

# Una demostración de Onoma, el conjugador en línea de verbos y neologismos verbales en español

*A system demonstration of Onoma, the online conjugator for Spanish verbs and verb neologisms.*

**Eduardo Basterrechea**

Molino de Ideas s.a.,  
Madrid  
ebaste@molinodeideas.es

**Luz Rello**

NLP & Web Research Groups,  
Universitat Pompeu Fabra,  
luz.rello@upf.edu

**Rodrigo Alarcon**

Molino de Ideas s.a.,  
Madrid  
ralarcon@molinodeideas.es

**Resumen:** Presentamos Onoma, un conjugador en línea para conjugar y analizar tanto los verbos existentes como los neologismos verbales en español. Su desarrollo es el resultado de un importante trabajo lingüístico sobre el sistema verbal español a través de una metodología innovadora. Tanto las evaluaciones obtenidas como su popularidad confirman su validez e interés.

**Palabras clave:** conjugador, neologismo verbal, herramienta en línea.

**Abstract:** We present Onoma, an online tool to conjugate and analyze Spanish verbs and verb neologisms. Its development results from of an important linguistic study in which the Spanish verbal system is approached with an innovative methodology. Both the evaluation and the noticeable success confirm its validity and interest.

**Keywords:** conjugator, verb neologism, online tool.

## 1. Introducción

Onoma es un conjugador de acceso libre en Internet que permite conjugar tanto verbos existentes, como verbos de creación reciente en español. La conjugación de los verbos que no están registrados en la base de conocimiento, se determina por medio de una cascada de transductores de estados finitos, determinando si un verbo nuevo es regular o irregular. El 26 % de los verbos nuevos registrados en nuestra base de datos son irregulares, por lo que esta capacidad para manejar cualquier tipo de verbo se convierte en uno de los puntos fuertes de Onoma. Otra ventaja de este sistema es la escasa cantidad de conocimiento que se requiere para entender la metodología en la cual se basa. Esta última es el resultado de un importante trabajo lingüístico sobre el sistema verbal español. La simplicidad de las reglas que componen esta metodología sobre las cuales se basan los transductores de estado finito del sistema, hace que ambos, tanto el sistema como la metodología, puedan utilizarse como una herramienta de aprendizaje en el campo del español como lengua extranjera.

A continuación daremos una breve descripción del funcionamiento interno del sistema (sección 2) y del sitio internet dedicado (sección 3). Finalmente, antes de concluir (sección 5), proporcionamos datos relevantes que confirman la validez y el interés de Onoma

ma y su metodología subyacente (sección 4).

## 2. Funcionamiento general del sistema

El sistema está compuesto por dos módulos que utilizan transductores de estados finitos. El hecho de basar todo el sistema en transductores permite su uso para el análisis a la vez que para la conjugación.

El primer módulo clasifica cada forma verbal en una de las siete clases y le añade la información necesaria para la generación de su paradigma. La clase 1 incluye los verbos regulares. La clase 2 incluye los irregulares afectados por cambios de letra en la última consonante de la raíz (e.g. *sequé* de *secar*). La clase 3 incluye los irregulares que sufren cambios de tilde (e.g. *crío* de *criar*). La clase 4 incluye los irregulares con alternancias de vocal o con diptongación de la vocal de su raíz (e.g. *sirvo* de *servir*). La clase 5 incluye los irregulares cuya raíz termina en vocal y sufren reglas de irregularidad heterogéneas (e.g. *leyo* de *leer*). La clase 6 incluye los más irregulares con reglas y patrones propios a esta clase (e.g. *decir*, *romper*). Finalmente, la clase 7 incluye *haber*, *ser*, *estar* y los de raíz no-silábica (*ir*, *dar* o *ver*).

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 lu-

gar, se genera un paradigma regular hipotético y, en el caso de tratarse de un verbo irregular, se procede a la modificación del paradigma hipotético. Aunque los patrones y reglas pueden pertenecer a varios conjuntos (salvo los propios a la clase 6), a cada clase de verbo se le asocia un conjunto propio. Una descripción más extendida se puede encontrar en Rello y Basterrechea (2010).

### 3. Sitio de internet

El sitio de internet<sup>1</sup> propone cuatro opciones.

- “Analiza” permite consultar el lema de formas verbales conocidas.
- “Conjuga” permite conjugar verbos conocidos del sistema.
- “Inventa” permite conjugar verbos desconocidos del sistema<sup>2</sup>.
- “Aprende” proporciona al usuario explicaciones sobre la metodología.

### 4. Datos de interés

#### 4.1. Evaluación

Ya que los conjugadores en línea cubren casi todos los casos conocidos, esta evaluación se realizó con neologismos ausentes en la base de datos y formados mediante prefijación, sufijación y parasíntesis (ver tablas 1 y 2). Esos neologismos, 4307 en total, fueron extraídos de un corpus *ad-hoc*, compuesto a partir de la Wikipedia española, así como de una colección diversa de tres millones de artículos periodísticos escritos en español. De los siete conjugadores en línea que hemos encontrado, tres conjugan algunos tipos de neologismos: el conjugador Reverso<sup>3</sup>, el conjugador del Grupo de Estructuras de Datos y Lingüística Computacional de la Universidad de Las Palmas de Gran Canaria<sup>4</sup> (Gedlc) y el conjugador Verbix v.2.0<sup>5</sup>.

#### 4.2. Datos generales

Aunque la mayoría de verbos nuevos sean regulares, el 26 % de los verbos nuevos registrados en nuestra base de datos después de la

<sup>1</sup><http://onoma.es/>, Mayo 2011

<sup>2</sup>Que luego pasan a ser conocidos.

<sup>3</sup><http://conjugador.reverso.net/conjugacionespanol.html>, Mayo 2010

<sup>4</sup><http://www.gedlc.ulpgc.es/investigacion/scogeme02/flexver.htm>, Mayo 2010

<sup>5</sup><http://www.verbix.com/languages/spanish.shtml>, Mayo 2010

Sistema	con prefijo	con sufijo	parasintético
Reverso	37,5 %	0 %	0 %
Gedlc	75 %	25 %	12,5 %
Verbix	50 %	87,5 %	100 %
Onoma	100 %	87,5 %	100 %

Cuadro 1: Precisión en la conjugación.

Sistema	Precisión
Reverso	37,5 %
Gedlc	100 %
Verbix	87,5 %
Onoma	100 %

Cuadro 2: Precisión en el análisis.

Clase	1	2	3	4	5	6	7
Num. neologismo	3.154	27	9	39	945	87	46

Cuadro 3: Clasificación de neologismos.

evaluación son irregulares (clase 2 a 7 en la tabla 3). La capacidad de un sistema de conjugación para manejarlos tiene por lo tanto una importancia notable. Onoma tiene registrado un total de 15.085 verbos de los cuales 4224 pertenecen a las clases irregulares.

Tanto el sitio de internet como el libro que formaliza la metodología (Basterrechea y Rello, 2010) han tenido un éxito notable: el año pasado, el sitio de internet tuvo 204 726 páginas visitadas mientras que el libro fue descargado 7211 veces.

### 5. Conclusión

Onoma es un conjugador de verbos del español que tiene la capacidad de procesar neologismos. En este sentido, su evaluación ha demostrado que compite ventajosamente con otros conjugadores en línea. La metodología que se ha desarrollado para la creación de la herramienta constituye una novedad y presenta, además, un interés pedagógico para la enseñanza del español. Estas características son probablemente la raíz de su popularidad.

#### 5.1. Estimación de duración

Se estima 5 minutos por interviniente.

### Bibliografía

Basterrechea, Eduardo y Luz Rello. 2010. *El verbo en español*. Molino de ideas S.A.

Rello, Luz y Eduardo Basterrechea. 2010. Onoma: un conjugador de verbos y neologismos verbales. En *Proceedings of the SEPLN 2010*, Valencia, Spain.