

# CAPÍTULO 7

## DETECCIÓN Y TRATAMIENTO DE FORMAS VERBALES

### 7.1 VERBOS

El español es una lengua que utiliza masivamente la flexión para la formación de palabras<sup>1</sup>. Se distinguen dos tipos de flexiones: la verbal o conjugación si se realiza con verbos; y la nominal o declinación, si se hace con formas que admiten género y/o número, como nombres, adjetivos, artículos, etc.[Goñi 95]. La flexión nominal da lugar a cuatro formas mientras que la verbal, combinando tiempo, modo, número y persona, hasta 53 formas flexivas simples diferentes<sup>2</sup>. Por esta razón el reconocimiento y categorización de las formas verbales lo hemos implementado de manera diferente al del resto de palabras que forman la frase.

---

<sup>1</sup> La flexión es el proceso en el que se unen morfemas flexivos a palabras o a temas (raíces de los vocablos que no constituyen palabras por sí mismas).

<sup>2</sup> El verbo se puede definir como aquella parte de la oración que contiene morfemas flexivos de número, persona, tiempo y modo [Santana 94].

El procesamiento de las unidades especiales (siglas, abreviaturas, nombres propios, etc.) se realiza comprobando que la palabra en cuestión se ajusta a la definición dada para cada caso concreto y con la ayuda de diccionarios específicos<sup>3</sup>.

Para las restantes palabras (sustantivos, adjetivos, preposiciones, ...) se dispone de diccionarios léxicos con todas las posibles formas derivadas que un vocablo pueda tener. Este sistema no puede aplicarse en el caso de los verbos. Disponer de un diccionario con una entrada para cada una de las posibles formas de cada verbo junto con información de la categoría gramatical supondría un consumo de memoria muy elevado (hay más de 400.000 formas verbales).

Para el procesamiento de los verbos empleamos el modelo **raíz-paradigma**. Las **raíces** o lexemas son las unidades mínimas dotadas de significado pleno y los **paradigmas** son grupos de palabras que sirven de modelo o clase, con las mismas particularidades morfológicas [Goñi 95]. Por ejemplo, si consideramos la forma verbal *pensamos*, la raíz sería *pens* y *-amos* pertenecería al paradigma regular de la 1ª persona del plural del presente de indicativo de la 1ª conjugación.

Con este modelo los diccionarios de verbos sólo contienen los infinitivos, reduciendo así la información necesaria y el tamaño de los diccionarios y, por consiguiente, el consumo de memoria.

El procesamiento a realizar consiste en:

1. comprobar la coincidencia entre la desinencia verbal y alguna de las formas del paradigma
2. identificar la forma del paradigma al que pertenece dicha raíz
3. reconocer la raíz de la forma verbal utilizando los diccionarios

El modelo **raíz-paradigma** tiene dos ventajas principales: flexibilidad, ya que permite dar cuenta de cualquier tipo de irregularidad, y eficiencia computacional. Modelos similares al nuestro ya han sido implementados comercialmente, como ejemplo podemos citar el *analizador morfosintáctico del Instituto de Lingüística Computacional de Pisa para el español* [Rat83], el *procesador morfológico de IBM para el español* [Rod90], *MORFOGEN* [Pen91] y el *analizador morfológico del proyecto Polyglot*.

---

<sup>3</sup> Véase Capítulo 5.

• **CLASIFICACIÓN DE LOS VERBOS**

Los verbos se pueden clasificar en tres grupos [Suances 98]:

1. Verbos **semirregulares**

Aquellos verbos que, en su conjugación, sufren solo pequeñas variaciones ortográficas, con el fin de mantener el mismo valor fonético en sus desinencias. Podemos decir que los semirregulares son verbos regulares con adaptaciones ortográficas en el desarrollo de sus tiempos, por así exigirlo un determinado patrón fonético.

Dentro de los semirregulares podemos diferenciar:

- Verbos **tíldicos**. Aquellos que en alguna de sus formas toman tilde o diéresis, o sufren cambios en su acentuación. Ejemplos: *actuar, prohibir, confiar*.
- Verbos **atíldicos**. Verbos que mantienen invariable el diptongo de la sílaba tónica del infinitivo. Ejemplos: *bailar, causar, adecuar*.

2. Verbos **regulares**

Aquellos que siguen el modelo o patrón de conjugación regular. Hay un patrón para la 1ª conjugación (verbos acabados en -ar), otro para la 2ª (verbos acabados en -er) y otro para la 3ª (verbos acabados en -ir). Ejemplos: *cantar, beber, existir*.

3. Verbos **irregulares**

Verbos que presentan irregularidades en su conjugación que pueden afectar a una o varias letras. O dicho de otro modo, aquellos que no se ajustan al modelo de conjugación regular en al menos una forma. Ejemplos: *colgar, hacer, salir*.

Semirregulares		Regulares	Irregulares
Tíldicos	Atíldicos		
actuar entrever reunir	bailar cambiar peinar	amar deber vivir	andar oler reír

Tabla 7.1 Ejemplos de las distintas clases de verbos

### 7.1.1 LOS DICCIONARIOS

Nuestro sistema dispone de dos diccionarios, uno para los verbos regulares (Infraere.ord) y otro para los irregulares (InfinIrregu.ord). El Infraere contiene 8511 infinitivos, 7586 de la 1ª conjugación, 442 de la 2ª y 483 de la 3ª. El InfinIrregu tiene 2624 infinitivos, 2065 de la 1ª conjugación, 279 de la 2ª y 280 de la 3ª.

Si comparamos nuestros diccionarios con el que utilizan [Goñi 95] en su *Plataforma léxica para el español*, que contiene 5200 verbos regulares y 2100 irregulares, ambos sistemas son del mismo orden de magnitud, pero menor que la *Base de Datos del Verbo Español* [Suances 98]. .

### 7.1.2 PARADIGMAS REGULARES

Se establece una clase paradigmática modelo para cada conjugación: una para la 1ª, otra para la 2ª y otra para la 3ª. De esta manera, con el diccionario de infinitivos regulares (Infraere.ord) y los tres paradigmas regulares se puede reconocer y categorizar cualquier forma verbal regular cuyo infinitivo se encuentre en dicho diccionario.

Para cada conjugación, modo y tiempo verbal tenemos una lista de desinencias verbales; junto a la desinencia aparece la categoría gramatical y la conjugación a la que pertenece<sup>4</sup>.

1ª conjugación	2ª conjugación	3ª conjugación
presente de indicativo	presente de indicativo	presente de indicativo
{"o", {"V..01I.O.."}, 1},	{"o", {"V..01I.O.."}, 2},	{"o", {"V..01I.O.."}, 3},
{"as", {"V..01U.O.."}, 1},	{"es", {"V..01U.O.."}, 2},	{"es", {"V..01U.O.."}, 3},
{"a", {"V..01H.O.."}, 1},	{"e", {"V..01H.O.."}, 2},	{"e", {"V..01H.O.."}, 3},
{"amos", {"V..01W.O.."}, 1},	{"emos", {"V..01W.O.."}, 2},	{"imos", {"V..01W.O.."}, 3},
{"áis", {"V..01Y.O.."}, 1},	{"éis", {"V..01Y.O.."}, 2},	{"ís", {"V..01Y.O.."}, 3},
{"an", {"V..01T.O.."}, 1},	{"en", {"V..01T.O.."}, 2},	{"en", {"V..01T.O.."}, 3},

<sup>4</sup> En las categorías del ejemplo aparecen puntos junto a letras y número, el punto es un comodín y significa cualquier carácter. La explicación detallada de las categorías gramaticales se da en el Anexo A.

### 7.1.3 PARADIGMAS IRREGULARES

Son modelos de conjugación para uno o varios verbos que presentan el mismo tipo de irregularidad. Por ejemplo, el paradigma **termin\_olar** es el modelo de conjugación de verbos como *aprobar*, *comprobar*, *consolar*, *probar*, *renovar*, *rodar*, *soñar* o *volar*. El **1** que aparece en el nombre del paradigma significa que en esa posición puede ir cualquier carácter; por ejemplo en el caso de *aprobar* el **1** es sustituido por una **b** y en el caso de *rodar* en el lugar del **1** se pone una **d**.

Un paradigma irregular se compone de:

- verbo o lista de verbos que compartan la irregularidad.
- modelo de conjugación. Para cada modo y tiempo verbal se establece la lista de desinencias, de manera que uniendo la raíz o lexema y la desinencia se obtiene la forma verbal deseada.

Todas las listas tienen la misma estructura:

desinencia	categoría	conjugación	paradigma
------------	-----------	-------------	-----------

Siguiendo con el ejemplo anterior, **termin\_olar**:

```

presente de indicativo
{"ue.o", {"V..01I.O.."}, 1, termin_olar},
{"ue.as", {"V..01U.O.."}, 1, termin_olar},
{"ue.a", {"V..01H.O.."}, 1, termin_olar},
{"o.amos", {"V..01W.O.."}, 1, termin_olar},
{"o.áis", {"V..01Y.O.."}, 1, termin_olar},
{"ue.an", {"V..01T.O.."}, 1, termin_olar},
    
```

Actualmente tenemos 100 paradigmas, con los que podemos conjugar 2625 verbos irregulares y semirregulares<sup>5</sup>.

<sup>5</sup> La lista de paradigmas irregulares se da en el Anexo B.

## 7.2 FORMAS VERBALES CON PRONOMBRES ENCLÍTICOS

El empleo de los pronombres enclíticos ha ido variando a lo largo del tiempo. Si nos remontamos a los siglos XV y XVI, el uso de formas verbales con pronombres enclíticos era muy frecuente (como ejemplo basta con leer un capítulo de una de las obras más ilustres y conocidas de la Literatura Española: *El Quijote*). Con el paso de los años su uso se ha ido reduciendo y en la actualidad se limita casi exclusivamente a infinitivos, imperativos y gerundios [Santana 94].

Un verbo puede llevar simultáneamente hasta tres pronombres átonos; en tal caso, la partícula **se** debe preceder al resto, el de segunda persona (**te, os**) adelanta siempre al de primera (**me, nos**) y cualquiera de estos dos antecede al de tercera (**le, la, lo, les, las, los**) [Santana 94].

Las formas verbales con pronombre enclítico no siempre resultan de la simple unión verbo y pronombre, sino que pueden producirse alteraciones que afectan tanto a caracteres del verbo como a la acentuación.

- PÉRDIDA DE CARACTERES

Estas alteraciones se producen con tres pronombres concretos: **nos, se** y **os**, y todos los derivados de ellos (**noslo, sela, osnos**, etc) y afectan siempre al último carácter del verbo.

- **nos**

1ª persona del plural del presente de subjuntivo + nos

1ª persona del plural de imperativo + nos

La alteración consiste en que la forma verbal pierde la **s** final:

comamos + nos => comámonos ( y no comámosnos)

Esta pérdida también se produce en otros tiempos del verbo pero apenas se usa el pronombre enclítico en ellos.

- **se**

2ª persona de plural + se

La forma verbal pierde la **s** final:

comamos + selo => comámoselo (y no comámosselo)

- **os**

2ª persona del plural de imperativo + os

La forma verbal pierde la **d** final:

comed + os => comeos (y no comedos)

A excepción del verbo ir:

id + os => **idos**

• **ACENTUACIÓN**

Las formas verbales con enclíticos siguen las reglas generales de acentuación del castellano cuando dicha forma es esdrújula o sobreesdrújula. Cuando a una forma verbal llana o esdrújula se le añade uno o más enclíticos, la forma resultante siempre lleva tilde, independientemente de si el verbo sin enclítico se acentúa o no. Así, por ejemplo, *decía*, *veía*, llevan tilde al igual que *decía-me-lo* y *veía-la*. En cambio *hablaba*, *mirando* no se acentúan y sin embargo *hablábase*, *mirándoos* sí llevan tilde por tratarse de palabras esdrújulas.

Cuando se unen dos enclíticos a una forma verbal aguda, incluyendo los monosílabos, la formación resultante siempre llevará tilde, aunque la forma verbal no se acentúe: *partió* y *partió-se-le*, *de* y *dé-se-la*, *da* y *dá-se-lo*, *pedir* y *pedír-me-la*. En los dos primeros ejemplos la forma verbal lleva tilde mientras que en los dos últimos no y sin embargo todas las formas con enclítico la llevan.

La acentuación de las formaciones con enclítico no sigue las reglas generales en los siguientes casos:

- formas verbales agudas seguidas de un enclítico. El verbo conserva su acento ortográfico: *da-le*, *decid-me*, *reír-se*, *partío-se*. Los dos primeros ejemplos siguen las reglas generales de acentuación, mientras que los dos últimos no, ya que las palabras llanas acabadas en vocal no llevan tilde.

- 2ª persona del plural del imperativo de los verbos reflexivos o en construcción reflexiva. Formas como *marcha-os*, *detene-os*, no llevan tilde a pesar de que el verbo es agudo y se agrupa con un solo enclítico. Sin embargo los verbos de la 3ª conjugación llevan tilde a causa del hiato: *partí-os*, *decí-os*. La única excepción la constituye el verbo ir: *id-os*. En este caso la forma verbal no pierde la d final y al no formar hiato no lleva tilde.

### 7.3 PREFIJOS Y FORMAS VERBALES

Dos de los procesos más significativos en la formación de palabras en castellano son la flexión<sup>6</sup> y la derivación [Goñi 95]. La derivación puede definirse como la formación de palabras a partir de lexemas adjuntando morfemas derivativos (sufijos y prefijos).

Basándonos en este concepto incorporamos un módulo de prefijación encargado de detectar prefijos en las formas verbales. Este módulo economiza recursos de almacenamiento, reduciendo el número de entradas verbales tanto de los diccionarios como de los paradigmas. Por ejemplo, a partir del verbo *vivir* podríamos obtener mediante prefijación:

con + vivir	=>	convivir
re + vivir	=>	revivir
sobre + vivir	=>	sobrevivir

Disponiendo del infinitivo *vivir* en el diccionario de verbos regulares (Infraere.ord) y del paradigma regular de la 3ª conjugación podríamos reconocer y categorizar los cinco verbos mencionados sin necesidad de tener una entrada para cada uno de ellos en el diccionario.

Este proceso es aplicable no sólo a los verbos regulares sino también a los irregulares. Considérese, por ejemplo, el verbo *poner*:

<sup>6</sup> Este concepto se define en el apartado 7.1 de este Capítulo.

ante + poner	=>	anteponer
dis + poner	=>	disponer
ex + poner	=>	exponer
pos + poner	=>	posponer
re + poner	=>	reponer
sobre + poner	=>	sobreponer

De esta manera con una tabla de prefijos<sup>7</sup> y el paradigma del verbo poner podemos conjugar siete verbos distintos. Como puede apreciarse, con este sistema se necesita una base de datos mucho menor.

## 7.4 PROCESAMIENTO DE FORMAS VERBALES

Para el reconocimiento y categorización de los verbos aprovechamos la estructura flexiva del castellano, de manera que a partir de un diccionario y de un conjunto de paradigmas podemos identificar cualquier forma verbal siempre que el sistema conozca el infinitivo asociado a ella.

El análisis de una forma verbal es el siguiente:

1. segmentación de la palabra en raíz y terminación
2. construcción de la forma principal (infinitivo) de la que proviene
3. búsqueda del infinitivo en el diccionario correspondiente en función de si el verbo es regular (Infraere.ord) o irregular (InfinIrreg.ord)
4. categorización de la forma verbal

El procesador verbal considera únicamente formas verbales simples; las formas compuestas como, por ejemplo, el pretérito perfecto de indicativo se tratan como dos formas simples independientes; es decir, la forma verbal compuesta **había amado** se analiza primero **había**, que sería reconocida como 1ª y 3ª persona del singular del pretérito perfecto simple del verbo haber y a continuación **amado**, que se categorizaría

<sup>7</sup> La lista de prefijos que utiliza el programa se da en el Anexo B.

como participio pasado singular masculino del verbo amar. Será en un nivel posterior a este módulo (en la categorización por reglas o en el análisis sintáctico) cuando se estudie la existencia de dos formas verbales seguidas para considerar el conjunto como una forma verbal compuesta.

#### 7.4.1 DESCRIPCIÓN DE LAS FUNCIONES

Las funciones para el tratamiento de las formas verbales se pueden dividir en dos grupos: por un lado, tendríamos las funciones para el manejo de los diccionarios de verbo; por otro, aquellas encargadas de la identificación y análisis de las formas verbales.

##### A) FUNCIONES PARA EL MANEJO DE LOS DICCIONARIOS

Las funciones para el manejo del diccionario de infinitivos regulares (Infrare.ord) aparecen descritas en el Capítulo 6, puesto que son las mismas que para el resto de diccionarios que usa el sistema, por lo que aquí sólo nos ocuparemos de las funciones relacionadas con el diccionario de infinitivos irregulares cuyo manejo sí presenta diferencias con los anteriores.

Todos los diccionarios descritos hasta el momento se encuentran en un fichero de texto y el programa se encarga de leer las palabras y las categorías (si la tienen) para formar la estructura *TDicc*<sup>8</sup> que es la que maneja. En el caso de los verbos irregulares no existe un fichero de infinitivos sino que esta información se encuentra en los paradigmas y el diccionario se obtiene a partir de la lista de paradigmas irregulares.

- **CargaIrregulares**

Sintaxis: **TDicc CargaIrregulares** (void)

Es la función encargada de reservar memoria y formar el diccionario de infinitivos irregulares. El proceso es el siguiente:

- calcula el tamaño del diccionario
- reserva memoria para cada una de las palabras-diccionario<sup>9</sup>

---

<sup>8</sup> En el apartado 6.8 del Capítulo 6 se describe esta estructura.

<sup>9</sup> Véase apartado 6.8 del Capítulo 6

- copia el infinitivo de cada uno de los paradigmas irregulares en el campo palabra de la palabra-diccionario y la categoría correspondiente en el campo categoría. En este caso todas las palabras del diccionario tienen la misma categoría gramatical puesto que todas ellas son infinitivos verbales.
- ordena alfabéticamente el diccionario por palabras  
Devuelve el diccionario con el formato adecuado para su manejo por parte del programa.

- LiberaIrregulares

Sintaxis: void **LiberaIrregulares** (TDicc \*Dicc)

Libera la memoria reservada para el diccionario de infinitivos irregulares.

## B) FUNCIONES PARA EL ANÁLISIS DE LAS FORMAS VERBALES

- es\_verb

Sintaxis: T Categorías **es\_verb** (TPalabraDic palIn, char infin[MAX\_NUM\_CAT][100],  
boolean no\_comprobar\_lista\_infin, int \*paradigma)

Es la función principal para el reconocimiento y categorización de las formas verbales.

Esta labor se realiza en varias etapas:

1. Se comprueba si se trata de una forma verbal regular

De esta tarea se ocupa la función **es\_termin\_regular**.

Sintaxis: T Categoría **es\_termin\_regular** (TPalabraDic palIn, char infinit[100], boolean  
no\_comprobar\_lista\_infin, int \*paradigma,  
TModoBusqueda modo)

Recorre las listas de paradigmas regulares para ver si la terminación de la palabra considerada se corresponde con alguna de ellas. En caso afirmativo, descompone la palabra en raíz verbal y morfema flexivo. A partir de la raíz forma el infinitivo al que pertenece la forma verbal reconocida y lo busca en el diccionario de infinitivos regulares (Infraere.ord). Para la búsqueda en este diccionario utiliza la

función *BuscaPalabra*, sin distinguir mayúsculas y minúsculas (*ignorarCase*) y se busca la primera ocurrencia de la palabra en el diccionario (*busquedaInicial*)<sup>10</sup>.

La función devuelve la categoría asociada a la terminación regular reconocida y el infinitivo, en caso de que haya sido identificada. El infinitivo se devuelve a través del parámetro de entrada *infin*.

A continuación se repite el proceso por si la palabra en cuestión pertenece a distintos tiempos verbales, modos o personas. Por ejemplo, la forma verbal **come** puede ser 3ª persona del singular del presente de indicativo del verbo comer o 2ª persona del singular del imperativo del citado verbo. Véase también el caso de los verbos regulares y algunos irregulares de la 2ª y 3ª conjugación; los cuales comparten la misma forma para 1ª y 3ª persona del singular del pretérito indefinido de indicativo (por ej. **yo vivía, él vivía**) o el de los verbos de la 1ª conjugación que coinciden en forma verbal para la 1ª y 3ª persona del singular del pretérito perfecto simple de indicativo (por ej. **yo amaba, él amaba**). Esta coincidencia de formas entre 1ª y 3ª persona del singular se da en la mayoría de los tiempos verbales tanto de la 1ª como de la 2ª y 3ª conjugación<sup>11</sup>.

## 2. Se comprueba si es una forma verbal irregular

La función encargada de realizar esta tarea es **es\_termin\_irregular**.

Sintaxis: TCategoria **es\_termin\_irregular** (TPalabraDic palIn, char *infin* [100],  
boolean *no\_comprobar\_lista\_infin*, int \*paradigma, TModoBusqueda modo)

El proceso es análogo al descrito para las formas regulares, solo que en este caso se comprueba si el sufijo de la palabra bajo estudio se encuentra en alguno de los paradigmas irregulares. En caso afirmativo se copia el infinitivo del paradigma correspondiente en *infin* y se devuelve la categoría gramatical asociada a la terminación reconocida. La diferencia con el proceso anterior es que ahora el infinitivo no se busca en ningún diccionario sino que se obtiene directamente del paradigma.

---

<sup>10</sup> El significado de estas variables y la función *BuscaPalabra* se describen en el Capítulo 6.

<sup>11</sup> En todo momento estamos considerando tiempos verbales simples. Como ya se indicó anteriormente, las formas compuestas se analizan como dos formas simples.

### 3. Se comprueba si es una forma regular de un verbo irregular

De esta labor se ocupa la función **es\_termin\_irregular\_regular**.

Sintaxis: TCategoria **es\_termin\_irregular\_regular** (TPalabraDic palIn, char infinit[100], boolean no\_comprobar\_lista\_infin, int \*paradigma, TModoBusqueda modo)

El proceso es el mismo que el descrito en el punto 1, pero ahora el infinitivo de la forma reconocida se busca en el diccionario de infinitivos irregulares (InfinIrregu.ord). En caso de encontrarlo se comprueba que la forma verbal reconocida como regular no se encuentra también entre las formas de los paradigmas irregulares, ya que no puede darse el caso de que una misma forma verbal sea regular e irregular al mismo tiempo.

Como en los casos anteriores, si la forma verbal analizada ha sido reconocida la función devuelve la categoría gramatical de la forma considerada y el infinitivo.

### 4. Formas verbales con enclíticos

El siguiente paso es estudiar si la palabra considerada es una forma verbal con uno o varios pronombres enclíticos<sup>12</sup>. El proceso de identificación se organiza en dos partes:

- a) se reconocen los posibles pronombres enclíticos.
- b) se analiza la forma verbal correspondiente

#### 4.1 Reconocimiento de los pronombres enclíticos

La función encargada de esta parte es **elimina\_enclítico**.

Sintaxis: int **elimina\_enclítico** (char \*palabra\_buscada, int &cont, int &doble\_busqueda)

En primer lugar comprueba si la palabra incorpora un pronombre de los definidos en la lista de pronombres enclíticos. Cada pronombre de la lista lleva asociado un número que servirá para su identificación. Para ello utiliza la función **termina\_en\_enclítico**.

Sintaxis: int **termina\_en\_enclítico** (char \*palabra\_buscada, int &i)

Esta función devuelve 0 si la palabra no contiene ningún enclítico o el número asociado al pronombre reconocido.

---

<sup>12</sup> En el Anexo B se da la lista de pronombres enclíticos que utiliza el procesador verbal.

Una vez identificado el pronombre enclítico se considera la conservación o corrección de la tilde de acuerdo con lo expuesto en el apartado 7.2. Para el estudio de la tilde se realiza la silabificación (separación en sílabas) de la palabra, determinando cuál es su sílaba y letra tónica. A partir de esta información y de las reglas de acentuación se determina si hay que cambiar de posición, conservar o hacer desaparecer la tilde de la sílaba tónica. Las reglas para tomar esta decisión son:

- Si el verbo con enclítico no lleva tilde y el pronombre no es *os*, el pronombre enclítico es monosílabo y la forma verbal sin enclítico será aguda. Luego hay que tildar la forma no pronominalizada siempre que no sea monosílabo.
- Si se trata de una forma verbal que lleva el pronombre enclítico *os* y no está tildada, la forma sin enclítico no llevará tilde. Por ejemplo: *comed*, *comeos*.
- Si el verbo con enclítico está acentuado de acuerdo con las reglas ortográficas del castellano, la forma verbal sin enclítico puede o no llevar tilde (por ejemplo, *comámonos* y *partióse* llevan tilde, sin embargo *comamos* no se acentúa mientras que *partió sí*). En estos casos se analiza la forma no pronominalizada con y sin acento.
- Si el verbo con enclítico no está tildado según las reglas generales de acentuación, la forma sin enclítico conserva la tilde.

#### 4.2 Análisis de la forma verbal sin enclítico

Una vez obtenida la forma no pronominalizada, se procede a su análisis de acuerdo con los casos expuestos en el apartado 7.2. Para este estudio se han implementado las siguiente funciones:

- *BuscaEnclitico\_conD*

Sintaxis: T Categorías **BuscaEnclitico\_conD** (TPalabraDic palIn, int encl, T Categorías cats, char infin[MAX\_NUM\_CAT][100], boolean no\_comprobar\_lista\_infin, int paradigma[MAX\_NUM\_CAT], int \*doble\_búsqueda)

Esta función estudia las irregularidades en ciertos verbos cuando se les añade el pronombre enclítico **os** y que consiste en que la 2ª persona del plural del imperativo pierde la **d** final (ej. *comed* + *os* => *comeos*).

Añade una d al final de la palabra y la forma resultante se analiza según lo expuesto en los pasos 1 2 y 3. La función devuelve el infinitivo y las posibles categorías asociadas a la forma verbal bajo estudio.

- *BuscaEnclítico\_conS*

Sintaxis: T Categorías **BuscaEnclítico\_conS** (TPalabraDic palIn, int encl, T Categorías cats, char infin[MAX\_NUM\_CAT][100], boolean no\_comprobar\_lista\_infin, int paradigma[MAX\_NUM\_CAT], int doble\_busqueda)

Esta función estudia las irregularidades en ciertos verbos cuando se les añade el pronombre enclítico **nos**, **se** o una de sus combinaciones (**sele/a/o/es/as/es**, **seme**, **sete**, **senos**, **seos**) y que consiste en que la forma verbal pierde la s final (ej. *comamos + selo => comámoselo*).

Añade una s al final de la palabra considerada y la analiza siguiendo los pasos 1, 2 y 3. Como en el caso anterior la función devuelve el infinitivo y las posibles categorías de la forma verbal analizada.

- *BuscaEnclítico\_sinAcento*

Sintaxis: T Categorías **BuscaEnclítico\_sinAcento** (TPalabraDic palIn, int encl, T Categorías cats, char infin[MAX\_NUM\_CAT][100], boolean no\_comprobar\_lista\_infin, int paradigma[MAX\_NUM\_CAT])

Esta función estudia las irregularidades en ciertos verbos cuando se les añaden pronombres enclíticos y que consiste en que la forma resultante lleva acento y sin embargo la forma verbal origen no lo llevaba. (ej. *hablaba => hablábase*).

Elimina la tilde de la forma verbal y se repiten los pasos 1, 2 y 3. Análogamente la función devuelve las posibles categorías de la forma verbal considerada junto con su infinitivo.

- *casoParticipio*

Sintaxis: bool casoParticipio (int numEncl, T Categoría cat)

Comprueba si la categoría verbal es la 2ª persona del plural del presente de imperativo y el pronombre enclítico es **os** porque en ese caso no se trata de una forma verbal con enclítico (ej. *desconcertados*, *despavoridos*).

Las funciones para el reconocimiento y categorización de las formas verbales sin pronombre enclítico ya existían anteriormente, por lo que nos hemos limitado a incorporarlas al programa. Lo que se ha añadido es el tratamiento de las formas verbales con enclíticos que sufren alteraciones ya sea en su grafía (pérdida de caracteres finales) o en su acentuación. En la versión anterior ya se trataban las formas verbales con enclíticos pero sin tener en cuenta las reglas expuestas en el apartado 7.2.

- `es_participio_irregular`

Sintaxis: TCategoria **es\_participio\_irregular** (TPalabraDic palln, char infinit[100])

Esta función se ocupa del análisis de los participios irregulares. En estos casos puede tratarse de verbos que son regulares en el resto de sus formas, como por ejemplo abrir, o de verbos que presentan irregularidades en otras formas, por ejemplo morir.

Para su reconocimiento y categorización se dispone de una lista donde junto al participio irregular aparece, por este orden, la categoría gramatical, la conjugación a la que pertenece y el infinitivo<sup>13</sup>.

Los participios irregulares que ya están implementados en los paradigmas irregulares se han excluido de esta lista para evitar redundancias.

<pre> {"abierto", {"V..46S.M.."}, 3, "abrir"}, {"abierta", {"V..46S.F.."}, 3, "abrir"}, {"abiertos", {"V..46P.M.."}, 3, "abrir"}, {"abiertas", {"V..46P.F.."}, 3, "abrir"},                 </pre>
--

La función comprueba si la palabra en cuestión se encuentra en esta lista y en caso afirmativo devuelve la categoría correspondiente y el infinitivo.

Todas las funciones para el procesamiento de las formas verbales se encuentran en la librería Dicc.lib<sup>14</sup>.

<sup>13</sup> La lista de participios irregulares se da en el Anexo B.

<sup>14</sup> Véase apartado 6.8 del Capítulo 6

## 7.5 CORRECCIÓN Y DEFINICIÓN DE NUEVOS PARADIGMAS IRREGULARES

### 7.5.1 CORRECCIÓN DE PARADIGMAS IRREGULARES

En la fase de Entrenamiento del sistema descubrimos que algunos paradigmas ya existentes contenían errores en una o varias formas y otros estaban incompletos.

- Corrección de paradigmas irregulares

Errores en alguna persona de un determinado modo y tiempo verbal.

Paradigmas	Forma verbal	Forma incorrecta	Forma correcta
termin_gar	1ª pers. sing. presente indicativo	go	gue
termin_haber	2ª pers. Sing. imperativo	ha	he
termin_caber	1ª/3ª pers. sing. presente indicativo	cabe/cabe	quepo/cabe

- Paradigmas incompletos

Ausencia de una persona de un modo y tiempo concreto, de un tiempo verbal entero o de un infinitivo en el grupo de infinitivos de un paradigma.

Paradigma	Forma o tiempo añadido
termin_haber	Hay (3ª pers. sing. presente indicativo)
termin_guir	Pretérito imperfecto y futuro imperfecto de subjuntivo
termin_o2ir	Pretérito imperfecto y futuro imperfecto de subjuntivo
termin_zar	Verbo rentabilizar <sup>15</sup>

### 7.5.2 PARADIGMAS PENDIENTES

En la versión anterior del conjugador verbal faltaban algunos paradigmas por implementar; se había definido el paradigma y los infinitivos que pertenecían al mismo

---

<sup>15</sup> Faltaba incluir este verbo en el paradigma termin\_zar.

pero no se había la lista de sufijos de cada tiempo verbal, por lo que decidimos completar la tarea. Los paradigmas pendientes figuran en la siguiente Tabla:

Paradigma	Verbos
termin_raer	raer
termin_errar	errar
termin_asir	asir
termin_yacer	yacer
termin_roer	roer
termin_placer	placer
termin_puar	puar
termin_rehusar	rehusar

Tabla 7.2 Paradigmas irregulares pendientes

### 7.5.3 NUEVOS PARADIGMAS IRREGULARES

Según Jaime Suances-Torres en la lengua española hay 103 modelos de conjugación, que se recogen en la Base de Datos del Verbo Español [Suances 98].

Tras conocer este trabajo decidimos comprobar si nuestro módulo para el tratamiento de los verbos contenía los 103 modelos que aparecen en el mencionado trabajo. El resultado fue que nos faltaban paradigmas para algunos verbos irregulares. Para la implementación de los nuevos paradigmas nos ha sido de gran ayuda el *conjugador verbal* del Grupo de Investigación en Estructuras de Datos de la Universidad de las Palmas de Gran Canaria, disponible de forma gratuita hasta el momento en Internet [Conjverb 99].

La Tabla 7.3 presenta los nuevos paradigmas irregulares incorporados así como ejemplos de verbos que pertenecen a esos paradigmas.

Paradigma	Verbos
termin_henchir	henchir
termin_gollar	degollar, regoldar
termin_eilar	cafeinar, descafeinar
Termin_u8nir <sup>16</sup>	counir, reunir
termin_embaír	embaír
termin avergonzar	avergonzar
termin_u9ir <sup>17</sup>	argüir
termin_cir	lucir, relucir
termin_guir	distinguir
Termin_ver2	antever, prever

Tabla 7.3 Nuevos paradigmas irregulares

En la actualidad podemos analizar 101 de los 103 modelos de conjugación. Los verbos para los que no disponemos de paradigmas son: heroizar y todaviizar; verbos que, por otra parte, no se encuentran en el Diccionario de la Real Academia Española.

## 7.6 FICHEROS DE DEPURACIÓN

### 7.6.1 VERBOS

Es el fichero de depuración donde se recogen todas las formas verbales encontradas en el texto procesado: regulares, irregulares, con y sin enclítico, con y sin prefijo. El formato de este fichero es el descrito en el apartado 5.2.4 del Capítulo 5.

aprovechó	V..41H.0..	1	infraere.ord	-> , aprovechó(aprovechar) un descuido
tenga	V..02H.0..	1	infraere.ord	-> país tenga(tener) un ejército
tenga	V..034.0..	1	infraere.ord	-> país tenga(tener) un ejército

Ejemplo del fichero **verbos.dep**

<sup>16</sup> El número 8 indica que la letra anterior lleva tilde.

<sup>17</sup> El número 9 indica que la letra anterior lleva diéresis.

### 7.6.2 ENCLÍTICOS

Este fichero incluye todas las formas verbales con enclítico reconocidas. Su formato es el mismo que el del caso anterior con la diferencia de que en el anterior se escribían todas las posibles categorías de la forma verbal considerada y ahora solo se escriben las que llevan información del pronombre enclítico.

recogerla	V..00N.007	1	diccInfinIrregu	->a recogerla(recoger) » ,
perdonárselas	V..00N.026	1	infraere.ord	->solían perdonárselas(perdonar) por su
detenerse	V..00N.002	1	infraere.ord	->ni detenerse(detener) , los

Ejemplo del fichero **enclíticos.dep**

### 7.6.3 VERBOS CON PREFIJO

Es el fichero que contiene todas las formas verbales reconocidas por la función *BuscaconPrefijo*<sup>18</sup>. El formato de este fichero es algo distinto al de los restantes ficheros de depuración:

Forma verbal	Prefijo, infinitivo	Categoría	Encontrada en diccionario	diccionario en el que se encuentra el infinitivo
--------------	---------------------	-----------	---------------------------	--

desconcentren	(des, concentrar)	V..03T.0..	1	infraere.ord
desconcentren	(des, concentrar)	V..02T.0..	1	infraere.ord
coprotagoniza	(co, protagonizar)	V..01H.0..	1	infraere.ord
coprotagoniza	(co, protagonizar)	V..02U.0..	1	infraere.ord

Ejemplo del fichero **verbos\_conPrefijo.dep**

<sup>18</sup> Esta función se describe en el apartado 6.8.2 del Capítulo 6.

## 7.7 RESULTADOS

### 7.7.1 RESULTADOS DEL PROCESADOR VERBAL

<b>TEXTOS EL MUNDO</b>			
	Verbos	Verbos con Enclítico	Verbos Enclítico/ Verbos <sup>19</sup>
Año 1994	22.60 %	0.2034 %	0.9088 %
Año 1995	21.92 %	0.2010 %	0.9174 %
<b>Total</b>	<b>22.26 %</b>	<b>0.2022 %</b>	<b>0.9131 %</b>

Tabla 7.4 Resultados del conjugador verbal  
Porcentajes sobre el número total de palabras procesadas  
(Datos medios mensuales)

<b>TEXTOS 860</b>			
	Verbos	Verbos con Enclítico	Verbos Enclítico/ Verbos <sup>20</sup>
Corpus1	27.94	0.6323	2.26
Corpus2	27.26	0.5371	1.97
Corpus3	27.01	0.6287	2.33
Corpus4	26.98	0.5894	2.18
Corpus5	27.33	0.6881	2.52
Corpus6	27.88	0.5721	2.05
Corpus7	27.52	0.6284	2.28
<b>Media</b>	<b>27.42</b>	<b>0.6109</b>	<b>2.23</b>

Tabla 7.5 Resultados del conjugador verbal  
Porcentajes sobre el número total de palabras procesadas

<sup>19</sup> Porcentaje de formas verbales con enclítico sobre el número total de verbos.

<sup>20</sup> Porcentaje de formas verbales con enclítico sobre el número total de verbos.

A la vista de los resultados, cabe destacar el bajo porcentaje de formas verbales con pronombre enclítico de los artículos de El Mundo (0.9131 %) con respecto al encontrado en los Textos 860 (2.23 %).

	<b>Textos El Mundo</b>	<b>Textos 860</b>
Palabras por frase	25	29
Verbos por frase	6	8
% Palabras de la frase que son verbos	23.24 %	27.4 %

Tabla 7.6 Datos medios sobre el total de frases procesadas

<b>TEXTOS 860</b>						
	Infinitivos	Gerundios	Participios	Indicativo	Imperativo	Subjuntivo
Corpus 1	20.51	2.72	27.30	42.09	0.046	7.32
Corpus 2	19.85	2.84	26.08	44.02	0	7.20
Corpus 3	20.28	2.81	26.36	43.81	0.023	6.71
Corpus 4	19.59	2.96	26.94	43.59	0.024	6.88
Corpus 5	20.18	2.39	26.22	43.53	0.102	7.57
Corpus 6	20.10	2.73	26.43	44.89	0	5.86
Corpus 7	20.79	3.11	24.98	44.21	0.047	6.85
<b>Media</b>	<b>20.18</b>	<b>2.79</b>	<b>26.33</b>	<b>43.73</b>	<b>0.034</b>	<b>6.91</b>

Tabla 7.7 Distribución por modos de las formas verbales en los Textos 860

(Porcentajes sobre el número total de verbos)

Como podemos comprobar, el modo indicativo es el más utilizado (43.73 %) seguido por el infinitivo (20.18 %). Por el contrario, el modo imperativo es el que se usa con menor frecuencia (0.034 %), cosa lógica ya que este tipo de modo verbal no es el más característico de este tipo de textos (jurídicos, documentos).

### 7.7.2 ESTUDIO DE LOS VERBOS MÁS UTILIZADOS EN EL CORPUS DE ENTRENAMIENTO

Los datos de este estudio se han sacado a partir de los ficheros de depuración *verbos.dep* y *enclíticos.dep* correspondientes a los artículos periodísticos de El Mundo del período 1994-95. Este estudio tiene como objetivo comprobar qué formas verbales, infinitivos y pronombres enclíticos se utilizan con mayor frecuencia. Los resultados obtenidos se muestran en las siguientes Tablas.

- FORMAS VERBALES E INFINITIVOS

TEXTOS EL MUNDO			
Forma verbal	Frecuencia	Infinitivo	Frecuencia
ha	14707	haber	18501
es	12432	ser	10445
fue <sup>21</sup>	9494	estar	9831
han	4985	poder	5949
era	4147	hacer	3503
está	2937	tener	2797
puede	2094	ir	2737
sido	1949	ver	2411
había	1759	volver	1547
sea	1557	deber	1511

Tabla 7.8 Diez primeras formas verbales e infinitivos más frecuentes en 12 meses (Datos medios mensuales)

<sup>21</sup> La forma verbal **fue** puede pertenecer tanto al verbo ser como al ir; en este nivel de análisis no es posible distinguir a cuál de los dos verbos pertenece, para ello sería necesario conocer el contexto en el que dicha forma aparece.

<b>TEXTOS 860</b>			
<b>Forma verbal</b>	<b>Frecuencia</b>	<b>Infinitivo</b>	<b>Frecuencia</b>
ha	129	ser	532
es	116	haber	282
ser	82	estar	243
puede	73	poder	191
sido	63	deber	137
debe	58	preguntar	99
han	55	tener	81
son	49	hacer	77
será	48	decir	59
sean	47	ver	57

Tabla 7.9 Diez primeras formas verbales e infinitivos más frecuentes en los Textos 860 (Datos medios por Corpus)

Forma verbal	fue	fueron	fuera	fuese	fuesen	fuere
Nº total	98	35	13	9	2	2
Verbo ser	96	32	13	9	2	2
Verbo ir	2	3	0	0	0	0

Tabla 7.10 Formas verbales comunes a los verbos *ser* e *ir* encontradas en los Textos 860

De las 159 formas verbales totales comunes a los verbos *ser* e *ir* encontradas en los Textos 860, el 98.11 % son del verbo *ser* y tan solo el 1.89 % pertenecen al verbo *ir*. Luego ante la duda, la probabilidad de que la forma considerada sea del verbo *ser* es mucho más elevada.

- ENCLÍTICOS

En primer lugar estudiamos qué pronombres enclíticos se añaden con mayor frecuencia a las formas verbales y después qué formas verbales con enclíticos son las más comunes en el Corpus de Entrenamiento.

Como puede comprobarse en la Tabla 7.11, el pronombre enclítico que más veces se usa con formas verbales es **se**, seguido por **lo** y **le**.

<b>TEXTOS EL MUNDO</b>					
Enclítico	Frecuencia absoluta	Enclítico	Frecuencia absoluta	Enclítico	Frecuencia absoluta
se	33	os	12	Mela	6
lo	23	sela	12	Senos	6
le	22	seles	11	Melas	5
la	19	nosla	10	tela	5
los	18	melos	9	telos	4
nos	18	selas	9	telas	3
las	17	selos	9	nosle	3
les	17	noslo	8	noslos	2
melo	15	seme	8	mele	1
selo	15	telo	8	noslas	1
te	14	sele	7	sete	1

Tabla 7.11 Empleo de pronombres enclíticos en los textos de El Mundo

<b>TEXTOS 860</b>			
Enclítico	Frecuencia absoluta	Enclítico	Frecuencia absoluta
se	734	nos	6
lo	38	me	5
las	34	selo	2
le	34	sele	2
los	25	sela	1
la	23	selos	1
les	18		

Tabla 7.12 Empleo de pronombres enclíticos en los Textos 860

<b>TEXTOS EL MUNDO</b>			
Forma verbal con enclítico	Frecuencia	Forma verbal con enclítico	Frecuencia
refiriéndose	751	preguntarse	387
acercarse	701	colocarse	302
retirarse	657	haciéndose	299
entrevistarse	487	olvidarse	294
enterarse	432	iniciarse	293

Tabla 7.13 Formas verbales con pronombres enclíticos más frecuentes en 24 meses (1994-95)

<b>TEXTOS 860</b>			
Forma verbal con enclítico	Frecuencia	Forma verbal con enclítico	Frecuencia
beneficiarse	30	considerarse	12
obtenerse	22	basándose	12
presentarse	16	realizarse	11
interpretarse	15	hacerlo	11
hacerse	15	adaptarse	11

Tabla 7.14 Formas verbales con pronombres enclíticos en los Textos 860

<b>TEXTOS 860</b>				
infinitivos	gerundios	indicativos	subjuntivos	imperativos
89.15 %	9.11 %	0.22 %	0.11 %	1.41 %

Tabla 7.15 Distribución de los tiempos verbales que incorporan pronombres enclíticos en los Textos 860

De acuerdo con los datos anteriores, el empleo de pronombres enclíticos en la actualidad se limita casi exclusivamente a infinitivos, gerundios e imperativos, hecho que ya apuntaba [Santana 94]. Es más, podemos decir, que en la mayor parte de los casos los pronombres enclíticos se incorporan a infinitivos.

**CAPÍTULO 7 DETECCIÓN Y TRATAMIENTO DE FORMAS VERBALES...139**

7.1	VERBOS .....	139
	• <i>CLASIFICACIÓN DE LOS VERBOS</i> .....	141
7.1.1	<i>LOS DICCIONARIOS</i> .....	142
7.1.2	<i>PARADIGMAS REGULARES</i> .....	142
7.1.3	<i>PARADIGMAS IRREGULARES</i> .....	143
7.2	FORMAS VERBALES CON PRONOMBRES ENCLÍTICOS .....	144
7.3	PREFIJOS Y FORMAS VERBALES .....	146
7.4	PROCESAMIENTO DE FORMAS VERBALES .....	147
7.4.1	<i>DESCRIPCIÓN DE LAS FUNCIONES</i> .....	148
7.5	CORRECCIÓN Y DEFINICIÓN DE NUEVOS PARADIGMAS IRREGULARES .....	155
7.5.1	<i>CORRECCIÓN DE PARADIGMAS IRREGULARES</i> .....	155
7.5.2	<i>PARADIGMAS PENDIENTES</i> .....	155
7.5.3	<i>NUEVOS PARADIGMAS IRREGULARES</i> .....	156
7.6	FICHEROS DE DEPURACIÓN.....	157
7.6.1	<i>VERBOS</i> .....	157
7.6.2	<i>ENCLÍTICOS</i> .....	158
7.6.3	<i>VERBOS CON PREFIJO</i> .....	158
7.7	RESULTADOS .....	159
7.7.1	<i>RESULTADOS DEL PROCESADOR VERBAL</i> .....	159
7.7.2	<i>ESTUDIO DE LOS VERBOS MÁS UTILIZADOS EN EL CORPUS DE ENTRENAMIENTO</i> .....	161