

«A Baptism of Fire»: Towards a Practical Hybrid Approach for the Lexicographic Indexation of Phraseological Units with Religious Lexical Components in English and Spanish

«Bautismo de fuego»: hacia un enfoque híbrido sistemático para la indexación lexicográfica de unidades fraseológicas con componentes religiosos en inglés y español

José Luis ROJAS DÍAZ

NHH Norwegian School of Economics

jose.diaz@nhh.no

Juan Manuel PÉREZ SÁNCHEZ

University of Antioquia

jmanuel.perez@udea.edu.co

Resumen: Tradicionalmente, los investigadores han demostrado un particular interés por el estudio de la relación entre fraseología y lexicografía [p. ej. Alonso Ramos (2006); Mellado Blanco (2008); Buendía Castro y Faber (2015); Paquot (2015); Nuccorini (2020)] hasta el punto de denominarlo «matrimonio científico» (Leroyer 2006). De igual manera, los académicos se han interesado de manera creciente por el análisis semántico de las unidades fraseológicas (en nuestro texto, PUs, por sus siglas en inglés) [e.g., Grčić Simeunović y de Santiago (2016) y Torijano y Recio (2019)]. Entre los problemas

que estos y otros estudios han señalado se encuentra la recurrente referencia a la inexactitud y la dificultad para indexar las PUs en los recursos lexicográficos. Aunque algunos investigadores consideran los enfoques onomasiológicos como un interesante punto de partida [e.g., Bosque (2017) y Siepmann (2008)], sigue siendo necesario establecer una metodología sistemática respecto a la fraseología que incluya tanto el análisis semántico de las entradas como su indexación.

Nos proponemos abordar esta necesidad analizando 242 modismos (199 en español y 43 en inglés) extraídos de una base de datos con 21.045 modismos que ha sido compilada a partir de dos diccionarios fraseológicos: el *Diccionario fraseológico documentado del español actual* (en adelante DFDEA) (Seco, Andrés *et al.*, 2004), y el Collins COBUILD Dictionary of Idioms (en adelante CCDI) (Sinclair y Moon 1997). Los criterios empleados para seleccionar las unidades resultantes con vistas a su análisis han sido: (i) debían incluir al menos un componente léxico relacionado con la religión, y (ii) el modismo tenía que ser nominal o verbal. El componente religioso fue identificado de manera semi-automática utilizando el UCREL's Semantic Analysis System (USAS) (Archer *et al.*, 2002).

Las aportaciones de este artículo son las siguientes: (i) presenta un análisis lexicográfico de la macroestructura y la microestructura de los dos recursos fraseológicos antes mencionados, (ii) ofrece un modelo de análisis semántico para las PUs con componentes relacionados con la religión, (iii) propone un método de indexación alternativa de las PUs en recursos lexicográficos que implican enfoques semasiológicos y onomasiológicos; y finalmente, (iv) propone una manera sistemática de utilizar la información semántica y pragmática para crear entradas semánticas para las PUs.

En conclusión, al examinar dicha serie de entradas fraseológicas, este estudio arroja luz sobre la composición semántica de las PUs. También sugiere un enfoque híbrido sistemático para su indexación lexicográfica en inglés y español.

Palabras clave: fraseología; lexicografía; religión; onomasiología; semasiología; indexación.

Abstract: Traditionally, researchers have had a particular interest in the study of the relationship between phraseology and lexicography [e.g., Alonso Ramos (2006); Mellado Blanco (2008); Buendía Castro and Faber (2015); Paquot (2015); Nuccorini (2020)] to the point of having labeled it a «scientific marriage» (Leroyer 2006). In addition, scholars have been increasingly interested in the semantic analysis of phraseological units (henceforth PUs) [e.g., Grčić Simeunović and de Santiago (2016) and Torijano and Recio (2019)]. Among the problems that these and several other studies have pointed out, there is the recurrent reference to inaccuracy and difficulty in indexing PUs in lexicographic resources. Although some scholars consider onomasiological approaches as an interesting starting point [e.g., Bosque (2017) and Siepmann (2008)], a systematic methodology in phraseology that includes both the semantical analysis of the entries and their indexation is still needed.

We intend to address that need here through the analysis of 242 idioms (199 in Spanish and 43 in English) extracted from a 21,045-idiom database that was compiled from two phraseological dictionaries: the *Diccionario fraseológico documentado del español actual* (henceforth DFDEA) (Seco, Andrés *et al.*, 2004), and the Collins COBUILD Dictionary of Idioms (henceforth CCDI) (Sinclair and Moon 1997). The criteria

employed to select the resulting analysis units were: (i) they had to include at least one lexical component related to religion, and (ii) the idiom had to be nominal or verbal. The religious component was identified semi-automatically by using the UCREL's Semantic Analysis System (USAS) (Archer *et al.*, 2002).

The contributions of this paper are as follows: (i) it presents a lexicographic analysis of the macrostructure and microstructure of the two phraseological resources previously mentioned, (ii) it offers a model of semantic analysis for PUs with religion-related components, (iii) it proposes an alternative indexation method of PUs in lexicographic resources involving semasiological and onomasiological approaches; and finally, (iv) it shows a systematic way to use semantic and pragmatic information in order to create semantic entries for PUs.

In conclusion, by closely examining said set of phraseological entries, this study sheds light on the semantic composition of Pus. It also suggests a systematic hybrid approach for their lexicographic indexation in English and Spanish.

Keywords: phraseology; lexicography; religion; onomasiology; semasiology; indexation.

1. INTRODUCCIÓN

Over the years, the study of phraseology within lexicography has been of interest to many scholars (e.g. Alonso Ramos 2006, Mellado Blanco 2008, Buendía Castro and Faber 2015, Paquot 2015, Nuccorini 2020). In most cases, these studies focus, among several other topics, on the type of units that are indexed in dictionaries, how phraseological units (Henceforth PUs) are labeled or tagged in lexicographic resources, and the procedures for indexing PUs in dictionaries. In their studies, phraseology scholars have developed interesting theoretical conceptions that, in our opinion, could have already been applied to lexicographic practices. For instance, the notion of «functional equivalence» put forward by Corpas Pastor (2000) and Dobrovol'skij and Piirainen (2005), as well as the concept of semantic *unicum* included in Gries' (2008) parameters for the definition of what a phraseological unit is, are just two of those useful-to-lexicography ideas derived from phraseological studies. However, only a few of these notions have been put into practice to enhance the indexation methods of PUs in new lexicographic resources, reprints, or renewed versions of pre-existing works.

One possible cause of the imbalance between the conclusions reached through phraseological studies and their application to the lexicographic practice might be related to the lexicographic conventions rooted in traditional lexicographic principles. Thus, Veisbergs (2020) points out that some of those lexicographic conventions deal with the marking/labeling of PUs as well as with the choice of equivalents and definitions. Although his study deals with bilingual dictionaries, Veisbergs's conventions also apply to the monolingual dictionaries. Later in his chapter Veisbergs (2020) warns his readers about one of the challenges that users face when looking for PUs in dictionaries, namely: most of the time, they do not know where to find PUs in dictionaries. Hence, the main objective of this article is to identify and analyze lexicographic and semantic information found in two phraseological monolingual dictionaries (one in English and

one in Spanish) as a previous step towards a lexicographic proposal that encompasses both semasiological and onomasiological approaches (see section 2). More precisely, the present study intends to examine (i) the headwords used to index PUs as well as the definitions offered by the dictionaries being studied and (ii) to provide an alternative for the indexation of PUs based on the semantic information extracted from their lexical components and their corresponding definitions. In order to carry out the analyses needed, a database with PUs containing religion-related lexical components taken from the Collins COBUILD Dictionary of Idioms (henceforth CCDI) (Sinclair and Moon 1997) and from the *Diccionario fraseológico documentado del español actual* (henceforth DFDEA) was built.

Deignan, Lima, and Lopez Mora (1998), referred to by Deignan (2005), states that «Spanish [...] draws on religion for a number of idiomatic expressions.» (p. 261). Likewise, Boers and Stengers (2008) compared two repertoires of idioms, both in English and in Spanish, and concluded that «idioms derived from religion and superstition made up an especially large segment of the Spanish idiom repertoire» (p. 368). In this line, a general frequency query applied to the DFDEA (Seco, Andrés *et al.*, 2004) has resulted in the finding that the most frequent noun among the lexical components of the PUs included in that work was the word Dios ('God'). As a result, it was decided to extract the PUs containing religion-related lexical components in order to create a sample database for this study, since it has been demonstrated that religion is an important source for the creation of PUs in Spanish and potentially in English.

The use of a database consisting of PUs with religion-related components shows that, as in many other types of PUs, the tenors —i.e., the figurative meanings of metaphorical lexical units— are not related to religion. However, several of those PUs' definitions are indeed semantically linked and can be grouped under a common concept. In consequence, this paper will be divided as follows: The second section aims to provide an overview of the relationship between phraseology and lexicography, as well as on the definitions of PU, macrostructure, and microstructure. In the third section, data, tools, and methods used in this study are described. In turn, the fourth section will deal with the different headwords, definitions, and tenors of the phraseological entries being analyzed. In the fifth section, the lexical, semantic, and lexicographic analyses employed in the study are applied to a lexicographic representation proposal that encompasses the several layers of information contained in a PU and in the lexical units that constitute it. Finally, in the sixth section, the conclusions resulting from this study are presented.

2. PHRASEOLOGY, LEXICOGRAPHY, AND SEMANTICS

According to García-Page, phraseology should be defined in terms of its object of study. This statement raises the question: «what is the object of study of phraseology?»

(2008), the answer to which is the same as offering a definition of PU. Authors like Burger have offered a classification of idioms in which he includes 'idioms,' 'partial idioms,' and 'non-idioms' (1998). Likewise, Mel'čuk, divides PUs into 'idioms,' 'collocations,' and 'clichés' (2012), while 'idioms,' in turn, are divided into 'full idioms,' 'semi-idioms,' and 'quasi-idioms.' The database compiled for the analysis intended in this study is composed by nominal and verbal idioms. The characteristics of these PUs meet what Mel'čuk proposes as a definition of 'full idiom' in the following terms: «an idiom AB is a full idiom if its meaning does not include the meaning of any of its lexical components: 'AB' $\not\supset$ 'A' and 'AB' $\not\supset$ 'B'» (Mel'čuk 2012). This last definition put forward by Mel'čuk will be the one applied to 'phraseological units' in this study.

As for the relationship between phraseology and lexicography, according to Leroyer (2006), this relationship should be considered a «scientific marriage» since they have been related for a long time. In his study, this author retrieved more than 1,700 reference entries from the EURALEX site containing both terms –phraseology and lexicography– in their keywords or title (Leroyer 2006). Nonetheless, just as any other marriage, it is unique and not exempt from different complications and obstacles. Phraseology has indeed been widely explored from a lexicographic point of view; however, most of the conclusions and observations resulting from research remain in the field of scholarly studies, and few of the outcomes from such studies have been introduced into the lexicographic practice. In line with this, such a rather dysfunctional marriage has been addressed by authors like Moon (2008) and Paquot (2015), who have carried out extensive studies regarding the representation and treatment of phraseology in dictionaries.

On the one hand, these authors have indicated the lack of consistency in how PUs have been either selected or indexed in dictionaries. For example, Moon (2008) gives evidence on how the PUs chosen to be indexed in dictionaries do not adopt the suggestions put forward by phraseological theories. She also states that dictionaries lack contextual information, particularly regarding the use of idioms (Moon 2008). This last issue is addressed by Paquot (2015) as well, although her ideas are focused on contextual information related to the use of collocations.

On the other hand, from the point of view of lexicography, Atkins and Rundell (2008) state that, although multiword expressions (henceforth MWE) are a «central part of the vocabulary in most languages,» it is not common that dictionaries identify and label them into specific categories. These authors argue that the difficulties in systematizing the indexation of MWE (PUs included) are due to the plethora of denominations, definitions, and taxonomies of MWE derived from the criteria employed in their identification (Atkins and Rundell 2008).

Nevertheless, and despite the difficulties pointed out by Atkins and Rundell (2008), these authors dedicate a whole section of their book to identifying different types of MWE, including PUs. Therefore, although the identification of PUs is a challenge in itself, it is possible indeed, and it should be carried out exhaustively within the lexicographic practice.

It is necessary to note that the issues that have been put forth by the studies related to both phraseology and lexicography apply to the lexicographic practice in many languages, including Spanish. In this way, González (2006) notes how the *Diccionario de la Real Academia Española* (DRAE) uses two different taxonomies for the indexation of PUs. In her study, González concludes that the selection criteria employed in the DRAE for the inclusion of collocations follow the classification system developed by Corpas Pastor (1996), while idioms are categorized by using the taxonomy proposed by Casares (1950). Regarding labeling in dictionaries, Ortega Ojeda and González Aguilar (2008) present the labels used in two general language dictionaries in Spanish. They conclude that both the labeling in those dictionaries and the criteria used for classifying PUs in both works were inaccurate.

As has already been stated, the indexation of PUs is considered as one of the challenges in the craft of lexicographic resources. Most of the repertoire of dictionaries and databases available for consultation have been built using a semasiological approach, which requires the users to know the form of the expression or lexical unit they are looking for (Kocjančič 2004). This approach poses a series of problems depending on the kind of user that consults those resources. On the one hand, for many novel users (e.g., language learners) the problem has to do with the headwords used to index MWEs in lexicographic resources. On some occasions, the headwords are not very intuitive, or the dictionary does not include the necessary information in the guidelines for users to know how to look for those headwords. On the other hand, advanced users (e.g., translators and linguistic mediators) will not use a semasiological dictionary because the kind of question they will be asking would be something like: ‘how can I express Y (concept) in a certain language?’ i.e., they take an onomasiological approach.

The proposal offered in this paper for the indexation of PUs in lexicographical resources, which will be explained in detail later (see section 5), combines both of these approaches and can be considered a hybrid approach. As shown in the following sections, after carrying out a semantic analysis of the definitions of the phraseological entries in the present study, an alternative indexation was considered feasible.

In the first place, the idea that PUs constitute a whole semantic unit is not new. Veisbergs states that «Some theorists [...] agree that words and idioms share a common trait in that they both exist as a single semantic unit» (2020). Additionally, he traces the semantic unit idea back to the work of Chafe (1968). Likewise, Cruse (2000) states that «all the meaning of the phrasal unit attaches to the phrase, and none to its constituents.» (Cruse 2000). This conception of PUs working as a semantic *unicum* taken from Gries’ criteria for identifying PUs (2008) can be used in lexicography not only for crafting definitions from a semasiological approach but for the indexation of PUs from an onomasiological perspective [as proposed by Bosque (2017) and Siepmann (2008)]. Additionally, this view of PUs being a semantic *unicum* is directly related to their meanings being mostly figurative and to the fact that those figurative meanings are developed through semantic mechanisms such as metaphor and metonymy. Therefore, it was deemed necessary to determine the type of semantic relationship that mediated in the

development of the figurative meanings of the PUs under study here. This information would provide both lexicographers and end-users with valuable information regarding the frequency of use of semantic mechanisms such as metaphor and metonymy, the type of *vehicles* —literal expressions and words— and *tenors* —figurative meanings or referents— that are preferred by speakers, or aspects such as the importance of what Warren 1992 called evaluative-attributive features —i.e., semantic features that are used «to indicate the encoder’s attitude towards the referent.»

In summary, all the concepts related to phraseology, lexicography, and semantics presented above were applied to the sample selected for the present study in order to outline a hybrid proposal for the lexicographic representation of PUs, as it will be explained in the following sections.

3. DATA, TOOLS, AND METHODS

For this study, two dictionaries were used: one in English (CCDOI) (Sinclair and Moon 1997) and one in Spanish (DFDEA) (Seco, Andrés *et al.*, 2004). This phraseological/lexicographic study derives from previous semantic and morphosyntactic analyses of lexicographic databases (Rojas Díaz and Pérez Sánchez 2019, Rojas Díaz 2020). These prior studies have shown that the indexation of PUs in dictionaries can be modified to make it easier for users to access information. One of the analyses that were absent from the studies by Rojas Díaz and Pérez Sánchez (2019) and Rojas Díaz (forthcoming) is the lexicographic and semantic analysis of both the headwords under which PUs were listed and the PUs’ definitions offered by the dictionaries.

4,285 entries (18,123 word forms) were identified and extracted from the CCDOI, and 16,760 (55,831 word forms) were obtained from the DFDEA. All the lexicographic entries in the CCDOI were idioms, while the entries in the DFDEA included both idioms and collocations. For this analysis, it was decided to apply a series of selection criteria (that will be explained in detail later) to create a subset of data that could be examined in detail. In order to establish said criteria, all word forms were POS-tagged with TreeTagger (Schmid 1994). Then, the tags were homogenized to be readable. Later, all word forms were assigned a semantic tag employing UCREL’s Semantic Analysis System (henceforth USAS). USAS is a POS and semantic tagger, the semantic tags of which are divided into 232 semantic categories based, in turn, on 21 discourse fields identified by McArthur (1981) (Archer *et al.*, 2002). After having all this linguistic data gathered, the data were analyzed, and their most salient features were chosen to be used as criteria to narrow down the sample.

The first criterion that was chosen was the word-form-number criterion (see figure 1). As the frequency analysis was carried out, it was evident that the range between three and five word form idioms amounts to more than 50% of all the entries in the dictionary (see table 1). Therefore, it was decided to compile the database for this study out of entries with three, four, and five word forms.

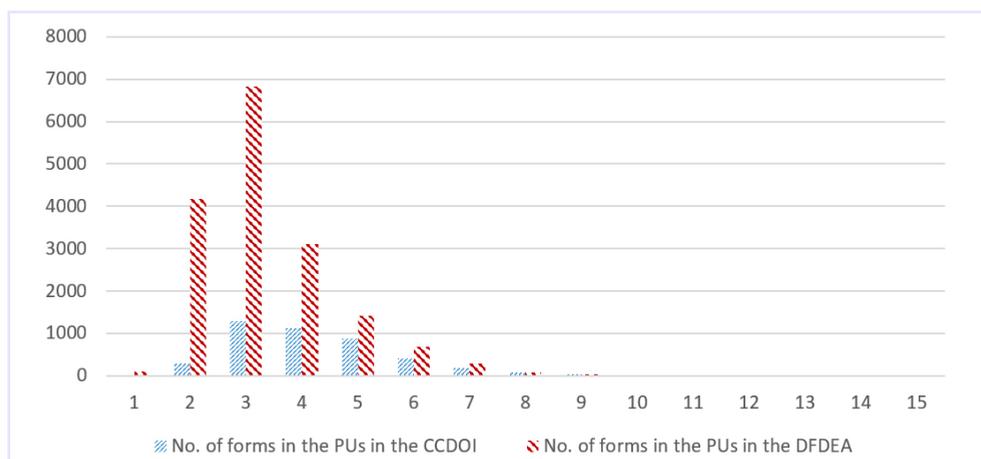


Figure 1. No. of word forms in the entries of the CCDOI and the DFDEA

The descriptive statistical analysis carried out in this study showed that most verbal idioms were on the top of the list in both dictionaries. The CCDOI ranked nominal idioms in second place, while the DFDEA ranked them fourth. Thus, verbal and nominal idioms were selected in order to create the subset for the analysis intended here.

This study also included a general frequency query applied to both dictionaries. This query showed that in the DFDEA, the most common noun was the word «Dios» (God) with 143 occurrences, while in the CCDOI, the most common noun was the word «head» with 66 occurrences. This interesting result can be related to previous works, such as the study carried out by Deignan, Lima *et al.*, (1998), referred to by Deignan (2005), in which the author notes that: «Spanish [...] draws on religion for a number of idiomatic expressions» (Deignan 2005). Similarly, the study by Boers and Stengers (2008), in which they compared two repertoires of idioms, one in English and one in Spanish, identified religion as a recurrent topic for idioms in Spanish. In light of that, it was decided to choose idioms that include religion-related words as a lexical criterion for selecting the sample for the present analysis. In order to isolate the entries containing religion-related words, the USAS' tag set labeled as S9, named 'religion and the supernatural,' was used as a filter.

Table 1. Number of PUs to be analyzed after the selection criteria were applied

Dictionary	No. of entries (PUs)	Number of PUs in the sample		
		PUs after word form filter	PUs after Verbal-Nominal filter	PUs after Semantic field (S9) filter
CCDOI	4,285	3,286	2,387	199
DFDEA	16,760	11,367	4,932	43
Database total				242

After the three previously mentioned filters were applied —*i.e.*, number of word forms, type of PU, and inclusion of religion-related word forms, respectively—, the result was an analysis sample containing 242 idioms of —*i.e.*, 199 in Spanish and 43 in English (see table 1) —. The reason behind such an unbalanced sample is directly related to the corresponding sizes of the original databases. Nevertheless, as this is not intended to be a cross-linguistic analysis, the balance of the sample will not affect the outcome of the study.

Table 2. Examples of PUs extracted for analysis

Nominal idioms		
No. of forms	CCDOI	DFDEA
Three-form PUs	holy of holies	<i>el Santo Grial</i> (‘the Holy Grail’)
Four-form PUs	a baptism of fire	<i>maná caído del cielo</i> (‘mana coming down from heaven’)
Five-form PUs	a devil of a job	<i>el espíritu de la contradicción</i> (‘the spirit of contradiction’)
Verbal idioms		
No. of forms	CCDOI	DFDEA
Three-form PUs	sell your soul	<i>ser el demonio</i> (‘to be the demon’)
Four-form PUs	take something as gospel	<i>poner en los altares</i> (‘to put in the altars’)
Five-form PUs	play merry hell with someone	<i>contar los pelos al diablo</i> (‘to count the hairs of the devil’)

Once the sample was selected (see table 2 for some examples), the entries and some aspects of the macrostructure and the microstructure of the dictionary were analyzed, as will be explained in the following section.

4. ANALYSES AND RESULTS

Since the objective of this study is to extract as much lexicographic and semantic information as possible from the phraseological database, two different analyses were carried out: (i) A lexicographic analysis of the PUs’ headwords and types of definition, and (ii) a semantic analysis of the definitions offered for each PU. Both types of analysis will be explained next.

4.1. Lexicographic analysis

The lexicographic analysis focuses on how PUs are indexed (lemmatization and headwords) in the two dictionaries being studied, as well as the type of information offered by each dictionary regarding the definitions of those PUs. Thus, this analysis

focuses on each work's macrostructure and microstructure. According to Hartmann and James, a dictionary's macrostructure is «the overall list structure which allows the compiler and the user to locate information» (1998). In addition, the microstructure is defined as «the internal design of a reference unit» (1998). Rojas Díaz (forthcoming) has already made a detailed analysis of the megastructure, the macrostructure, and the microstructure of both the CCDOI and the DFDEA. Nevertheless, this article intends to show how certain information offered in both structures could be altered to display phraseological information in an innovative manner.

As explained above, the database for this analysis contains only nominal and verbal idioms (see table 2). In both dictionaries, PUs are indexed under headwords extracted from the word components of the PUs. As a result, the 199 PUs extracted from the DFDEA were indexed under 69 different headwords, while in the case of the CCDOI, the 43 extracted PUs were indexed under 21 headwords. Interestingly, yet expectedly, some of the headwords were not words or expressions related to religion (see table 3).

Table 3. Distribution of headwords in the database

Headwords		
	Religious headwords	Non-religious headwords
CCDOI	hell [12], ghost [3], gospel [3], soul [2], devil [2], baptism [1], prayer [1], angel [1], heaven [1], holy [1], blessing [1], curate [1], ark [1], church [1], fate [1]	gift [3], body [2], tin [2], feast [2], cow [1], belfry [1]
DFDEA	<i>alma</i> [24], <i>gracia</i> [17], <i>demonio</i> [11], <i>santo</i> [10], <i>cielo</i> [8], <i>altar</i> [8], <i>Dios</i> [8], <i>espíritu</i> [7], <i>hostia</i> [5], <i>diablo</i> [5], <i>infierno</i> [4], <i>señor</i> [4], <i>advenimiento</i> [3], <i>ángel</i> [3], <i>Cristo</i> [2], <i> bendición</i> [2], <i>sambenito</i> [2], <i>maná</i> [1], <i>tierra</i> [1], <i>fuego</i> [1], <i>pájaro</i> [1], <i>trapo</i> [2], <i>misa</i> [2], <i>satanás</i> [1], <i>sacramento</i> [1], <i>religión</i> [1], <i>musa</i> [1], <i>ídolo</i> [1], <i>oremus</i> [1], <i>duende</i> [1], <i>vía crucis</i> [1], <i>sagrado</i> [1], <i>ánima</i> [1], <i>teología</i> [1], <i>grial</i> [1], <i>virgen</i> [1], <i>pascua</i> [1], <i>pecado</i> [1] ¹	<i>pena</i> [4], <i>artículo</i> [4], <i>edad</i> [4], <i>mandar</i> [3], <i>tiro</i> [2], <i>presencia</i> [2], <i>año</i> [2], <i>agua</i> [2], <i>auto</i> [2], <i>visita</i> [2], <i>baile</i> [2], <i>piel</i> [2], <i>ir</i> [2], <i>cuento</i> [2], <i>palma</i> [2], <i>pobreza</i> [1], <i>mano</i> [1], <i>canto</i> [1], <i>vida</i> [1], <i>rueda</i> [1], <i>golpe</i> [1], <i>ave</i> [1], <i>cara</i> [1], <i>pelo</i> [1], <i>seno</i> [1], <i>estado</i> [1], <i>placer</i> [1], <i>caza</i> [1], <i>pobre</i> [1], <i>pan</i> [1], <i>voz</i> [1] ²

Table 3 (above) presents the distribution of headwords in the sample, specifying the number of PUs found under the corresponding headword.

1. Translation of religious Spanish headwords in order of appearance: soul, grace, demon, saint, heaven, altar, God, spirit, host, devil, inferno, lord, coming, angel, Christ, blessing, stigma, mana, mass, Satan, sacrament, religion, muse, idol, pray, imp, way of the cross, sacred, soul in purgatory, theology, grail, virgin, Easter, sin.

2. Translation of non-religious Spanish headwords in order of appearance: sorrow, article, age, send, shot, presence, year, water, act, visit, dance, skin, go, story, palm, poverty, hand, Earth, fire, bird, rag, song, life, wheel, hit, bird, face, hair, core, state, pleasure, hunt, poor, bread, voice.

Both dictionaries provide detailed user guidelines. Those guidelines state that the headwords are lemmatized and indexed in alphabetical order. However, there are some differences between the two dictionaries in question regarding the selection of headwords. On the one hand, the CCDOI offers information regarding how headwords are chosen out of the corresponding PUs:

Generally, the word we choose as headword is a noun: for example, *rock the boat* is under the noun **boat** as headword, and *sit on the fence* is under **fence**. If there are two nouns, then the headword is the first noun: for example, *it's raining cats and dogs* is under the headword **cats** and *cost an arm and a leg* under **arm**. If the idiom contains no nouns, then an adjective is chosen: for example, *go easy on someone* is under the headword **easy** and in *black and white* is under **black**. If the idiom contains no nouns or adjectives, then the headword will be either a verb or an adverb. (Sinclair and Moon 1997)

The CCDOI also includes a series of exceptions to the general rules. These rules and their exceptions allow the users who know the components of a certain PU to find the information that they are looking for accurately. On the other hand, unlike the CCDOI, the DFDEA does not include any information regarding the order or priority for choosing headwords in its extensive user guidelines. Both dictionaries incorporate a 'consultation guide' that includes all the PUs indexed. These consultation guides mark each word corresponding to the headword under which each PU is indexed in bold. Although both dictionaries state that the PUs are listed alphabetically under their corresponding headwords, each one of them does it differently.

On the one hand, the CCDOI only considers the PUs' lexical components for applying that alphabetical order. On the other hand, the DFDEA takes the «whole sequence of letters» that composes a PU to index it under the headword (Seco, Andrés *et al.*, 2004). An example of this is presented in table 4.

Table 4. Example of alphabetical indexation of PUs under headwords

	Alphabetical order	
	CCDOI	DFDEA
Headword	Hell	<i>Infierno</i> ('Hell')
PU	all hell breaks loose	<i>al infierno</i> (‘to hell’)
	all hell breaks out	<i>al infierno con</i> (‘to hell with’)
	come hell or high water	<i>bajada a los infiernos</i> (‘descent to hell’)
	through hell and high water	<i>bajar a los infiernos</i> (‘go down to hell’)
	from hell	<i>el quinto infierno</i> (‘the fifth hell’)
	give someone hell	<i>irse al infierno</i> (‘go to hell’)
	go through hell	<i>mandar al infierno</i> (‘send to hell’)

Although both dictionaries (CCDOI and DFDEA) analyzed in the present work provide users with information to help them navigate the dictionary, the CCDOI uses a rather counterintuitive alphabetical indexation method of PUs under the headwords. In contrast, the DFDEA does not include any information about how headwords were chosen. These issues could probably lead to inaccurate searches in dictionaries, as indicated by Atkins and Varantola (1998).

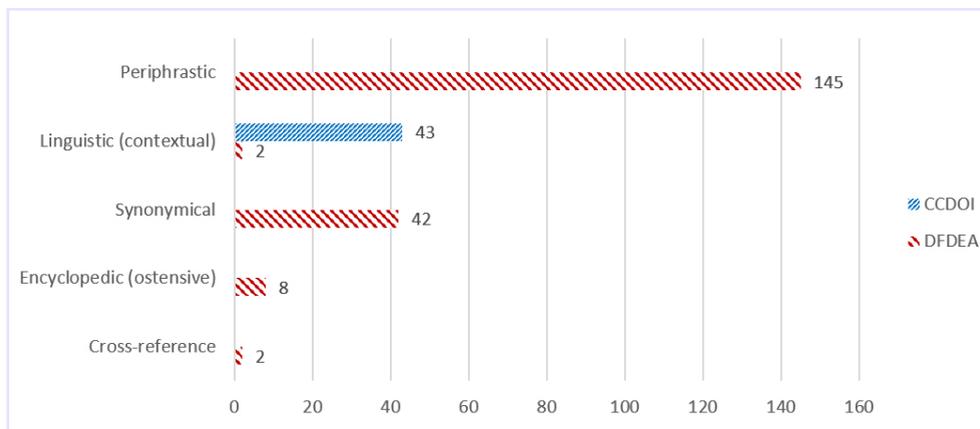


Figure 2. Distribution of lexicographic definitions in the CCDOI and in the DFDEA

With regards to definitions, Porto Dapena defines four different types, namely: (i) encyclopedic, (ii) linguistic or metalinguistic, (iii) synonymic, and (iv) periphrastic (2002) (see table 5). A difference was found concerning the type of definition used by both the CCDOI and the DFDEA (see figure 2). The CCDOI resorts exclusively to what Porto Dapena has named 'linguistic (contextual)' definition (Porto Dapena 2002). In contrast, the DFDEA includes all four types of definitions put forth by Porto Dapena, and cross-references when a PU is a variant of another one.

Table 5. Types of lexicographic definitions found in the CCDOI and the DFDEA following Porto Dapena's classification (2002)

Lexicographic definition	
Type of definition	Definition
Encyclopedic (ostensive)	«Consists of offering the referent instead of the <i>definiens</i> or as a component of the latter.» ³
Linguistic (contextual)	«Those that mainly express the applicability of the word-entry to a certain type of reality.» ⁴
Synonymic	«Those in which the <i>definiens</i> is constituted [...] by a synonym of the <i>definiendum</i> .» ⁵
Periphrastic	This definition has two objectives, either «to answer the question of what is the <i>definiendum</i> ?» or to establish the relationship of the <i>definiendum</i> with other words of the language.» ⁶

As presented in table 5, the combination of periphrastic and synonymic definitions amount to 93.9 % of the definitions used in the DFDEA in this dataset. These two types of definitions are classified into what Porto Dapena (2002) has named 'conceptual definitions,' which, in his own words, are the «most common type of definition used in traditional monolingual lexicography.» Table 6 (below) includes examples of how the different types of definition were used in the dictionaries being analyzed:

Table 6. Examples of definitions in the CCDOI and the DFDEA

Lexicographic definition		
CCDOI		
Entry	Type of definition	Definition
a baptism of fire	Linguistic (contextual)	If your first experience of a new situation is very difficult or unpleasant, you can describe it as a baptism of fire .
DFDEA		
Entry	Type of definition	Definition
<i>alma en pena</i> (*soul in sorrow')	Linguistic (contextual)	<i>Se usa frec en constr de sent comparativo, designando a la pers que anda errante solitaria y melancólica.</i> (*it is used frequently in comparative constructions to designate a person that is alone and melancholic')
<i>la palma del martirio</i> (*the palm of martyrdom')	Encyclopedic (ostensive)	<i>el título de mártir</i> (*the title of martyr')
<i>rendir el alma</i> (*surrender the soul')	Synonymic	<i>morir</i> (*to die')
<i>el quinto infierno</i> (*the fifth hell')	Periphrastic	<i>un lugar muy lejano</i> (*a faraway place')
<i>ánima en pena</i> (*soul in purgatory in sorrow')	Cross-reference	<i>alma en pena</i> (*soul in sorrow')

3. Original in Spanish: «*consiste en colocar el referente en lugar del **definiens** o como componente de este*».

4. Original in Spanish: «*en las que, en principio, se habla de la aplicabilidad de la palabra-entrada a un determinado tipo de realidad*».

5. Original in Spanish: «*en el que el **definiens** está constituido [...] por un sinónimo del definido*».

6. Original in Spanish: «*que intenta responder a la pregunta «qué es el **definiendum**?» o «la relación capaz de establecer el definido con otra palabra de la lengua*».

4.2. Semantic analysis

The second type of analysis presented here deals with the semantic evaluation of the definitions of the PUs in the sample. As shown in table 6 (above) and section 4.1., five different strategies for defining the PUs in question were identified. From a complementary perspective, the 242 PUs that make up the database were grouped into general concepts encompassing the PUs' corresponding definitions. To do so, each definition was analyzed individually and assigned a one-word semantic descriptor corresponding to the general idea or concept that the PU intends to transmit (e.g., dislike, undesirable, unpleasantness), as shown in figure 4. The assigning of such semantic condensations to each PU's definition made it clear that an alternative methodology for indexing PUs in dictionaries was possible. Such methodology had several features in common with works such as the Longman Language Activator (Summers 1993), in English, or the *Diccionario de ideas afines* (Corripio 2007) in Spanish. However, perhaps the main difference between our methodology and the ones applied in said two lexicographic works is that, in our approach, the indexation method could be used in general language dictionaries and in specialized ones. At that point, however, it was impossible to determine if those semantic condensations were related to one another. Therefore, it was necessary to match manually one of the semantic tags included in the tagset offered by UCREL's Semantic Analysis System (USAS) with each semantic condensation. In turn, this semantic tagging allowed for the grouping of said semantic condensations, which would now work as the entries of the dictionary. Henceforth, those semantic condensations of the PUs' definitions will be named 'semantic entries.'

USAS offers two semantic analysis levels, i.e., it groups semantic fields (Archer *et al.*, 2002) under discourse fields (McArthur 1981). In other words, McArthur's discourse fields are hypernyms of the semantic fields included in USAS' tagset. Following this idea, the grouping of semantic entries under hypernym semantic fields would also allow to identify other semantic entries that could be considered or listed as co-hyponyms (see figure 3).

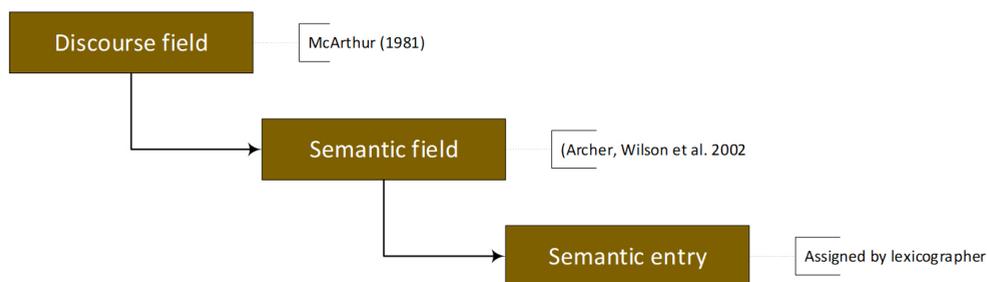


Figure 3. Semantic map: distribution of hypernyms and hyponyms

This type of indexation could also be used as a tool by lexicographers to observe the distribution of lexical or phraseological entries in a semantic or conceptual map.

After assigning one of the tags included in USAS's tagset to each semantic entry, it was possible to identify eight discourse fields in the sample extracted from the CCDOI and fifteen in the one extracted from the DFDEA (see table 7).

Table 7. Distribution of discourse fields in the samples

Discourse field distribution		
Discourse field (alphabetic order)	No of PUs (CCDOI)	No of PUs (DFDEA)
Architecture, housing, and the home	0	1
Emotion	6	34
Food and farming	0	1
General and abstract terms	21	32
Government and public	1	8
Language and communication	0	8
Life and living things	2	23
Movement, location, travel, and transport	0	2
Numbers and measurement	0	7
Psychological actions, states, and processes	4	29
Social actions, states, and processes	5	37
Substances, material, object, and equipment	0	4
The body and the individual	1	6
Time	3	5
World and environment	0	2
Total	43	199

As shown in table 7, the distribution of discourse fields among the definitions of the PUs being analyzed is uneven. One conclusion that might be drawn from these data is that although all the PUs in the study contained at least one religion-related lexical unit, not all of them had a religious connotation. In other words, and taking into consideration that most of the expressions in the study were metaphorical in nature, the use of religious *vehicles* —the metaphor studies' term for the literal meaning of a word or expression— did not necessarily entail a religious *tenor* —the metaphor studies' term for the figurative meaning of a word or expression—. Take, for example, the distribution of the discourse field 'emotion' (see figure 4), in which it becomes apparent that all the PUs included in the figure have at least one religion-related component word. However, all their meanings or referents are related to emotions such as dislike or unpleasantness, and none to religion.

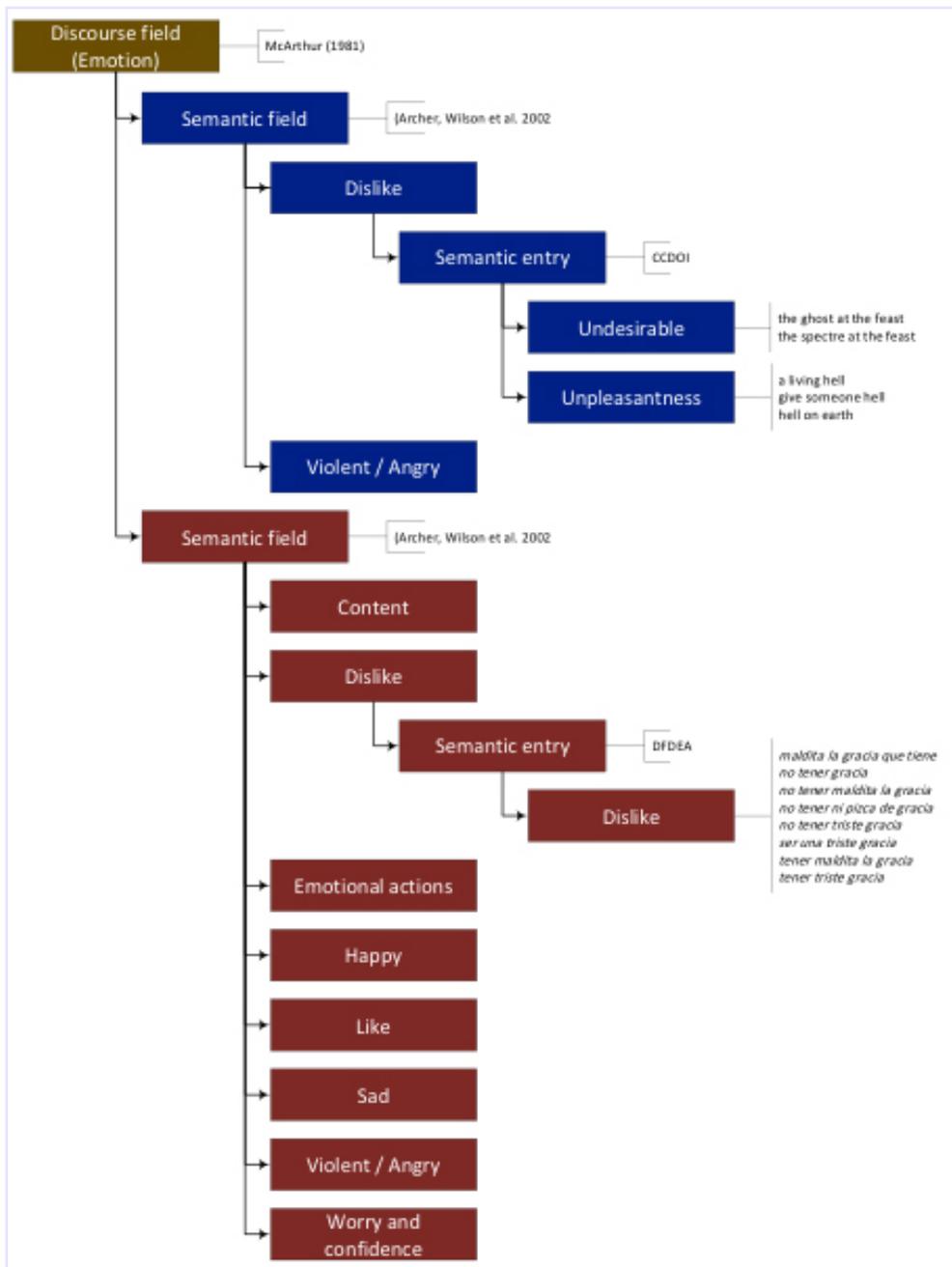


Figure 4. Example of a semantic map in the study database

The use of USAS' tagset allowed us to discover that the semantic field 'religion and the supernatural' —tagged as S9 under the discourse field 'social actions, states, and processes' (S)— was present in 12 PUs, under six different semantic entries (see table 8). Those 12 PUs were found in the Spanish dictionary, and they amounted to 6.03% of the PUs in the Spanish dataset and 4.95% of the whole database. In other words, only 4.95% of the PUs that included at least one religion-related component word referred to a religion-related concept.

Table 8. Semantic classification of religion-related tenors in the database

Religion-related tenors in the database		
Semantic map	PU example and definition	No. of PUs
Discourse field: social actions, states, and processes		37
Semantic field: religion and the supernatural		12
Semantic entry: fate	<i>estar de Dios</i> (‘to be of God’) To be determined by the destiny or the providence	1
Semantic entry: martyrdom	<i>la palma del martirio</i> (‘the palm of martyrdom’) The title of a martyr	2
Semantic entry: specter	<i>ángel de la guarda</i> (‘guardian angel’) Celestial being that has been assigned by God to take care of each person	2
Semantic entry: worship	<i>comerse los santos</i> (‘to eat the saints’) To attend the church frequently	1
Semantic entry: holiness	<i>elegir a los altares</i> (‘raise to the altars’) To beatify or canonize	5

Finally, as explained in Section 2 (above), the entries under study were also analyzed and classified according to the type of semantic relationship established between the PUs' literal and figurative meanings. This analysis and classification —shown in table 9— resulted in metaphor being the most common type of semantic relationship, followed by metaphonymy (Goosens 1990) and metonymy, respectively.

Table 9. Frequency of semantic relationships in each dictionary

Semantic relationship	Frequency		
	CCDOI	DFDEA	Grand total
Metaphor	41	119	160
Metaphonymy	0	37	37
Metonymy	2	27	29
None [literal meaning]	0	11	11
Simile	0	3	3
Metaphor based on another metaphor	0	2	2
Grand Total	43	199	242

Additionally, this analysis and classification also allowed for the identification of what Warren 1992 has named evaluative-attributive features —i.e., the type of evaluation (either positive, negative) that the encoder intentionally transfers from the literal word or expression to the figurative referent—. Thus, entries were assigned a label —positive, negative, or neutral— according to the type of evaluative-attributive feature transferred from the original meaning to the figurative one.

The lexicographic and semantic analyses presented thus far offer a general idea about how an alternative to the traditional semasiological approach could be useful for the indexation of PUs in lexicographic resources, an idea that somehow resembles the onomasiological conception of lexicographic resources that have been put forth in works such as the Longman Language Activator and the *Diccionario de ideas afines*. However, as it will be explained next, the approach presented here allows for a hybrid type of indexation, where both onomasiology and semasiology are combined in order to provide different users with several options for finding the type of information they look for in a lexicographic resource. Thus, in the next section, an example of the application of this approach to the microstructure of an entry, both in paper and electronically, will be presented in detail.

5. INDEXATION PROPOSAL: AN APPLIED EXAMPLE

As pointed out in the previous section, selecting certain PUs containing religion-related word forms does not guarantee to find PUs with religion-related meanings. This conclusion is in line with the idea that if users do not know the exact expression of their interest, finding a definition in a semasiological dictionary would be challenging. In contrast, when users know the concept they want to convey, it is possible to find an expression —or a series of expressions, if there is more than one— that transmit the meaning they want to express.

Take the example of two concepts peripheral to the semantic field of ‘religion and the supernatural’ that were classified by the USAS tagset as part of the semantic field ‘life and the living things,’ namely: ‘alive’ and ‘dead.’

In the database, it was possible to identify two PUs related to ‘life and the living things’ in the CCDOI, while, in the DFDEA, 23 PUs were registered in that discourse field. The following example will show the definitions offered by the dictionaries that were analyzed. The semantic map for those entries is shown in figure 5:

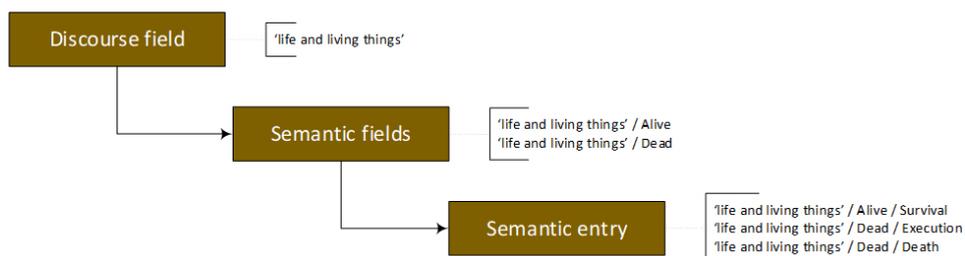


Figure 5. Semantic map of the discourse field ‘life and living things’ in the database

However, as stated before in this section, should a user intend to find PUs related to a certain concept (in this example, the concepts would be ‘death’ and ‘execution,’ in Spanish; and ‘survival,’ in English), the lexicographic resource would have to offer a conceptual index for said user to identify the concept they are looking for. Based on the works by McArthur (1981) and Archer *et al.*, (2002), the index could somehow be similar to the example presented in figure 6.

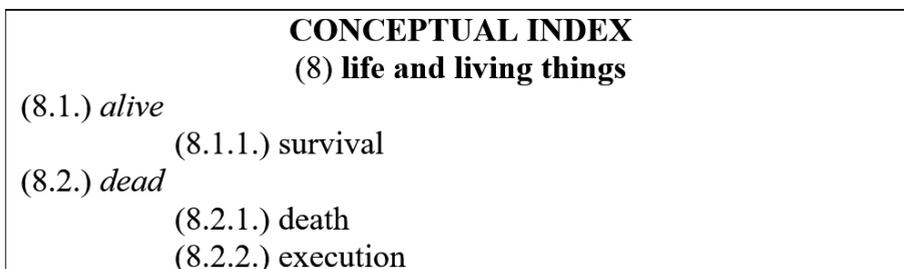


Figure 6. Example of a conceptual index

This conceptual index could be used in an onomasiological dictionary. However, and as suggested in section 2, this model could be used in a hybrid approach in which the monolexical or phrasal semantic entries could be indexed alphabetically and the PUs will be included according to their meaning (concept) and not their word forms. Therefore, this concept index can offer the users the opportunity to look for expressions on co-hyponym conceptions like the case of ‘death’ and ‘execution’ (co-hyponyms) under the semantic field of ‘dead’ (hyperonym). These indexes provide accuracy within a particular semantic or discourse field (see figure 6).

The lexicographic article can contain as much information as the publisher needs or allows. The following (see figure 7) examples in English (from the CCDOI) and in Spanish (from the DFDEA) show how would it look under this hybrid indexation method:

<p>(8.1.1.) SURVIVAL</p> <ul style="list-style-type: none"> ◆ verbal idiom <ul style="list-style-type: none"> ■ definition: if you do something because it is the only way you can earn enough money to buy the basic things that you need to live. You are doing something to... <ul style="list-style-type: none"> ▶ hold body and soul together ⊖ <i>For a while he held body and soul together by working as a migrant laborer.</i> ▶ keep body and soul together ⊖ <i>20-year-old Rafael says he's selling firewood to keep body and soul together.</i> 	<p>(8.2.1.) MUERTE</p> <ul style="list-style-type: none"> ◆ locución verbal <ul style="list-style-type: none"> ■ definición: morir <ul style="list-style-type: none"> ▶ dar su alma ⊕ [literario] <i>Esto dijo Don Victorino..., dando el alma a Dios.</i> ▶ exhalar su espíritu ⊖ [literario] <i>Y Jesús, clamando de nuevo con gran voz, entregó su espíritu.</i> ▶ irse al infierno ⊖ [coloquial] <i>Si os movéis de aquí, si aparece la policía o si me caigo al suelo, hago estallar este petardo y os vais todos al infierno.</i>
--	---

Figure 7. Example of a semantic entry in (1) English and (2) Spanish for a paper lexicographic resource

As shown in figure 7, a semantic entry could include the concept indexed, the type of PU(s) that would express that concept, and the definition that would apply to all the PUs in the entry. Additionally, the entry could also include a mark for expressing the attributive-evaluative feature (see section 2) [positive (), negative (), and neutral ()] — associated with the PUs use, a diaphasic label (e.g., formal, informal, colloquial, literary, etcetera), and an eventual diatopic label (e.g., British English).

Likewise, this indexation approach can be applied to electronic resources, namely electronic or web-based databases. An example of such an application can be seen in figure 8, which presents a query carried out through an application for the consultation of databases called *onofrasikon*, which is still under development at NHH Norwegian School of Economics.

Semantic entry

SEMANTIC ENTRY

Semantic Entry: Survive

Discourse field: life and living things

Semantic field: Alive

Entry: hold body and soul together

Type of PU: Verbal idiom

Sem. Relationship: Metonymy Attr. Ev. Feature: Neutral

Diatopic: Diaphasic:

Morphosyntactic Pat.: V N Conj N Adv

Morphosemantic Pat.: M B Z S S

Definition: If you do something to keep body and soul together or to hold body and soul together, you do it because it is the only way you can earn enough money to buy the basic things that you need to live.

NHH EFMD EQUIS ACCREDITED CEMS ONOFRASIKON

Ok

Figure 8. Example of a semantic entry in English and Spanish in an electronic lexicographic resource

As observed in both figures 7 and 8 (above), the indexation methodology presented here can be applied to either physical or electronic lexicographic resources. Nevertheless, it is necessary to continue reflecting on how to apply the long-standing dialogue between phraseology and lexicography to an effective lexicographic representation of PUs, which is an enterprise that has already been undertaken by authors like Fraile Vicente (2008) and Nuccorini (2020), among several others.

6. CONCLUSIONS

This study shows that PUs can be indexed using a different approach from the traditional, either purely semasiological or onomasiological ones, for the indexation and creation of lexicographic entries. Such an approach entails the semantic analysis and tagging of PUs in order to determine their hypernym concept (semantic field), which would serve as an anchor for the indexation of PUs under a hybrid approach.

The semantic analysis of PUs prior to their indexation under the hybrid approach proposed here is one of its most important contributions. That analysis not only allows for the identification of semantic fields, the grouping of PUs under those semantic fields, and a more effective and useful indexation, but it also provides important information such as the semantic relationships between the original and the figurative meanings of PUs, or the evaluative-attributive features being transferred from the original referents to the figurative ones.

In summary, the approach proposed here combines a great deal of information extracted from PUs by following phraseological theories of analysis and applying that information to the indexation and elaboration of phraseological entries in lexicographic resources, thus consummating the long-praised ‘scientific marriage’ between phraseology and lexicography.

BIBLIOGRAPHY

- ALONSO RAMOS, M., Ed. (2006). «Diccionarios y Fraseología». Anexos de Revista de Lexicografía.
- ARCHER, D., A. WILSON and P. RAYSON (2002). *Introduction to the USAS category system*.
- ATKINS, B. T. S. and M. RUNDSELL (2008). *The Oxford Guide to Practical Lexicography*. New York, Oxford University Press.
- ATKINS, B. T. S. and K. VARANTOLA (1998). «Language Learners». *Using dictionaries: The Final Report on the EURALEX / AILA Research Project on Dictionary Use Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*. B. T. S. Atkins. Tübingen, Max Niemayer Verlag.
- BOERS, F. and H. STENGERS (2008). «A quantitative comparison of the English and Spanish repertoires of figurative idioms». *Cognitive linguistic approaches to teaching vocabulary and phraseology* F. Boers and S. Lindstromberg. Berlin, Mouton de Gruyter: 355-373.

- BOSQUE, I. (2017). «Posibilidades de un diccionario onomasiológico de modismos para estudiantes de ELE». *La adquisición de la lengua española: aprendizaje, enseñanza, evaluación. Estudios en Homenaje a Marta Baralo Ottonello*. M. C. Ainciburu and C. Fernández Silva. Buenos Aires, Autores de Argentina: 26-37.
- BUENDÍA CASTRO, M. and P. FABER (2015). «Phraseological units in English-Spanish legal dictionaries: a comparative study». *Fachsprache* (3-4): 161-175.
- BURGER, H. (1998). *Phraseologie: Eine Einführung am Beispiel des Deutschen*. Berlin, Schmidt.
- CASARES, J. (1950). *Introducción a la lexicografía moderna*. Madrid, CSIC.
- CHAFE, W. (1968). «Idiomacity as an Anomaly in the Chomskyan Paradigm». *Foundations of Language* 4: 109-127.
- CORPAS PASTOR, G. (1996). *Manual de fraseología española*. Madrid, Gredos.
- CORPAS PASTOR, G. (2000). *Acerca de la (in)traducibilidad de la fraseología. Las lenguas de Europa, estudios de fraseología, fraseografía y traducción*. G. Corpas Pastor. Granada, Comares: 438-522.
- CORRIPIO, F., Ed. (2007). *Diccionario de ideas afines*. Barcelona, Herder.
- CRUSE, A. (2000). *Meaning in language*. Oxford, Oxford University Press.
- DEIGNAN, A. (2005). *Metaphor and Corpus Linguistics*. Amsterdam, John Benjamins Publishing Company.
- DEIGNAN, A., A. LIMA and E. LOPEZ MORA (1998). *Metaphor, culture and the classroom. IATEFL 1998 Manchester Conference Selections*. P. Grundy. Whistable, IATEFL.
- DOBROVOL'SKIJ, D. and E. PIIRAINEN (2005). *Figurative Language: Cross-cultural and Cross-linguistic Perspectives*. Amsterdam, Elsevier.
- FRAILE VICENTE, E. (2008). «Sugerencias para mejorar el tratamiento de las expresiones idiomáticas en los diccionarios fraseológicos en inglés y español». *TRANS Revista de traductología* (12): 123-148.
- GARCÍA-PAGE, M. (2008). *Introducción a la fraseología española*. Barcelona, Anthropos.
- GONZÁLEZ, M. I. (2006). *La definición lexicográfica de las unidades fraseológicas: la aplicación de modelos formales. Diccionarios y fraseología*. M. Alonso. La Coruña, Universidade da Coruña: 221-233.
- GOOSENS, L. (1990). «Metaphonymy: the Interaction of Metaphor and Metonymy in Expressions for Linguistic Action». *Cognitive Linguistics* 1: 323-340.
- GRČIĆ SIMEUNOVIĆ, L. and P. DE SANTIAGO (2016). *Semantic approach to Phraseological Patterns in Karstology. Proceedings of the XVII Euralex International Congress*, Ivane Javakhishvili Tbilisi State University.
- GRIES, S. (2008). *Phraseology and linguistic theory: a brief survey. Phraseology: an interdisciplinary perspective*. S. Granger and F. Meunier. Amsterdam, John Benjamins: 6.
- HARTMANN, R. R. K. and G. JAMES (1998). *Dictionary of Lexicography*. New York, Routledge.
- KOCJANČIČ, P. (2004). «Acerca de la macroestructura y la microestructura en el diccionario bilíngüe.» *Verba Hispanica* 12(1): 171-185.
- LEROYER, P. (2006). «Dealing with phraseology in business dictionaries: focus on dictionary functions - not phrases». *Linguistik online* 27(2): 183-194.
- McARTHUR, T. (1981). *Longman Lexicon of Contemporary English*. Harlow, Longman.
- MEL'ČUK, I. (2012). *Phraseology in the language, in the dictionary, and in the computer. Yearbook of Phraseology*. K. Kuiper. New York, De Gruyter Mouton. 3: 31-56.
- MELLADO BLANCO, C., Ed. (2008). *Colocaciones y fraseología en los diccionarios*. Frankfurt am Main, Peter Lang.

- MOON, R. (2008). «Sinclair, phraseology, and lexicography». *International Journal of Lexicography* 21(3): 243-254.
- NUCCORINI, S. (2020). «Digging into a Thesaurus for Treasures: Phraseological Conceptual Maps in the Longman Language Activator». *Contrastive Phraseology: Languages and Cultures in Comparison*. P. Cotta Ramusino and F. Mollica. Newcastle upon Tyne, Cambridge Scholars Publishing: 349-375.
- ORTEGA OJEDA, G. and M. I. GONZÁLEZ AGUILAR (2008). *La técnica fraseográfica: el DRAE-2001 frente al DEA-1999. Colocaciones y fraseología en los diccionarios*. C. Mellado. Frankfurt am Main, Peter Lang: 232-245.
- PAQUOT, M. (2015). *Lexicography and Phraseology. The Cambridge Handbook of English Corpus Linguistics*. D. Biber and R. Reppen. Cambridge, Cambridge University press: 460-477.
- PORTO DAPENA, J.-Á. (2002). *Manual de Técnica Lexicográfica*. Madrid, Arco/Libros.
- ROJAS DÍAZ, J. L. (2020). «“From head to toe”: a lexical, semantic, and morphosyntactic study of idioms in phraseological dictionaries in English and Spanish». *MonTI* (Special issue 6), 287-326. doi: <http://dx.doi.org/10.6035/MonTI.2020.ne6.9>
- ROJAS DÍAZ, J. L. and J. M. PÉREZ SÁNCHEZ (2019). «You Took the Word Out of My Mouth»: A Morphosyntactic and Semantic Analysis of a Phraseological Lexicon of Colombian Spanish. *Computational and Corpus-Based Phraseology - Third International Conference, Europhras 2019 Malaga, Spain, September 25–27, 2019 Proceedings*. G. M. Corpas Pastor, Ruslan, Springer: 375-390.
- SCHMID, H. (1994). «Probabilistic Part-of-Speech Tagging Using Decision Trees». *Proceedings of the International Conference on New Methods in Language Processing*, Manchester.
- SECO, M., O. ANDRÉS and G. RAMOS, Eds. (2004). *Diccionario fraseológico documentado del español actual*. Madrid, Aguilar.
- SIEPMANN, D. (2008). «Phraseology in learners' dictionaries: What, where and how?». *Phraseology in Foreign Language Learning and Teaching*. F. G. Meunier, Sylviane. Amsterdam, John Benjamins Publishing Company: 185-202.
- SINCLAIR, J. and R. MOON, Eds. (1997). *Collins COBUILD Dictionary of Idioms*. Glasgow, Harper-Collins publishers.
- SUMMERS, D., Ed. (1993). *Longman language activator*. Harlow, Longman.
- TORIJANO, J. A. and M. Á. RECIO (2019). «Translating Emotional Phraseology: A Case Study». *Computational and Corpus-Based Phraseology*. R. Mitkov and G. Corpas Pastor, Springer: 391-403.
- VEISBERGS, A. (2020). «Phraseology in General Bilingual Dictionaries: Idioms as Equivalents of Single Words». *Contrastive Phraseology: Languages and Culture in Comparison*. P. Cotta Ramusino and F. Mollica. Newcastle upon Tyne, Cambridge Scholars Publishing: 331-343.