

Hacia la elaboración de un diccionario de onomatopeyas en español

Aroa Orrequia-Barea, Cristian Marín-Honor
Universidad de Cádiz

Onomatopoeias have always been a controversial matter in linguistics due to two main reasons: on the one hand, it has been frequently assimilated to the interjections word class; on the other, because of the translation-related problems they pose crosslinguistically. These two problems and the rather scarce literature on onomatopoeias have encouraged the Gabinete de Lingüística de Corpus of the ILA at the Universidad de Cádiz to elaborate an onomatopoeias dictionary of Spanish. The aim of this article is to deal with the steps that have been followed to create this dictionary as well as the difficulties and the future applications that it may have.

Keywords: onomatopoeia, dictionary, lexicography, corpus linguistics, parallel corpus

1. Introducción

El objetivo de este artículo es explicar la metodología seguida para la elaboración de un diccionario de onomatopeyas en español. Desde el Gabinete de Lingüística de Corpus del ILA de la Universidad de Cádiz se ha impulsado la creación de este diccionario por dos motivos fundamentales. Por un lado, existe escasa bibliografía sobre el tema de las onomatopeyas, hecho que da lugar a problemas en el ámbito de la traductología (Albrecht y Métrich 2016); por otro, las onomatopeyas plantean importantes problemas de análisis y categorización en los estudios lingüísticos. De manera general, las onomatopeyas se han incluido dentro de la categoría de las interjecciones (Bally 1965: 129 y Kleiber 2006: 10-11), pues una gran mayoría de estas últimas tienen un origen onomatopéyico. Blanco Escoda (2013: 18) considera que algunos casos de interjecciones, y como consecuencia de onomato-

peyas, son *pragmatemas*, esto es, “enunciados fijos situacionalmente condicionados” (Blanco Escoda 2014: 14). En este artículo, sin embargo, vamos a centrarnos en las onomatopeyas que hacen referencia a ruidos y sonidos no humanos. En este caso, según Barnas (2017: 14) citando a Blanco y Mejri (en prensa), no se pueden considerar pragmatemas, pues no están limitadas semánticamente por la situación de enunciación. No obstante, el objetivo de este artículo no es la definición de las onomatopeyas como unidad lingüística. Lo que pretendemos es que la creación de este diccionario arroje luz sobre las cuestiones traductológicas que las onomatopeyas plantean.

Para la elaboración del diccionario se han seguido tres fases: la primera ha sido la compilación de un corpus de onomatopeyas¹ de la lengua española mediante el estudio y análisis de cómics. La segunda fase, en la que nos encontramos actualmente, pasa por el uso de *Sketch Engine*, una página web que permite buscar las onomatopeyas obtenidas en corpus compilados de Internet. La última fase consiste en la creación de una interfaz web para albergar el diccionario.

2. Objeto de estudio

Generalmente, las onomatopeyas han estado asociadas al mundo animal, sin embargo, este estudio ha pretendido dar un paso más allá, y centrarse en aquellas que se refieren a los ruidos y sonidos no humanos, ya sean de origen natural (*plof*) o artificial (*tictac*). El motivo de esta elección viene derivado de la ausencia de estudios para estas últimas frente a las producidas por animales, las cuales ya han sido estudiadas en profundidad (Barbérís 1992).

En español solo existe un diccionario de onomatopeyas del cómic (Gasca y Gubern 2008) que se centra únicamente en los ruidos y sonidos que aparecen en ellos, siendo este el problema que encontramos, pues se restringe a este género literario. Aunque resulta obvio que los cómics son un discurso perfecto para ver el uso y variedad de las onomatopeyas, en este trabajo ampliaremos el corpus empleando otros géneros discursivos. Esta elección se hace por dos motivos fundamentales: en algunas ocasiones las onomatopeyas son inventadas por el autor y, de manera general, la onomatopeya necesita del apoyo visual para ser interpretada y entendida.

¹ Este corpus fue presentado como comunicación en el XXXVI Congreso Internacional de la Asociación Española de Lingüística Aplicada celebrado en Cádiz del 19 al 21 de abril de 2018.

3. Metodología

3.1 Compilación de muestras

Para la elaboración de nuestro corpus era necesaria la compilación de una considerable cantidad de onomatopeyas que fueran representativas de la lengua española. En un primer intento, se buscaron en Internet listas de onomatopeyas que ya estuvieran compiladas, sin embargo, pronto nos dimos cuenta de que la mayoría estaban copadas por los sonidos emitidos por los animales y que realmente no tenían base científica. Por este motivo, decidimos acudir a la literatura, concretamente a los cómics, pues son el género de las onomatopeyas por excelencia.

Nuestro corpus está compuesto por distintos números de 15 populares cómics escritos en español: *Zipi y Zape* de José Escobar, *Mortadelo y Filemón* y *Rompetechos* de Francisco Ibáñez, *El doctor Cataplasma* de Martz Schmidt, *Sir Tim O'Theo* de Raf, *Benito Boniato, estudiante de bachillerato* de Carlos Fresno Crespo y Luis Fresno Crespo, *Superlópez* de Juan López Fernández, *Agamenón* de Nené Estivill, *Los señores de Alcorcón y el holgazán de Pepón* y *Rigoberto Piciaporte, solterón de mucho porte* de Roberto Segura, *Don Berrinche* de José Peñarroya, *Las hermanas Gilda* de Manuel Vázquez Gallego, *El repórter Tribulete* de Guillermo Cifré, *El loco Carioco* de Carlos Conti y *Hug, el troglodita* de Jorge Gosset Rubio.

En el análisis de los cómics se han extraído manualmente alrededor de 300 onomatopeyas, excluyendo variaciones de estas, tales como *crash* y *crashh*.

Estos cómics podrían no considerarse una muestra de la lengua actual, pues fueron escritos entre los años 40 y 80. Sin embargo, creemos que constituyen una base de onomatopeyas bastante sólida, ya que el estudio de cómics más actuales, como *El invierno del dibujante* de Paco Roca y *Blacksad* de Juan Díaz Canales, ha revelado que actualmente el uso de las onomatopeyas ha disminuido bastante y que no varían con respecto a los mencionados arriba.

3.2 Extracción de concordancias

En numerosas ocasiones se han encontrado muestras que han sido inventadas por el autor de la obra, por ejemplo, *ñiaiiiiiiiiick scrioaiiiiiiiiiick*, usada por Ibáñez para hacer referencia al freno del coche. Por este motivo y con el fin de tener una ejemplificación más clara del uso de las onomatopeyas, hemos acudido a una herramienta de lingüística de corpus llamada *Sketch Engine*. Esta página web tiene diversas aplicaciones entre las que destacan la administración de corpus, así como el análisis de textos o la minería de datos.

El uso de *Sketch Engine* tiene varias ventajas para nuestro trabajo. Por un lado, la búsqueda en un corpus extraído de la web permite que nuestro corpus no esté restringido al género de los cómics, sino que también contenga ejemplificación de otros géneros. Esto nos permite dar unos resultados más concretos no solo en cuanto a contextos en los que pueden aparecer la onomatopeya, sino también a la hora de afinar con su significado. Por otro lado, el hecho de que *Sketch Engine* extraiga los datos de Internet, nos permite saber qué onomatopeyas se emplean en la lengua real, descartando así las que hayan podido quedar desfasadas en la actualidad.

Para la búsqueda de las concordancias, hemos utilizado el corpus del español Spanish Web 2011 (esTenTen11) que contiene unos nueve mil millones de palabras. Para facilitar la búsqueda y aglutinar los casos de variantes de onomatopeyas, en la opción “Concordance” hemos usado *CQL* (*Corpus Query Language*) que nos permite buscar usando expresiones regulares, esto es, una serie de caracteres que forman un patrón de búsqueda. Los resultados de la búsqueda se han descargado en formato XML.

Hemos obtenido resultados de 107 onomatopeyas, ya que muchas son inventadas por el autor. Esto ocurre con bastante frecuencia en el caso de los instrumentos musicales. De hecho, distintos autores utilizan distintas onomatopeyas para el mismo instrumento, como ocurre por ejemplo para la trompeta: *tuuut* o *tatatitataitaiiii*.

En una fase posterior es necesario un filtrado manual, pues en algunas concordancias encontramos onomatopeyas que se han lexicalizado, tales como *baby boom*, *el crack del 29* u otras que coinciden con nombres propios o apellidos, como *Blops* o *Tromp*. Actualmente nos encontramos trabajando en este filtrado. El corpus inicial constaba de un millón de palabras, sin embargo, después del filtrado el corpus contiene alrededor de 500 000 palabras.

3.3 Elaboración del diccionario

La idea es que el diccionario de onomatopeyas esté albergado en una plataforma online para facilitar su consulta al usuario. Antes de crear esta página web, es necesario establecer la estructura que va a tener la base de datos que lo conforma, donde se almacenan todas las onomatopeyas. Por este motivo, el primer paso para implementar la interfaz es la creación del *universo del discurso*, esto es, una descripción de la información que va a formar parte de la base de datos y un esquema sobre cómo estos datos se relacionan entre ellos.

En este caso, el universo que necesitamos consta de un número de onomatopeyas que se van a relacionar con diferentes concordancias según el ruido al que hace referencia.

El segundo paso necesario para la implementación de la interfaz es la organización de los datos en tablas con sus correspondientes campos. En este caso, utilizaremos tres tablas: una de onomatopeyas, otra de las concordancias, esto es, ejemplos de *Sketch Engine* donde aparezca la onomatopeya en su contexto, y una tercera de los ruidos y sonidos a los que hacen referencia dichas onomatopeyas.



Figura 1. Tablas que conforman la base de datos

Las tablas de la Figura 1 van a establecer una relación entre una onomatopeya y uno o más ruidos a los que representa. De igual modo, un ruido puede estar representado por una o más onomatopeyas. Esta relación onomatopeya-ruido va a tener asociadas unas concordancias. Por ejemplo, la onomatopeya *pum* puede representar el ruido de un disparo, como la concordancia del ejemplo (1), o el ruido de llamar a una puerta, como en el ejemplo (2):

- (1) Las perdices se posan en otro árbol, llega el cazador y: pum, pum, pum. Nada. Las perdices se vuelven a posar en otro árbol, llega el cazador y lo mismo: nada, ni una.
- (2) Tras un interrogatorio con la policía de Nueva York, que le acusaba de un atentado que causó varios muertos. ¡Pum, pum!, ¿quién es? ¡la policía!

La página web que alberga el diccionario está hecha mediante *Django*, un *framework* de desarrollo web escrito en el lenguaje de programación *Python* y constará de dos partes principales:

- a. Área de administración, accesible solo para sus creadores, donde se pueden crear y administrar los usuarios encargados de introducir los datos relativos a las onomatopeyas, sus concordancias y los ruidos a los que se refieren.

- b. Área de usuario que permite buscar las onomatopeyas que se encuentren introducidas en la web. La búsqueda tiene dos modalidades: por onomatopeyas o por ruidos y sonidos. En el primer caso, al introducir la onomatopeya, obtenemos como resultado los ruidos a los que se refieren dicha onomatopeya. La segunda opción permite buscar el sonido y el diccionario devuelve todas las onomatopeyas que se asocian a este.

La posibilidad de una doble búsqueda se hace necesaria dada la funcionalidad que pretendemos del diccionario, esto es, como herramienta de apoyo para la traducción. En este sentido, será muy útil para los traductores, pues podrán buscar un sonido o ruido y así seleccionar de una lista, cada uno con un contexto que la ejemplifique claramente, la onomatopeya que más se adecue a sus necesidades.

4. Conclusiones

En primer lugar, creemos que la creación de un diccionario de onomatopeyas era necesario, ya que los existentes carecen de una base científica tan específica. De esta forma, nuestro trabajo ofrece una nueva perspectiva lexicográfica que amplía las apariciones en el DRAE y en los diccionarios citados en el cuerpo del artículo. Nuestro diccionario no pretende solo dar el significado, sino también mostrar ejemplos o usos reales de cada una de las onomatopeyas, así como sus variantes. Todo esto es posible gracias a una base de onomatopeyas bastante amplia recopilada de los cómics más populares de la lengua española. El hecho de que estos cómics fueran escritos en otra época no constituye un problema, pues se han comprobado cómics más actuales y las onomatopeyas, en su mayoría, se mantienen invariables. Aun así, el uso de *Sketch Engine*, una herramienta que extrae corpus de la web, nos ha permitido descartar aquellas que ya han quedado en desuso.

En segundo lugar, constatamos que la creación de este tipo de diccionario se constituye como una herramienta muy útil para traductores y editores. La herramienta *Sketch Engine* nos ha permitido la recopilación de un corpus de 500 000 palabras que constituye la base de nuestro diccionario. Con este corpus hemos podido analizar y estudiar los significados y usos reales y actuales de las onomatopeyas en tanto en cuanto los resultados han sido extraídos de páginas webs. Esto, además, ha descartado las onomatopeyas que habían sido inventadas por los autores de los cómics e incluso algunas que ya habían quedado desfasadas por el paso del tiempo. *Sketch Engine*, por tanto, nos permite tener un diccionario actualizado con muestras reales de la lengua.

Por último, una de las aplicaciones más inminentes que nos planteamos es la ampliación del diccionario a otras lenguas tales como inglés y francés, pues sería el instrumento perfecto para solucionar los problemas traductológicos que plantean las onomatopeyas por su diferente realización morfofonológica en cada uno de los idiomas.

Referencias bibliográficas

- Albrecht, J. y Métrich, R. (eds) 2016. *Manuel de traductologie*. Walter de Gruyter
- Bally, C. 1965. *Linguistique générale et linguistique française*. Berna: Francke.
- Barbérís, M.-J. 1992. Onomatopée, interjection: un défi pour la grammaire? *L'information grammaticale* 53: 52-57.
- Barnas, M. 2017. Les pragmatèmes dans les dialogues dans les romans de Marc Lévy. Memoria de máster. Université Grenoble Alpes.
- Blanco Escoda, X. 2013. Les pragmatèmes: définition, typologie et traitement lexicographique. *Verbum* 4: 17-25.
- Blanco Escoda, X. 2014. Microestructuras lexicográficas para unidades frásticas: los pragmatemas. *Káñina. Revista Artes y Letras, Universidad de Costa Rica XXXVIII*: 13-18.
- Gasca, L. y Gubern, R. 2008. *Diccionario de onomatopeyas del comic (Sonido e imagen)*. Madrid: Cátedra.
- Kleiber, G. 2006. Sémiotique de l'interjection. *Langages* 16: 10-23.
- Sketch Engine. <http://www.sketchengine.eu> (último acceso 24 agosto de 2018).