

# **Reconocedor y generador de formas verbales con pronombres enclíticos**

Autores: Z. Hernández, G. Rodríguez, J. Pérez, F. Carreras, S. Bogliani.

Investigador Principal: O. Santana.

Grupo de Investigación en Estructuras de Datos

Departamento de Informática y Sistemas

Universidad de las Palmas de Gran Canaria

## **Resumen**

El propósito del presente trabajo consiste en: a) reconocer las diferentes formas conjugadas de un verbo, identificando su infinitivo, tiempo, número y persona y b) generar una forma verbal a partir de su infinitivo y conjugación; en ambos procesos se consideran las modificaciones que son debidas a la presencia de pronombres enclíticos. El reconocimiento y la generación de las formas verbales se realizan por medio de una misma estructura de datos de manera bidireccional.

## **0.– Introducción.**

Con la popularización y el uso generalizado de los ordenadores surge la necesidad de facilitar la comunicación entre las personas y las máquinas; este objetivo se cubre en mayor medida cuando las interfaces se acercan al lenguaje natural. Se plantean en esta línea problemas de consultas a bases de datos documentales, generación y análisis de textos escritos, problemas de corrección ortográfica, etc. El análisis morfológico constituye una base sólida sobre la que es posible aproximarse con mayor rigor al sintáctico e incluso al semántico.

Este trabajo se enmarca en un conjunto de herramientas orientadas a servir de ayuda en la elaboración de documentos escritos; unas están dedicadas a analizar el texto tales como frecuencia de aparición de palabras, empleo de formas verbales, detección de aliteraciones, etc. y otras ofrecen facilidades para la creación literaria como los diccionarios de sinónimos o ideológicos, búsquedas en texto libre, etc. Se ha desarrollado un instrumento con la capacidad de: a) identificar el infinitivo y la conjugación que presenta una forma verbal dada y b) generar una determinada forma verbal a partir del infinitivo y la conjugación que se desee; se entiende como parte de la conjugación la incorporación de pronombres enclíticos.

En la sección 1 se hace un estudio sobre la morfología del verbo, las características de los pronombres enclíticos se exponen en la sección 2, en la sección 3 se plantea de forma detallada el problema del reconocimiento y se concreta su realización, la sección 4 trata la generación de formas verbales, los resultados experimentales se presentan en la sección 5 y las conclusiones y perspectivas se muestran en la sección 6.

## **1.— Morfología del verbo.**

Por sus características formales, el verbo es aquella parte de la oración que contiene *morfemas* flexivos de número, persona, tiempo y modo. Suele aplicarse la denominación de *desinencias* a los morfemas de número y persona, y la de *características* a los de modo y tiempo. Si se suprime de una forma verbal su desinencia y característica queda el *lexema* o *raíz*. La unión de la raíz con la característica se denomina *tema*. Cualquier forma verbal se construye básicamente con una raíz y una *terminación* constituida por los morfemas.

Si se exceptúa la posición del *acento de intensidad*, la mayor parte de los verbos son regulares y mantienen invariable la raíz al conjugarse. Entre otros tipos de anomalías los verbos irregulares presentan variaciones vocálicas, consonánticas o

mixtas en su raíz. Las irregularidades afectan siempre a más de un tiempo aunque pueden agruparse en un máximo de tres los grupos de tiempos que comparten la misma irregularidad.

## 2.— Pronombres enclíticos.

El empleo de los pronombres enclíticos ha ido reduciéndose con el tiempo, y ahora se usan casi exclusivamente cuando el verbo está en infinitivo, imperativo o gerundio.

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, les, la, las, lo, los.**

Al unirse los pronombres enclíticos con el verbo, se producen algunas alteraciones que afectan al último carácter del verbo:

– delante del enclítico **nos** se pierde la **s** de la primera persona del plural del subjuntivo e imperativo

**comamos + nos ⇒ comámonos**

esta pérdida se produce también en otros tiempos del verbo, pero el pronombre enclítico es muy poco usual en ellos. En la segunda persona del plural seguida del enclítico **se** desaparece una **s**

**comamos + selo ⇒ comámoselo**

– la forma reflexiva de la segunda persona del plural del imperativo pierde la **d**

**comed + os ⇒ comeos**

### 2.1.— El acento ortográfico.

El acento ortográfico de las formaciones con enclíticos está siempre de acuerdo con las reglas generales cuando dicha formación es esdrújula o sobresdrújula. Cuando una forma verbal llana o esdrújula se agrupa con uno o más enclíticos, la vocal prosódicamente acentuada del verbo lleva siempre tilde, lo exija o no cuando se emplea sin enclíticos: **decía-me-lo, veía-la, oía-lo** —de acuerdo con las formas verbales **decía, veía, oía—, hablaba-se, mirando-os, quisiera-lo** —en contraste con **hablaba, mirando, quisiera**. Y cuando una forma verbal aguda —incluyendo las monosilábicas— se agrupa con dos enclíticos, la vocal prosódicamente acentuada del verbo se escribe siempre con tilde, lo requiera o no cuando se emplea sola: **partió-se-le, oír-se-lo, dé-se-la** —de acuerdo con **partió, oír, dé—, dá-se-lo, dí-me-lo, decid-nos-lo, pedir-me-la** —en contraste con **da, di, decid, pedir**.

El acento ortográfico de las formaciones con enclíticos deja de estar de acuerdo con las reglas generales en algunos casos en que la formación resulta con acentuación llana. En las formas verbales agudas —incluyendo las monosilábicas— seguidas de un sólo enclítico: **da-le, fui-me, decid-me, reír-se, oír-lo; dé-le, salí-me, partió-se** el verbo conserva su acento ortográfico originario; solamente en los cinco primeros ejemplos la formación se atiene a las reglas generales del uso ortográfico (**dale** como **sale**, **fuime** como **fuiste**,...) y en los tres últimos ejemplos diverge de las reglas generales, ya que una palabra llana terminada en vocal normalmente no lleva tilde.

Aparecen también sometidos a un régimen ortográfico especial los imperativos plurales de los verbos reflexivos, o en construcción reflexiva, tras la pérdida de la desinencia **d**. Formas como **marcha-os, detene-os** han de emplearse sin tilde, a pesar de que la forma verbal es aguda y de que se agrupa con sólo un enclítico. Sin embargo los verbos de la tercera conjugación llevan tilde a causa del hiato como **partí-os**, salvo la forma **id-os** del verbo **ir** en el que **í-os** es un arcaísmo.

### **3.– El reconocimiento de formas verbales.**

El proceso de identificación se organiza en dos fases: primero se reconocen los posibles pronombres enclíticos y luego la forma o formas verbales correspondientes; la meta es el infinitivo (o posibles infinitivos) acompañado de los tiempos, números y personas (y en su caso de la combinación de pronombres enclíticos) que dan lugar a la forma original.

– Las formas verbales afectadas por pronombres enclíticos necesitan deshacerse de tales accidentes, luego se retoca la forma no pronominalizada si estuviera afectada por alguna de las alteraciones señaladas en la sección 2. Se considera la desaparición o conservación de la tilde en virtud de que debe mantenerse la sílaba tónica. El procedimiento que se encarga del estudio de la tilde emplea un separador silábico y un comprobador de las reglas ortográficas de acentuación que han sido desarrollados al efecto; permite realizar la separación en sílabas de una palabra, determinar cuál es su sílaba tónica y cambiar de posición, situar o hacer desaparecer la tilde dada la sílaba tónica.

– Como por un lado el número de raíces diferentes de cada verbo es bastante menor que el de sus formas verbales y por otro las terminaciones posibles son limitadas y sabidas, resulta factible afrontar la solución en tres pasos: 1) fragmentar la forma verbal en raíz y terminación, 2) relacionar la raíz con su infinitivo y 3) asociar la terminación con la flexión que le corresponde. A fin de tener un número óptimo de raíces, se opta por un criterio de corte basado en terminaciones que no siempre da lugar a la raíz lingüística; si la forma coincide con una terminación se toma como raíz parte del principio de la terminación y al extender esta norma a raíces reducidas a una letra, se aminoran ambigüedades. A cada raíz se aplica una función discriminadora que permite acceder de forma eficiente a un registro con una lista de infinitivos a los que pertenece, acompañados de los grupos de terminaciones que admite y para cada

uno se comprueba con facilidad qué flexión representa la terminación dada en caso de que la admita.

#### **4.– La generación de formas verbales.**

Generar una forma verbal dados el infinitivo y la conjugación que se pretende obtener es una tarea resuelta mediante la definición de un conjunto de reglas y excepciones que atienden a unos modelos de conjugación, [San93a].

Al disponer de una estructura de datos que permite conocer para cada verbo cuál es su infinitivo, qué raíces tiene, qué terminaciones admite cada raíz y qué tiempo, número y persona representa cada una de ellas, es posible generar una forma conjugada a partir del infinitivo y del tiempo, número y persona que se desee con poco esfuerzo. Basta con añadir la capacidad de acceder a las distintas raíces que un determinado verbo admite en su conjugación y detallar cómo se debe llevar a cabo la unión entre raíces y terminaciones.

El proceso de generación se organiza en dos fases: primero se genera la forma verbal conjugada y luego se unen los posibles pronombres enclíticos; el resultado será la forma (o posibles formas) que se obtienen a partir del infinitivo al conjugarlo en el tiempo, número y persona dados y añadirle la combinación de pronombres enclíticos especificada.

– A partir del infinitivo, el proceso recorre las raíces del verbo y recoge los grupos de terminaciones asociados para determinar qué raíces y con qué terminaciones dan lugar al tiempo, número y persona requeridos.

– Cuando se haya de añadir una combinación de pronombres enclíticos a una forma verbal, han de tenerse en cuenta las reglas de unión expuestas en la sección 2 en cuanto a desaparición o simplificación de consonantes y a la posible aparición de la tilde. El procedimiento que se encarga del estudio de la tilde es el comentado en la sección 3.

## **5.– Resultados experimentales.**

La aplicación se ha desarrollado en C++ bajo Windows. Las estructuras de datos utilizadas por el programa ocupan un total de 1,01 Mb en disco. Sobre un procesador Pentium a 100 Mhz con 16 Mb de memoria RAM y un disco con velocidad de transferencia de 1.120 Kilobytes/segundo, se generan las 732.588 formas conjugadas derivadas de 11.830 infinitivos a una velocidad de 536 formas por segundo y se identifican 266 formas por segundo. Los infinitivos y gerundios –típicamente incorporan pronombres enclíticos– pronominalizados en sus formas más frecuentes, junto con los imperativos en sus formaciones reflexivas se generan a razón de 310 formas por segundo y se identifican 180 formas por segundo. Todos los casos, tanto de generación como de reconocimiento, se han aleatorizado para evitar los efectos, favorables o desfavorables, de seguir un orden.

## **6.– Conclusiones.**

Se ha logrado una herramienta que permite identificar el tiempo, la persona, el número y los pronombres enclíticos que pueda poseer una forma verbal conjugada cualquiera y que además es capaz de generar las formas verbales a partir de un infinitivo, una conjugación y una combinación de pronombres enclíticos. Próximamente se incluirán, tanto para su identificación como para su generación, las flexiones correspondientes a la sustantivación, adjetivación y adverbialización de formas verbales. Es factible también su extensión a la flexión de palabras en general que permitiría análisis de textos de una gran profundidad y nivel de detalle. Resulta de interés su inclusión en la realización de búsquedas complejas en las que se desee actuar con independencia de la flexión que afecte a la forma considerada.

## Agradecimientos.

Queremos agradecer al profesor Dr. Manuel Alvar Ezquerra del Departamento de Filología Española I de la Universidad de Málaga su colaboración en cuantas consultas le hemos formulado a lo largo del desarrollo del presente trabajo.

## Referencias:

- [Als90] Alsina, R.: "Todos los Verbos Castellanos Conjugados". 17ª Edición. Teide. 1990.
- [Alv93] Alvar Ezquerra, M.: "La formación de palabras en español". Ed.: Arco/Libros. Madrid. 1993.
- [Dic90] "Diccionario Actual de la Lengua Española. VOX". Ed.: Biblograf, S.A.. Barcelona. 1990.
- [Gil85] Gili Gaya, S.: "Curso superior de sintaxis española" Vox. Biblograf. Barcelona, 1985.
- [Góm91] Gómez Torrego, L.: "Manual de Español Correcto". Arco/Libros, S. A. Madrid, 1991.
- [Góm92] Gómez Torrego, L.: El buen uso de las palabras. Arco/Libros, S. A. Madrid, 1992.
- [Lyo75] Lyons, J.: "Nuevos horizontes de la lingüística". Alianza Editorial. Madrid, 1975.
- [Mar78] Martinet, A.: "Elementos de lingüística general". Gredos. Madrid, 1978.
- [Rea89] Real Academia Española: "Esbozo de una nueva gramática de la lengua española". Espasa Calpe. Madrid, 1989.
- [Rod93] Rodríguez, A.; Hernández, Z.; Santana, O.: **Agrupaciones de Tiempos Verbales en un Texto**. Anales de las II Jornadas de Ingeniería de Sistemas Informáticos y de Computación, Quito (Ecuador). 132/137. Abril, 1993.



- [**San93a**] Santana, O.; Hernández, Z. J.; Rodríguez, G.: **Conjugaciones Verbales**. Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). Boletín N° 13: 443-450. Febrero, 1993.
- [**San93b**] Santana, O.; Rodríguez, J.C.; González, J.D.: **FRECTEXT: Una Aplicación de Ayuda a la Elaboración de Documentos**. Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). Boletín N° 13: 451-462. Febrero, 1993.
- [**Sec91**] Seco, M.: "Diccionario de dudas y dificultades de la lengua española". 9ª Edición. Espasa Calpe. Madrid, 1991.