

Onoma: un conjugador de verbos y neologismos verbales*

Onoma: a conjugator tool for verbs and verb neologisms

Luz Rello y Eduardo Basterrechea

Molino de Ideas s.a.

Nanclares de Oca, 1F, P F1 28022 Madrid

lrello, ebaste@molinodeideas.es

Resumen: Este artículo presenta el funcionamiento de *Onoma*, un conjugador en línea que es capaz de conjugar y analizar tanto los verbos existentes como los verbos nuevos en español. El sistema, basado en diferentes módulos y en transductores de estados finitos, se ha evaluado con otros conjugadores en red, además de con una base de datos que contiene 15.367 verbos, de los cuales 4.307 constituyen neologismos.

Palabras clave: conjugador, neologismo verbal, transductores de estados finitos.

Abstract: In this paper we introduce *Onoma*, a new conjugating tool which is available online. *Onoma* generates and analyses both existing verbs and Spanish verb neologisms. This application of finite state transducers was evaluated against other online conjugators and also on a database consisting of 15.367 verbs containing 4,307 neologisms.

Keywords: conjugator, verb neologism, finite state transducers.

1. Introducción

Onoma es un conjugador de acceso libre en Internet,¹ basado en diferentes módulos de estados finitos, que permite conjugar tanto los verbos existentes como los verbos nuevos en español.

La novedad del sistema reside en la posibilidad de conjugar posibles verbos pero que no están registrados en los diccionarios y en la capacidad de determinar, por medio de una serie de transductores de estados finitos (TEF), si un verbo nuevo es regular o irregular. Esta información es relevante ya que, aunque la mayoría de verbos nuevos sean regulares, el 26 % de los verbos nuevos registrados en nuestra base de datos son irregulares (Véase la Sección 5.1.1).

El método utilizado identifica tanto los verbos nuevos como los ya existentes en español y los clasifica en una de las siete clases propuestas: una clase para los verbos regulares y seis clases para los verbos irregulares. Este procedimiento de clasificación se realiza mediante seis TEF. Una vez que la forma verbal está clasificada, en función de su clase se realizan una serie de operaciones mediante otros TEF para su análisis o para la generación de su paradigma verbal.

Una de las ventajas de este sistema es la

escasa cantidad de conocimiento que se precisa para llevar a cabo la clasificación de los neologismos, ya que las reglas están basadas en características formales del verbo y, en raros casos, es necesario conocer la palabra de la que deriva el verbo.

Otra ventaja que presenta este modelo es la naturaleza de las reglas utilizadas en los TEF que son susceptibles de ser aprendidas y son fáciles de recordar, lo que hace que el sistema pueda utilizarse como una herramienta de aprendizaje en el campo del español como lengua extranjera. Por un lado se reduce el número de modelos de paradigmas de conjugación que deben aprenderse ya que éstos suelen superar la centena ((Gomis y Segura, 1998), (Santana et al., 1998)) y, por otro, se cubre una de las lagunas de los métodos actuales de aprendizaje ((Puebla, 1995), (Mateo, 2008)), ya que no cuentan con una metodología que pueda predecir, en la mayoría de los casos, cuándo un verbo es irregular.

El resto del artículo se estructura de la siguiente manera: en la sección que sigue se realiza un estado de la cuestión, mientras que en la Sección 3 se expone el corpus utilizado. El apartado 4 está dedicado a explicar la clasificación de los verbos, tanto si se trata de neologismos o de verbos ya registrados en los diccionarios, y la sección 5 se centra en las operaciones necesarias para su análisis o para la generación de sus paradigmas. En la

* Agradecimientos

¹*Onoma* puede consultarse en:
<http://conjugador.onoma.es>

sección 6 realizamos una evaluación comparativa con siete conjugadores mientras que se ofrecen las conclusiones en el apartado 7.

2. Estado de la cuestión

Otras aproximaciones al procesamiento de la morfología verbal del español ofrecen una alta cobertura de los verbos españoles registrados en los diccionarios ((Puebla, 1995), (Mateo, 2008)) o en bases de datos propias más amplias (Santana et al., 1998) pero sin llegar a la predicción de todos los posibles neologismos verbales, mientras que *Onoma* cuenta con reglas especiales para verbos que no se han dado en español pero que, sin embargo, son posibles.

Aunque la documentación sobre el funcionamiento de los conjugadores accesibles en la web es escasa, algunos están basados en memoria² y otros en máquinas de estados finitos (Santana et al., 1997)³. Otros trabajos cuyo fin es el procesamiento de la morfología verbal de español y de la de otras lenguas también hacen uso de una aproximación por reglas ((Tzoukermann y Liberman, 1981), (Görz, 1988), (Gasser, 2009)). Existen otros estudios que recurren a estrategias de aprendizaje automático para inferir patrones de comportamiento de la morfología verbal ((Anick y Artemieff, 1992), (Albright y Hayes, 2002), (Creutz y Lagus, 2004), (Parkes, Malek, y Marcus, 2007)); sin embargo, hasta donde alcanza nuestro conocimiento, esta aproximación no se ha llevado a cabo para el español.

Se ha observado (Beesley y Karttunen, 2000) que mediante el uso de TEF se puede obtener como resultado una implementación robusta de sistemas generadores o/y analizadores morfológicos en procesos basados en la concatenación de unidades. El uso de transductores de estados finitos ofrece la posibilidad de generar paradigmas verbales además de realizar el proceso inverso: el análisis de la forma verbal (Kaplan y Kay, 1994). Esta estructura también posibilita la implementación ordenada de reglas (Karttunen, Kaplan,

y Zaenen, 1992). *Onoma* emplea diferentes módulos para extraer la información necesaria sobre la forma verbal y de transductores de estados finitos en cascada para (1) la determinación de la naturaleza del verbo (si es irregular o no, sea o no un neologismo) y (2) para el análisis o generación del paradigma verbal de una forma verbal determinada.

Además del tratamiento que se da a los neologismos, nuestro sistema difiere del resto de las aproximaciones que utilizan TEF en la arquitectura (Santana et al., 1997)⁴ y en el mismo diseño de los transductores, que no están basados, excepto uno, en reglas de concatenación (Tzoukermann y Liberman, 1981)⁵. Nuestro método genera un paradigma hipotético regular para cada verbo que se requiere conjugar y luego, en caso de tratarse de un verbo irregular, determina sus irregularidades y aplica las modificaciones correspondientes sobre el hipotético paradigma regular para proporcionar las formas correctas de la conjugación irregular correspondiente.

3. Creatividad en la morfología verbal del español

La creación de verbos nuevos en español es especialmente productiva debido a la riqueza de esquemas morfológicos de la lengua (Almela, 1999). Un neologismo verbal en español se puede deber o bien a procesos morfológicos aplicados sobre palabras existentes o bien a la incorporación de extranjerismos verbales (*digitalizar* de *to digitalize*.) Entre los procedimientos morfológicos se pueden distinguir la prefijación, la sufijación y la parasíntesis. Mediante la prefijación un prefijo se añade a un verbo ya existente en la lengua (*des-* en *desagrupar*), mientras que mediante la concatenación de sufijos se crean verbos nuevos a partir de palabras que pertenecen a otra categoría (*-izar* en *superficializar*). Finalmente, en la parasíntesis se añaden en combinación un sufijo y un prefijo a la vez (*en-* y *-ecer* en *enlanguidecer*). Aun siendo poco común, el procedimiento de la parasíntesis es frecuente en los verbos si se compara su aparición en el

²Conjugador del Diccionario de la Real Academia de la (RAE, 2001). Disponible en: <http://buscon.rae.es/draeI/>

³El conjugador desarrollado por el Grupo de Estructuras de Datos y Lingüística Computacional (GEDLC) de la universidad de Las Palmas de Gran Canaria. Disponible en: www.gedlc.ulpgc.es/investigacion/scogeme02/fluxver.htm

⁴Este lematizador/flexionador (FLAVER) está basado en un programa de segmentación; en tres listas que contienen prefijos, desinencias y pronombres; y en dos módulos, uno para las desinencias verbales y otro para extraer información externa del verbo en cuestión.

⁵Este método aplica una desinencia específica para 62 tipos de conjugación, dando como resultado 150 modelos de conjugación

resto de las categorías gramaticales (Serrano, 1999). En este estudio consideramos como la base de derivación el constituyente inmediato anterior que sufre uno de los procedimientos morfológicos ya mencionados.

Un verbo español se forma a partir de una raíz, de las desinencias de tiempo y modo y de las de persona y número⁶. El paradigma de un verbo en español, teniendo en cuenta sus 19 tiempos verbales (simples y compuestos), las siete personas gramaticales para cada tiempo, exceptuando el presente de imperativo e incluyendo las dos formas de infinitivo y gerundio y las variaciones de género y número del participio, cuenta con 140 formas verbales para cada verbo.

La dificultad del procesamiento de la morfología verbal española radica en cuatro puntos: (1) las desinencias de tiempo y modo y las desinencias de persona y número pueden estar realizadas en el mismo segmento morfológico; (2) la raíz puede sufrir variaciones; (3) la rica productividad debida a los procesos creativos en los que se añaden prefijos o/y sufijos; y (4) la irregularidad del verbo, que tiene como consecuencia que tanto la raíz como las desinencias se aparten del paradigma hipotético regular.

4. Base de datos

El corpus utilizado en este estudio consiste en una base de datos creada *ad hoc*. Consta de 15.367 verbos en infinitivo junto con sus paradigmas verbales y contiene, además de los verbos documentados en el diccionario de la Real Academia de la Lengua (11.060 verbs) (RAE, 2001), considerados como verbos registrados en el diccionario; incluye también los paradigmas verbales de 4.307 verbos, considerados como neologismos, que se han extraído de la Wikipedia española así como de una colección de tres millones de artículos de diferentes periódicos escritos en español⁷.

5. Arquitectura

El sistema está compuesto por dos módulos que utilizan máquinas de estados finitos. El primer módulo es un clasificador (TEF) que tiene la finalidad de reconocer la forma verbal, clasificarla en una de las siete clases y

⁶Se considera la vocal temática como parte de la raíz.

⁷Los periódicos con mayor representación en el corpus son: *El País*, *ABC*, *Marca*, *Público*, *El Universal*, *Clarín*, *El Mundo* y *El Norte de Castilla*

añadir información formal sobre cada forma verbal (Véase la Sección 5.1).

Esta información se utiliza en el segundo módulo, compuesto a su vez por varias máquinas de estados finitos, donde se crea el paradigma verbal en dos pasos: en primer lugar, se genera un paradigma regular hipotético y, en segundo lugar, en el caso de tratarse de un verbo irregular, se procede a la modificación del paradigma regular siguiendo un sistema de patrones y reglas (Véase la Sección 5.2).

5.1. Clasificación de verbos y de neologismos verbales

La clasificación se realiza en seis pasos, recurriendo a dos tipos de conocimiento: (1) información formal sobre la forma verbal y (2), en muy raros casos, conocimiento de la base de derivación. Adicionalmente diferentes procesos utilizan los siguientes módulos: (1) un módulo que extrae el patrón acentual de la forma verbal; (2) otro que divide la palabra en sílabas; (3) un tercer módulo, se determina, si es necesario, la base de derivación de la forma verbal; y (4) otro que coloca la tilde sobre las formas verbales que lo precisen.

El algoritmo se implementa con una estructura de selección que gestiona la activación de seis módulos. La operación de cada módulo es simple, excepto la del módulo 4, que se implementa como un transductor de cascada y cuyo output, constituye el input de los módulos siguientes (5 y 6). Cada módulo recibe como input la forma verbal que, si no resulta clasificada, pasa al módulo siguiente, y así sucesivamente.

Mediante la implementación de estos TEF se consigue que la forma verbal quede clasificada en alguna de las siguiente clases:

Clase 1: Verbos regulares.

Clase 2: Verbos irregulares afectados por los cambios de letra en la última consonante de la raíz (*sequé* de *secar*).

Clase 3: Verbos irregulares que sufren cambios de tilde (*crió* de *criar*).

Clase 4: Verbos irregulares con alternancias de vocal o con diptongación de la vocal de su raíz (*sirvo* de *servir* y *hielo* de *helar*).

Clase 5: Verbos irregulares cuya raíz termina en vocal y sufren reglas de irregularidad heterogéneas (*leyó* de *leer*, *oigo* de *oír*).

Clase 6: Verbos Magníficos⁸ que comparten reglas de irregularidad y patrones propios de su clase.

Clase 7: Verbos irreductibles: el verbo auxiliar (*haber*), los copulativos (*ser* y *estar*) y los verbos monosílabos (*ir*, *dar* o *ver*).

La implementación de este módulo asegura que la forma verbal quede clasificada y según la clase a la que pertenezca el verbo, su paradigma se verá afectado por una serie de reglas y patrones (Véase la Sección 5.2) a no ser que se trate de un verbo irreductible, en cuyo caso sus formas verbales se hallan almacenadas en la memoria.

5.1.1. Transductores de clasificación

El **TEF 1** comprueba si la forma verbal es uno de los seis verbos irreductibles, un verbo Magnífico o un verbo formado por cualquiera de los anteriores más un prefijo. Mientras que los verbos irreductibles se buscan en memoria, los verbos Magníficos o los prefijados se verán afectados por patrones y reglas de irregularidad propias (clase 6). En el caso de que la forma verbal esté compuesta por un prefijo más un verbo irreductible, se añade dicho prefijo a las formas verbales flexionadas del verbo. En este caso, se revisan las reglas de acentuación de las formas generadas.

El **TEF 2** comprueba si el verbo termina en *-quirir* (*adquirir*) o pertenece a la siguiente lista: *dormir*, *errar*, *morir*, *oler*, *erguir* o *desosar*; si es así, la forma es reconocida como irregular y se trata mediante las reglas de irregularidad de la clase 4.

El **TEF 3** identifica si el infinitivo termina en vocal, en cuyo caso comprueba la conjugación a la que pertenece la forma verbal. En caso de pertenecer a la primera conjugación (-ar) (*criar*), el verbo se clasifica como verbo irregular de la clase 3 –es decir, verbos que sufren cambios de tilde–, mientras que si la forma verbal pertenece a la segunda (-er) o tercera (-ir) conjugación (*leer*, *oír*), el verbo se adscribe a la clase 5.

Existen tres casos de posibles verbos con la raíz terminada en vocal que no se han dado en nuestra base de datos: se trata de los verbos terminados en *-ier*, *-uer* y *-iir*. No obstante, se han creado reglas hipotéticas de ge-

neración de los paradigmas para que el algoritmo dé cobertura a todos los verbos posibles aunque no necesariamente existentes.

El **TEF 4** reconoce si la forma verbal pertenece a la primera conjugación y la raíz termina en *-c*, *-z*, *-g* o *-gu* (*secar*, *trazar*, *delegar*) o si pertenece a la segunda o tercera conjugación y la raíz termina en *-c*, *-g*, *-gu*, *-qu*, *-ll* o *-ñ* (*conocer*, *corregir*, *seguir*). Las formas verbales que cumplan alguna de estas condiciones pertenecen a la clase 2 y constituyen el imput de los transductores 5 y 6. Asimismo, las formas verbales que no hayan sido interceptadas por ninguno de estos transductores se analizan mediante los transductores siguientes.

El **TEF 5** relaciona la vocal de la raíz del verbo con otras categorías gramaticales de etimología común. Las formas verbales de la primera conjugación cuya vocal en la raíz sea una *e* (*helar*) o una *o* (*contar*) y la palabra relacionada posea un diptongo (*hielo*, *cuento*), así como las formas verbales de la segunda y de la tercera conjugación que presenten una *e* (*servir*) en la raíz, pertenecen a la clase 4.

Por último, el **TEF 6** corrobora la existencia de un diptongo en la raíz (*européizar*) y clasifica tales casos en el grupo 3.

De este modo, si la forma verbal casa con algunas de las condiciones de los transductores, puede tratarse de un verbo irregular y se conjuga siguiendo unas reglas y unos patrones determinados; si, por el contrario, no coincide con ninguna condición, la forma verbal posee un paradigma regular.

A pesar de que generalmente se cree que los neologismos verbales son regulares, lo cierto es que los datos demuestran que un 26,8% de los verbos nuevos son irregulares (Rello y Basterrechea, 2010). En el Cuadro 1 se muestra el número de neologismos que cada transductor asigna a cada clase de irregularidad.

5.2. Generación de los paradigmas verbales

El módulo de generación del paradigma verbal se compone de dos pasos. En primer lugar, cuenta con un módulo que genera el paradigma verbal regular mediante la concatenación de unidades. Para los verbos regulares el procedimiento de generación finaliza aquí. En cambio, los verbos irregulares pasan a un segundo módulo, compuesto por diferen-

⁸Denominamos Magníficos a los siguientes verbos: *traer*, *valer*, *salir*, *tener*, *venir*, *poner*, *hacer*, *decir*, *poder*, *querer*, *saber*, *caber*, *andar* y los verbos terminados en *-ducir* (Basterrechea y Rello, 2010).

Tipo de neologismo	TEF	Número de neologismos
regular		3.154
irregular	TEF 1	27
irregular	TEF 2	9
irregular	TEF 3	39
irregular	TEF 4	945
irregular	TEF 5	87
irregular	TEF 6	46
Total		4.307

Cuadro 1: Clasificación de neologismos verbales

tes máquinas de estados finitos, para que se apliquen sobre las hipotéticas formas regulares generadas las alteraciones correspondientes a las reglas de irregularidad en función de la clase en la que se haya clasificado la forma verbal.

5.2.1. Patrones y reglas de irregularidad

Una *regla de irregularidad* es un cambio que se aplica sobre la hipotética forma regular conjugada, mientras que un *patrón de irregularidad* se corresponde con el conjunto de las personas gramaticales del paradigma verbal que sufren una regla de irregularidad determinada. Se emplea un total de 40 reglas de irregularidad, agrupadas en tres tipos; existen, además, siete patrones, a los que se suman los dos patrones adicionales que siguen los verbos Magníficos

Cada patrón, además de estar compuesto por un conjunto de personas gramaticales afectadas por una misma regla, se correlaciona con una característica formal de la forma conjugada, lo que permite predecir las personas gramaticales que son irregulares en el paradigma verbal de cualquier verbo irregular. A continuación se describen los patrones y las condiciones formales que las formas verbales deben satisfacer para pertenecer a un patrón determinado:

- (1) Patrón **To**: Verbos cuya sílaba tónica aparece en la raíz.
- (2) Patrón **Te**: Verbos cuya sílaba tónica aparece en la raíz y cuya desinencia comienza por la vocal *e*.
- (3) Patrón **Dei**: Verbos cuya desinencia empieza por la vocal *e* o *i*.

- (4) Patrón **Dao**: Verbos cuya desinencia empieza por la vocal *a* o *o*.
- (5) Patrón **Di**: Verbos cuya desinencia es tónica, pero empieza por *i* átona.
- (6) Patrón **Dti**: Verbos cuya desinencia empieza por *i* tónica.
- (7) Patrón **Dt-i**: Verbos cuya desinencia es tónica, pero comienza por cualquier vocal excepto por *i*.

En función del patrón que posea el verbo y de la clase a la que el verbo irregular pertenezca, se aplica una determinada regla de irregularidad. Por ejemplo, el Patrón **Dei** activa las reglas de irregularidad que afectan a la tercera persona de singular y a la primera y tercera persona de plural del presente de imperativo, a todas las personas del presente de subjuntivo y a la primera persona de singular del pretérito perfecto simple⁹. Así, el verbo *escenificar* se aplica una regla de irregularidad de sustitución de la letra *c* por *qu* en las personas afectadas por el patrón **Dei** como sucede en la forma *escenifique*, primera persona de singular del presente de subjuntivo.

Los verbos Magníficos poseen, como ya se ha señalado, dos patrones específicos:

- (8) Patrón **Fc**: en todas las personas gramaticales de los tiempos de futuro y condicional del modo indicativo.
- (9) Patrón **I4**: en todas las personas gramaticales del pretérito perfecto simple de indicativo y de los pretéritos perfectos y el futuro simple de subjuntivo.

Las reglas de irregularidad pueden presentar uno de los tres tipos de alteración siguientes:

- **Sustitución**: por ejemplo, *z* se sustituye por una *c* en el patrón **Dei**, para crear, en el verbo *trazar* la forma *trace*, primera persona de singular del presente de subjuntivo.
- **Adición**: por ejemplo, la letra *z* se añade entre la raíz y las desinencias en las formas afectadas por el patrón **Dao**, como sucede en la forma *conozco* del verbo *conocer*.

⁹Las personas gramaticales afectadas por cada patrón se detallan en (Basterrechea y Rello, 2010).

- **Elisión:** por ejemplo, la vocal *i* se elimina de las formas flexionadas que componen el patrón **Di**, como se observa en *taño*, primera persona de singular del presente de indicativo de *tañer*.

Las 40 reglas de irregularidad se dividen en cinco grupos, que corresponden a las clases 2-6 de verbos irregulares que ofrecen los TEF del primer módulo.

- **Cambios ortográficos de consonante en la raíz (Clase 2):** constituidos por 9 TEF que modifican la consonante final de la forma verbal flexionada con el fin de mantener su pronunciación. Estas reglas de irregularidad siguen los patrones **Dei**, **Dao** y **Di**. Por ejemplo, en el verbo *seguir* se realiza una sustitución de *gu* a *g* en las formas afectadas por el patrón **Dao**, como en *sigo*, primera persona de singular de presente de indicativo.
- **Cambios ortográficos de tilde (Clase 3):** constituidos por 2 TEF que se activan en el patrón **To** acentuando la vocal de la raíz verbal: *vacío* de *vaciar*.
- **Alternancias de vocal o diptongación de la vocal de la raíz (Clase 4):** constituidos por 8 TEF que operan sobre la vocal de la raíz, que puede diptongarse o sustituirse por otra vocal. Estas reglas de irregularidad se presentan en los patrones **To** y **Dti**. Por ejemplo, la sustitución vocálica en *servir* de *servir* en el patrón **Dti**.
- **Verbos irregulares cuya raíz termina en vocal (Clase 5):** compuestos por 8 TEF que aplican reglas de irregularidad heterogéneas afectando a aquellos verbos cuya raíz termina en vocal en los patrones **To**, **Te**, **Dei**, **Di** y **Dti**. Por ejemplo, en *oír* se añade la letra *y* después de la raíz en las formas flexionadas correspondientes al patrón **Te**.
- **Reglas específicas de los verbos Magníficos:** comprendidas por 13 TEF activados en los patrones **Fc**, **I4**, **Dao** y **To**. Por ejemplo, en el verbo magnífico *tener* se cambia la raíz (*tuve*) en el patrón **I4**; se modifican las formas flexionadas añadiendo la letra *g* después de la raíz en el patrón **Dao** (*tengo*); y, finalmente, se añade una *d* detrás de la raíz en las formas reconocidas en el patrón **Fc** (*tendré*).

Las reglas de irregularidad deben aplicarse en orden, ya que es común encontrar paradigmas verbales que combinan varias reglas de irregularidad y su aplicación ordenada garantiza la correcta generación del paradigma.

Por ejemplo, en el verbo *dormir* primero se realiza una sustitución de la vocal *o* por la vocal *u* en las formas flexionadas del patrón **Dti** (*durmáis*) y después se aplica la diptongación de la vocal de la raíz en el patrón **To** (*duermo*).

6. Evaluación

El sistema se ha evaluado y acomodado a nuestra base de datos, que contiene 15.367 verbos, utilizando un método de prueba y error.

Además, se ha llevado a cabo una evaluación comparativa del procesamiento de neologismo por parte de *Onoma* y de siete conjugadores¹⁰ accesibles en Internet.

La evaluación se realizó con neologismos –ausentes de la base de datos empleada en este trabajo– formados mediante (1) prefijación, (2) sufijación y (3) parasíntesis (Véase el Anexo A). De los siete conjugadores evaluados en comparación con nuestro sistema, solo tres conjugan algunos tipos de neologismos (conjugadores 2, 5 y 7).

Cuando se trata de neologismos formados a partir de un prefijo el conjugador 7 los reconoce como verbos, pero no los conjuga correctamente en el 50% de los casos. Por ejemplo, al conjugar el neologismo *entresalir* propone *entresalo* como primera persona de singular del presente del indicativo en lugar de *entresalgo*.

Una de las razones que explican el alto rendimiento de *Onoma* en la generación de

¹⁰Los conjugadores evaluados son:

1- Conjugador de la Real Academia de la Lengua:

<http://buscon.rae.es/drael/>

2- Conjugador Reverso:

<http://conjugador.reverso.net/conjugacion-espanol.html>

3- Conjugador de WordReference:

<http://www.wordreference.com/conj/ESverbs.asp>

4- Conjugador de la Universidad de Oviedo:

<http://www6.uniovi.es/dic/conjuga.html>

5- Conjugador del Grupo de Estructuras de Datos y Lingüística Computacional de la Universidad de Las Palmas de Gran Canaria:

<http://www.gedlc.ulpgc.es/investigacion/scogeme02/flexver.htm>

6- Conjugador SpanishDict:

<http://www.spanishdict.com/conjugate/>

7- Conjugador Verbix v.2.0:

<http://www.verbix.com/languages/spanish.shtml>

Sistema	neologismo con prefijo	neologismo con sujifo	neologismo parasintético
Conjugator 2	37,5 %	0 %	0 %
Conjugator 5	75,0 %	25,0 %	12,5 %
Conjugator 7	50,0 %	87,5 %	100,0 %
Onoma	100,0 %	87,5 %	100,0 %

Cuadro 2: Evaluación de la precisión en la conjugación de neologismos.

Sistema	Precisión análisis
Conjugator 2	37.5 %
Conjugator 5	100 %
Conjugator 3	87.5 %
Onoma	100 %

Cuadro 3: Evaluación de la precisión en el análisis de neologismos.

neologismos sufijales es el hecho de que nuestro sistema, en el momento en el que el usuario inventa un verbo, solicita información sobre el mismo que el usuario puede introducir. La información solicitada es o bien la palabra o bien el verbo del cual el neologismo deriva.

Sin embargo, nuestro sistema posee una limitación que deberá ser tratada en un futuro próximo. A la hora de analizar formas verbales, *Onoma* presenta no sólo el análisis morfológico de las formas flexionadas, sino también las formas correspondientes con los neologismos introducidos por los usuarios y, en ocasiones, estos neologismos no son verbos usados o con un significado interpretable. Por ejemplo, la forma *fui* que puede corresponder a la tercera persona de singular del pretérito perfecto simple del indicativo tanto del verbo *ir* como del verbo *ser*, es detectada por nuestro sistema también como una forma verbal posible del hipotético verbo *fuir*. En el caso de considerar estos neologismos como falsos positivos, la precisión *Onoma* para el análisis morfológico sería de un 87.5%. Para la evaluación del análisis morfológico de las formas verbales, se han comparado los mismos conjugadores, utilizando para el análisis formas verbales ambiguas (Véase el Anexo A). Entre los sistemas evaluados, únicamente los conjugadores 2, 3 y 5 realizan análisis morfológicos.

7. Conclusiones

Onoma es un conjugador de verbos del español que tiene, entre uno de sus principales rasgos, la posibilidad de procesar neologismos.

En este sentido, la evaluación realizada ha mostrado que *Onoma* puede compararse favorablemente con otros conjugadores en línea, sobre todo en lo que se refiere al reconocimiento y a la generación de paradigmas de neologismos irregulares.

El modelo lingüísticamente motivado que se ha desarrollado para la creación de la herramienta constituye una novedad y presenta, además, un interés pedagógico en lo que se refiere a la enseñanza de la conjugación del español.

Por otra parte, puesto que los transductores en los que se basa *Onoma* están implementados en un sistema de gestión de bases de datos, es posible modificarlos con facilidad y con independencia del resto del programa. Una de las futuras tareas consistirá en la integración del algoritmo de *Onoma* en un procesador morfológico para el español, que pueda también tratar adecuadamente el resto de las categorías léxicas que constituyen clases abiertas.

Bibliografía

- Albright, A. y B. Hayes. 2002. Modeling English Past Tense Intuitions with Minimal Generalization. *Proceedings of the 6th Workshop of the ACL Special Interest Group in Computational Phonology (SIGPHON)*, páginas 58–69.
- Almela, R. 1999. *Procedimientos de formación de palabras en español*. Ariel, Barcelona.
- Anick, P. y S. Artemieff. 1992. A high-level morphological description language exploiting inflectional paradigms. *Proceedings of COLING 1992*, páginas 67–73.
- Basterrechea, E. y L. Rello. 2010. *El verbo en español. Construye tu propio verbo*. Molino de Ideas, Madrid.
- Beesley, K.R. y L. Karttunen. 2000. Finite-State Non-Concatenative Morphotactics. *Proceedings of the 5th Workshop of the ACL Special Interest Group in Computational Phonology (SIGPHON)*, páginas 1–12.

- Creutz, M. y K. Lagus. 2004. Modeling English Past Tense Intuitions with Minimal Generalization. *Proceedings of the 7th Meeting of the ACL Special Interest Group in Computational Phonology: Current Themes in Computational Phonology and Morphology*, páginas 43–51.
- Gasser, M. 2009. Semitic Morphological Analysis and Generation Using Finite State Transducers with Feature Structures. *Proceedings of the 12th Conference of the European Chapter of the ACL*, páginas 309–317.
- Gomis, P. y L. Segura. 1998. *Vademécum del verbo español*. SGEL. Sociedad General Española de Librería, Madrid.
- Görz, G. 1988. A Finite State Approach to German Verb Morphology. *Proceedings of COLING 1988*, páginas 212–215.
- Kaplan, R.M. y M. Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics*, 20:331–378.
- Karttunen, L., R.M. Kaplan, y A. Zaenen. 1992. Two-level morphology with composition. *Proceedings of COLING 1992*, páginas 141–148.
- Mateo, F. 2008. *Bescherelle. Les verbes espagnols*. Hatier, Paris.
- Parkes, C.H., A.M. Malek, y M.P. Marcus. 2007. Towards Unsupervised Extraction of Verb Paradigms from Large Corpora. *In Proceedings of the 6th Workshop on Very Large Corpora*, páginas 110–117.
- Puebla, J. 1995. *Cómo conjugar todos los verbos del español*. Playor, Madrid.
- RAE. 2001. *Diccionario de la lengua española, 22 edición*. Espasa, Madrid.
- Rello, L. y E. Basterrechea. 2010. Automatic conjugation and identification of regular and irregular verb neologisms in Spanish. *Proceedings of the NAACL 2010, Workshop on Computational Approaches to Linguistic Creativity, CALC-10*.
- Santana, O., F.J. Carreras, Z.J. Hernández, J.R. Pérez, y G. Rodríguez. 1998. *Manual de la conjugación del español. 12 790 verbos conjugados*. Arco Libros, Madrid.
- Santana, O., J.R. Pérez, Z.J. Hernández, F.J. Carreras, y G. Rodríguez. 1997. FLAVER: Flexionador y lematizador automático de formas verbales. *Lingüística Española Actual*, 19(2):229–282.
- Serrano, D. 1999. La derivación verbal y la parasíntesis. *Gramática descriptiva de la lengua española*, 2:4683–4756.
- Tzoukermann, E. y M.Y. Liberman. 1981. A Finite-State Morphological Processor for Spanish. *Proceedings of the 13th Conference on Computational Linguistics*, páginas 277–282.

A. Anexo 1: Verbos evaluados

Neologismos formados a partir de prefijación: *autotraer, autodestruir, cocopilotar, cohacer, entreleer, entresalir, pseudoandar y readquirir*.

Neologismos formados a partir de sufijación: *caperuzar, googlear, insomniar, monitorear, submitir, triquiñolear, verborrear y verdular*.

Neologismos formados a partir de parasíntesis: *enazulear, atorrijar, empirifollar, endormecer, enmariposar, enmarronar, ennegrir y sonrojecer*.

Formas flexionadas ambiguas: *amase* de los verbos *amasar* y *amar*; *ase* de los verbos *asir* y *asar*; *di* de los verbos *decir* y *dar*; *era* de los verbos *ser* y *erar*; *fui* de los verbos *ser* e *ir*; *paren* de los verbos *parar* y *parir*; *sé* de los verbos *ser* y *saber*; y *viste* de los verbos *vestir* y *ver*.