

# A Spanish e-dictionary of collocations

**María Auxiliadora Barrios**

Universidad Complutense de Madrid  
auxibarrios@filol.ucm.es

**Igor Boguslavsky**

Universidad Politécnica de Madrid /  
IITP, Russian Academy of Sciences  
iboguslavsky@fi.upm.es

## Abstract

We present a new e-dictionary of Spanish (in progress) called *Diretes* (Diccionario RETicular de ESpañol). It contains descriptions of collocations by means of Lexical Functions (LFs), both standard and non-standard, in the sense of the Meaning – Text Theory by Igor Mel’čuk. At present, *Diretes* contains about 50,000 collocations. This paper concentrates on the collocations in which the collocate is an adjectival or an adverbial phrase. These collocations are mostly extracted from the *Práctico* combinatorial dictionary of modern Spanish. We explain the structure of the e-dictionary, the types of information it contains and the way it is presented. We also show how the LF-interpreted collocations can be used in NLP applications. We demonstrate it with the SemETAP semantic analyzer, in which LFs are used to normalize semantic structures and make inferences.

## 1 Introduction

This paper presents a Spanish e-dictionary called *Diretes*. It has several sources. The first of them is the BADELE.3000 database (Barrios and Bernardos, 2007; Barrios, 2010), which contains 25,000 collocations described by means of Lexical Functions (LFs) of the Meaning-Text Theory (MTT) (Mel’čuk, 1996, 2014). Recently, we built a new database and doubled the number of collocations, so that now *Diretes* totals about 50,000 items. An important source of data is the EsTenTen corpus (SketchEngine, <https://www.sketchengine.eu/estenten-spanish-corpus>). Our next step consists in incorporating the data of *Práctico* – a well-known dictionary of Spanish collocations. We aim at interpreting the *Práctico* collocations in terms of LFs, as we did in previous portions of *Diretes*. Lexical Functions have been proposed in MTT as a tool for the formalization of lexical relations and classifying collocations and some paradigmatic relations (such as synonymy, antonymy and semantic derivatives). However, standard MTT LFs cannot cover the whole material of *Práctico*. A significant part of uncovered material is presented by expressions containing adjectives and adverbs, which are our primary concern in this paper. To bring some order into this group of collocations, we widely use non-standard LFs and a set of semantic features.

This paper is structured as follows. In section 2, we summarize some relevant characteristics of two Spanish combinatorial dictionaries particularly useful for our task. In section 3, we present the Spanish e-dictionary we are building, *Diretes*. In section 4, we relate lexical resources such as *Diretes*, that store LFs, to NLP applications, that make use of LFs. Drawing on the example of the SemETAP semantic analyzer we show that LFs can be effectively used for the normalization of semantic structures and for drawing inferences. Finally, we present our conclusions and outline future work.

## 2 *Redes* and *Práctico*, two Spanish combinatorial dictionaries

*Redes* and *Práctico* (Bosque, 2004, 2006) are two Spanish combinatorial dictionaries containing a carefully selected set of collocations, which constitute the collocational knowledge of educated native speakers of Spanish. *Redes* contains 7,115 entries and *Práctico* 14,000 (Barrios, 2007). Both of them are relevant sources of semantic information, particularly *Redes*, a dictionary that offers a detailed lexical and

semantic analysis. The combinatorial data are presented in this dictionary by means of lexical classes, each one described by semantic features. For instance, the entry of the adjective *férreo* ‘(referring to) iron’ reflects first of all the primary meaning of the adjective (‘made of iron’) and then its figurative meanings, in which it modifies action nouns such as *control férreo* ‘iron grip’, nouns of physical objects used in figurative sense, as *mano férrea* ‘iron fist’, phrases as *regla férrea* ‘iron rule’, etc. For each collocation there is a real example taken from a corpus of more than 250 million words. *Redes* is a dictionary mostly intended for research purposes.

*Práctico* is conceived as a dictionary for practical purposes. It includes all the collocations from *Redes* and many more. It is useful, first of all, for native speakers interested in perfecting their mastery of language, for authors, translators and language learners. It gives fewer examples than *Redes* and does not use the explicit semantic classification of *Redes*, but it preserves its semantic structure. In both dictionaries, each entry contains a large number of collocations: for instance, the *Práctico* entry of the adjective *aromático* ‘aromatic’ shows thirteen nouns (such as the Spanish equivalents of *plant*, *herb*, *drink*, *wine*, *oil*, etc.) but not *flor* ‘flower’ nor *rosa* ‘rose’, even though in the real world flowers in general and roses in particular are aromatic often enough to expect the existence of these collocations.

*Redes* and *Práctico* are valuable sources of combinatorial information. As opposed to collocational material extracted from large collections of texts automatically, which often contains a lot of rubbish, materials offered by *Redes* and *Práctico* are a result of thorough individual research and exhibit the highest standard of quality. What they lack is some degree of formalization, which could render them more useful for applications. This is what we are trying to achieve in the *Diretes* project.

### 3 *Diretes*: A Spanish e-dictionary supplied with Lexical Functions

Electronic dictionaries are structured sets of lexicographic data in numerical form accessible in different ways and having multiple functionalities (De Schryver, 2003). Some of them are targeted at humans and some are machine-readable, which means that they are useful not only for humans but also for computers that can read their contents (Dziemianko, 2017). The problem that arises here is that even if an e-dictionary is machine-readable, its contents are designed for human consumption: in many cases, text understanding requires inferences of diverse kinds, which is still unfeasible for the machines; some of these inferences need to be based on dictionary definitions, some others are not linguistic but pragmatic or cultural (Barrios, in press).

On the other hand, NLP tools reuse different linguistic resources, such as dictionaries. In recent years, many NLP researchers are actively developing practices oriented to sharing data on the web, which are called *linked data* (Bizer et al, 2011). Different models to represent linguistic linked data have been proposed, some of them focused on lexical resources, and some others on ontologies, catalogues of linguistic data or even corpora models (Bosque-Gil et al, 2016, 2018).

Many of the new electronic dictionaries are human-oriented: collocations and meanings of lexical units are explained in a natural language rather than in a formalism suitable for machines. What we propose in *Diretes* is to create contents accessible to machines, in a way similar to some other dictionaries within the Meaning – Text approach, such as the French dictionaries Dicouèbe<sup>1</sup> and DiCoInfo (L’Homme, 2008), the English and Russian ETAP-4 dictionaries<sup>2</sup> and the Spanish dictionary of emotions DICE<sup>3</sup> and DiCoEnviro (Ortego Antón, 2011). At present, *Diretes* contains about 50,000 collocations. Among them, there are 551 adjectival and adverbial collocations extracted from *Práctico* and *Redes* (beginning with the letter *a*). In this paper, we concentrate on these collocations.

---

<sup>1</sup> <http://olst.ling.umontreal.ca/dicouebe/index.php>

<sup>2</sup> <http://cl.iitp.ru>

<sup>3</sup> <http://www.dicesp.com/paginas>. The Spanish dictionary DICE provides LFs for the semantic field of emotions, but emotions is only one of the 664 semantic fields of *Diretes*. DICE contains 200 entries, and all of them belong to the field of emotions; *Diretes* contains 372 entries for emotions, which provide 7737 collocations.

*Diretes* assigns a large amount of LF information to words. First of all, one should distinguish between standard and non-standard LFs (Mel'čuk, 2014: 173-174). As for standard LFs, adjectives and adverbs can act as values of several of them, including semantic derivatives  $A_i$  and  $Adv_i$ , plus a number of syntagmatic LFs, such as Magn (meaning 'very, to a high degree', such as *infinite* in *infinite patience*), Ver (meaning 'such as should be', e.g. *legitimate* in *legitimate demand*), Bon (meaning 'good', such as *fruitful* in *fruitful analysis*), Pos (meaning 'positive evaluation', e.g. *favourable* in *favourable opinion*), Epit (meaning 'redundant clichéd modifier', such as *sweet* in *sweet dream*). All of them are useful when formalizing not only adjective collocations but also adverbial ones. All of these LFs can combine with the LF Anti (meaning 'opposite'): if the expression *to pay an arm and a leg* is covered by Magn, then *to pay a mere trifle* and *to cost peanuts* are covered by AntiMagn.

In *Diretes*, we widely use the conceptual relation TypeOf, which denotes hypernymy (similar to LF Gener of MTT). To make the description more precise, we introduced several semantic variants of the TypeOf relation: TypeOf–form (Sp. Tipo de–forma), TypeOf–function (Sp. Tipo de–función), TypeOf–print (Sp. Tipo de–estampado) and some others.

In Fig. 1, one can see a fragment of *Diretes* illustrating some collocations of this class. Here the first column shows the identification number of each lexical relation; the second one, the name of this relation; the third, the argument of the lexical relation, its grammatical features and its semantic label (which is the name of the hypernym); the fourth shows the value of the lexical relation, its grammatical features and its semantic label; the fifth signals if this lexical relation was automatically inherited from the relation between the hypernym and the value; the sixth one is filled in manually if the automatically inherited lexical relation is incorrect (such as *ponerse el bolso* 'to put on one's bag'); the seventh suggests of the level at which this lexical relation could be learnt by students of Spanish as a second language; and the last one shows a real example of use taken from *SketchEngine*.

Finally, there is a large portion of collocations formalized by means of non-standard LFs. We classified them using some of the most productive semantic features shown in Table 1.

Id-RS	Id-FA	rY	Id-Argumento	Id-Valor	Heredada	Rechazada	ELE	Ejemplo
160553	Tipo de-estampado		camisa (s. f. sg.) 1 - ropa	a cuadros (loc. adj. SA SA) 1 -	No	No	A	
164021	Tipo de-estampado		falda (s. f. sg.) 1 - ropa	a cuadros (loc. adj. SA SA) 1 -	No	No	S	
233633	Tipo de-función		bolso (s. m. sg.) 1 - Compleme	a cuestras (loc. adj. --) 1 - Sin	No	No	B	
205515	Tipo de		vestido (s. f. sg.) 1 - ropa	a la moda (loc. adj. --) 1 - Sin	No	No	B	
205819	Tipo de		traje (s. m. sg.) 1 - ropa	a medida (loc. adj. --) 1 - Sin	No	No	B	
206287	Tipo de		vestido (s. f. sg.) 1 - ropa	a medida (loc. adj. --) 1 - Sin	No	No	B	
160849	Tipo de-estampado		blusa (s. f. sg.) 1 - ropa	a rayas (loc. adj. SA SA) 1 - Sin	No	No	B	
160541	Tipo de-estampado		camisa (s. f. sg.) 1 - ropa	a rayas (loc. adj. SA SA) 1 - Sin	No	No	A	^Se puso una camisa a rayas.
164013	Tipo de-estampado		falda (s. f. sg.) 1 - ropa	a rayas (loc. adj. SA SA) 1 - Sin	No	No	S	
184253	Tipo de-estampado		jersey (s. m. sg.) 1 - ropa	a rayas (loc. adj. SA SA) 1 - Sin	No	No	B	
182949	Tipo de-estampado		prenda (s. f. sg.) 1 - ropa	a rayas (loc. adj. SA SA) 1 - Sin	No	No	B	
163893	Tipo de-función		faja (s. f. sg.) 1 - Ropa interior	abdominal (adj. c. c.) 1 - Sin a	No	No	S	Desarrollan un programa de tonifica
153433	Tipo de-forma		bota (s. f. sg.) 1 - Calzado	abierto (adj. c. c.) 1 - Rasgo fís	No	No	B	
160721	Tipo de-forma		camisa (s. f. sg.) 1 - ropa	abierto (adj. c. c.) 1 - Rasgo fís	No	No	A	Viste pantalones pirata, camisa abie
163417	Tipo de-forma		chaqueta (s. f. sg.) 1 - ropa	abierto (adj. c. c.) 1 - Rasgo fís	No	No	S	
153985	Tipo de-forma		sandalia (s. f. sg.) 1 - Calzado	abierto (adj. c. c.) 1 - Rasgo fís	No	No	C	Utiliza sandalias abiertas y cómodas
163981	Tipo de-forma		falda (s. f. sg.) 1 - ropa	abombado (adj. c. c.) 1 - Sin a	No	No	S	
153993	Tipo de-forma		sandalia (s. f. sg.) 1 - Calzado	abotinado (adj. c. c.) 1 - Sin as	No	No	C	Completó su vestuario con unas san
162297	Tipo de		camisa (s. f. sg.) 1 - ropa	abotonado (adj. c. c.) 1 - Sin a	No	No	B	El cuarentón de la camisa abotonada
206119	Tipo de		trenca (s. f. sg.) 1 - ropa	abotonado (adj. c. c.) 1 - Sin a	No	No	C	
161205	Tipo de-forma		blusa (s. f. sg.) 1 - ropa	abotonado (adj. c. c.) 1 - Sin a	No	No	B	Vendemos blusas abotonadas de mi
163425	Tipo de-forma		chaqueta (s. f. sg.) 1 - ropa	abotonado (adj. c. c.) 1 - Sin a	No	No	S	Esos hombres están sudando pero s
164097	Tipo de-forma		falda (s. f. sg.) 1 - ropa	abullonado (adj. c. c.) 1 - Sin e	No	No	S	Lo más característico de este vestid
206419	Tipo de-función		agenda (s. f. sg.) 2 - Sin asignar	académico (adj. c. c.) 1 - Sin a	No	No	C	He traído la agenda académica del o
164129	Tipo de-forma		falda (s. f. sg.) 1 - ropa	acampanado (adj. c. c.) 1 - Sin	No	No	C	Hay algunos modelos que vienen co

Fig.1. Some adjectival collocations formalized by means of TypeOf relations in *Diretes*

Semantic Feature	Adjective/adverbial expression	Example of use
Material	<i>abonado</i> 'fertilised'	<i>tierra abonada</i> 'potting soil'
Appearance	<i>abierto</i> 'open'	<i>mente abierta</i> 'open mind'

Place	<i>aéreo</i> ‘aerial’	<i>tráfico aéreo</i> ‘air traffic’
Manner	<i>a boca jarro</i> ‘point-blank’	<i>decir algo a boca jarro</i> ‘to say something bluntly’
Cause	<i>abrasador</i> ‘burning’	<i>sol abrasador</i> ‘blazing sun’
Able to	<i>accesible</i> ‘accessible’	<i>lugar accesible</i> ‘accessible place’
Quantity	<i>a partes iguales</i> ‘in equal parts’	<i>dividir a partes iguales</i> ‘divide in equal parts’
Time	<i>anual</i> ‘annual’	<i>convocatoria anual</i> ‘annual call’
Recurrence	<i>asiduo</i> ‘regular’	<i>orador asiduo</i> ‘regular guest speaker’
Speed	<i>a toda máquina</i> ‘at full speed’	<i>trabajar a toda máquina</i> ‘to work at full speed’

Table 1. Semantic features used to characterize non-standard LFs.

In *Diretes* the words are organized as a net, not as a hypernym/hyponym hierarchy. In the latter case, a “mother” may have several “children”, while a “child” may have only one mother. In *Diretes* this is not so: a “child” may have several “mothers”. For instance, a word such as *reloj* ‘watch’ is labeled as belonging both to the class ‘artifact’ and ‘accessory’. One of the salient features of *Diretes* is that the database was designed to implement the LF Domain Principle (which is similar but not identical to the lexical inheritance principle of (Mel’čuk & Wanner, 1996: 229)). According to this principle, most words sharing a hypernym usually develop similar collocations (Barrios, 2009; Barrios, 2010; Barrios, Bernardos 2007). Below, we present the structure of the database and then we illustrate the LF Domain Principle.

In *Diretes* the data are organized in several tables. The four most important tables are: a) lemmas; b) the hierarchy of semantic labels; c) semantic predictions; and d) semantic and lexical relations.

In the first table, the lemmas of the dictionary are tagged by semantic labels (i.e. hypernyms); for instance, *camisa* (shirt) is labeled as ‘piece of clothing’ and *calcetín* (sock) as ‘underwear’.

In the second table, the semantic labels are structured in a hierarchy of nine levels; for instance, ‘*ropa y accesorios*’ (‘clothing and accessories’) is the “mother” of ‘*ropa*’, ‘*zapatos*’ and ‘*accesorios*’ (‘clothing’, ‘shoes’ and ‘accessories’); and ‘*ropa*’ (‘clothing’) is in its turn the “mother” of ‘*ropa interior*’ (‘underwear’).

In the third table we predict some relations that can be inherited from “mothers” to “children”; for instance, ‘*ropa y accesorios*’ (‘clothing and accessories’) is related to four verbs and its Lexical Functions are: *llevar (puesto)* ‘to wear’ (Real<sub>1</sub>), *ponerse* ‘to put sth on’ (IncepReal<sub>1</sub>), *quitarse* ‘to take sth off’ (FinReal<sub>1</sub>), *estropearse* ‘to get damaged’ (Degrad). The semantic label ‘*ropa*’ (‘clothing’) inherits these verbs and then we add manually thirteen new verbs or verbal phrases, such as *sentar bien* ‘look good’ (BonFact<sub>1</sub>), *sentar mal* ‘look bad’ (AntiBonFact<sub>1</sub>), *arreglar* ‘to fix’ (CausPredPlusVer), etc. Some of them are inherited by the “grandchild” label ‘*ropa interior*’ ‘underwear’, and we add some new particular verbs such as *quedar apretado* ‘be tight’ (AntiBonFact<sub>1</sub>). To sum up, in this particular case, the table of semantic predictions contains forty-six collocations represented by lexical functions to be inherited by nouns such as *camisa* ‘shirt’, *calcetín* ‘sock’, *anillo* ‘ring’ or *reloj* ‘watch’.

In the fourth table, we collect all the collocations and semantic relations attached to each corresponding Lexical Function. To conclude with the ‘clothing and accessory’ example, all the nouns labeled as ‘clothing’, ‘shoes’, ‘accessories’, ‘complements’, and ‘underwear’ inherit all the verbs from the third table; then, in the fourth table there are 4989 collocations for nouns such as *camisa* ‘shirt’, labeled as ‘clothing’ and *calcetín* ‘sock’, labeled as ‘underwear’ (2567 were automatically inherited and 2422 were manually added); 909 collocations for nouns such as *botas* ‘boots’ labeled as ‘shoes’ (539 were automatically inherited); 1060 collocations for nouns such as *anillo* ‘ring’ labeled as ‘accessory’ (626 were automatically inherited). There are also 987 collocations for nouns such as *billetera* ‘wallet’ labeled as ‘complements’ (151 were automatically inherited).

The LF Domain Principle says that some collocations can be predicted on the basis of the LF domain, i.e. the list of words likely to be keywords of this LF. For example, we can predict that all words denoting fruits, vegetables and objects made from organic materials can be keywords of Degrad (which means ‘to become permanently worse or bad’); and all words denoting artifacts can be keywords of CausFunc<sub>0</sub>;

the domain of CausFunc<sub>0</sub> is the set of nouns denoting things that can be created. Semantic labels for each domain allow us to predict groups of collocations, such as *to build* for ‘building’ (*temple, tower, concert hall, castle*, etc.) and ‘housing’ (*apartment, flat, duplex*, etc.); *to compose* for ‘text’ (*poem, novel, essay*, etc.) and ‘music’ (*symphony, melody, sing*, etc.); *to make* for ‘clothes’ (*shirt, trousers, coat*, etc.) and ‘food’ (*cake, paella, soup*, etc.). This is applicable to many other LFs, such as LiqueFunc<sub>0</sub> (‘to cause something to not exist anymore’), IncepFunc<sub>0</sub> (‘to start existing’), FinFunc<sub>0</sub> (‘to finish existing’), CausFact<sub>0</sub> (‘to cause something to start to work’), LiqueFact<sub>0</sub> (‘to cause something to finish working’), IncepPredPlus (‘to increase’), FinPredPlus (‘to decrease’), Son (‘to emit a characteristic sound’) (Barrios and Goddard, 2013).

The implementation of the LF domain principle, as well as the lexical inheritance principle, allows us to generate automatically thousands of collocations, and consequently it is possible to complete the lexicographic task in less time. Once both principles are applied, we can analyze the meaning in a deeper way, by means of dimensions of meaning, as Mel’čuk and Wanner propose, or even by means of primes and molecules, as Barrios and Goddard proposes for the LF Degrad after analysing some English and Spanish collocations related to this LFs: “The intuition behind the Degrad function is that there is a common semantic core to all the verbs (...) First, all the Lexico-Syntactic Frames include the following pair of components: something bad happens to something for some time; because of this, after this, this something is not like it was before. Second, with one partial exception, the explications all share the following component in the Process section: when it happens, it happens slowly, people can’t see it. These three components are (arguably) enough to capture a serviceable core or ‘prototype’ for the intuition behind the ‘Degrad’ notion” (Barrios and Goddard, 2013: 239).

#### **4 Adjectival and adverbial Lexical Functions in semantic analysis**

In this section, we show that LFs stored in e-dictionaries such as *Diretes* can be effectively used in NLP applications. Let us recall that the interest that LFs aroused in the community from their very inception was largely motivated by the fact that they can be useful for different tasks, both lexicographic and related to computational linguistics. To name but a few publications, early attempts of using LFs in NLP are described in (Arsentjeva et al., 1969; Streiter, 1996; Wanner, 1996; Polguère, 1998; Mel’čuk, Wanner, 2001). Apresjan et al. (2007) explains how LFs can be used in language learning. Apresjan et al. (2002) presents LFs included in the electronic combinatorial dictionaries of Russian and English. In these dictionaries, about 50,000 Russian and 25,000 English words are supplied with LFs. It is shown that LFs can improve lexical and syntactic disambiguation during parsing, idiomatic translation in machine translation and synonymous paraphrasing. The latter task is described in detail in (Apresjan, Tsinman, 2002). In (Lambrey, Lareau, 2015) LFs are used in language generation. Formalization of LFs carried out in (Jousse, 2010; Fonseca et al., 2016) can be used for the development and efficient consulting of lexical databases.

Here, we present yet another application in which LFs can be put to good use. It is semantic analysis as implemented in the SemETAP system (Boguslavsky, 2011). The task of the semantic analyzer is to represent the meaning of the text in an explicit and unambiguous way. Two levels of semantic structure are distinguished: Basic Semantic Structure (BSemS) interprets the text in terms of ontological concepts; Enhanced Semantic Structure (EnSemS) extends BSemS by means of a series of inferences. LFs are used in SemETAP at two stages: in constructing and normalizing BSemS and in drawing inferences thereof. Below, we will illustrate both of these types.

We will call a syntactic derivative of word L such a word, or phrase, L’ that has the same (or very close) meaning as L, but belongs to a different syntactic category and hence displays a different behavior. Actantial syntactic derivatives (S<sub>i</sub>, A<sub>i</sub>, Adv<sub>0</sub>, Adv<sub>i</sub>) are in some way oriented towards one of the actants of the keyword. Such derivatives are supplied with a numerical index, which corresponds to the number of this actant.

In BSemS all predicates should be brought to the normalized form, which means that syntactic derivatives should be replaced by their keyword. In case of actantial derivatives, normalization also requires that the

i-th actant of the keyword be explicitly established. Here are some examples of actantial derivatives and normalizing operations they trigger:  $A_1(\textit{fear}) = \textit{fearful, frightened}$  ( $\approx$  ‘such that fears something’),  $A_2(\textit{fear}) = \textit{fearsome}$  ( $\approx$  ‘such that is feared’);  $Adv_1(\textit{hurry}) = \textit{hastily}$  ( $\approx$  ‘hurrying’),  $Adv_2(\textit{permit}) = \textit{with the permission}$  ( $\approx$  ‘being permitted’).

$A_1$ : The *child was fearful* <*frightened*>  $\implies$  ‘the child feared something’

$A_2$ : The *consequences were fearsome*  $\implies$  ‘one could fear the consequences’

$Adv_1$ : *He said good bye hastily*  $\implies$  ‘he said good bye; while saying it he was hurrying’

$Adv_2$ : *The evidence was examined by the experts with the permission of the court*  $\implies$  ‘the evidence was examined by the experts; the court permitted the experts to examine the evidence’.

Some examples of other types of LFs that also trigger inferences:

$Real_1(\textit{promise}) = \textit{fulfil}$ : *He fulfilled his promise to help me*. Inference: ‘he helped me’.

$CausFunc_0(\textit{crisis})$ : *bring about (a crisis)*. Inference: ‘the crisis takes place’.

$LiquFunc_0(\textit{beard})$ : *shave off (one's beard)*. Inference: ‘the beard exists no longer’.

## 5 Future work

We presented a project which aims at compiling a new e-dictionary of Spanish supplied with Lexical Functions and other information. In its current state, it contains about 50,000 lexical relations: 20,000 cover the most frequent collocations of Peninsular Spanish; that is, the collocations that any student of B2 level should master (based on frequency corpus data, Barrios, 2010). 30,000 other collocations correspond to the domain of the body and body parts, emotions, clothing and accessories. We are working now on the domain of the house, and in the next months we will work on artifacts, food and evaluation domains. Our goal is to obtain a database of 75,000 collocations described in terms of Lexical Functions by the end of 2020. Another immediate task is to significantly enlarge the set of adjectival and adverbial non-standard LFs. We have a large number of collocations in our database that are still lacking adequate description in terms of LFs. We are also planning to bring our adjectival semantic classification closer to WordNet standards.

## Acknowledgements

This work is developed within the project DiRetEs (Diccionario Reticular Español: Diccionario analógico y relacional con acceso en red desde el sentido y desde la forma) and was partially supported by the Ministerio de Ciencia, Innovación y Universidades (FFI2017-83293) and the RSF grant No. 16-18-10422-P, which is gratefully acknowledged.

## References

- Apresjan, Ju., I. Boguslavsky, L. Iomdin, and L. Tsinman. 2002. Lexical Functions in NLP: Possible Uses. *Computational Linguistics for the New Millennium: Divergence or Synergy? Festschrift in Honour of Peter Hellwig on the occasion of his 60th Birthday*, edited by Manfred Klenner and Henriette Visser, 55–72. Frankfurt/M., Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang.
- Apresjan, Ju., L. Tsinman. 2002. Formal'naja model' perifrazirovanija predlozhenij dlja system pererabotki tekstov na estestvennyx jazykax, *Russkij jazyk v nauchnom osveschenii* 2 (4):102-146. (“A formal model of paraphrasing for NLP systems”; In Russian)
- Apresjan, Ju., P. Diachenko, A. Lazurskiy, L. Tsinman. 2007. O kompjuternom uchebnike leksiki russkogo jazyka. *Russkij jazyk v nauchnom osveschenii* 2 (14):48-112. (“Computer manual of the Russian vocabulary”; In Russian)
- Arsentjeva, N., N.Balandina, A.Krasovskaja. 1969. O mashinnoj realizatsii sistemy perifrazirovanija. Institut prikladnoj matematiki AN SSSR, preprints 25, 26, 27. Moscow. (“Computer implementation of the paraphrasing system”; In Russian)
- Barrios Rodríguez, M<sup>a</sup> Auxiliadora. 2007. Diccionarios combinatorios del español: diferencias y semejanzas entre Redes y Práctico. *RedELE. Red electrónica de Didáctica del Español como Lengua Extranjera*, 11, 1 - 14.

- Barrios Rodríguez, María Auxiliadora. 2009. Domain, domain features of lexical functions and generation of values by analogy according to the MTT approach. Proceedings of Fourth International Conference on Meaning-Text Theory. <http://olst.ling.umontreal.ca/pdf/ProceedingsMTT09.pdf>.
- Barrios Rodríguez, María Auxiliadora. 2010. *El dominio de las funciones léxicas en el marco de la Teoría Sentido-Texto. Estudios de Lingüística del español (ELiEs)*, vol 30. <http://elies.rediris.es/elies30/index30.html>.
- Barrios Rodríguez, M.A. (in press) ¿Aún queda alguien para quien no exista un diccionario? *Diretes*, un diccionario electrónico apto para máquinas. In M.C. Cazorla, M.A. García Aranda, & P. Nuño (Eds.), *Homenaje a Manuel Alvar*. Lugo: Axac.
- Barrios, M<sup>a</sup> Auxiliadora; Bernardos, Socorro. 2007. BaDELE3000: An implementation of the lexical inheritance principle. In Reuther, Tillman; Wanner, Leo. (eds.) *Wiener Slawistischer Almanach. Sonderband 69*. 68 - 77.
- Barrios, M<sup>a</sup> Auxiliadora; Goddard, Cliff. 2013. ‘Degrad verbs’ in Spanish and English Collocations. *Lexical Functions and contrastive NSM semantic analysis. Functions of Language*, 20:2, 219-249.
- Bizer, Christian, Tom Health and Tim Berners-Lee. 2011. In Amit P. Sheth (ed.), *Linked Data: The Story so Far: Semantic Services, Interoperability and Web Applications: Emerging Concepts*. Ringgold Inc, Portland, 1–23.
- Boguslavsky I. 2011. Semantic analysis based on linguistic and ontological resources. *Proceedings of the 5th International Conference on Meaning-Text Theory (MTT'2011)*. Barcelona, September 8–9, 2011. pp. 25–36.
- Bosque I. 2004. REDES. Diccionario combinatorio del español contemporáneo. Las palabras en su contexto. Ediciones SM, Madrid.
- Bosque I. 2006. Diccionario combinatorio PRÁCTICO del español contemporáneo. Las palabras en su contexto. Ediciones SM, Madrid.
- Bosque-Gil, Julia, Jorge Gracia, Elena Montiel-Ponsoda y Guadalupe Aguado-de-Cea. 2016. Modelling multilingual lexicographic resources for the web of data: the K dictionaries case. In Kenerman, Ilan, Iztok Kosem, Simon Krek y Lars Trap-Jensen (eds.), *Proceedings GLOBALEX Lexicographic Resources for Human Language Technology* [http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-GLOBALEX\\_Proceedings-v2.pdf](http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-GLOBALEX_Proceedings-v2.pdf)
- Bosque-Gil, Julia, Jorge Gracia, Elena Montiel-Ponsoda y Asunción Gómez-Pérez. 2018. Models to represent linguistic linked data, *Natural Language Engineering*, 24(6), pp. 811-859. doi:10.1017/S1351324918000347
- De Schryver, G.-M. 2003. Lexicographers’ Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16(2):143–199.
- Dziemianko, Anna. 2018. Electronic dictionaries. In Pedro A. Fuertes Oliveras (ed.), *The Routledge handbook of Lexicography*. Routledge, Oxon.
- Fonseca, A., Sadat, F., and Lareau, F. 2016. A lexical ontology to represent lexical functions. In Proceedings of LangOnto’16 at LREC, Portoroz’.
- Jousse, A.-L. 2010. Modèle de structuration des relations lexicales basée sur le formalisme des fonctions lexicales. Ph.D. thesis, Université de Montréal/Université Paris 7.
- L’Homme , M.C. (2008). Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés. *Traduire*, 217, 78-103.
- Mel’čuk, I. 1996. Lexical Functions: A Tool for the Description of Lexical Relations in Lexicon. *Lexical Functions in Lexicography and Natural Language Processing*, edited by L. Wanner, 37–102. Amsterdam: John Benjamins Publishing Company.
- Mel’čuk I. 2014. *Semantics. From meaning to text*. Volume 3. John Benjamins Publishing Company. Amsterdam/Philadelphia.

- Mel'čuk I., and L. Wanner. 1996. Lexical Functions and Lexical Inheritance. *Lexical Functions in Lexicography and Natural Language Processing*, edited by L. Wanner, 209-278. Amsterdam: John Benjamins Publishing Company.
- Mel'čuk I., and L. Wanner. 2001. Towards a Lexicographic Approach to Lexical Transfer in Machine Translation (Illustrated by the German-Russian Language Pair). *Machine Translation* 16:21–87.
- Ortego Antón M.T. 2011. La compilación de DiCoEnviro en español. *Las TIC: Presente y futuro en el análisis de corpus*. Universitat Politècnica de València
- Polguère, A. 1998. Pour un model stratifié de la lexicalisation en génération de texte. *Traitement Automatique des Langues* 39 (2):57–76.
- Streiter, O. 1996. Linguistic Modeling for Multilingual Machine Translation. Aachen: Shaker Verlag.
- Wanner, L. 1996. *Lexical Functions in Lexicography and Natural Language Processing*, edited by L. Wanner. Amsterdam: John Benjamins Publishing Company.