

FREECONJ-TIP: UN CONJUGADOR VERBAL PARA EL ESPAÑOL BAJO LICENCIA GNU

FRANCISCO JAVIER CARRERAS-RIUDAVETS

DAVID MARTÍN-GONZÁLEZ

GUSTAVO RODRÍGUEZ-RODRÍGUEZ

CARLOS AARÓN VILLANUEVA-ROCA

Universidad de Las Palmas de Gran Canaria

RESUMEN

El programa FreeConj-TIP se ha desarrollado en el lenguaje de programación C++, bajo licencia GNU General Public, y automatiza la conjugación de verbos del español. FreeConj-TIP resuelve la conjugación aplicando reglas elementales de la ortografía española y de conjugación. La conjugación de los verbos altamente irregulares se resuelve en base a datos. Este diseño, que hace uso por un lado de reglas poco complejas y por otro de datos, ha permitido que el código de FreeConj-TIP sea sencillo y fácil de incorporar a otras aplicaciones de procesamiento del lenguaje natural. Igualmente, al trabajar con ficheros de texto plano, para almacenar los datos, se facilita la incorporación de nuevos infinitivos y sus formas conjugadas.

PALABRAS CLAVE: Conjugador, verbos, morfología, lingüística computacional

ABSTRACT

The FreeConj-TIP software has been developed in the programming language C++, under the GNU General Public. This program automates the Spanish verb conjugation. FreeConj-TIP solves the conjugation applying basic rules of conjugation and Spanish orthography. The conjugation of the highly irregular verbs is resolved based on data. This design, which uses on the one hand some simples rules and on the other data, has allowed FreeConj-TIP code to be simple and easy to incorporate into other applications of natural language processing. Furthermore, since FreeConj-TIP works with plain text as a format for saving their data, it is easy to add new infinitives and their conjugate forms.

KEYWORDS: Conjugator, verbs, morphology, computational linguistic

1. INTRODUCCIÓN

En la actualidad las tecnologías de la información y comunicación juegan un papel cada vez más importante en el mundo de las comunicaciones y en los avances científicos. Sin embargo, las reservas siguen siendo muchas a la hora de publicar en Internet ciertos trabajos o datos. Aunque la mentalidad de los investigadores está evolucionando en este aspecto, existe un vacío de ciertos recursos dentro del área de la lingüística computacional.

Existen herramientas y recursos lingüísticos en Internet destinados a usuarios finales, donde no es necesario poseer conocimiento sobre ingeniería informática para usarlos, pero aún no hay suficientes recursos gratuitos de ingeniería de la lengua para desarrolladores de programas.

En el caso concreto de los verbos, son varios los conjugadores disponibles en Internet (Provencio, 2006), sin embargo, no es posible acceder al código fuente de dichos conjugadores para su incorporación en otras aplicaciones de procesamiento del lenguaje natural. Esto permitiría el desarrollo de nuevas soluciones de forma más rápida, ya que parte del trabajo relacionado con la conjugación estaría resuelto.

FreeConj-TIP pretende ser un primer paso en la realización de un conjunto de programas de apoyo en el avance del tratamiento lingüístico para el español, bajo licencia GNU General Public, y dirigidos a desarrolladores de software, que puedan incorporar estas aplicaciones en otros proyectos relacionados con el procesamiento del lenguaje natural.

2. ESTADO DEL ARTE

FreeConj-TIP es el resultado del estudio en profundidad de las reglas de conjugación de los verbos españoles y sus irregularidades basado en la información de la Real Academia Española (2009). El grupo Text & Information Processing TIP de la Universidad de Las Palmas de Gran Canaria ha hecho una clasificación exhaustiva de las irregularidades en la conjugación de los verbos españoles que ha dado

lugar al libro “La conjugación de verbos en español y su morfología” (Carreras Riudavets y otros, 2010). En este estudio se contabilizan, además de los tres modelos regulares, 42 modelos irregulares y 15 modelos con defectividad de la primera conjugación, 35 modelos irregulares y 16 modelos defectivos de la segunda conjugación, 45 modelos irregulares y 11 modelos defectivos de la tercera conjugación. Se han clasificado 22 reglas de irregularidad morfológica y 12 reglas de cambio ortográfico en la conjugación. Se describen 14 reglas de irregularidad en los participios. Se tiene en cuenta la conjugación propia de zonas geográficas, como Río de la Plata, que usan el pronombre *vos* y de otras zonas que usan *ustedes*, en lugar de *vosotros*, como en Canarias.

3. CONJUGADOR FREECONJ-TIP

El objetivo principal del proyecto ha sido construir un conjugador verbal con una implementación sencilla, pero sin que por ello existan verbos que no puedan ser conjugados de acuerdo a todos los modelos de conjugación que les resulten aplicables. Para alcanzar este objetivo se ha utilizado como referencia la información sobre modelos de conjugación contenida en el conjugador online del grupo de investigación “Text & Information Processing - TIP” de la Universidad de Las Palmas de Gran Canaria y el libro *La conjugación de verbos español y su morfología* (Carreras-Riudavets y otros, 2010).

El desarrollo de FreeConj-TIP resulta relativamente inusual, debido a la necesidad de ofrecer un código sencillo a la par de ser capaz de conjugar todos los verbos del español. Esto ha propiciado que se utilice un enfoque mixto en la construcción del programa. En lugar de conjugar los verbos exclusivamente por reglas escritas en el código del programa o de utilizar solo información contenida en archivos y medios externos, FreeConj-TIP utiliza reglas para los modelos de conjugación más simples y frecuentes del español, mientras que para aquellos verbos más complejos de conjugar, así como para los menos frecuentes, se utilizan datos contenidos en ficheros de texto plano adicionales.

Esta forma de operar del FreeConj-TIP permite que el conjugador sea fácilmente actualizable y que cualquier usuario, incluso sin conocimientos de programación, sea capaz de especificar que un verbo pueda conjugarse de acuerdo con un cierto modelo de conjugación ya incluido en la documentación del programa o incluso proponer un modelo de conjugación completamente nuevo.

3.1. Características

Una característica propia del español es que un verbo puede tener varios modelos de conjugación simultáneamente, ya que puede conjugarse de distintas maneras dependiendo de su significado. FreeConj-TIP tiene en cuenta este aspecto y permite obtener todas las conjugaciones posibles para cada verbo.

Si se desea añadir a FreeConj-TIP nuevos verbos para que se conjuguen según ciertos modelos de conjugación, sólo debe editarse los ficheros de texto plano que incluyen la información de los modelos.

En el fichero *models.txt* se encuentran todos los modelos, agrupados según su modelo de conjugación. En el fichero *cod_flex.txt* se recogen los códigos asignados a cada una de las flexiones, empezando por el infinitivo con el código 1 y terminando con el código 98 que le corresponde a la segunda persona del singular del imperativo (vos). La conjugación de los tiempos compuestos y las voces pasivas no ha sido implementada en FreeConj-TIP. Sin embargo, obtener estas formas es trivial, ya que solo debe combinarse las formas del verbo haber con el participio del verbo que se desea conjugar y pueden estar disponibles en una próxima versión.

3.2. Reglas implementadas

Las reglas que se han implementado en FreeConj-TIP se clasifican en cuatro tipos: reglas ortográficas, morfológicas, acentuación y defectividad.

Se considera una variación ortográfica cuando es necesario modificar las últimas letras del lexema para adecuarse a las reglas ortográficas del español y mantener el sonido original del infinitivo.

Las variaciones morfológicas son las que se producen cuando en algunas formas se realizan cambios en el lexema del verbo. Las variaciones de acentuación son aquellas en las que se añade una tilde a la vocal débil del lexema cuando la acentuación recae sobre la raíz, como en *rehúso*, del verbo *rehusar*. Por último, FreeConj-TIP tiene en cuenta que algunos verbos admiten en ciertos casos una conjugación defectiva y elimina, por tanto, las formas correspondientes.

Los distintos tipos de variaciones no son excluyentes. Un verbo puede necesitar los cuatro tipos de variaciones descritos para ser conjugado correctamente. FreeConj-TIP tiene en cuenta las distintas variaciones que se deben aplicar para cada modelo de conjugación y las emplea para todos los verbos que sigan un cierto modelo.

4. IMPLEMENTACIÓN

FreeConj-TIP se suministra como una clase del lenguaje de programación C++, y ofrece la conjugación completa de un verbo de acuerdo a todos los modelos de conjugación que se le puedan aplicar. El programador sólo debe crear un objeto de esta clase y pasar como argumento una *string* que contenga el infinitivo del verbo que se desea conjugar. Las funciones suministradas permiten obtener una forma en particular del verbo o, si se desea, un puntero a una matriz de tres dimensiones, donde el primer índice indica el modelo de conjugación aplicado; el segundo, el tiempo, y el tercero, la persona (o, en el caso de las formas no personales, es decir, infinitivo, participio y gerundio, cuál de estas formas se desea obtener).

FreeConj-TIP verificará que la entrada es un infinitivo correcto y comprobará si existen modelos de conjugación para el verbo especificado en el fichero *models.txt*. Siempre que un verbo no aparezca en este fichero, será conjugado como si fuera un verbo completamente regular. Una vez que FreeConj-TIP ha encontrado los modelos de conjugación que el verbo puede utilizar, aplicará las desinencias regulares si el modelo actual las utiliza, o bien, si estamos ante un verbo completamente irregular, leerá las desinencias del

fichero correspondiente que contiene la conjugación completa de ese verbo.

Una vez que FreeConj-TIP ha leído las desinencias, aplica las reglas que sean necesarias para conjugar el verbo, prestando atención a las variaciones ortográficas, morfológicas y de acentuación que han sido indicadas anteriormente.

Para añadir un nuevo modelo que sea completamente irregular y que no esté ya implementado en el programa, basta con modificar el fichero *models.txt* y, a continuación, crear un fichero *.txt* cuyo nombre será el del modelo añadido y que se colocará en la carpeta llamada *irregular*. Para el verbo *hacer*, se debe crear el fichero *hacer.txt* dentro de la carpeta *irregular*.

Cada fichero de modelo completamente irregular especifica una desinencia por línea, excepto la primera línea, que indicará el número de caracteres que se debe eliminar del infinitivo para obtener la parte del lexema que permanece invariable en todas las formas. Para el verbo *hacer*, este número debería ser 4, ya que la parte invariable del lexema de todas las formas de cada verbo de este modelo se consigue eliminando los cuatro últimos caracteres del infinitivo *-acer*; así para el verbo *hacer* sería *h*, y para *satisfacer*, *satisf*. A continuación, debe especificarse la desinencia del participio de este verbo y, en la tercera línea, la del gerundio.

De la línea 4 en adelante, debe indicarse las desinencias de todas las formas personales (yo, tú, vos, usted, él, nosotros, vosotros, ustedes y ellos) de todos los tiempos en el orden descrito en la documentación.

Si una forma no se utiliza en un cierto modelo de conjugación por ser éste defectivo o por tratarse de las formas no utilizadas del imperativo, en su lugar se pondrá el símbolo almohadilla.

Por lo que respecta a los verbos cuyos modelos están programados dentro de FreeConj-TIP y se conjugan mediante reglas, también es posible agregar nuevos modelos, aunque esto requeriría editar el código del programa para establecer las normas que se deben aplicar a los verbos que se conjuguen según el nuevo modelo.

Cabe señalar la posibilidad de que un verbo, además de la conjugación completamente regular, admita alguna que no lo sea, porque tenga una variación morfológica o de acentuación o sea

defectivo según su significado. En este caso, será necesario especificar que el verbo admite el modelo de conjugación regular y que también admite su conjugación siguiendo un modelo concreto.

Para los verbos cuya única irregularidad se encuentra en el participio, hay una línea con un formato especial, que comienza por #participios. En ella, y detrás de cada infinitivo, se pondrán los participios que acepta, separados por guiones. Por ejemplo, sabiendo que el verbo *imprimir* acepta los participios *imprimido* e *impreso* y *abrir*, solamente *abierto*, sería:

```
#participios:imprimir-imprimido-impreso,abrir-abierto
```

5. PRESTACIONES

Para comprobar la fiabilidad de las reglas implementadas en FreeConj-TIP, se han conjugado unos 14613 verbos y se han contrastado las formas resultantes de cada conjugación con las que ofrece el Conjugador TIP (Rodríguez, 2009), un programa que trabaja exclusivamente en base a datos sin aplicar ninguna regla y cuyas conjugaciones han sido contrastadas con la Real Academia Española. La prueba se ha llevado a cabo sin incluir ninguna conjugación en base a datos en FreeConj-TIP.

De los 14613 verbos conjugados, 10564 tiene una conjugación regular y no son defectivos. Los 4049 restantes poseen algún tipo de irregularidad o defectividad en su conjugación o se les debe aplicar alguna variación ortográfica. Todos los verbos que se ajustan a las reglas implementadas en el programa han sido conjugados correctamente (3673 verbos). Los 376 verbos restantes poseen una alta irregularidad o alguna irregularidad poco frecuente que no ha sido implementada en FreeConj-TIP. Estos 376 verbos pertenecen a 43 modelos distintos, por lo que sólo es necesario incluir 43 verbos (1,17%) en la base de datos para que FreeConj-TIP funcione correctamente para el 100% de los verbos testeados. Las reglas implementadas en FreeConj-TIP cubren el 89,76% de los verbos irregulares (incluyendo los que tienen alguna variación ortográfica).

6. CONCLUSIONES

FreeConj-TIP es programa desarrollado en el lenguaje de programación C++, bajo licencia GNU General Public, que realiza la conjugación de verbos del español. Si bien es verdad que existen muchos conjugadores en Internet para usuarios finalistas, no se ha puesto a disposición de la comunidad científica ni de desarrolladores de recursos lingüísticos, un código de programación abierto que realice esa tarea. FreeConj-TIP resuelve la conjugación aplicando reglas elementales de la ortografía española y de conjugación para verbos con una cierta irregularidad en un 97,42% de un total de 14613 verbos, y el 2,57% restante, correspondiente a verbos muy irregulares, se resuelve en base a datos. Este diseño, que hace uso por un lado de reglas poco complejas y por otro de datos, ha permitido que el código de FreeConj-TIP sea reducido, comprensible, sencillo de modificar y fácil de incorporar a otras aplicaciones de procesamiento del lenguaje natural.

Por otro lado, la sencillez no debe estar sólo del lado programático, sino también de su potencialidad y configuración, por ello, el almacén de la información necesaria para su funcionamiento se ha dispuesto en ficheros de texto plano con un formato simple y comprensible. Estas características hacen de FreeConj-TIP un programa fácil de incorporar a otras aplicaciones, fácil de configurar y admite la posibilidad de incorporar tantos infinitivos como sean necesarios para una explotación óptima.

REFERENCIAS BIBLIOGRÁFICAS

- Croft W., Cruse D. A. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Evans V., Green M. 2006. *Cognitive Linguistics. An Introduction*. Edinburgh: Edinburgh University Press.
- Eynon T. 2002. "Cognitive linguistics", *Advances in Psychiatric Treatment*, 8: 399-407.

- Provencio, H. 2006. “Materiales didácticos y Aplicaciones de Tecnología Lingüística en red para la enseñanza / aprendizaje del verbo”, *Revista Electrónica de Didáctica ELE*, N° 6, [en línea]. Disponible en <http://www.educacion.gob.es/redele/revista6/HerminiaProvencio.pdf>
- Carreras-Riudavets, F. J., Hernández-Figueroa, Z., Rodríguez-Rodríguez, G. 2010. *La conjugación de verbos español y su morfología*. Editorial LULU.
- Real Academia Española. 2009. *Nueva gramática de la lengua española*. Editorial Espasa-Calpe
- Rodríguez-Rodríguez, G., Carreras-Riudavets, F., Hernández-Figueroa, Z. 2009. *Conjugador de verbos español TIP*. Disponible en <http://tip.dis.ulpgc.es>