	Futuro simple o Futuro caminaré caminarás caminará caminaremos caminaréis / caminarán caminarán	Presente camine camines camine caminemos caminéis / caminen caminen
Pretérito imperfecto o Copretérito caminaba caminabas caminaba caminábamos caminabais / caminaban caminaban	Condicional simple o Pospretérito caminaría caminarías caminaría caminaríamos caminaríais / caminarían caminarían	Pretérito imperfecto o Pretérito caminara o caminase caminaras o caminases caminara o caminase camináramos o caminásemos caminarais o caminaseis / caminaran o caminasen caminaran o caminasen
Pretérito perfecto simple o Pretérito caminé caminaste caminó caminamos caminasteis / caminaron caminaron		Futuro simple o Futuro caminare caminares caminare camináremos caminareis / caminaren caminaren
IMPERATIVO		
camina (tú) / caminó (vos) caminad (vosotros) / caminen (ustedes)		

Es una tabla, dividida en tres secciones. La primera corresponde a las formas no personales del verbo, la segunda a las conjugaciones en los distintos tiempos correspondientes al modo indicativo y subjuntivo y en la última se encuentra la conjugación correspondiente al modo imperativo.

Se observó que las dos primeras secciones siempre existen, pudiendo tener alguna fila/columna vacía pero la existencia de la tercera sección depende de la existencia de conjugaciones en el modo imperativo.

Como fue mencionado en el proceso de generación de etiquetas para la captura de esta página se utilizan predicados de las bibliotecas de SWI-Prolog que devuelven una estructura con los elementos que la componen.

La estructura que devuelve el predicado utilizado (`load_html_file`) es una `lista_contenido`, una lista de element donde cada uno de ellos tiene la siguiente estructura:

```
element(nombre, lista_atributos, lista_contenido)
```

nombre : es el nombre del elemento

lista_atributos: es una lista de parejas name=valor

lista_contenido: define el contenido para el elemento.

En el caso de las páginas correspondientes a las conjugaciones de los verbos de la RAE la estructura es una lista de dos elementos, en la que en el segundo componente de la lista de contenido del primer elemento se encuentra la información relevante con respecto a las conjugaciones de los verbos (La Tabla), el resto corresponde a información de formato de la página.

```
[element (_,_,[_element(_,_,Tabla)], element(estado_bien,_,_))]
```

La parte correspondiente a la Tabla tiene la siguiente estructura:

```
[element(_,_,[element(_,_,E)])]
```

y en la posición marcada con E se encuentra los datos correspondientes al cuerpo de la tabla, en el resto se encuentra información de formato de la tabla.

El cuerpo de estas tablas esta formado por una lista de filas:

- fila 1 - Titulo Formas no Personales
- fila 2 - Las formas no personales
- fila 3 - Titulo Indicativo / subjuntivo
- fila 4 - Primera fila de estos tiempos (Presente,Futuro,Presente)
- fila 5 - segunda fila de estos tiempos (Preterito Imperfecto,Condicional, ..)
- fila 6 - tercer fila de estos tiempos
- fila 7 - Titulo Imperativo
- fila 8 - Valores Imperativo

Existen como mínimo las seis primeras filas, en los casos correspondientes a verbos que tienen el modo imperativo definido existen los 8 elementos.

Formas no personales

La estructura correspondiente a las Formas No Personales es la siguiente:

```
element(_,_,[FP1,_FP2,_FP3])
```

Empíricamente se comprobó que esta lista siempre tiene 3 elementos (correspondientes a las 3 columnas) aunque alguno de ellos puede estar vacío.

Cada uno de ellos tiene la siguiente estructura:

```
element(_,_,[_element(_,_,Valor)])
```

En la posición Valor se encuentra el dato correspondiente a la forma no personal. En caso de no existir se encuentra un blanco

En principio la ubicación de las distintas formas no personales es la siguiente:

- Columna 1 → Infinitivo
- Columna 2 → Participio
- Columna 3 → Gerundio

En el programa desarrollada se manejaron con un conjunto de hechos que permiten ubicar la posición de cada forma personal en la fila de la tabla correspondiente a estas formas, en caso de que se intercambien las posiciones entre sí sólo sería necesario modificar estos hechos:

```
ubFNP(forma_no_personal, Nro_columna)
```

Conjugaciones

En principio la ubicación de las distintas conjugaciones correspondientes a los distintos modos/tiempos es la siguiente:

Fila 4

Columna 1
Presente del modo indicativo
Columna 2
Futuro del modo indicativo
Columna 3
Presente del modo subjuntivo

Fila 5

Columna 1
Pretérito imperfecto del modo indicativo
Columna 2
Condicional simple del modo indicativo
Columna 3
Pretérito imperfecto del modo subjuntivo

Fila 6

Columna 1
Pretérito perfecto simple del modo indicativo
Columna 2
Vacía
Columna 3
Futuro del modo subjuntivo

En el programa desarrollado estas posiciones se manejaron con un conjunto de hechos de la forma:

```
ubicacion(modo,tiempo,fila,columna)
```

por lo que en caso de que se produzcan cambios entre ellas sólo sería necesario modificar estos hechos y todo el resto no se vería afectado.

La estructura correspondiente a cada una de las filas de conjugaciones es la siguiente:

```
element(_,_,[element(_,_[ ' '])]) (Fila de conjugaciones vacía)  
element(_,_[X,Y,Z])
```

La lista puede tener uno, dos o los tres elementos

Se comprobó empíricamente que en caso de tener uno corresponde a la primera columna, en caso de tener 2 corresponden efectivamente a la primera y segunda columna; de lo contrario tiene 3 elementos pudiendo ser alguno de ellos vacío.

Cada uno de estos elementos (X,Y,Z) tienen la siguiente estructura donde la lista de conjugaciones efectivamente se encuentra en la posición marcada como L

```
element(_,_[_element(_,_[L])])
```

Modo imperativo

La estructura correspondiente a las conjugaciones en modo imperativo es la siguiente:

```
element(_,_[element(_,_[element(_,_[LC])])])
```

LC es la posición donde se encuentran los valores correspondientes a las conjugaciones en modo imperativo.

ANEXO B – ETIQUETAS EAGLES

Fragmentos extraídos de Introducción a las etiquetas Eagles (v.2.0)

www.lsi.upc.es/~nlp/freeling/parole-es.html#verbos

Las etiquetas se forman por una secuencia de 7 caracteres, teniendo un significado especial y una codificación especial cada una de las posiciones

En el siguiente cuadro se describe para cada posición qué representa, los posibles valores que pueden tener y cómo se codifica ese valor.

VERBOS			
Pos.	Atributo	Valor	Código
1	Categoría	Verbo	V
2	Tipo	Principal	M
		Auxiliar	A
		Semiauxiliar	S
3	Modo	Indicativo	I
		Subjuntivo	S
		Imperativo	M
		Infinitivo	N
		Gerundio	G
		Participio	P
4	Tiempo	Presente	P
		Imperfecto	I
		Futuro	F
		Pasado	S
		Condicional	C
		-	0
5	Persona	Primera	1
		Segunda	2
		Tercera	3
6	Número	Singular	S
		Plural	P
7	Género	Masculino	M
		Femenino	F

- El lema del verbo ha de ser siempre el infinitivo.
- Etiquetamos las formas del verbo *haber* como auxiliares (VA), las del verbo *ser* como semiauxiliares (VS) y las restantes como principales (VM).
- El atributo de *Género* sólo afecta a los participios, para el resto de formas este atributo no se especifica (0).
- Para las formas de infinitivo y gerundio no se especifican los atributos de *Tiempo*, *Persona*, *Número* y *Género*, por lo que su valor será 0.