

María Auxiliadora Barrios Rodríguez

## DIRETES, A SPANISH MONOLINGUAL DICTIONARY BASED ON LEXICAL-SEMANTIC RELATIONS

**Abstract** *Diretes* is a Spanish monolingual e-dictionary based on Lexical-Semantic Relations which are formalized by Lexical Functions, a formal tool explored within the Meaning-Text Theory. This dictionary consists of a relational database which aims to reflect the cognitive links of the lexicon through a network of semantic and lexical associations. Currently it contains more than 100,000 collocations and semantic relations, and includes dictionary entries for simple words, compound words, idioms and formulaic expressions. The majority of the definitions of each entry consist of two parts: a minimal definition (written with basic words) and an expanded definition (which collects world knowledge data). The structure of the entries of the dictionary for formulaic expressions, however, are more complex as they also include a pragmatic function, a prototypical scenario, and certain other specific information. Semantic relations include synonymy, antonymy, hyponymy and other productive relations not utilized to date in other dictionaries. Lexical Functions allow the user of the dictionary to consult the entries, including idioms, through both onomasiological and semasiological approaches. This paper summarizes the purpose of the creation of *Diretes* and the idiosyncrasies of this dictionary.

**Keywords** relational database; lexical functions; collocations; semantic relations; idioms; formulaic expressions

### 1. Why a New Spanish Monolingual Dictionary?

Spanish Lexicography is an ancient discipline with its inception in 1492 via the first *Spanish-Latin Lexicon* of Antonio Nebrija. For a century, all the published dictionaries in Spain were bilingual, amongst them the *Spanish-Arabic Lexicon* of Pedro Alcalá (1505), the *Spanish-French Lexicon* of Nebrija (1516), and the *Spanish-Italian Lexicon* of Casas (1570).<sup>1</sup> Particularly relevant were the bilingual Lexicons developed by the Spanish missionary in America, some of them under the influence of the book of grammar and the dictionary of Nebrija, such as the *Lexicon of the Mexican Language* of Andrés de Olmedo (1547) or that from de Tzendal and other languages, de Francisco de Cepeda (1560).

The first two Spanish monolingual dictionaries were published in 1611 by Covarrubias and Rosal. Throughout the seventeenth century several bilingual dictionaries were published with no other Spanish monolingual dictionary coming into existence until 1729 when the Spanish Royal Academy published the first of twenty eight different versions of its Spanish monolingual dictionary. After almost three centuries it can be asserted without doubt that this dictionary has become the most popular dictionary

<sup>1</sup> All the dictionaries mentioned in this paragraph can be referenced in the academic website <https://apps.rae.es/nttle/SrvltGUISalirNtle>.

amongst Spanish speakers both in Spain and abroad. Its influence was such that over two centuries all of the Spanish monolingual dictionaries were a slightly modified copy of the last version of this academic dictionary.

Over the last century and a half, however, some notable monolingual dictionaries have been published by lexicographers with their own particular layout and original definitions: the *DCR* by Cuervo, which is focused on grammatical features;<sup>2</sup> the *DCECH* by Corominas and Pascual, which is a historical dictionary;<sup>3</sup> the *DUE* by María Moliner, which was the first standard dictionary containing some collocations;<sup>4</sup> the *DEA* by Seco et al., the authors of which applied the criteria of substitution when writing the definition;<sup>5</sup> and the two combinatorial dictionaries *Redes* and *Práctico*, by Bosque (2004; 2006), focused on collocations, which do not contain definitions but meanings based on semantic relations: their content and approach see them unique in the current landscape of international lexicography.<sup>6</sup>

Despite the five aforementioned dictionaries breaking new ground there are some potential lexicographic innovations that have not yet been applied: neither Spanish nor other languages dictionaries have yet to be completed following the Meaning-Text Theory criteria. This linguistic model has been viewed as the “only really new thing that has appeared in recent dictionaries” (Simone, 2016, p. 29). The Spanish lexicographic tradition of the two previously mentioned centuries formed the basis for these innovative dictionaries of the last century and a half. By the same token, we take into consideration the fact that the richness of the unique Spanish combinatorial dictionaries, *Redes* and *Práctico*, has significantly facilitated the creation of *Diretes*. The reason being our dictionary is based on data derived from various sources, the most important being the collocations and semantic structures present in both dictionaries.<sup>7</sup> Due to their great depth, *Diretes* aspires to be the first monolingual Spanish dictionary to be completed within the MTT model, although, as will be shown, we have added some particularities which are not present in canonical dictionaries of the model.

The next two sub-sections concentrate on demonstrating the relevance of this model when compiling dictionaries (§ 1.1), and particularly why a dictionary such as *Diretes* would be necessary (§ 1.2). Section 2 concentrates on the definitions in *Diretes*. Section 3 presents the methodology followed when working with semantic relations. Section 4 summarizes the methodology for the formalization of the collocations. Section 5 portrays the novelties around the idioms. And finally, Section 6 presents the conclusions.

<sup>2</sup> The *Diccionario de construcción y régimen de la lengua castellana* by Rufino Cuervo (1886), <https://www.rae.es/archivo-digital/diccionario-de-construccion-y-regimen-de-la-lengua-castellana>.

<sup>3</sup> The *Diccionario crítico etimológico castellano e hispánico* by Joan Corominas and José A. Pascual (1980).

<sup>4</sup> The *Diccionario de Uso del Español* by María Moliner (1966).

<sup>5</sup> The *Diccionario del Español Actual*, by Manuel Seco, Olimpia Andrés, Gabino Ramos (1999).

<sup>6</sup> Some other non-academic lexicography has been developed over the last thirty years (for further information, see Bosque & Barrios, 2017).

<sup>7</sup> Lack of space prevents us from mentioning all the sources we use. Most of them are open access corpora rich in collocational information, such as the esTenTen, a Spanish corpus from the Sketch Engine, <https://www.sketchengine.eu/estenten-spanish-corpus/>.

## 1.1 Dictionaries Within the Meaning-Text Theory

The Meaning-Text Theory (hereinafter MTT) is a lexicological and lexicographic model which proposes a systematic methodology for dictionaries based on formal tools referred to as Lexical Functions (Mel'čuk 1984; 1996; 2014; 2023; Mel'čuk et al, 1995; 2007; 2018; Apresjan 1992; 2009; Polguère 2000; 2003; 2007). Lexical Functions (LFs) facilitate the classification of lexical relations by means of the formalization not only of collocations but also of paradigmatic relations, such as synonymy, antonymy, and semantic derivatives. For instance, within the MTT the semantic relation between *grain* and *wheat* is similar to that between *drop* and *water*: both are associated to one LF called Sing, which means '(the minimal) part of'. This LF allows us to formalize both collocations as shown in (1). Similarly used is the relation between *flock* and *sheep*, and *string* and *garlic*, both of them related to an LF called Mult, which means 'group of', as shown in (2):

1. Sing(wheat) = grain (of); Sing(water) = drop (of)
2. Mult(sheep) = flock (of); Mult(garlic) = string (of)

The first MTT dictionary was a monolingual French print dictionary (Mel'čuk, 1984), which was followed by the *Lexique actif du français* (Mel'čuk & Polguère, 2007), a printed Lexicon containing around 20,000 semantic relations and collocations. The leap from the printed dictionary to the on-line dictionary for the French language was made with the *Réseau Lexical du Français* (Polguère, 2014), accessible via website (<https://spiderlex.atilf.fr/fr>), which was based on the French dictionary *Dicouèbe* (<http://olst.ling.umontreal.ca/dicouebe/index.php>).

The Spanish language relies on the *DICE*, an on-line dictionary, the vocabulary of which was focused on nouns of emotions in its first version (Alonso Ramos, 2004). The last version contains collocations, with its entries being nouns from all lexical fields (<http://www.dicesp.com/paginas>).

There is also an automatic Russian-English translator (<http://cl.iitp.ru/etap3>) based on LFs and some dictionaries for Russian. Although most of the dictionaries developed within the MTT (as the recently appointed) make use of the LFs, there are a few exceptions. A case in point *Russian Synonyms Dictionary* which was directed by Jury Apresjan (2018) as well as a monolingual Russian dictionary, *The Active Dictionary of Russian* (an ongoing project of the Russian Language Institute also directed by Apresjan). The last one includes the definitions of all senses of words, the government patterns, the constructions in which the word is used, and certain other linguistic information.<sup>8</sup> The definitions are "systemic, comprehensive, non-tautological, and explanatory of co-occurrence properties of lexical items" (Apresjan, 2011).

Particularly prolific are the applications of the MTT to terminology (see, amongst others, L'Homme, 2012; 2018; L'Homme et al, 2016), such as the *Dictionnaire fondamental de l'informatique et de l'Internet* ([https://olst.ling.umontreal.ca/?page\\_id=53](https://olst.ling.umontreal.ca/?page_id=53)), a legal

<sup>8</sup> I would like to express my gratitude to Igor Boguslavsky for the oral information about this dictionary.

terminology database ([https://olst.ling.umontreal.ca/?page\\_id=1357](https://olst.ling.umontreal.ca/?page_id=1357)), the *Dictionnaire fondamental de l'environnement* ([https://olst.ling.umontreal.ca/?page\\_id=214](https://olst.ling.umontreal.ca/?page_id=214)); and two projects that apply the MTT and the semantic frames to the terminology: the *Framed Dicoenviro* ([https://olst.ling.umontreal.ca/?page\\_id=2364](https://olst.ling.umontreal.ca/?page_id=2364)), and the *Framed DicoInfo* ([https://olst.ling.umontreal.ca/?page\\_id=3243](https://olst.ling.umontreal.ca/?page_id=3243)).

## 1.2 Innovations in Diretes

*Diretes* is a Spanish monolingual e-dictionary based on Lexical-Semantic Relations which are formalized by LFs. The name of the dictionary is an acronym of *Dlccionario RETicular del ESpañol* (Spanish Reticular Dictionary) which aims to express the deliberate intention of reflecting the cognitive links of the lexicon through a network of semantic and lexical associations (Polguère, 2014; Barrios, 2020). As with most of the modern dictionaries, *Diretes* is a relational database (Atkins & Rundell, 2008). Currently it contains more than 100,000 collocations and semantic relations (Barrios, 2022a). Lexical inheritance (Mel'čuk & Wanner, 1996; Barrios, 2010) provided the base of a third part of the collocations, meaning that the formalization of more than 25,000 collocations in the database was performed automatically (as we will see in Section 4). A pilot test of *Diretes* is available at <http://diretes.es/>.

The main particularities of the dictionary, amongst others, are the following:

- a. *Diretes* includes dictionary entries not only for words but also for idioms, compound words and formulaic expressions.
- b. Definitions of words, idioms and compound words consist of two parts: a minimal definition (written in basic words) and an expanded definition (which collects world knowledge data). This structure allows us to offer concomitantly definitions corresponding both to monolingual dictionaries of Spanish and to encyclopedic dictionaries. In fact, minimal definitions are linguistic definitions based on semantics whilst expanded definitions are based on data from knowledge of the world (Roche, 2015). Both definitions provide ontological features because the minimal definition relies on ontological relations by means of the semantic label, whilst the expanded one relies on them by means of the expanded information. The latter is particularly useful for Terminology, Second Language Teaching and other applied fields of linguistics (Frawley, 1997).<sup>9</sup>
- c. The structure of the entries of the dictionary for formulaic expressions are more complex as they also include a pragmatic function (Barrios, 2022b), a prototypical scenario (Blanco, 2013), and certain other information. For instance, for the Spanish expression *¡Pide un deseo!* (lit. *Make a wish!*) the entry presents the following data: a) the pragmatic function (to generate empathy); b) typical scenarios (such as after blowing out the candles on the cake); c) the attitude of the speaker (more a childish game than of a

<sup>9</sup> In *Diretes* we have yet to work with terminology, with the exception being for some words from the Geographic field.

superstition); d) the expectation with respect to the listener (closing the eyes while mentally making a wish). Formulaic expressions include pragmatemes (Mel'čuk, 1995; Blanco, 2013), stereotypical speech acts (Kauffer, 2018) and productive structures (Ovejas, 2022).

- d. Semantic relations not only include synonymy, antonymy and hyponymy, but also some productive relations not utilized so far in other dictionaries such as: a) “of an animal”, which allows us to reflect the relationship between *hand/claw*, *mouth/snout*, *lip/beak*, *arm/thigh*, etc.; b) “profession or work”, that shows the relationship between pairs of words such as *hand/chiropractor*, *feet/runner*, *bread/baker*, *wood/carpenter*, etc.
- e. Collocations are collected by Lexical Functions. This means that, considering a simple word, such as *hand*, *Diretes* offers all the verbs that combine with *hand* organized in a semantic way: a) verbs expressing the idea that ‘to cause (the hand) to not work anymore’, such as *tie*, *handcuff*; b) verbs expressing the idea ‘it does not work properly’, such as *get stiff*, *shake*; c) verbs expressing actions with a hand in a figurative sense, such as *hold out your hand*, *offer your hand*, etc.
- f. Idioms are collected following both onomasiological and semasiological approaches: for instance, *malas lenguas* (‘gossiping tongues, evil tongues’) are attached to the noun *falsedad* (‘falseness’); similarly *becerro de oro* (‘golden calf’) and *buenas palabras* (lit. good words). Consequently, if someone does not remember an idiom related to *falsedad*, this person can go from the *falsedad* entry to consulting all entries of the three idioms.

## 2. Definitions

When writing definitions in the dictionary we considered two preliminary ideas. The first is the habitual risk of circularity in definitions. In order to avoid this, we were inspired by a previous work which attempts to show the advantages of the so-called *minimal and inverse definitions*: “*minimal* because I used to write only the minimal set of features necessary to understand the meaning of the word, and *inverse* because I wanted people to be able to decipher the word when reading the meaning” (Barrios, 2019, p. 196). This kind of definition must be written within the Natural Semantic Metalanguage model (NSM) (Goddard, 2012; 2016; Goddard & Wierzbicka, 2014; Wierzbicka, 1985; 1996). In particular, we resorted to two of their most relevant tools: primes (universal and simple words that cannot be decomposed) and molecules (universal words that can be explained by primes). As the methodological requirement for both tools exceed the temporal possibilities of our project in *Diretes* we did not use primes and molecules but simple words, meaning words that allow us to avoid circularity (for instance, if we write that a *table* is an *object used to put things, to work or to eat*, we consider that the meaning of the words *object*, *to use*, *to put*, *thing*, *to work* and *to eat* do not include the meaning of the word *table*).

The second point regarding definitions relates to the fact that we consider the meaning of a word includes not only semantic features but also world knowledge features: “I



no longer believe, again contrary to an age-old doctrine in word semantics, that it is possible and useful to separate core meaning from residual meaning and to consider the core as more important” (Kövecses, 1993, p. 247). As cultural features and data collected from referents can be relevant when attempting to understand the meaning of many words, we made the decision to divide the *Diretes*’s definitions into two parts: the first being a linguistic definition (written with basic words, as explained above); the second attempting to reflect cognitive links caused by extra-linguistic factors. We aim to illustrate both parts of these definitions in § 2.1 and § 2.2.

## 2.1 Definitions of Simple Words, Compound Words, and Idioms

Simple words, compound words, and idioms are described by means of minimal and expanded definitions, as explained above.

Illustration 1 shows both parts of the definition for the idiom *hombre anuncio* (sandwich-board man). Minimal definition (*definición mínima*) says (we translate to English the Spanish definition): “Person announcing something with a sign on the chest and another one on the back.” In this case, the minimal definition is considerably longer because this is the only way we found to present the meaning of this expression: it is necessary to include the features of *person*, *sign*, *chest* and *back*. Expanded definition (*definición expandida*) says: “This job is more than two centuries old, although in some cities, such as Madrid, it is prohibited (in this case, since 2008).” Accordingly, the second part of the definition adds some cultural features related to the use of this word in Spain (as *Diretes* describes the uses of Spanish from Spain):

### Hombre-anuncio

#### Acepción núm 1:

**Definición mínima:** Persona que anuncia algo, con un cartel sobre el pecho y otro sobre la espalda.

**Definición expandida:** Es un trabajo con más de dos siglos de antigüedad, aunque en algunas ciudades, como Madrid, está prohibido (en este caso, desde 2008).

**Etiqueta semántica:** Persona con algún vínculo social o profesional.

**Fig. 1:** Definition for *hombre anuncio* (‘sandwich-board man’) in *Diretes*

Under the expanded definition, the entry contains the semantic label (*etiqueta semántica*), a type of hyperonym which belongs to a hierarchy of some nine levels of semantic labels. That particular hierarchy structures all the entries of *Diretes*. In this unequivocal case, the semantic label is “Person with a social or professional link.” Although it has not yet been applied in *Diretes*, the entries as in that for *hombre anuncio* will be illustrated through one picture to illustrate a more explicit meaning.

Before closing this section, it is necessary to mention a few important issues related to the definitions: a) the definitions have not been partially extracted automatically, although, in future work, we will attempt to automate the inheritance of shared semantic features; b) no particular corpora have been used to extract the encyclopaedic definitions, except for the data from Bosque (2004; 2006) and the sources mentioned in footnote 7.

## 2.2 Description for Formulaic Expressions

Formulaic expressions of *Diretes* are collected in a separate part of the database as they do not necessarily contain any meaning except that of a communicative intention. In fact, formulaic expressions portray information about a pragmatic function: for instance, when someone says *¡a buenas horas, mangas verdes!* (lit. ¡at good times, green sleeves!), the listener should understand that this expression is attached to the pragmatic function of reproaching: the speaker thinks the listener arrives too late to do what was supposed to be done, resembling an English speaker who says, in a similar situation, *too little too late!*

Regarding the formulaic expressions, the entry of *Diretes* contains a minimal and an expanded definition (in a similar way to the one explained in the §2.1). Minimal definition for this expression is translated to English: “A person who is in one place tells another person who is arriving at that time at that place that they are too late to do what they had to do together” (see Figure 2). Expanded definition adds some cultural information: “During the period of the Catholic Monarchs, there was a guard corps (called “Holy Brotherhood”) who wore green sleeved shirts as part of their uniform and had a reputation for being late when given notice.”

As this kind of expression does not have meaning, it makes no sense to assign a semantic label but a pragmatic function (such as to reproach, *reprochar*), as shown in Figure 2 (see *función pragmática*). In order to understand this function, these kinds of entries also offer a natural context with an example of use (see *situación típica* (typical situation) in Figure 2), for instance: “There is a family dinner. Several people in the family have arranged to meet a few hours before the dinner to prepare everything, but one of them arrives when they have already finished, and someone says to him: “A buenas horas, mangas verdes”.

## ¡A buenas horas, mangas verdes!

### Acepción núm 1:

**Definición mínima:** Una persona que está en un lugar, le dice eso a otra persona que está llegando en ese momento para que entienda que llega demasiado tarde para hacer lo que tenían que hacer juntos.

**Definición expandida:** La expresión tiene su origen en la época de los Reyes Católicos, quienes reformaron un cuerpo, llamado Santa Hermandad, que tenía como objetivo perseguir a los delincuentes. Llevaban unas camisas con mangas verdes y solían llegar cuando ya se habían dispersado los causantes del altercado.

**Función pragmática:** Reprochar

**Situación típica:** Hay una cena familiar. Varias personas de la familia han quedado unas horas antes de la cena para prepararlo todo, pero uno de ellos llega cuando ya han terminado de preparar toda la cena, y le dicen: ¡A buenas horas, mangas verdes!

Fig. 2: Entry for the expression ¡A buenas horas, mangas verdes! (equivalent to *Too little too late*)

### 3. Semantic Relations

Semantic Relations in *Diretes* collect more than the usual relations (hyponymy, synonymy, etc.) As has been revealed, LFs do not cover all the semantic relations, hence we created new functions for them. Figure 3 illustrates this concept with a small group of words associated to *hand*:

Id-FA	Id-Argumento	rY	Id-Valor	He
⊕ _ABROGATIO	mano ( s. f. sg.) 1 - Extremidad superior		muñón ( s. m. SA) 1 - Sin asignar	No
⊕ _ABROGATUS	mano ( s. f. sg.) 1 - Extremidad superior		manco (adj. c. c.) 1 - Defectuoso	No
⊕ _AFIJOIDE	mano ( s. f. sg.) 1 - Extremidad superior		quiro- (morf. -) 1 - Parte de una palabra	No
⊕ _ANIMAL	mano ( s. f. sg.) 1 - Extremidad superior		garra ( s. m. sg.) 1 - Parte de un animal	No
⊕ _ANIMAL	mano ( s. f. sg.) 1 - Extremidad superior		pezuña ( s. m. sg.) 1 - Parte de un animal	No
⊕ _ANIMAL	mano ( s. f. sg.) 1 - Extremidad superior		zarpa ( s. m. sg.) 1 - Parte de un animal	No
⊕ _ARS	mano ( s. f. sg.) 1 - Extremidad superior		prestidigitación ( s. f. SA) 1 - Ocupación	No
⊕ _ARS	mano ( s. f. sg.) 1 - Extremidad superior		quiropática (SA f. sg.) 1 - Profesión	No
⊕ _ARTIFEX	mano ( s. f. sg.) 1 - Extremidad superior		prestidigitador (adj. c. c.) 1 - Persona	No
⊕ _ARTIFEX	mano ( s. f. sg.) 1 - Extremidad superior		quiropático ( s. m. sg.) 1 - Profesional	No
⊕ _AUCTUS	mano ( s. f. sg.) 1 - Extremidad superior		manaza ( s. f. SA) 1 - Sin asignar	No
⊕ _COMPOSITUM	mano ( s. f. sg.) 1 - Extremidad superior		maniatar ( v. -) 1 - Sin asignar	No
⊕ _COMPOSITUM	mano ( s. f. sg.) 1 - Extremidad superior		maniblanco (adj. c. c.) 1 - Animal	No
⊕ _COMPOSITUM	mano ( s. f. sg.) 1 - Extremidad superior		manicorto (adj. c. c.) 1 - Cualidad	No
⊕ _COMPOSITUM	mano ( s. f. sg.) 1 - Extremidad superior		manigordo ( s. m. sg.) 1 - Animal	No
⊕ _COMPOSITUM	mano ( s. f. sg.) 1 - Extremidad superior		manilargo (adj. c. c.) 2 - Cualidad	No
⊕ _COMPOSITUM	mano ( s. f. sg.) 1 - Extremidad superior		maniobrar ( v. -) 1 - Acción humana	No
⊕ _COMPOSITUM	mano ( s. f. sg.) 1 - Extremidad superior		manirroto (adj. c. c.) 1 - Cualidad	No
⊕ _COMPOSITUM	mano ( s. f. sg.) 1 - Extremidad superior		manivacio (adj. c. c.) 1 - Cualidad	No
⊕ _CORPORIS CURA	mano ( s. f. sg.) 1 - Extremidad superior		manicura ( s. f. SA) 1 - Sin asignar	No

Fig. 3: Some of the semantic relations related to *mano* (hand)



As there is a productive relation, related to “(like the one) of an animal”, we created a function called ANIMAL,<sup>10</sup> which allows us to show the relation between *hand/claw*, *mouth/snout*, *lip/beak*, *arm/thigh*, etc. There is also a productive semantic relationship attached to something that is cancelled, which was labelled as ABROGATIO, meaning ‘abolition’, and covers, amongst other things, the relationship between *mano/muñón* (‘hand/stump of the arm’); ABROGATUS (‘who was cancelled’), covers the relationship between *mano/manco* (‘hand/one-handed’). ARTIFEX (meaning ‘artist’, ‘craftsman’, ‘expert’) covers the relationship between *hand/chiropractor*, *prestidigitator*, and ARS (‘art’) is used for the relationship between *hand/chiropractic*, *prestidigitation*.

Although Figure 3 only shows a small part of the set of these functions, to date a total of only ninety-seven functions naming semantic relations have been generated. Lack of space prevents us from presenting a complete list, however we mention some other functions such as A<sub>0</sub>POSSESSOR (adjective for the owner of something), which covers *left-handed*, *right-handed*, *ambidextrous* for *hand*; A<sub>1</sub>VITIUM (adjective for the first argument of something defective) such as *albino*, *hairless* for *hair*; CORPORIS CURA (body care) such as *manicure* for *hand* and *pedicure* for *foot*; CAEDO (trying to kill) such as *hang*, *decapitate* and *strangle* for *neck*, etc.

#### 4. Collocations

Collocations of the dictionary are formalized in the usual manner within the MTT, that being by means of LFs. As mentioned before, some of our collocations were obtained automatically (which is a novelty in the MTT scenario, for further information see Barrios, 2010): we will try to illustrate it in the following lines.

As we work with lexical fields, we can predict some of the regular collocations shared by co-hyponyms, such as the words for joints: *elbow*, *wrist*, *knee* and *ankle*. As these words share semantic features, we were able to incorporate some collocations attached to the following LFs:

- a) FinReal<sub>1</sub>, which means ‘not to do work anymore’ and covers the verbs *dislocar/desencajar* (‘to dislocate’), *fracturarse* (‘to fracture’), *romperse* (‘to break’);
- b) IntentarCaus<sub>1</sub>PlusBonFact<sub>0</sub>, which means ‘(the first argument of the verb) try in order that something works in a better way’, and covers the verb *operarse* (‘to operate on’);
- c) IntentarCausPlusBonFact<sub>0</sub>, which means ‘to try to make something work in a better way’ and covers the verb *escayolar* (‘to put a cast on’) and *inmovilizar* (‘to immobilize’);
- d) S<sub>0</sub>AntiBonFact<sub>0</sub>, which means ‘noun that expresses that something does not work properly’ and covers the nouns *esguince* (‘sprain’), *lesión* (‘lesion’), *torcedura* (‘strain’), *fractura* (‘fracture’), etc.

<sup>10</sup> In order to distinguish the functions between the LFs of the MTT, we use a capital letter: our non-canonical LFs are functions we created *ad hoc* to facilitate the retrieval of information in an onomasiological sense. We also chose Latin words, as MTT researchers made with LFs.

Figure 4 shows some of these collocations related to *tobillo* ('ankle') (the last column portrays which collocations were obtained automatically through a yes/no marc):

Relaciones Semánticas				
Id-FA	Id-Argumento	rY	Id-Valor	Here
_MORBUS	tobillo (s. m. sg.) 1 - Articulación	sinovitis (s. f. SA) 1 - Sin asignar	Sí	
_PROTECTOR	tobillo (s. m. sg.) 1 - Articulación	tobillera (s. f. SA) 1 - Sin asignar	No	
A0AntiBonFact0	tobillo (s. m. sg.) 1 - Articulación	hinchado (adj. c. c.) 1 - Sin asignar	Sí	
A0AntiBonFact0	tobillo (s. m. sg.) 1 - Articulación	inflamado (adj. c. c.) 1 - Sin asignar	Sí	
A0AntiBonFact0	tobillo (s. m. sg.) 1 - Articulación	lesionado (adj. c. c.) 1 - Sin asignar	Sí	
AntiBonFact0	tobillo (s. m. sg.) 1 - Articulación	hinchar (v. -) 1 - Sin asignar	Sí	
AntiBonFact0	tobillo (s. m. sg.) 1 - Articulación	inflamar (v. -) 1 - Sin asignar	Sí	
AntiBonFact1	tobillo (s. m. sg.) 1 - Articulación	doler (v. -) 1 - Sin asignar	Sí	
AntiBonReal1	tobillo (s. m. sg.) 1 - Articulación	entumecer (v. -) 1 - Sin asignar	Sí	
AntiBonReal1	tobillo (s. m. sg.) 1 - Articulación	lesionar (v. -) 1 - Sin asignar	Sí	
AntiBonReal1	tobillo (s. m. sg.) 1 - Articulación	torcer (v. -) 1 - Acción	No	
CausPredPlusBon	tobillo (s. m. sg.) 1 - Articulación	fortalecer (v. -) 1 - Sin asignar	Sí	
FinReal1	tobillo (s. m. sg.) 1 - Articulación	desencajar (v. -) 1 - Sin asignar	Sí	
FinReal1	tobillo (s. m. sg.) 1 - Articulación	fracturar (v. -) 1 - Sin asignar	Sí	
FinReal1	tobillo (s. m. sg.) 1 - Articulación	romper (v. -) 1 - Sin asignar	Sí	
IntentarCaus1PlusBonFact0	tobillo (s. m. sg.) 1 - Articulación	operar (v. -) 1 - Sin asignar	Sí	
IntentarCausPlusBonFact0	tobillo (s. m. sg.) 1 - Articulación	escayolar (v. -) 1 - Sin asignar	No	
IntentarCausPlusBonFact0	tobillo (s. m. sg.) 1 - Articulación	inmovilizar (v. -) 1 - Sin asignar	Sí	
Parte de	tobillo (s. m. sg.) 1 - Articulación	articulación (s. f. sg.) 1 - Estructura	No	
Real1	tobillo (s. m. sg.) 1 - Articulación	doblar (v. -) 1 - Sin asignar	Sí	
SOAntiBonFact0	tobillo (s. m. sg.) 1 - Articulación	dolor (s. m. sg.) 1 - Sentimiento negativo	No	
SOAntiBonFact0	tobillo (s. m. sg.) 1 - Articulación	esguince (s. m. SA) 1 - Sin asignar	Sí	
SOAntiBonFact0	tobillo (s. m. sg.) 1 - Articulación	lesión (s. f. sg.) 1 - Efecto físico en animados	No	
SOAntiBonFact0	tobillo (s. m. sg.) 1 - Articulación	torcedura (s. f. SA) 1 - Sin asignar	No	
SOFinReal1	tobillo (s. m. sg.) 1 - Articulación	fractura (s. SA SA) 1 - Sin asignar	Sí	
			No	

Fig. 4: Some of the collocations related to *tobillo* (ankle)

### 5. Idioms

Idioms have their own entry in *Diretes*. The novelty being, thanks to the LFs and the functions mentioned in the § 4, idioms are accessible both by semasiological and onomasiological senses. Hereinafter, we illustrate this by an example of a concrete case, the Spanish idioms related to *false* and *falseness*.

The first one is an adjective associated with certain idioms, such as *figura decorativa* ('figurehead'), *caballo de Troya* ('Trojan horse'), *castillo de naipes* ('house of cards'), *cuento chino* (lit. Chinese tale, meaning tall story), *flor de un día* (lit. flower of one day, meaning short-lived), *música celestial* (lit. heavenly music, meaning music to my ears), *un quiero y no puedo* (lit. a want and I can't, meaning will o' the wisp). We applied several grammatical tests to verify that all these idioms function as an adjective (see Barrios, 2022a) and then we associated these idioms with the adjective *falso* ('false'). In this way, if someone does not remember some of these idioms, they can consult the entry *falso* and all the entries specific to each idiom will be accessible by clicking.

The semasiological sense is present, as always, when writing the idiom in the search engine of the dictionary. Instead of *false*, the user can write *falsedad* ('falseness'). The idioms related to this word function as a noun, therefore *Diretes* will offer information about *becerro de oro* ('golden calf'), *buenas palabras* (lit. good words, meaning nice but false words) and *malas lenguas* (lit. bad tongues, meaning gossip, added a nuance of falsehood).

Figure 5 shows the results for *falso* ('false'):

**falso**  
Palabras relacionadas

Acepción núm: 1

Definición mínima: Que no es lo que se podría pensar que es.  
Definición expandida: Se dice de cosas (piel falsa), de palabras (lo que dices es falso), de resultados (el resultado de la suma es falso), etc.  
Etiqueta semántica: Cualidad

**falso como argumento**

**Locuciones nominales**

Propiedad de personas en relación con personas	falso	<a href="#">figura decorativa</a>	nivel ELE: S
Propiedad de personas, entidades o situaciones	falso	<a href="#">caballo de Troya</a>	nivel ELE: S
Propiedad de personas, entidades o situaciones	falso	<a href="#">castillo de naipes</a>	nivel ELE: S
Propiedad de personas, entidades o situaciones	falso	<a href="#">cuento chino</a>	nivel ELE: S
Propiedad de personas, entidades o situaciones	falso	<a href="#">flor de un día</a>	nivel ELE: S
Propiedad de personas, entidades o situaciones	falso	<a href="#">música celestial</a>	nivel ELE: S
Propiedad de personas, entidades o situaciones	falso	<a href="#">un quiero y no puedo</a>	nivel ELE: S

Fig. 5: Idioms related to *falso* (false)

## 6. Conclusions

*Diretes* is an ongoing project with considerable information still to be added as well as problems to be solved. The majority of the difficulties found relate to the following aspects: a) correctly writing a minimal definition with simple words; b) selecting the most appropriate pragmatic function for formulaic expressions; c) avoiding overproduction of functions created *ad hoc* for semantic relations; d) selecting the correct LFs for collocations; e) selecting the most accurate words (often more than one) for the onomasiological search of idioms.

Other problems are related to questions that were impossible for us to cover in this paper, such as the selection of the examples appropriate to different levels of Spanish learning, or the debugging of errors, a very important task in *Diretes* due to the automatization of the procurement of collocations and the LFs attached to them.

Although many years away from being completed, we aspire in the long run to be able to refine the working methodology in order to improve its efficiency.

## References

- Alonso Ramos, M. (2004). Elaboración del Diccionario de Colocaciones del español y sus aplicaciones. In P. Bataner, & J. de Cesaris (Eds.), *De lexicografía. Actes del I Simposium Internacional de Lexicografía* (pp. 149–162). IULA & Edicions Petició.
- Apresjan, J. (1974). Regular polysemy. *Linguistics* 142/5, 5–32.
- Apresjan, J. (1992). Systemic Lexicography as a Basis of Dictionary-making. *Dictionaries: Journal of the Dictionary Society of North America* 14, 79–87.
- Apresjan, J. (2000). *Systemic Lexicography*. Oxford University Press.
- Apresjan, J. (2009). The Theory of Lexical Functions: An Update. In D. Beck et al. (Eds.), *Fourth International Conference on Meaning-Text Theory* (pp. 1–14). Observatoire de linguistique Sens-Texte.
- Apresjan, J. (2018). *New Explanatory Dictionary of Russian Synonyms*. PubMix.com.
- Apresjan, V. (2011). Active dictionary of the Russian language: theory and practice. In I. Boguslavsky, & L. Wanner (Eds.), *Proceedings of the 5th International Conference on Meaning-Text Theory* (pp. 13–24). Universitat Pompeu Fabra.
- Atkins, S. B. T., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford University Press.
- Barrios, M. A. (2010). El dominio de las Funciones Léxicas en el marco de la Teoría Sentido-Texto. *Estudios de Lingüística del Español* 30, 1–477.
- Barrios, M. A. (2010). Minimal and inverse definitions: a semi-experimental proposal when building a Spanish dictionary with semantic primes and molecules. In P. Bert, K. Mullan, & L. Sadow (Eds.), *Studies in Ethnopragsmatics, (Cultural) Semantics, and Minimal English* (pp. 191–212). Springer.
- Barrios, M. A. (2020). The making of the Diretes dictionary. In Z. Gavriilidou, M. Mitsiaki, & A. Fliatouras (Eds.), *Lexicography for Inclusion. EURALEX Proceedings. XIX Congress of the European Association for Lexicography* (pp. 13–22).
- Barrios, M. A. (2022a). *DiRetEs*: Diccionario Reticular del Español. Resultados de la primera fase de un proyecto. In L. Iomdin, J. Milićević, & A. Polguère (Eds.), *Lifetime Linguistic Inspirations. To Igor Mel'čuk from Colleagues and Friends for his 90th Birthday. Wiener Slawistischer Almanach Sonderband* 101 (pp. 73–84).
- Barrios, M. A. (2022b). Meaning and function of Spanish formulemes and pragmatemes vs. illocutionary verbs. *Russian Journal of Linguistics* 26/4, 1031–1049. doi: 10.22363/2687-0088-31597
- Blanco, X. (2013). Les pragmatèmes: définition, typologie et traitement lexicographique. *Verbum* 4, 17–25.
- Frawley, W. (1997). New forms of specialized dictionaries. *International Journal of Lexicography* 1(3), 189–213.



Goddard, C. (2012). Semantic Primes, Semantic Molecules, Semantic Templates: Key Concepts in the NSM Approach to Lexical Typology. *Linguistics* 50(3), 711–743. <https://doi.org/10.1515/ling-2012-0022>

Goddard, C. (2016). Semantic molecules and their role in NSM lexical definitions. *Cahiers de lexicologie*, 109, 13–34.

Goddard, C., & A. Wierzbicka (2014). *Words and Meanings. Lexical Semantics across Domains, Languages and Cultures*. Oxford University Press.

L’Homme, M. C. (2012). Using ECL (Explanatory Combinatorial Lexicology) to discover the lexical structure of specialized subject fields, In Apresjan, J. et al. (Eds.), *Words, Meanings and other Interesting Things. A Festschrift in Honour of the 80th Anniversary of Professor Igor Alexandrovic Mel’cuk* (pp. 378–390). RCK.

L’Homme, M. C. (2018). Maintaining the balance between knowledge and the lexicon in terminology: a methodology based on frame semantics. *Lexicography* 4(1), 3–21.

L’Homme, M. C., B. Robichaud, & C. Subirats. (2016). A Proposal for combining “general” and specialized frames. In *Proceedings of the Workshop on Cognitive Aspects of the Lexicon* (pp. 156–165).

Kauffer, M. (2018). Qu’est-ce qu’un ALS ? *Verbum* XL/1, 35–50.

Mel’cuk, I. (1984). Le Dictionnaire Explicatif et Combinatoire du Français Contemporain. Prol. *Dictionnaire Explicatif et Combinatoire du Français Contemporain. Recherches lexicosémantiques I* (pp. 3–15). Les Presses de l’Université de Montréal.

Mel’cuk, I. (1995). Phrasemes in Language and Phraseology in Linguistics. In E. Everaert, J. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms. Structural and Psychological Perspectives* (pp. 167–232). Lawrence Erlbaum Associates.

Mel’cuk, I. (1996). Lexical Functions: A Tool for the Description of Lexical Relations in Lexicon. In L. Wanner (Ed.), *Lexical Functions in Lexicography and Natural Language Processing* (pp. 37–102). John Benjamins.

Mel’cuk I. (2014). *Semantics. From meaning to text*. Volume 3. John Benjamins.

Mel’cuk I. (2023). *General Phraseology. Theory and Practice*. John Benjamins.

Mel’cuk, I., Clas, A., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Editions Duculot.

Mel’cuk, I., & Polguère, A. (2007). *Lexique actif du français – l’apprentissage du vocabulaire fondé sur 20.000 dérivations sémantiques et collocations du français*. Boeck & Larcier.

Mel’cuk, I., & Polguère, A. (2018). Theory and Practice of Lexicographic Definition. *Journal of Cognitive Science* 19/4, 417–470. 10.17791/jcs.2018.19.4.417

Polguère, A. (2000). Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. In U. Heid, S. Evert, E. Lehmann, & C. Rohrer (Eds.), *Proceedings of the Ninth EURALEX International Congress* (pp. 517–527). Stuttgart.



Polguère, A. (2003). Collocations et fonctions lexicales: pour un modèle d'apprentissage. In F. Grossmann, & A. Tutin (Eds.), *Les Collocations: analyse et traitement* (pp. 117–141). Éditions de Werelt.

Polguère, A. (2007). Lexical Function Standardness. In L. Wanner (Ed.), *Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In Honour of Igor Melčuk* (pp. 43–96). John Benjamins.

Polguère, A. (2014). From writing dictionaries to weaving lexical networks. *International Journal of Lexicography* 27/4, 396–418. <https://doi.org/10.1093/ijl/ecu017>

Roche, C. (2015). Ontological definitions. In H. J. Kockaert, & F. Steurs (Eds.) *Handbook of Terminology: Volume 1* (pp. 128–152). John Benjamins.

Wierzbicka, A. (1985). *Lexicography and Conceptual Analysis*. Karoma.

Wierzbicka, A. (1996). *Semantics: Primes and Universals*. Oxford University Press.

## Acknowledgements

This work was carried out within the framework of the project DiRetEs (Second Phase) and was supported by the Ministry of Science and Innovation of Spain (PID2021-122894-NB-I00) and the FEDER Program (European Union). I would like to express my gratitude to the anonymous reviewers for their helpful observations, to Svetlana Antropova and Deborah Paton for all her suggestions on the English revision of the manuscript and to Begoña Sanromán for her outstanding help. Any remaining errors are the sole responsibility of the author.

## Contact information

**María Auxiliadora Barrios Rodríguez**

Universidad Complutense de Madrid

auxibarrios@filol.ucm.es