

Building a parallel corpus of literary texts featuring onomatopoeias: ONPACOR

Aroa Orrequia-Barea^a – Cristian Marín-Honor^b
University of Jaén^a / Spain
University of Cádiz^b / Spain

Abstract – Onomatopoeias constitute a much neglected subject in linguistics. The rather scarce literature on onomatopoeias is derived from a lack of reliable empirical data on the topic. In order to bridge this gap, we have compiled a parallel corpus of literary texts featuring onomatopoeias: the *Onomatopoeia Parallel Corpus* (ONPACOR). The corpus consists of onomatopoeias in English, Spanish and French extracted from comics and representative corpora of each language. ONPACOR has been built on the basis of existing translations to the languages of reference. This article describes the methodology used to compile the corpus, as well as the applications that it can have.

Keywords – corpus linguistics; lexicology; onomatopoeias; ONPACOR; parallel corpus; web application

1. INTRODUCTION¹

The fact that onomatopoeias have not been thoroughly studied in linguistics is probably due to the lack of agreement regarding their categorial status (Barbérís 1992: 52). Admittedly, there have been several attempts to compile dictionaries of onomatopoeias such as Gasca and Gubern (2008), a dictionary of onomatopoeias used in Spanish comics, or Enckell and Rézeau (2003), a dictionary of onomatopoeias in French. There have also been attempts to include onomatopoeic forms in general dictionaries such as the *Diccionario de la Lengua Española* (2020) in Spanish or the *Merriam Webster*

¹ We wish to express our gratitude to María Luisa Mora Millán (University of Cádiz) and José María García Núñez (University of Cádiz) for providing insight and assistance during this project. We must also thank the anonymous reviewers and the editors for their comments, which have greatly improved the manuscript. Finally, we are also immensely grateful to Encarnación Almazán Ruiz (University of Jaén) and Alejandro Alcaraz Sintés (University of Jaén) for their invaluable help during the revision process.



Dictionary in English. However, these collections are not very thorough, since these dictionaries are not specifically devoted to onomatopoeias.

This lack of multilingual resources on onomatopoeias has led us to compile the *Onomatopoeia Parallel Corpus* (ONPACOR). This parallel corpus consists of a set of onomatopoeic words in English, Spanish and French retrieved from representative corpora in each language, namely, *The British National Corpus* (BNC) for English, the *Corpus de Referencia del Español Actual* (CREA; Real Academia Española 2020) for Spanish and the *FRANTEXT* corpus (ATILF 2019) for French. ONPACOR has been built on the basis of existing translations of the three languages of reference. For this reason we have restricted our search to literary texts.

The article is organised as follows. Section 2 offers a review of the literature, establishes the distinction between interjections and onomatopoeic expressions, and dwells on the translation problems onomatopoeias pose. Section 3 describes the methodology used to build our parallel corpus. In Section 4, we mention some applications and further developments of the corpus. Section 5 provides information on further research. Finally, Section 6 offers a summary and some conclusions.

2. REVIEW OF THE LITERATURE

As already mentioned, there has been little linguistic research on onomatopoeias. There are several reasons for this. On the one hand, there is no real consensus as regards their definition. Additionally, the grammatical status of onomatopoeias is uncertain, which leads to an incorrect consideration of onomatopoeias and interjections as belonging to the same class (Melnikienė 2016: 169). On the other hand, although some authors, such as de Saussure (2011: 69), claim that the number of onomatopoeic words in languages is very small, in actual fact the prevalence of onomatopoeias is language-specific.

Before launching into the corpus compilation process, it seemed essential to define both onomatopoeias and interjections as clearly as possible. According to the *Merriam Webster Dictionary*, an onomatopoeia is “the naming of a thing or action by vocal imitation of the sound associated with it (such as *buzz*, *hiss*).” In other words, they are part of the “creative process of the conventional language” (Arboleda and Arce-Lopera 2017: 172) in order to reproduce human, animal, natural or artificial sounds (de la Rosa Regot 2015: 2), that is, an onomatopoeia constitutes “an imitative-driven

transformation of a sound of nature into a word” (Assaneo *et al.* 2011). According to de Saussure (2011: 69), onomatopoeic words are one of the exceptions to the so-called arbitrariness of the linguistic sign, insofar as their phonological forms are not arbitrary, but clearly seem to be linked to their meaning. For these reasons, some authors (Sugahara 2011: 2) do not consider them to be, strictly speaking, parts of speech.

The fact that onomatopoeias are imitations of sounds may lead to the conclusion that their representation in language does not vary crosslinguistically. Thus, “one could ideally expect that the imitation of a simple noise should be a single speech sound, the closest one from an acoustical point of view” (Assaneo *et al.* 2011), since sounds are identical regardless of the language system. However, contrary to this intuition, every language has its own onomatopoeic words made up of the consonants and vowels belonging to the phonological system of each language, with clearly different and distinctive properties. As a consequence, onomatopoeias are usually assimilated to the phonological system of that language, causing variations in the shape of words from one language to another in their attempt to represent the same sound (Assaneo *et al.* 2011). An obvious, and much cited, example of this is the onomatopoeic word representing the rooster’s crow in different languages. In Spanish it is *quiquiriqui*, but in French it is *cocorico* and in English *cock-a-doodle-do* (de Buron-Brun 2006: 768).

Nevertheless, the imitation can vary from an almost complete match between the sound and the word representing it—where the human vocal possibilities are exploited so as to obtain the closest imitation to a non-human sound, as is the case in animal onomatopoeias—to a more imprecise representation, in which the mapping between the sound and the word is just an approximation (Rhodes 2010: 279). This approximate imitation supports the theory of ‘Sound Symbolism’, which posits that speakers invest certain phonemic sounds with particular meanings. Therefore, certain phonemes seem to embody certain non-acoustic properties, such as the use of, for example, the front closed high vowel *i*, which conveys the idea of a non-human sharp high-pitched noise, as in *clink* or *click*,² while at the same time it seems to trigger the idea of a diminutive sense (Rhodes 2010: 284). Rhodes (2010: 280), in fact, states that structured sound symbolism “deals with regularities of relationships between forms and meaning that do not readily appear in traditional morphological analysis.”

² In the *Merriam Webster Dictionary*, *click* is defined as “a slight sharp noise” and *clink* as “a slight sharp short metallic sound.”

Even though onomatopoeias in all languages are an extremely tiny set of words as compared to other classes, e.g. verbs, there seem to exist languages —such as Japanese— which are more onomatopoeia-prone than others, such as English or Spanish (Sugahara 2011: 1). In any case, onomatopoeias are frequently used in people's daily life to “complement and enrich their verbal or written communication” (Arboleda and Arce-Lopera 2017: 172) and to express feelings and emotions, which cannot always be expounded by using just words. In fact, Arboleda and Arce-Lopera (2017: 172) studied the widespread use of onomatopoeic words to express speakers' experience with food and concluded that “compared to adjectives (e.g., *soft*, *smooth*), onomatopoeias can better describe a texture because they use symbolic sounds to represent a sensory experience.”

As far as interjections are concerned, these are words related to the expression of feelings (Melnikienė 2016: 177). According to the *Merriam Webster Dictionary*, an interjection is “an ejaculatory utterance usually lacking grammatical connection: such as (a) a word or phrase used in exclamations (such as *Heavens! Dear me!*); (b) a cry or inarticulate utterance (such as *Alas! Ouch! Phooey! Ugh!*) expressing an emotion.” Melnikienė (2016), drawing on Barbéris (1992) and Swiatkowska (2000), distinguishes between two different types of interjections: ‘modal’ and ‘dictal’. On the one hand, modal interjections refer to fixed words belonging to any word class which express emotions, attitudes or feelings (Melnikienė 2016: 177). For example, the Spanish interjection *ay* can be used either in a context of surprise or in a context of pain. As Barbéris (1992: 52) states, the distinguishing feature of this type of interjections is that they do not reflect any aspect of the world, but the emotions of the subject (Melnikienė 2016: 177). On the other hand, dictal interjections refer to the imitation or adaptation of natural sounds and, as a result, this type of words are impersonal and do not express emotions or feelings (Melnikienė 2016: 178).

The main problem to distinguish between onomatopoeias and interjections lies in that most onomatopoeias are considered interjections. However, there are many interjections that do not have an onomatopoeic origin (Kleiber 2006: 11). This can lead to confuse them, basically because some onomatopoeias can be used as interjections. A clear example is the Spanish onomatopoeic expression of laughter *ja*, which is also used as an interjection to show disagreement or disbelief (Husillos Ruiz 2018: 18). The main difficulty arises with onomatopoeias that imitate human sounds, which is where the

boundaries between both words become blurred. This is the reason why these two types of concepts seem to intersect in some contexts (Kleiber 2006: 10).

Despite the aforementioned problem, there are clear differences between interjections and onomatopoeias. First, interjections are considered an independent word class, a part of speech, whereas onomatopoeias are not. Both of them are invariable words and independent as far as their syntactic properties are concerned, but only interjections are able to express meaningful speech acts. In fact, interjections can be the head of syntactic groups or even make up locutions, whereas onomatopoeias cannot. For Kleiber (2006: 12), interjections are “phrases with an implicit predication,” which means that they have a semantic import. For instance, the French modal interjection *pouah!* is used by speakers to convey their strong repugnance at something disgusting (Kleiber 2006: 12).

Regarding morphology, apart from the word class of interjections proper, other word classes, such as nouns or adjectives, can also do duty as interjections. For example, in Spanish, the noun *caracoles* can be used as an interjection to show anger, surprise or admiration (Husillos Ruiz 2018: 21). Nevertheless, all these words have something in common: they are used to communicate the speaker’s natural sounds of feelings and emotions (de Buron-Brun 2006: 768).

From a pragmatic and semantic point of view, onomatopoeias tend to lack semantic content because they are restricted to the imitation of sounds. However, they are influenced by extralinguistic factors or the context in which they are used, sometimes even in visual material, such as comics. For this reason, it is very difficult to associate each onomatopoeia to just one meaning. In most cases, one onomatopoeia may represent more than one sound. To give a simple example, the Spanish word *pum* can refer to either a gunshot or a door knock (Orrequia-Barea and Marín-Honor 2018: 97). Certainly, attempts have been made to tackle this problem by means of lists of onomatopoeias or glossaries, though most often lacking reliability (Sugahara 2011: 2). Besides, certain scholars, such as Sugahara (2011), have already focused on the onomatopoeia translation problem. Additionally, de la Rosa Regot (2015) provides strategies for translating onomatopoeias between English and Spanish. Similarly, Husillos Ruiz (2018) compiled a multilingual glossary and conducted research in translation strategies in advertising. Also, there have been attempts to compile dictionaries of onomatopoeias, such as Gasca and Gubern (2008) in Spanish, and

Enckell and Rézeau (2003) in French. Despite this body of research, little consideration has been given to crosslinguistic differences, and there is a dramatic lack of multilingual resources. The compilation of a multilingual corpus or even a multilingual dictionary focused on onomatopoeic words may help solve translation-related problems among languages. Interestingly, onomatopoeias are not just words which imitate sounds, but they are also used to help language users complement and enhance their communication. For this reason, it is necessary to translate them because not doing so may cause a loss of meaning. Since one single non-human sound may be conveyed by means of different onomatopoeic expressions, it seems necessary to have a multilingual resource containing all these equivalent words or expressions. We believe that the *Onomatopoeia Parallel Corpus* (ONPACOR) constitutes a major step in that direction.

3. METHODOLOGY

In this section we discuss the process of compilation of ONPACOR. We have followed four main steps: the compilation of a list of onomatopoeias, the search for the concordances, the translation phase and the creation of the interface. These steps will be explained in what follows.

3.1. *Compilation of onomatopoeias*

Two different strategies were implemented to make the compilation based on the availability of reference corpora. Our first idea was to extract onomatopoeias from corpora of each language since we wanted to have empirical evidence that those onomatopoeic forms were actually used in the language. For this reason, we intended to download the corpora to look for onomatopoeias using regular expressions to get as many onomatopoeic forms as possible without restricting them to the most common ones. However, we could only follow this procedure with the BNC, since it was the only corpus that could be downloaded. For Spanish and French, the CREA and *FRANTEXT* corpora were not downloadable, so that we had to follow a different process, namely manually extracting onomatopoeias from comics. Nonetheless, we extracted as many instances from comics as we found in the BNC so that the sample would be balanced.

For Spanish, we chose CREA, which includes oral and written texts produced in every Spanish-speaking country from 1975 until 2004. The main reason for the selection of this corpus was that 90% of the texts are written, whereas only 10% represent the spoken language. As for French, the corpus chosen is *FRANTEXT*, which basically contains literary and philosophical texts from 1180 to 2009, which suited the purpose of our corpus. Finally, regarding English, we worked with the BNC, of which 90% belong to written texts from the late twentieth century. These three corpora were suitable for the purpose of our project, since they consist of written texts, which made it easier to find the translations.

3.1.1. English

Firstly, we made an analysis of the form of the onomatopoeic words in the BNC in order to compile them. The purpose of this analysis was to find patterns in the formation of onomatopoeias. Therefore, we systematised some typical combinations found in onomatopoeias in three groups: 1) vowels, 2) consonants and 3) endings of words (Kwon 2015: 39–71).³

1. Long vowels tend to be used to represent slow movements while short tend to represent quick ones. Additionally, onomatopoeias often consist of similar words with a change of vowels to represent two-phased movements, such as *ding-dong* or *flip-flop*. Vowels in ablaut-like alternation are usually repeated twice or more, sometimes even three or four times.
2. Consonants are usually found in pairs to represent different movements and ways of doing things. Most typically we find the combinations: consonant + *-l* or *-r*.
3. Endings of words are usually made up of two-consonant clusters or repeated vowels.

All the above-mentioned systematisations were captured by means of regular expressions, which are patterns that are frequently used in text editors to look for,

³ Kwon's (2015: 40) research is focused on 'phonaesthemes', that is, "recurrent pairings of sound and meaning." Although we do not deal with phonaesthemes here, Kwon's study is used as a basis to systematise some patterns of sound symbolism that are expected to be found in onomatopoeias.

substitute and replace a sequence of characters. This sequence has to fulfil the criteria set out by the regular expression.⁴

As the main purpose was to find most of the onomatopoeias in the BNC, the following regular expressions, based on the previous patterns of formation, were used:

1. To find consonants that were repeated at least three times: `[bc-df-hj-np-tv-z]{3}`. This regular expression yielded onomatopoeias such as *zzz*.
2. To find the pattern of up to two consonants plus vowels, repeated at least twice, followed optionally by an indefinite number of consonants: `[bc-df-hj-np-tv-z]{0,2}vowel{2,}[bc-df-hj-np-tv-z]{0,}`. We typed each of the five different vowel graphemes in the vowel slot. Some of the results were: *craark*, *beep*, *riing*, *boom* or *uuummm*.

3.1.2. Spanish and French

Comic books are known to be a very rich repository of onomatopoeias. That is the reason why we manually compiled a list of onomatopoeic words from the most popular comics in each language. For Spanish, we chose a variety of issues from some of the most popular comics in Spain, among others, *Zipi y Zape* by José Escobar, *Mortadelo y Filemón* and *Rompetechos* by Francisco Ibáñez, *El doctor Cataplasma* by Martz Schmidt and *Sir Tim O'Theo* by Raf. As for French, we chose such popular comics as *Astérix et Obélix* by René Goscinny and Albert Uderzo, *Spirou et Fantasio* by Rob-Vel and *Les Aventures de Tintin et Milou* by Georges Remi. As a result, we achieved a compilation of 500 onomatopoeias.

3.2. Concordances

After compiling the three lists of onomatopoeias, the next step was to check whether there were concordances with those onomatopoeias in each corpus. As a matter of fact, the motivation to select each corpus was based on the existence of written literary texts, so that we could find the translations in the following step. Although onomatopoeias are frequently used in dialogues and in the spoken language, it was not until the nineteenth

⁴ For further information on regular expressions see <https://regexr.com/>.

century and the birth of Realism that onomatopoeias started to appear in novels, in an attempt by writers to reproduce the colloquial language (Bueno Pérez 1994: 15).

The process of looking for concordances was basically the same for the three languages. We searched for each onomatopoeia in the chosen corpus and then stored all its concordances. We also stored some interesting metadata, which was included in each concordance, such as the year of publication, the author and the title of the work. In the case of the CREA interface, there is a book filter which makes it easier to find the translation. Since we downloaded the BNC, we used *AntConc* (Anthony 2019) for the retrieval of the data.

3.3. Translations

In order to be able to find existing translations of the texts, at this stage we needed to use corpora mainly containing written or literary texts. For this purpose, we used the website *Index Translationum*, a database of book translations promoted by UNESCO.⁵ The user can set up the search criteria for any given book, so that the database provides a record of the translations in all different languages. In order to restrict the search, we provided just the author's name and the title. As previously mentioned, this information was obtained in the second step. The list of records contains such information about each translation as the title of the work in the target version, the language used in the translation, the translator, the place and the year of publication, among others. This is illustrated in (1) and (2), which show a search for *The Colour of Magic* by Terry Pratchett, which returned 36 hits.

(1) Pratchett, Terry: *El color de la magia* [Spanish] / Macía, Cristina / México, D.F.: Roca [México], 1989. 224 p. *The Colour of Magic* [English]

(2) Pratchett, Terry: *La huitième couleur* [French] / Marcel, Patrick / Nantes: l'Atalante [France], 1993. 283 p. *The Colour of Magic* [English]

This database was used to filter out the titles which lacked translations in the other languages. Once the translation of the book was found, the exact excerpt—namely the section which included the onomatopoeia—was extracted.

⁵ <http://www.unesco.org/xtrans/>.

3.4. Implementing the interface

ONPACOR will be hosted on an online platform so that users can access it easily.⁶ This platform is actually a web application, which can be operated in a browser. To implement this platform, we needed to design a database to host the texts featuring the onomatopoeias in the three different languages. For the design of the database, we used a conceptual model (Coronel and Morris 2016: 71–274), because it represents a comprehensive picture of the information, as the users are going to see it, ignoring implementation details as well as the structure of the information, which makes it more understandable for researchers in the Humanities. To shape this information, we used an ‘entity-relationship model’ (Chen 1976), which is the most widely used conceptual model. It is made up of a set of concepts which allows to describe reality by means of a set of linguistic and graphic representations, presenting a natural vision of the real world. Moreover, this model has a number of advantages. First, it only reflects the existence of the information. Secondly, it does not depend on any particular database or operating system. Thirdly, it is open thus allowing the system to be updated as much as possible, which is a great advantage for an on-going project such as this one. The major steps taken to create the database are presented in the following paragraphs.

The first step was to generate a universe of discourse (Boole 1854), namely, a description of the information, the collection of objects that will be included in the database as well as a schema about how these data are related. In ONPACOR, the universe of discourse is made up of the onomatopoeias, which were extracted in the compilation step. These onomatopoeias are related to the concordances that were extracted from the corpus, and, at the same time, they are related to the translations in the other two languages.

The second step was to build the database, for which purpose we used a standard database management system, *MySQL*, for two main reasons: on the one hand, it is one of the most commonly used open source databases and, on the other hand, it is used for relational databases, which fits perfectly with ours. To build the database, we used two tables, one for the concordances and another for the onomatopoeia. As can be seen in Figure 1, some information is required in the concordance table, namely, the language, the concordance itself and each translation. Likewise, it is possible to highlight whether

⁶ A similar platform has already been implemented to host the online dictionary of onomatopoeias in Spanish (Orrequia-Barea and Marín-Honor 2018).

one particular extract belongs to the original version. The information that is going to be introduced into this database may be restricted by using the parameters found in *MySQL*. This includes a ‘Variable chain of characters’ (varchar) or a ‘Boolean operator’ (bool) to indicate whether the excerpt is original or not, as shown in Figure 1.

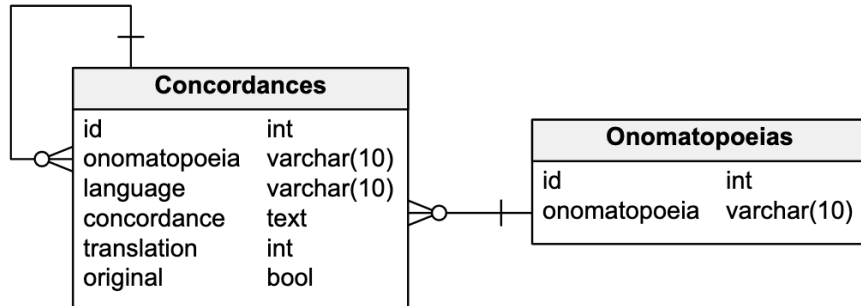


Figure 1: Structure of the database of ONPACOR

The two tables in Figure 1 establish a relationship between an onomatopoeia and one or more concordances, since the onomatopoeic word sometimes coincides in the three languages. Likewise, the concordances are able to establish relationships between themselves, so that the three translations can be related in order to align the extracts in the parallel corpus.

The third step in the implementation of the interface was to set up the website. The site hosting ONPACOR has been created by using *Django*, a Python framework for web development. This website displays two main sections: the administration area and the users’ area. The former is used to input the information that is going to be displayed and is accessible only to the creators. A number of safeguards have been put into place when creating a new entry. For example, each language can be added only once, so that no mistake can be made by the administrator. In addition, if the original language option has already been selected, the system will not allow the administrator to add more original languages to that concordance. In the user area, there is a query box to type in the onomatopoeias to be searched for. When typing an onomatopoeic word, the system displays the matches already available in the database to help the user. Afterwards, by clicking on the search button, queries yield concordances in the three languages with a 100-word context, which helps see their meaning and use. The box displaying the excerpt in the original version is highlighted in bold typeface and a thicker frame line. This feature is particularly useful for translators, since it helps them identify clearly

which is the original excerpt and which is the translation. In Figure 2, the original text is the French one, extracted from *Les Trois Mousquetaires* by Alexandre Dumas.

ONPACOR

Hush
Why, by your marchioness, your duchess, your princess. She must have a long arm. - Hush! said Porthos, placing a finger on his lips.
English

Chut
Mais par votre marquise, votre duchesse, votre princesse; elle doit avoir le bras long. - Chut! dit Porthos en mettant un doigt sur ses lèvres
French

Chis
Pues por medio de vuestra marquesa, vuestra duquesa, vuestra princesa; debe de tener largo el brazo. -¡ Chis! Dijo Porthos poniendo un dedo sobre sus labios.
Spanish

Figure 2: Screenshot of the web app ONPACOR

4. APPLICATIONS OF ONPACOR

In our view, ONPACOR has three main applications. Firstly, when finished, it will constitute a huge compilation of literary texts that not only translators but also lexicographers and writers in general can use. Secondly, ONPACOR will become a multilingual resource available for the research community. Thirdly, it will allow

researchers to conduct comparative studies of onomatopoeias in different languages and account for the different mechanisms employed by translators.

One of the main reasons for the compilation of ONPACOR was to compare onomatopoeias crosslinguistically. There is evidence that different onomatopoeic words represent the same sound, as the example of the rooster's crow mentioned in Section 2 clearly showed. However, this crosslinguistic variation does not always take place. Generally speaking, the most common strategy used in the translation of onomatopoeias is that of 'equivalence' (Mayoral Asensio 1992: 139). In this sense, the translator uses an onomatopoeic word available in the target language, which is equivalent in meaning to the original one. This is illustrated in the examples provided in Figure 2 (see Section 3.4) where the translators use *hush!* for English, *¡chis!* for Spanish and *chut!* for French.

Besides, another common strategy is the use of loanwords, that is, to use the onomatopoeia in its original form without translating it. Even though the use of loanwords is very common in comics, according to Mayoral Asensio (1992: 139), translators should avoid this strategy which impoverishes the target language and causes onomatopoeic words to disappear due to lack of use.

It is undeniable that, due to their use in comics and because English is the most widely used language on the Internet, onomatopoeias are more frequently attested in English than in Spanish or French. What is more, some of the English onomatopoeic forms have been assimilated in languages such as Spanish and French, as is the case of *boum* or *pum* (de Buron-Brun 2006: 770). This way of proceeding has been found in the translations of *Le Rouge et le Noir* by Stendhal —with the onomatopoeia *bah!*— and in those of the comics of *Astérix et Obélix* —where only the laughter bubbles of *ha ha* were translated into Spanish as *ja, ja, ja*. As the onomatopoeia of laughter coincides in both English and French, its translation was not necessary. Alternatively, as some onomatopoeic expressions are difficult to translate because of lack of correspondence, translators tend to omit them in the target text. Sometimes the omission is caused by the presence of other elements, such as images, which make the onomatopoeias redundant. In our corpus, we found this mechanism, among others, in *The Colour of Magic* by Terry Pratchett, where the onomatopoeic word is found in the English version, but not in the Spanish and the French versions.

5. FURTHER RESEARCH

Once this large trilingual database has been compiled, the next step will be the compilation of a multilingual dictionary of onomatopoeias. The idea is to make use of all the collected data to create an electronic tool that will display the examples and the meaning of each onomatopoeic form. In fact, a tentative version of this dictionary has already been implemented, though only for Spanish (Orrequia-Barea and Marín-Honor 2018). In this dictionary, the user can search either for the onomatopoeic word, for example, *pum*, or for the real sound or noise it represents, which is a gunshot. These two options are necessary, since onomatopoeic expressions are usually associated with more than one sound or vice versa. Apart from the onomatopoeia and its meaning, the dictionary also displays the concordances so that users may see the onomatopoeic forms in context. We are currently working on the English and French versions, but we do not discard the possibility of widening this project to other languages, such as Italian and German.

As regards the translations, we have chosen just one translation for each original book. However, it would be interesting to look into different translations of the same work by different translators. This may help improve our understanding of the mechanisms that are used by translators as well as include variations or different renderings of the same onomatopoeic word.

In addition, in the future the corpus and derived dictionaries may well be complemented by the creation of an extension for messaging apps allowing users to include onomatopoeias in their conversations. Given the widespread use of text messaging today, people need to convey as much accurate information as possible in their conversations to put forward messages properly. It is true that, to a certain extent, this need is already catered for by emojis, stickers and gifs but it is undeniable that these resources do not always serve our communicative needs. That is why we are considering the implementation of an onomatopoeia extension that would allow users to introduce them in the messages without having to actually type them. The idea is to use the previous compilation of onomatopoeic words to create what may be termed ‘onomojis’, as we have coined them, that is, emojis consisting of an onomatopoeia. Their design could be similar to that used in comics, with the typical font, shape and colours of onomatopoeias in this literary genre.

6. CONCLUSIONS

The analysis of onomatopoeias has been neglected in linguistics. Adding to the scarce literature on the topic, it has been shown that there is a lack of multilingual resources related to these types of words. For this reason, ONPACOR proves to be a useful digital tool, which can shed light on many onomatopoeias related issues. ONPACOR constitutes a really useful resource for translators, filling the gap that currently exists in the multilingual dimension of onomatopoeias.

For the sake of accuracy, and to include just onomatopoeias in ONPACOR, a thorough study has been carried out regarding the differences and similarities between onomatopoeias and interjections. This means that translators are going to have a solid compilation of onomatopoeias which is going to help them in their translations. The interface displays the onomatopoeic words in the three languages, and provides information about the context which, in turn, may also help the translator decide whether the onomatopoeia fits in the translation. Additionally, the concordances for the immediate context and the translations have been carefully selected so that they represent evidence of real language in use. Furthermore, the interface will allow translators and linguists to conduct comparative studies between different languages and to carry out research into the strategies deployed by translators. Although the web app has already been created, this is still an on-going project and we are currently introducing new onomatopoeias and concordances as the project progresses.

REFERENCES

- Anthony, Laurence. 2019. *AntConc*. Tokyo: Waseda University.
- ATILF. *Base Textuelle Frantext*. 2019. ATILF-CNRS and Université de Lorraine. <https://www.frantext.fr/>
- Arboleda, Ana M. and Carlos Arce-Lopera. 2017. The French, German and Spanish sound of eating fresh fruits and vegetables. *Food Research International* 102: 171–175.
- Assaneo, María Florencia, Juan Ignacio Nichols and Marcos Alberto Trevisan. 2011. The anatomy of onomatopoeia. *PLoS ONE* 6/12. <https://doi.org/10.1371/journal.pone.0028317> (20 April, 2020.)
- Barb  ris, Jeanne-Marie. 1992. Onomatop  e, interjection: Un d  fi pour la grammaire. *L'Information Grammaticale* 53/1: 52–57.
- Boole, George. 1854. *An Investigation of the Laws of Thought. On which are Founded the Mathematical Theories of Logic and Probabilities*. New York: Dover Publications.
- Bueno P  rez, Mar  a Lourdes. 1994. La onomatopeya y su proceso de lexicalizaci  n: Notas para un estudio. *Anuario de Estudios Filol  gicos* 17: 15–26.

- Chen, Peter Pin Shan. 1976. The entity-relationship model: Towards a unified view of data. *ACM Transactions on Database Systems (TODS)* 1/1: 9–36.
- Coronel, Carlos and Steven Morris. 2016. *Database Systems: Design, Implementation and Management*. Boston: Cengage Learning.
- Corpus de Referencia del Español Actual*. 2020. Real Academia Española: Banco de datos (CREA). <http://www.rae.es>.
- de Buron-Brun, Bénédicte. 2006. La onomatopeya, ¿mucho ruido para pocas nueces o un rompecabezas para el traductor? In Manuel Bruña Cuevas, María de Gracia Caballos Bejano, Inmaculada Illanes Ortega, Carmen Ramírez Gómez and Anna Raventós Barangé eds. *La Cultura del Otro: Español en Francia, Francés en España*. Sevilla: Universidad de Sevilla, 768–784.
- de la Rosa Regot, Nuria. 2015. *Translating Sounds: The Translation of Onomatopoeia between English and Spanish*. Barcelona: The University of Barcelona Degree Dissertation.
- de Saussure, Ferdinand. 2011. *Course in General Linguistics*. New York: Columbia University Press.
- Diccionario de la Lengua Española*. 2020. Real Academia Española. <https://dle.rae.es>.
- Enckell, Pierre and Pierre Rézeau. 2003. *Dictionnaire des Onomatopées*. Paris: Puf.
- Gasca, Luis and Román Gubern. 2008. *Diccionario de Onomatopeyas del Cómic*. Madrid: Ediciones Cátedra.
- Husillos Ruiz, Araceli. 2018. *Las Pequeñas Palabras y su Traducción: Glosario Trilingüe de Onomatopeyas*. Valladolid: The University of Valladolid Degree Dissertation.
- Kleiber, Georges. 2006. Sémiotique de l'interjection. *Langages* 40/161: 10–23.
- Kwon, Nahyun. 2015. *The Natural Motivation of Sound Symbolism*. Australia: Queensland University PhD Dissertation.
- Mayoral Asensio, Roberto. 1992. Formas inarticuladas y formas onomatopéyicas en inglés y español. Problemas de traducción. *Sendebat: Revista de La Facultad de Traducción e Interpretación* 3: 107–140.
- Melnikienė, Danguolė. 2016. Le statut grammatical des onomatopées dans la linguistique moderne. *Verbum* 6/6: 168–187.
- Merriam Webster Dictionary Online*. 2020. <https://www.merriam-webster.com>
- Orrequia-Barea, Aroa and Cristian Marín-Honor. 2018. Hacia la elaboración de un diccionario de onomatopeyas en español. *CHIMERA: Romance Corpora and Linguistic Studies* 5/1: 93–99.
- Rhodes, Richard. 2010. Aural images. In Leanne Hinton, Johanna Nichols and John J. Ohala eds. *Sound Symbolism*. Cambridge: Cambridge University Press, 276–292.
- Sugahara, Takashi. 2011. *Onomatopoeia in Spoken and Written English: Corpus and Usage-based Analysis*. Japan: Hokkaido University PhD Dissertation.
- Swiatkowska, Marcela. 2000. *Entre Dire et Faire. De l'Interjection*. Krakow: Wydawnictwo Uniwersytetu Jagiellonskiego.
- The British National Corpus* 3.0. 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium.

Corresponding author

Aroa Orrequia-Barea

Univesity of Jaén

Department of English Studies

Building D2, office 241

Campus Las Lagunillas

23071, Jaén

Spain

E-mail: orrequia@ujaen.es

received: February 2020

accepted: June 2020