

# Learnary: Herramienta para el Aprendizaje de Idiomas mediante Datos Enlazados

## *Learnary: Language Learning Tool using Linked Data*

Salud María Jiménez-Zafra<sup>1</sup>, María Teresa Martín-Valdivia<sup>1</sup>,  
Carolina Cubillo-García<sup>2</sup>, Miguel Ángel García-Cumbreras<sup>1</sup>,  
Daniel García-Baena<sup>3</sup>

<sup>1</sup>Departamento de Informática, SINAI, CEATIC, Universidad de Jaén, Jaén, España  
{sjzafra, maite, magc}@ujaen.es

<sup>2</sup>Renfe, Madrid, España  
carolinacubillologarcia@gmail.com

<sup>3</sup>I.E.S. San Juan de la Cruz, Úbeda (Jaén), España  
daniel.gbaena@gmail.com

**Resumen:** *Learnary* es una herramienta web para facilitar el aprendizaje de idiomas, en concreto inglés y español, que hace uso de distintas fuentes de datos enlazados, como BabelNet, Wikipedia o WordNet. Por medio de los datos enlazados, este sistema proporciona más información de la que se puede obtener en un traductor sobre una palabra, como por ejemplo imágenes, descripciones, sinónimos, antónimos, etc., con el objetivo de ayudar a los usuarios en su comprensión.

**Palabras clave:** Datos enlazados, aprendizaje de idiomas, BabelNet, Wikipedia, WordNet, Babelfy, Yandex.

**Abstract:** *Learnary* is a web tool to facilitate language learning, specifically English and Spanish, that makes use of different linked data sources, such as BabelNet, Wikipedia or WordNet. By means of linked data, this system provides more information about a word than can be obtained in a translator, such as images, descriptions, synonyms, antonyms, etc., in order to help users understand it.

**Keywords:** Linked data, language learning, BabelNet, Wikipedia, WordNet, Babelfy, Yandex.

## 1 Introducción

El aprendizaje de idiomas es fundamental en la sociedad actual. Cada vez son más los escenarios en los que resulta imprescindible dominar diferentes lenguas. Las ventajas de conocer más de un idioma son muy diversas, desde incrementar la posibilidad de conseguir un trabajo hasta enriquecer la vida social y cultural. Según el Instituto Nacional de Evaluación y Calidad del Sistema Educativo, en España, más de cinco millones de alumnos estudian idiomas, el 82 % de los mismos, inglés. Además, científicamente se ha estudiado el impacto en nuestro cerebro cuando decidimos aprender idiomas distintos al materno y se ha demostrado que (Craik, Bialystok, y Freedman, 2010; Bialystok, Craik, y Luk, 2012; Yang et al., 2015):

- Mejora nuestra capacidad de concentración y nuestra memoria.
- Retrasa al menos cinco años la posible

aparición del alzhéimer.

- Aumenta nuestra capacidad de reacción a la hora de tomar decisiones.
- Mejora la agilidad mental.

¿Cómo pueden ayudar los datos enlazados en el aprendizaje de un idioma? En primer lugar, definiremos qué son los datos enlazados (Bizer, Heath, y Berners-Lee, 2011). Los datos enlazados son datos estructurados que se interrelacionan con otros datos para facilitar el acceso a la información, de manera que se pueda ir navegando entre datos relacionados de diferentes fuentes.

Los datos enlazados permiten que se muestre mucha más información de una palabra que la que se puede obtener en un diccionario o traductor, como por ejemplo acceder a más de una definición de una palabra, consultar la transcripción u obtener sinónimos y antónimos.

En este trabajo presentamos *Learnary*,

una demo cuyo objetivo es facilitar el aprendizaje de idiomas, en concreto español e inglés, haciendo uso de distintas fuentes de datos enlazados, como BabelNet (Navigli y Ponzetto, 2012), Wikipedia (Völkel et al., 2006) o WordNet (Miller, 1995).

## 2 Motivación

Aprender un idioma no es fácil y es una tarea cuya dificultad aumenta sobre todo con la edad. Esto se debe a que los lenguajes son sistemas extensos y complejos de simbolización. Sus fonemas, vocabulario y sus reglas gramaticales difieren de los de la lengua materna. Además, cuando se aprende un idioma y se lee un texto a menudo cuesta entender su vocabulario. En este punto los datos enlazados pueden ofrecer una gran ayuda, permitiendo que un usuario que no entiende una palabra o un conjunto de ellas, pueda obtener más información de las mismas (imágenes, descripciones, sinónimos, antónimos, traducciones, etc.).

## 3 Descripción del sistema

*Learnary* es una aplicación web capaz de procesar textos en inglés y en español, ofreciendo un análisis detallado de las palabras de estos textos para facilitar el aprendizaje de idiomas. Esta herramienta se ha construido haciendo uso de los lenguajes de programación Python, HTML, CSS, y JavaScript, y de los frameworks Flask, Bootstrap y JQuery. La plantilla de colores de la misma se ha seleccionado realizando un estudio de anomalías visuales (protanopia, deuteranopia, tritanopia, acromatopsia, protanomalia, deuteranomalia, tritanomalia y acromatomalia) para asegurar que cualquier persona pueda ver correctamente el contenido de la página.

### 3.1 Entrada del texto

En la Figura 1 se muestra la página principal de *Learnary*. Está disponible en dos idiomas, inglés y español (por defecto en español). En la parte derecha de la misma se describen las fuentes de datos enlazados y los recursos utilizados:

- BabelNet (Navigli y Ponzetto, 2012): red semántica multilingüe que integra conocimiento lexicográfico y enciclopédico de WordNet y Wikipedia. Está compuesta por unos 15 millones de Babel synsets. Cada Babel synset es una entrada de BabelNet que representa un significado y

contiene todos los sinónimos que expresan ese significado en diferentes idiomas. Cada Babel synset se identifica por un id.

- Babelfy (Moro, Raganato, y Navigli, 2014): algoritmo que vincula entidades y desambigua el sentido de las palabras seleccionando interpretaciones semánticas de alta coherencia. Se basa en la red semántica multilingüe de BabelNet.
- Wikipedia (Völkel et al., 2006): enciclopedia libre, políglota y editada de manera colaborativa.
- WordNet (Miller, 1995): base de datos léxica que agrupa palabras en conjuntos de sinónimos llamados synsets, proporcionando definiciones cortas y generales y almacenando la relaciones semánticas entre las palabras (sinónimos, hipónimos y merónimos).
- Yandex<sup>1</sup>: servicio gratuito multilingüe de traducción automática, desarrollado y proporcionado por Yandex, para traducir texto, voz, imágenes o vídeos en tiempo real de un idioma a otro.

En la parte izquierda, aparece una caja de texto donde el usuario puede escribir, o copiar y pegar, el texto que desea analizar.

### 3.2 Procesamiento inicial del texto

Una vez que el usuario ha introducido el texto y ha pulsado el botón “Analizar”, el sistema se encarga de procesarlo.

En primer lugar, se realiza una desambiguación de los términos haciendo uso de la API HTTP de Babelfy<sup>2</sup> para obtener el identificador asociado a cada término (*synset id*). Este procesamiento inicial permite identificar los términos sobre los cuales hay información en las fuentes de datos enlazados.

Tras procesar el texto, en la parte izquierda de la herramienta, aparecerá dicho texto con algunas palabras marcadas en color azul y con otras palabras en color negro. Las palabras marcadas en azul son las que se podrán consultar, es decir, aquellas para las cuales se ha obtenido información (Figura 2).

<sup>1</sup><https://translate.yandex.com/>

<sup>2</sup><http://babelfy.org/guide>

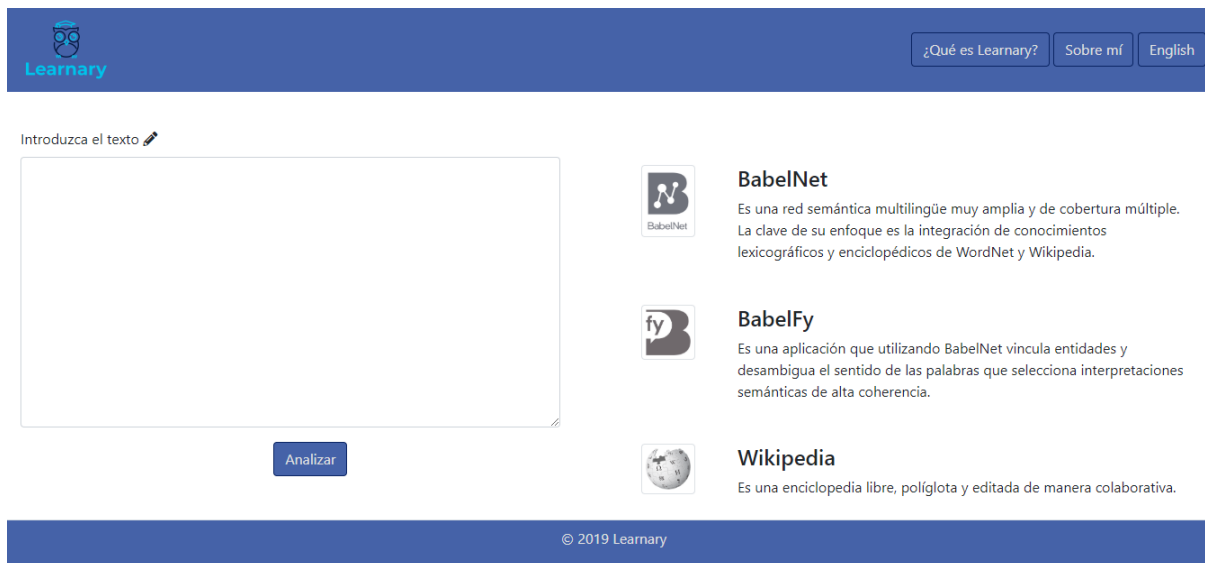


Figura 1: *Learnary* - Página principal.



Figura 2: *Learnary* - Procesamiento inicial.

### 3.3 Consulta de una palabra y recuperación de información

Para consultar las palabras sobre las que existe información (palabras azules), se deberá hacer *click* encima de ellas. Cuando el usuario pulse sobre una palabra azul, se recuperará toda la información disponible sobre ella en las distintas fuentes de datos (BabelNet, WordNet, Wikipedia y Yandex). A continuación, se especifica qué información se extrae de cada fuente y cómo se relacionan entre ellas (Figura 3). El lema de la palabra, la imagen, la categoría gramatical, la descripción y la transcripción se obtienen consultando la API HTTP de BabelNet<sup>3</sup> con el *synset*

*id* asociado a la palabra, obtenido en el procesamiento inicial con BabelFy. Además, el lema obtenido se utiliza para consultar la API MediaWiki<sup>4</sup> de Wikipedia, para ofrecer al usuario otras definiciones de la palabra diferentes a la obtenida con BabelNet. Los sinónimos y antónimos se extraen consultando la base de datos WordNet con el *synset id* de la palabra. Por último, la traducción se ofrece en español/inglés, francés, italiano y alemán por medio de la API del traductor Yandex<sup>5</sup>.

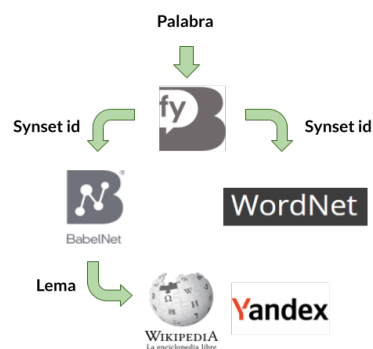


Figura 3: Información extraída y relación entre las fuentes de datos.

Finalizado este proceso, en la parte derecha de la pantalla, se mostrará toda la información disponible para la palabra seleccionada: nombre, categoría gramatical, descripción, imagen, sinónimos, antónimos, transcripción, traducción y otras definiciones. En la Figura 4 se puede ver el resultado de con-

<sup>3</sup><https://babelnet.org/guide>

<sup>4</sup>[https://www.mediawiki.org/wiki/API:Main\\_page/es](https://www.mediawiki.org/wiki/API:Main_page/es)

<sup>5</sup><https://yandex.com/dev/translate/>

Learnary

¿Qué es Learnary? Sobre mí English

Terminó agosto y llegó septiembre con sus tardes menos largas y las mañanas más frescas. Los días transcurrían lentos sobre el ajeteo de La Luneta. La gente entraba y salía de las tiendas, los cafés y los bazares, cruzaba la calle, se detenía en los escaparates y charlaba con conocidos en las esquinas. Mientras contemplaba desde mi atalaya el cambio de luz y aquel dinamismo imparables, era plenamente consciente de que yo también necesitaba cada vez con más urgencia ponerme en movimiento, iniciar una actividad productiva para dejar de vivir de la caridad de Candelaria y comenzar a juntar los duros destinados a solventar mi deuda. No daba, sin embargo, con la manera de hacerlo y, para compensar mi inactividad y mi nula contribución a la economía de la casa, me esforzaba al menos en colaborar en lo posible para aligerar las tareas domésticas y no ser sólo un bulto tan improductivo como un mueble arrumbado. Pelaba patatas, ponía la mesa y tendía la ropa en la azotea. Ayudaba a Jamila a pasar el polvo y a limpiar cristales, aprendía de ella algunas palabras en árabe y me dejaba osequiar por sus eternas sonrisas. Regaba las macetas, sacudía las alfombras y anticipaba pequeñas necesidades de las que antes o después alguien tendría que encargarse. En sintonía con los cambios de temperatura, la pensión se fue también preparando para la llegada del otoño y yo cooperé en ello. Cambiamos las camas de todos los cuartos; mudamos sábanas, retiramos las colchas de verano y bajamos los cobertores de invierno de los altillos. Me di cuenta entonces de que gran parte de aquella lencería necesitaba un repaso urgente, así que dispuse un gran cesto

**casa**  
Sustantivo

Descripción: Una casa es una edificación destinada para ser habitada.

Sinónimos

Antónimos

Pronunciación

Traducción

Otras definiciones

© 2019 Learnary

Figura 4: *Learnary* - Consulta de una palabra.

sultar la palabra “casa”. Para visualizar la información de un tipo concreto de información, ej. sinónimos, se deberá hacer clic en el elemento azul con este nombre y se desplegará toda la información relacionada.

#### 4 Conclusiones y trabajo futuro

En este trabajo se ha presentado *Learnary*, un prototipo para facilitar el aprendizaje de idiomas, en concreto inglés y español, mediante el uso de datos enlazados.

Como posible trabajo futuro nos planteamos algunos aspectos en los que podría mejorarse la aplicación:

- Ampliación a más idiomas.
- Tratamiento de la coherencia léxica para ofrecer definiciones más acordes al contexto.
- Inclusión de más información sobre la palabra buscada, por ejemplo, su pronunciación.
- Introducción de texto por voz.

#### Agradecimientos

Este trabajo ha sido parcialmente financiado por el Fondo Social Europeo, la Administración de la Junta de Andalucía (DOC\_01073), el Fondo Europeo de Desarrollo Regional (FEDER) y el proyecto LIVING-LANG (RTI2018-094653-B-C21) del Gobierno de España.

#### Bibliografía

Bialystok, E., F. I. Craik, y G. Luk. 2012. Bilingualism: consequences for mind and

brain. *Trends in cognitive sciences*, 16(4):240–250.

Bizer, C., T. Heath, y T. Berners-Lee. 2011. Linked data: The story so far. En *Semantic services, interoperability and web applications: emerging concepts*. IGI global, páginas 205–227.

Craik, F. I., E. Bialystok, y M. Freedman. 2010. Delaying the onset of alzheimer disease: Bilingualism as a form of cognitive reserve. *Neurology*, 75(19):1726–1729.

Miller, G. A. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Moro, A., A. Raganato, y R. Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics (TACL)*, 2:231–244.

Navigli, R. y S. P. Ponzetto. 2012. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial intelligence*, 193:217–250.

Völkel, M., M. Kröttsch, D. Vrandečić, H. Haller, y R. Studer. 2006. Semantic wikipedia. En *Proceedings of the 15th international conference on World Wide Web*, páginas 585–594.

Yang, J., K. M. Gates, P. Molenaar, y P. Li. 2015. Neural changes underlying successful second language word learning: An fmri study. *Journal of Neurolinguistics*, 33:29–49.