EURALEX XIX

Congress of the European Association for Lexicography

Lexicography for inclusion

7-11 September 2021

Ramada Plaza Thraki

Alexandroupolis, Greece

www.euralex2020.gr

Proceedings Book
Volume 1

Edited by Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

EURALEX Proceedings

ISSN 2521-7100

ISBN 978-618-85138-1-5

Edited by: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris

2020 Edition

The Making of the *Diretes* Dictionary: how to develop an e-dictionary based on automatic inheritance

Barrios M. A.

Complutense University, Spain

Abstract

DiRetEs is a Spanish monolingual e-dictionary that contains around 100,000 collocations and semantic relations formalized by means of Lexical Functions (LFs). LFs are different formulas, each one appropriate for a different group of collocations or lexical-semantic relations. This dictionary is based on *BADELE.3000* database. The peculiarities of this database are: a) it was built based on a map of semantic labels, a sort of hyperons of the lemmas; and b) it was designed to implement two principles: the principle of lexical inheritance which claims that most of the words sharing a hyperonym (such as *emotion*) could be present in similar collocations (*to feel joy, sadness, envy,* etc.); and the principle of the domain of LFs which claims that the analysis of LFs domains (which means, the set of words this LF was created for) is useful to predict collocations. The combination of both principles in the design of the database allows the lexicographer to automatically obtain new sets of collocations described by means of LFs; up till now it was applied to only one database, *BADELE*, in only one language, that being Spanish. In this paper we will present the methodological problems in connection with the automatic inheritance we face right now: predicting collocations by semantic labels and rewriting the map of semantic labels.

Keywords: e-dictionary; lexical inheritance; Lexical Functions

1 Introduction: automatic inheritance and dictionaries

DiRetEs is a Spanish monolingual e-dictionary (Barrios 2019) based on *BADELE.3000* database (Barrios and Bernardos 2007). This database includes not only collocations but also paradigmatic lexical relations, such as synonyms, ontological relations, such as *part of*, and speech acts, such as formulaic routines. Theoretical guidelines of the project are based on the Meaning-Text Theory principles (Mel'čuk 1996, 2014; Polguère 1997, 2014; Apresjan 2000) and inspired by some lexicographers works (Atkins and Rundell 2008; Fontenelle 2008; Hanks 2009, 2013; Granger 2012), however the central point is the automatic inheritance of some lexical relations (Mel'čuk and Wanner 1996; Barrios 2010), a concept which we will attempt to explain in the following lines.

Combinatorial Dictionaries (Bosque 2004, 2006) and Explanatory Combinatorial Dictionaries (Mel'čuk et al 1995; Mel'čuk and Polguère 2008), are dictionaries specializing in lexical co-occurrence of words. Lexical inheritance was a very promising proposal of Mel'čuk and Wanner (1996) for these kinds of dictionaries: a way to describe productive lexical relation in a particular lexical field. The authors claim that most of the words sharing a hyperonym (such as 'emotion') could be present in similar collocations (*to feel joy, sadness, envy,* etc.). There is a Lexical Function (LF) attached to these collocations, which means that there is a formal way to express this productive lexical relation: in the Meaning-Text Theory (MTT) these collocations are attached to Oper, a LF useful for light verbs, such as *feel*, *do, make* or *have*. The Mel'čuk and Wanner proposal saw a complementary approach in Reuther (1996) and a lengthy explanation in Milićević (1997). All of them offered a powerful theoretical approach, however they do not apply it in a large dictionary.

Domain inheritance was proposed by Barrios (2009, 2010) as a way to predict productive lexical relations attached to some LFs. It can be understood as a second step in the Mel'čuk and Wanner way. The analysis of LFs domains (which means, the set of words for which this LF was created) is useful to predict collocations: for instance, we can predict that many abstract nouns are related to some light-verbal collocations. If we create the list of entries of our dictionary and first of all classify them by lexical fields, we can see that, as in English and German, the Spanish verb *sentir* (to feel) is present in the emotions field. But we can also predict that *dar* (to give) is productive in the affection field (*dar un beso,* give a kiss; *dar un abrazo*, give a huge) and hit field (*dar una patada,* to kick; *dar un puñetazo*, to punch; *dar un codazo*, to nudge); and *tener* (to have) is present in the disease field (*tener fiebre*, to have a fever; *tener diabetes*, to have diabetes; *tener cáncer*, to have cancer), etc. Therefore the preliminary question is, not only how many lexical fields are predictable for the verb *to feel* and the LF Oper, but primarily how many other lexical fields are predictable in general for light verbs, and consequently, for the LF Oper. Similarly, as we will see, we can reflect on the other LFs which are not related to -light verbs collocations.

Automatic inheritance was proposed as a methodological task based on both principles by Barrios (2010) and implemented when designing *BADELE.3000*. Actually the design of this database was produced to prove that both the principle of lexical inheritance and the principle of domain of LFs were valid. It contained 3,300 nouns (the most frequent in the Spanish spoken in Spain) and 20,700 collocations formalized by means of LFs: around 9,000 lexical relations were obtained automatically and 11,700 were added manually (Barrios, 2010). Up till now the database has grown presently contains 19,845 words and multi-words and 101,988 lexical and semantic relations described by means of Lexical Functions; approximately a third of them, 32,948, were automatically obtained.

At present we are working on a new dictionary based on *BADELE.3000*, called *Diretes*. As far as we know, both

BADELE.3000 and *Diretes* are the only dictionaries that implement the automatic inheritance, there is no other dictionary in any other language in the world that does this. There are some other dictionaries and tools developed in the MTT with LFs, such as *Le Réseau Lexicale du Française*[1] and *Dicouèbe*,[2] the specialized dictionaries *DicoInfo, DicoEnviro* and *JuriDico*,[3] the English and Russian *ETAP-4* dictionaries[4] and the Spanish dictionary of emotions *DICE*.[5] Most of these dictionaries were developed following the MTT proposals, and some of them bring some new perspective to the theoretical model: *Le Réseau Lexicale du Française* is a hand-crafted net where all kinds of lexical relations are presented in a cognitive way (Polguère 2014), developed by a knowledge based lexicographic editor (Nabil et al 2012); *DicoEnviro* allows the implementation of the FrameNet methodology (L'Homme 2016), and ETAP-4 applies some Moscow School techniques to the automatic translation (Apresjan et al 2002). However, as mentioned before, none of them implement the automatic inheritance.

From our point of view, the automatic inheritance allows us not only to develop both manually and automatically an e-dictionary but also to verify the validity of both principles. From a practical perspective, implementing both principles allows the lexicographers to save time, to the point that while in 2010 *BADELE*.3000 contained 20,700 collocations, *Dicoèbe* (the French dictionary coetaneous) contained around 24,000 collocations: considering that the Spanish database was developed for one only researcher in one year and the French version was developed for a team during some years, we conclude that the automatic inheritance is convenient for any lexicographic task related to collocations and lexical relations. The only condition necessary is that the database needs to be designed with the ability to produce the inheritance automatically (Barrios and Bernardos 2007).

In this paper we will present two methodological problems that we face in connection with the automatic inheritance: a) the necessity of an accurate prediction by hyperons; and b) the necessity of rewriting a map of semantic labels on which the dictionary is based. We will focus on one theoretical question associated with both problems: the concept of the semantic label with which we work. The *Diretes's* project is on course and the first phase is scheduled to be concluded in one more year. Each phase is distributed around different lexical fields. In this phase we are working on the fields of food, clothes and professions. Lack of space does not allow us to explain in detail the tables of the database, however we will present three of them: the table of Lexical-Semantic Relations, the table of Semantic Predictions and the table of Semantic Labels.

This paper is organized as follows: Section 2 presents how we implement automatic inheritance in *Diretes*; Section 3 is consecrated to the problems we face right now: predicting collocations (3.1), predicting inheritance within the table of predictions (3.2), and the maps of semantic labels and its revision (3.3); Section 4 shows some results; and finally Section 5 summarizes our conclusions.

2 Implementing the automatic inheritance in *Diretes*

Automatic inheritance is present in a high proportion in *Diretes*. As previously mentioned, presently we have 32,948 lexical relations not only automatically obtained but also automatically formalized. We must emphasize that this inheritance is possible for collocations, not for paradigmatic or ontological relations. In order to visualize how the automatic inheritance in the dictionary is applied, we will present an extract from the table of the Lexical Relations.

Figure 1 shows several cases of different words associated to *pan* (bread). The first column of the table of Lexical Relations shows an internal registration number attached to each collocation of the dictionary.

The three following columns contain the most significant information. Indeed, the second one shows the LF associated with each one of the collocations. As we have explained before, the LF is a function proposed by the MTT for productive lexical and semantic relations; for instance, the second row below, underlined in red, contains the LF CausFunc_0, which means "to cause something to exist". The third column contains the lemmas and its grammatical information: respectively, word class and morphological features, for instance *pan, s., masc., sg.* (bread, noun, masc. sg.). It also shows the semantic label, which is a sort of hyperon, such as *producto para comer* (product to eat). The fourth column shows the value of the Lexical or Semantic relation: for the second row below, the relation between the LF CausFunc_0 and *pan* (bread) is expressed by *hacer* (to make), which is a verb automatically inherited. Then, adding the values of these columns we deduce that there is a collocation meaning 'to cause the bread to exist', that is *hacer pan* (to make bread).

The last three columns provide additional information and they are quite useful in the final process of each phase of the project, which is the revision process (it is necessary to revise each formalization for each collocation or semantic relation). The fifth column is related to the lexical automatic inheritance: *sí* (yes) means that the collocation was automatically inherited; *no* means that it was manually added. The information in this column is obtained automatically. The said example underlined in red, *hacer pan* (to make bread), was automatically obtained, and also *elaborar* (to elaborate) and *cocinar pan* (cook bread), underlined in green. However, the collocation with the verb *cocer* (bake), underlined in blue, was manually added, as shown by the word *no*.

The sixth column offers the possibility of rejecting the inheritance (if some default or mistake is detected). Let us imagine that the first expression underlined in green, *cocinar pan* (to cook bread) sounds quite unfamiliar for the reviewer of the dictionary and the fifth column shows that it was automatically inherited (see the results underlined in yellow). What should be the next step? This person must check the collocation in the corpus we use: the dictionaries, some corpus of

[1] https://lexical-systems.atilf.fr/spiderlex/
[2] http://olst.ling.umontreal.ca/dicouebe/index.php
[3] http://olst.ling.umontreal.ca/?page_id=335
[4] http://cl.iitp.ru/
[5] http://www.dicesp.com/paginas

Sketch Engine and other search engines. If the corpora prove that it is a frequent collocation, there is nothing to change as all the values of the column by default are marked with *no*, which means "not reject".
For this concrete case, the data prove that *cocinar pan* is a frequent collocation, however if the reviewer happens to find any mistake, then he should select the *sí* option. All of the *sí* results will be omitted in the web page of the dictionary, however are present in the database. At the end of each phase of the project, we will analyse the set of errors detected in the automatic inheritance: it constitutes rich information for the research on cognitive knowledge of the lexicon. Indeed, it shows how the language distinguishes features of objects or concepts that we do not distinguish consciously (Bosque 2004; Barrios 2010): one example is provided by the false inherited collocation *#ponerse un bolso* (#to put a handbag on), which sounds quite odd in Spanish and English, versus *ponerse una mochila* (to put a backpack on) (more details and a possible explanation in Barrios 2013).

	Id-FA	Id-Argumento	Id-Valor	Here	Rec	ELE
)685	A0Degrad	pan (s. m. sg.) 1 - Producto para comer	enmohecido (adj. c. c.) 1 - Sin a	No	No	B
)684	AntiBon	pan (s. m. sg.) 1 - Producto para comer	correoso (adj. c. c.) 1 - Sin asigr	No	No	B
!758	AntiBon	pan de Calatrava (p. c. - -) 1 - Postre	empalagoso (adj. c. c.) 1 - Sin a	Sí	No	S
;874	AntiBonFinFact0	panificadora (s. f. sg.) 1 - Pequeño electrodoméstico	averiar (v. - -) 1 - Sin asignar	Sí	No	S
;872	AntiBonFinFact0	panificadora (s. f. sg.) 1 - Pequeño electrodoméstico	estropear (v. - -) 1 - Sin asignar	Sí	No	S
.817	AntiBonFinFact0	panificadora (s. f. sg.) 1 - Pequeño electrodoméstico	romper (v. - -) 1 - Sin asignar	No	No	B
)126	AntiMagn	panadería (s. f. sg.) 1 - Local comercial	pequeño (adj. c. c.) 1 - Rasgo fí	No	No	A
)135	AntiMagn-temp	panadería (s. f. sg.) 1 - Local comercial	moderno (adj. c. c.) 1 - Sin asigr	No	No	A
)149	AntiReal1	panadería (s. f. sg.) 1 - Local comercial	incendiar (v. - -) 1 - Sin asignar	No	No	C
;087	AntiReal2	panadería (s. f. sg.) 1 - Local comercial	robar (v. - -) 1 - Sin asignar	Sí	No	S
)689	Bon	pan (s. m. sg.) 1 - Producto para comer	crujiente (adj. c. c.) 1 - Sin asigr	No	No	S
!764	Bon	pan de Calatrava (p. c. - -) 1 - Postre	delicioso (adj. c. c.) 1 - Sin asigr	Sí	No	S
!769	Bon	pan de Calatrava (p. c. - -) 1 - Postre	exquisito (adj. c. c.) 1 - Sin asigr	Sí	No	S
!762	Bon	pan de Calatrava (p. c. - -) 1 - Postre	sabroso (adj. c. c.) 1 - Sin asigna	Sí	No	S
;139	CausAntiBonFact0	panadería (s. f. sg.) 1 - Local comercial	robar (v. - -) 1 - Sin asignar	Sí	No	S
;869	CausDenuevoFact0	panificadora (s. f. sg.) 1 - Pequeño electrodoméstico	arreglar (v. - -) 1 - Acción	Sí	No	S
.818	CausDenuevoFact0	panificadora (s. f. sg.) 1 - Pequeño electrodoméstico	reparar (v. - -) 1 - Sin asignar	No	No	B
'429	CausFact0	panadería (s. f. sg.) 1 - Local comercial	abrir (v. - -) 1 - Acción	Sí	No	S
;483	CausFact1	filete empanado (s. m. sg.) 1 - Carne	poner (v. - -) 1 - Sin asignar	Sí	No	S
)952	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	amasar (v. - -) 1 - Acción	No	No	S
)826	CausFunc0	empanada (s. f. sg.) 1 - Sin asignar	cocer (v. - -) 1 - Sin asignar	No	No	B
)953	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	cocer (v. - -) 1 - Sin asignar	No	No	S
)820	CausFunc0	empanada (s. f. sg.) 1 - Sin asignar	cocinar (v. - -) 1 - Sin asignar	No	No	B
)954	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	cocinar (v. - -) 1 - Sin asignar	Sí	No	S
;373	CausFunc0	panadería (s. f. sg.) 1 - Local comercial	construir (v. - -) 1 - Sin asignar	Sí	No	S
)821	CausFunc0	empanada (s. f. sg.) 1 - Sin asignar	elaborar (v. - -) 1 - Sin asignar	No	No	B
)955	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	elaborar (v. - -) 1 - Sin asignar	Sí	No	S
!760	CausFunc0	pan de Calatrava (p. c. - -) 1 - Postre	elaborar (v. - -) 1 - Sin asignar	Sí	No	S
)817	CausFunc0	empanada (s. f. sg.) 1 - Sin asignar	hacer (v. - -) 1 - Sin asignar	No	No	A
)956	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	hacer (v. - -) 1 - Sin asignar	Sí	No	S
!768	CausFunc0	pan de Calatrava (p. c. - -) 1 - Postre	hacer (v. - -) 1 - Sin asignar	Sí	No	S

Figure 1: Extract from the table of Lexical-Semantic Relations with some lexical relations associated to *pan* (bread)

The last column is manually added and shows the level of Spanish that is appropriate for a student to teach this lexical or semantic relation: A, B and C levels follow the *European Framework of Reference for Languages*. We have added three additional levels in this last column: E (that means for experts), which is the level adequate for terminology; V (that means vocabulary), a level for unfamiliar words for many native speakers which constitutes rich vocabulary present in literature and some books; and S, which means "*sin asignar"* (not assigned), which is a temporal mark (automatically present by default) prior to the selection of the level.

3 Problems arising when implementing automatic inheritance

3.1 Predicting collocations

The first problem that arises when working with automatic inheritance is that all the predictions should be applied automatically before working manually: for instance, the value of $CausFunc_0$ for 'prepared food' is *to make*, for 'music' is *to compose*; and for 'literature' is *to write*. That means that all the nouns that could be labelled as 'prepared food' would combine with *to make*, such as *bread*, *salad, paella* or *soup*. Similarly *to compose* combines with *symphony, song,* and *melody*; and *to write* with *novel, poem* and *essay*. As we will see in section 3.3, *Diretes* has the same map of labels as *BADELE.3000*. In order to save time and effort all the predictions (the relation between $CausFunc_0$ and 'prepare food'/*to make*, etc.) and the inheritances should be done before starting with the manual addition of some other collocations. These predictions are a result of the introspection of the lexicographer via whom data can at times be found in the combinatorial dictionaries (Bosque 2004, 2006). Unfortunately, not all of these predictions are necessarily accessible for all the components of the team working on the dictionary: some predictions demand experience and a strong knowledge

of the MTT model. We will extrapolate on this point.
There are two possible ways to work with the predictions, each one of them aligned with a different difficulty level. The easiest way to predict productive relations is quite similar to the methodology applied by Mel'čuk and Wanner (1996): suppose the team is working with words such as *potato, tomato* and *cucumber*, and they observe that *to plant* is a productive verb for these nouns. As we can predict this collocation for all the set of vegetables, we write *to plant* in the table of predictions and we describe the appropriated LF ($CausFunc_0$, because, as we stated before, it means "to cause something to exist"). After the prediction, we apply the automatic inheritance and we obtain the fifty corresponding collocations formalized by means of this LF. We have avoided writing them manually. This methodology is useful for non MTT-experts and applicable to different lexical fields and different LFs. The second methodology, however, is more difficult. Applying the prediction by domains of LFs (Barrios 2010) as a first step, as we have said previously, demands not only a strong knowledge of the MTT model but also the ability to go from abstraction (the meaning of some LFs) to the lexicon (the potential domain for each one of these LFs). We will attempt to explain the process.
Firstly we should think about each one of the LFs and their potential meaning. Consider the case of $CausFunc_0$ and its meaning, 'to create'. We should calculate how many lexical fields could be the potential domain for this LF. In order to reach the answer we connect the extra-linguistic knowledge with the linguistic knowledge, and we conclude that if we can create objects, tools, food, leisure products, etc., at least one verb necessarily exists that expresses the meaning 'to create' for all the words naming these realities. In Spanish we describe 164 predictions for this LF, as Figure 2 shows; the relation between the second and fourth columns could be literally translated into English as *fruit/cultivate*, *animal cabin/ build*, *building/raise, rule/dictate, theoretical principle/discover, energy/produce*, etc. The case of *pescado, capturar/pescar* (*fish, catch*) is underlined in red as in Spanish we can accept this is a particular case of 'creation': the word for the animal *pez* (fish) differs from the word for the food *pescado* (fish), similar to the English words *pig/pork*. That is the reason we associate $CausFunc_0$ to *pescar un pez* (to catch a fish), as it means "to create the food fish".
For some other lexical relations the LF $CausFunc_0$ is not adequate, such as the cases of *consensuar/ negociar una norma jurídica* (to agree on a legal rule) or *trazar una obra pública* (to plot a public work). When some people agree on a legal rule, these people do not create a new rule but the conditions by which this rule can be dictated. A similar situation involves the action of plotting a public work. We cannot use $CausFunc_0$ however, is there any other way to formalize these lexical relations? When there is no LF adequate for any productive relations, in the MTT model it is possible to provide a new way to formalize them: if any researcher should discover a new productive relation that could be understood as a LF, he can propose a new LF which will be classified as a non-standard LF. There is a non-standard LF fairly close to $CausFunc_0$, called EssayerCausFunc0, which means "to try to cause something to exist" (Essayer was proposed by Polguère 2007). We translate the French verb *essayer* to the equivalent Spanish one, *intentar*; consequently, as the examples underlined in blue (Figure 2) show, we work with $IntentarCausFunc_0$.
All the examples underlined in red and blue in the Figure 2 exemplify how the lexicographer should have not only a high level of MTT knowledge in order to predict the domain of standard LFs, but also familiarity with non-standard LFs:

Relaciones semánticas | Predicciones semánticas

Id-PS	Id-ES	Id-FA	Id-A	Heredado
628	Norma jurídica	CausFunc0	promulgar (v. - -) 1 - Sin asignar	No
649	Norma jurídica	IntentarCausFunc0	negociar (v. - -) 1 - Sin asignar	No
648	Norma jurídica	IntentarCausFunc0	consensuar (v. - -) 1 - Sin asignar	No
627	Norma jurídica	CausFunc0	emitir (v. - -) 1 - Sin asignar	No
625	Norma jurídica	CausFunc0	aprobar (v. - -) 1 - Acción	No
677	Obra Pública	CausFunc0	construir (v. - -) 1 - Sin asignar	No
686	Obra Pública	IntentarCausFunc0	trazar (v. - -) 1 - Sin asignar	No
2614	Pequeño electrodoméstico	CausFunc0	montar (v. - -) 1 - Sin asignar	Sí
717	Percepción	CausFunc0	causar (v. - -) 1 - Sin asignar	Sí
718	Percepción	CausFunc0	desprender (v. - -) 1 - Sin asignar	No
723	Pescado	CausFunc0	pescar (v. - -) 1 - Sin asignar	No
722	Pescado	CausFunc0	capturar (v. - -) 1 - Sin asignar	No
745	Pescado azul	CausFunc0	pescar (v. - -) 1 - Sin asignar	Sí
744	Pescado azul	CausFunc0	capturar (v. - -) 1 - Sin asignar	Sí
767	Pescado blanco	CausFunc0	pescar (v. - -) 1 - Sin asignar	Sí
766	Pescado blanco	CausFunc0	capturar (v. - -) 1 - Sin asignar	Sí
1278	Pieza de bisutería o joyería	CausFunc0	diseñar (v. - -) 1 - Sin asignar	No
24004	Postre	CausFunc0	preparar (v. - -) 1 - Sin asignar	Sí
23194	Postre	CausFunc0	elaborar (v. - -) 1 - Sin asignar	Sí
24003	Postre	CausFunc0	hacer (v. - -) 1 - Sin asignar	Sí
831	Proceso	CausFunc0	causar (v. - -) 1 - Sin asignar	Sí
835	Proceso humano	CausFunc0	causar (v. - -) 1 - Sin asignar	Sí
838	Proceso médico	CausFunc0	causar (v. - -) 1 - Sin asignar	Sí
1292	Producto cinematográfico	CausFunc0	dirigir (v. - -) 1 - Sin asignar	No
1294	Producto cinematográfico	CausFunc0	producir (v. - -) 1 - Sin asignar	No
1293	Producto cinematográfico	CausFunc0	hacer (v. - -) 1 - Sin asignar	No
851	Producto energético	CausFunc0	producir (v. - -) 1 - Sin asignar	No
1241	Producto para comer	CausFunc0	cocinar (v. - -) 1 - Sin asignar	No
1243	Producto para comer	CausFunc0	hacer (v. - -) 1 - Sin asignar	No
1244	Producto para comer	CausFunc0	preparar (v. - -) 1 - Sin asignar	No
1242	Producto para comer	CausFunc0	elaborar (v. - -) 1 - Sin asignar	No

Registro: 1 de 156 | Filtrado | Buscar

Figure 2: Extract from the table Semantic Predictions, some data prior to the inheritance of $CausFunc_0$ ('to create')

The last rows of Figure 2 (underlined in green) show the collocations predicted for $CausFunc_0$ and *producto para comer* (product to eat): *cocinar, hacer, preparar* and *elaborar el pan* (to cook, to make, to prepare and to produce the bread). Now we can return to Figure 1 and check that *bread* has been labelled as *producto para comer* (product to eat), so the collocations underlined in red and green in Figure 1, *hacer, elaborar* and *cocinar el pan* (to make, to produce and to cook the bread) are also present in Figure 2, because they were automatically obtained from the Table of Semantic Predictions; while *cocer* (bake) (underlined in blue in the Figure 1) is not, because it was manually added directly to the Table of Lexical-Semantic Relations.

In *Diretes*, only for the LF $CausFunc_0$, we have 90 different semantic labels and 156 inheritable collocations (see the number underlined in pink, Figure 2), all of them predicted by the lexicographer's introspection. After the inheritance, we obtained 2,447 collocations related to $CausFunc_0$. If we consider that the total number of collocations for this LF is 4,901, we observe that almost 50 percent was automatically obtained. Once again the data show that the automatic inheritance saves time and effort. However, on the other hand, this small example proves that any project applying the automatic inheritance demands lexicographers with a strong knowledge of the MTT model and of the Lexicology and Semantics of the natural language object of the dictionary. We will comment on the examples underlined in yellow in the next section.

3.2 Predicting the automatic inheritance within the table of Semantic Predictions

There are some Semantic Predictions that can be inherited within the table of Semantic Prediction, as the last column in Figure 2 shows. That column contains two examples underlined in orange with the value *no*, and four examples underlined in yellow with the value *yes*. If we look at the preceding columns, we see that the value *no* is attached to *pescado* (fish, in the second column underlined in red) and the value *yes* is attached to *pescado azul* (blue fish) and *pescado blanco* (white fish), underlined in yellow. That means that the verbs *capturar, pescar* (*fish/catch*) were added manually for 'fish', and automatically inherited by 'white fish' and 'blue fish' within the table of semantic predictions. Thus we can produce not only inherited collocations but also inherited predictions.

At this point, we should say that we attempt to collect mostly linguistic information and that we also attempt to differentiate between linguistic items and ontological information. Subsequently, the question that arises is: is the relation between *fish* and *blue fish* linguistic or extra-linguistic? Is it any piece of information of real life or is it an expression we should work with?

In the MTT the relation between concepts such as 'fish' and 'blue fish' is close to the LF Gener, which means 'generic concept'. This LF is conceptually close to a hyperonym but it is not a hyperonym, because it does not form explicit semantic relations (such as the hypernym does) but a lexical relation, such as the one between *republic* and *state* (we can say *republican state*), or *liquid* and *substance* (we say *liquid substance*), or *process* and *regeneration* (we say *process of regeneration*) (examples taken from Mel'čuk 2015: 194). Then, in order to know if a word such as *fish* and its relation with *salmon* is a candidate for Gener, we attempt to build an expression for both words, such as *the fish salmon*. As it does work, we could formalize this relation such as (1) shows:

(1) Gener (salmon) = fish

As the word *blue* is a predicate that combines with *fish* it cannot be a value of Gener. As far as we know, within the MTT model, the LF Gener is only explored at the lemma level, that means describing words such as *salmon, sardine, hake* or *sea bass*, and its relation with *fish*, as the French Dictionary *Le reseau lexicale* (*LRL*) does (see the French entries *saumon, sardine, carpe, loup de mer*, etc., which contain formalizations like the one proposed in (1).

In the next Section we will analyse with more detail this LF and its relation with the concept of a semantic label. We will also attempt to answer the mentioned question, is the distinction between blue fish and white fish linguistic or extra-linguistic?

In *Diretes* we do not work with the LF Gener but with a concept close to this LF called *semantic label* (Milicevik 1997; Polguère 2003, 2011). A semantic label is a descriptive tool, equivalent to a hyperon and to the genus in the Aristotelic terminology. Milicevik points out that semantic labels are useful and well known in Artificial Intelligence, but there is no theoretical linguistic approach in this area except for technical applications. The semantic label of a word is usually the central meaning of this word (Milicevik 1997: 36-37), and can be taken from the definitions of good dictionaries (Polguère 2011). Milicevik (1997: 38-39) points out that there are three conditions for any semantic label: a) it takes up a central position in the meaning of the word (such as the meaning 'emotion' in *joy*); b) it reflects sufficiently enough the co-occurrence of this word (*to feel an emotion, to suffer an emotion*; *to enjoy an emotion*, etc.); c) it is useful to label a group of words (such as 'emotion' and *joy, anger, fear, envy*, etc.).

One particularity of the French *LRL* that Polguère develops is that the semantic labels are not directly present: they work with the LF Gener and with a sort of short paraphrase which expresses a central meaning of the word. Polguère (2011) explains that while WordNets works with the concept of synsets, which combines meaning and grammatical information, the hierarchy of semantic labels he proposes should be attached only to the meaning. Curiously enough, these paraphrases are expressed necessarily attached to a grammatical role, such as shown by some *LRL* examples: *admirateur* (admirer) and *admiratif* (admired) share the meaning 'who shows a feeling' (from this point we will translate the French *LRL* paraphrases into English). Compare this paraphrase with the one of *admirable*, 'which has a particular feature'; *admiration*, 'feeling', and *admirer* (to admire), 'to feel a feeling'. These examples and the other lemmas of the *LRL* demonstrate that the paraphrases expressing the semantic label have been redacted according to the syntactic function of the word described.

As we will explain in Section 3.3, we work with a slightly different concept of Milicevik's semantic label.

3.3 Revising the table of semantic labels

As mentioned in the previous Section, regarding the semantic labels, the French dictionary *LRL* shows some differences with *Diretes*. The *LRL* works with the relation between the word *fish*, Gener and *salmon* (see (1), and with paraphrases such as 'relatif à un animal' (related to an animal) (see the entry *saumon* in the LRL). *Diretes*, as Figure 2 shows (come back to the examples underlined in yellow), contains not only 'fish' as a semantic label, but also 'blue fish' and 'white fish'. What is the reason for this? The answer is that we attempt to obtain a higher granularity in our description in order to exploit as much as possible the automatic inheritance.

Figure 3 presents an extract from our table of semantic labels. The first column corresponds to what we call *raíz* (root) (underlined in green), which is the first distinction between words attached to entities (labelled as '*ser'*, being) and words attached to predicates or abstract nouns (labelled as '*concepto'*, concept). We have a count of nine levels in our table of semantic labels (see the column underlined in pink). This table was present in *BADELE.3000* however we are adding some new semantic labels, although presently no great changes affect its structure. The first levels respond to conceptual distinctions, and the last ones contain semantic labels defined by linguistic features: we will try to explain this distinction via the example of the words naming different types of food and sweets.

From the lowest level to the highest, the following table shows how we classify different types of labels. In the original database we counted on the label '*dulces y postres'* (sweets, created for *cake*, *ice-cream*, etc.) and the label '*platos preparados*' (prepared dishes, for *paella*, *croquetas* etc.). Both labels are labelled as '*alimento preparado'* (prepared food), which in turn is labelled as '*producto de consumo*' (consumed product), which in turn is labelled as 'producto' (product) (see the examples underlined in red in Figure 3).

Ver Etiquetas Semánticas

Raíz	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7	Nivel 8	Nivel 9
Ser	Ente	Producto	Producto artístico	Tipo de obra de teatro, novela o pe					
Ser	Ente	Producto	Producto artístico	Tipo de obra literaria					
Ser	Ente	Producto	Producto de consumo	Aderezo					
Ser	Ente	Producto	Producto de consumo	Alimento	Ahumados				
Ser	Ente	Producto	Producto de consumo	Alimento	Alimento vegetal	Cereal			
Ser	Ente	Producto	Producto de consumo	Alimento	Alimento vegetal	Legumbre			
Ser	Ente	Producto	Producto de consumo	Alimento	Alimento vegetal	Verdura			
Ser	Ente	Producto	Producto de consumo	Alimento	Carne				
Ser	Ente	Producto	Producto de consumo	Alimento	Embutido				
Ser	Ente	Producto	Producto de consumo	Alimento	Fiambre				
Ser	Ente	Producto	Producto de consumo	Alimento	Fruto	Baya			
Ser	Ente	Producto	Producto de consumo	Alimento	Fruto	Fruta			
Ser	Ente	Producto	Producto de consumo	Alimento	Fruto	Fruto seco			
Ser	Ente	Producto	Producto de consumo	Alimento	Marisco				
Ser	Ente	Producto	Producto de consumo	Alimento	Pasta				
Ser	Ente	Producto	Producto de consumo	Alimento	Pescado	Pescado azul			
Ser	Ente	Producto	Producto de consumo	Alimento	Pescado	Pescado blanco			
Ser	Ente	Producto	Producto de consumo	Alimento	Pulpos y calamares				
Ser	Ente	Producto	Producto de consumo	Alimento preparado	Dulces y postres	Dulce	Dulce cocido		
Ser	Ente	Producto	Producto de consumo	Alimento preparado	Dulces y postres	Dulce	Dulce congelado		
Ser	Ente	Producto	Producto de consumo	Alimento preparado	Dulces y postres	Dulce	Dulce frito		
Ser	Ente	Producto	Producto de consumo	Alimento preparado	Dulces y postres	Dulce	Dulce horneado		
Ser	Ente	Producto	Producto de consumo	Alimento preparado	Dulces y postres	Postre			
Ser	Ente	Producto	Producto de consumo	Alimento preparado	Plato preparado				
Ser	Ente	Producto	Producto de consumo	Alimento preparado	Refrigerio				
Ser	Ente	Producto	Producto de consumo	Bebida	Bebida alcohólica				
Ser	Ente	Producto	Producto de consumo	Bebida	Infusión				
Ser	Ente	Producto	Producto de consumo	Producto para la alimentación					
Ser	Ente	Producto	Producto de consumo	Producto que genera adicción					
Ser	Ente	Producto	Producto de consumo	Salsa y condimentos	Condimento	Especia			
Ser	Ente	Producto	Producto de consumo	Salsa y condimentos	Condimento	Hierba			
Ser	Ente	Producto	Producto de consumo	Salsa y condimentos	Producto de aderezo				
Ser	Ente	Producto	Producto de consumo	Salsa y condimentos	Salsa				
Ser	Ente	Producto	Producto de innovación						
Ser	Ente	Producto	Producto de la actividad del homl						
Ser	Ente	Producto	Producto de limpieza						

gistro: 1 de 355 Sin filtro Buscar

Figure 3: Extract from the table of Semantic Labels: preliminary data that explains the inheritance

'Producto' ('product') (see the third column in Figure 3) is an ontological label, and it can be attached to any object in the world which was produced by any person. Even if we think about the verb *producir* (to produce) as a candidate for collocations with the nouns naming these realities, it does not necessarily work on the linguistic level: some nouns that could be labelled as 'product' combine with this verb (*producir pizza*, to produce pizza; *producir mesas*, to produce tables; *producir petróleo*, to produce petrol) and some others do not (#*producir una infusion*, to produce an infusion; #*producir una escultura*, to produce a sculpture; #*producir agua sucia*, #to produce dirty water). This apparently incoherent behaviour points out however a coherent rule: any concept (such as that the Spanish concept '*producir'*, to produce) can take on a different role to its equivalent word. As the Spanish verb *producir* is attached to the context of a business production, when someone makes a tea at home, even if this person is producing a tea in same way, do not use the verb *producir*. Something similar happens within the fourth column: the label 'consume product' can be attached to the verb *consumir* (consume), however in Spanish, not all the words that could be labelled as '*producto de consumo'* (consumable product) necessarily combine with this verb.

However, within the fifth column, the label '*alimento preparado*' (prepared food, underlined in red in Figure 3) was created for words such as *paella*, *croqueta*, etc. All of them combine with the verb *preparar* (to prepare). Similarly, there are different collocations for the following semantic labels.

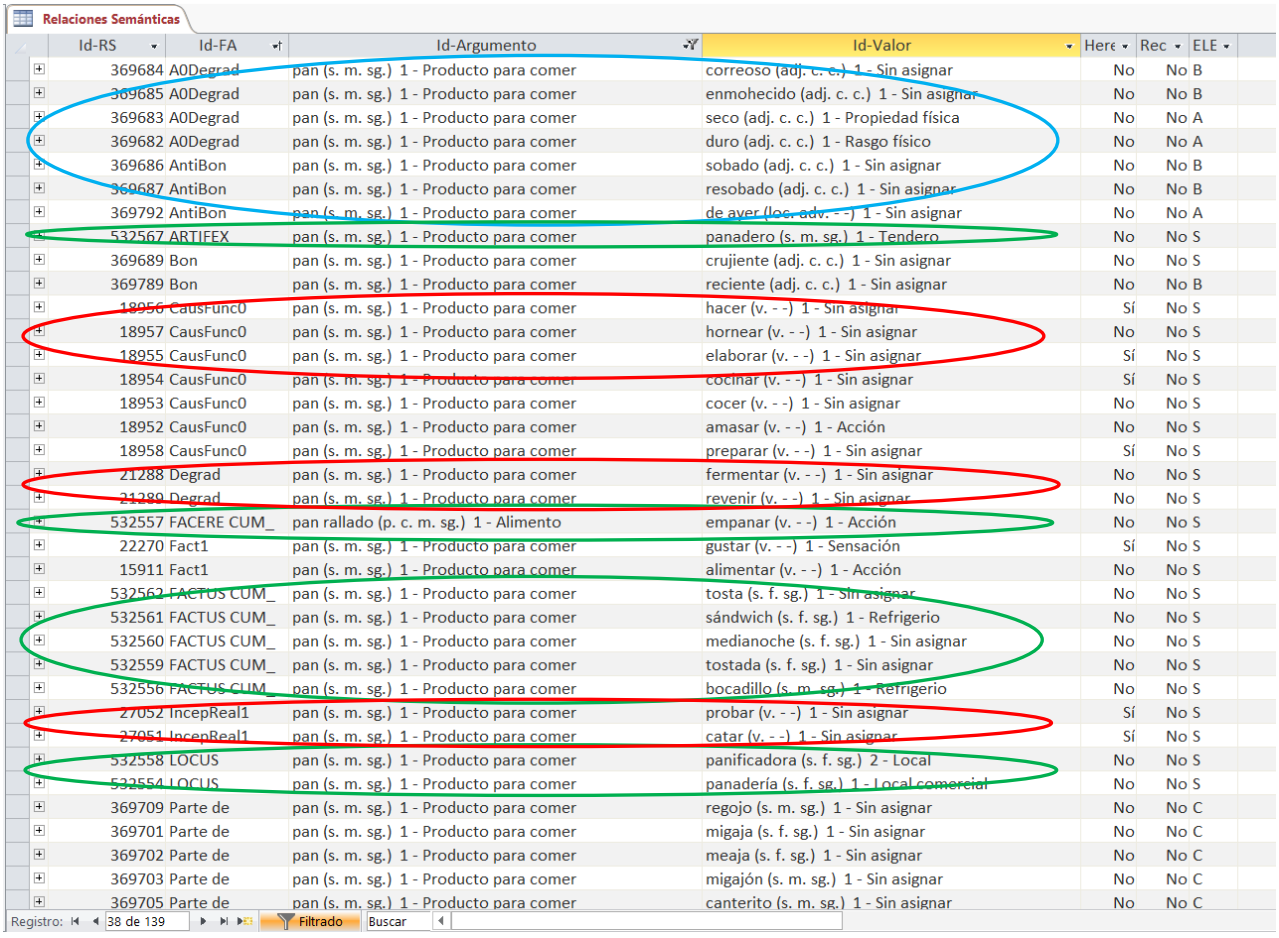
Let us come back now to the case of 'fish' (underlined in blue in Figure 3). We check in our corpus and see that *salmon* combines with *graso* (fatty) (we say *el salmon es graso*, the salmon is fatty). This combination is attached to the LF Pred (which means "to be") and we also observe that it is productive not only for *salmon* but for any blue fish. We conclude

that there is at least one collocation liable to be inherited (*graso*, fatty plus some nouns of fish), and we create a new semantic label for this set of nouns, which is 'blue fish'. The remaining nouns of fish will be labelled as 'white fish' (for them the combination with *graso* (fatty) is unusual in Spanish, and consequently, it will not be inherited).
As a result, our rule is quite simple: if there is at least one collocation productive for a group of words, we create a semantic label for them. This methodology allows us: a) to implement the automatic inheritance; and b) painting a map of semantic labels (some of them, as mentioned in section 2, unknown for our linguistic conscience) which is different for any ontology as it is partially based on concepts and mostly on linguistic behaviour.
The reason for this mixed organization is that we require our database to be useful not only for dictionaries but also for terminology and for ontologies. As is well known, working with terms implies working with concepts, because terms use to be monosemic, and at this level meanings overlap with concepts. Then, from level 1 to level 3 of our table of semantic labels, we work mostly with semantic labels attached to concepts. From level 5 to 8 we find semantic labels defined mostly based on linguistic features. In level 4, presently, we find semantic labels mixed (some of them are attached to concepts, some are based on linguistic behaviour). Finally level 9 is preserved for the future work on terminology.
Many of the semantic labels we work with were present in *BADELE.3000* but, the more we work on a particular lexical field, the more detailed is the semantic description of the words described. That implies that at times we discover new semantic labels and we add them to the table of the semantic labels of the e-dictionary. That was the case of *sweets*: in Spanish we use the word *dulce* (sweet) for *pasteles* (pie), *bizcochos* (cake), *galletas* (cookies), *natillas* (custard), etc. Some of them are baked, some of them cooked, some of them fried and some others are made without heat, but *a priori* we did not distinguish them because we do not use any Spanish expression equivalent, for instance, to the English expression *baked sweet*. We could create a semantic label such as 'dulce hecho con calor' ('sweet made with heat') but this potential semantic label raises two problems: on one hand, paraphrases such as this one are less intuitive than any word or expression in Spanish; on the other hand, we cannot apply the inheritance of the three verbs, *cocer* (to cook), *honear* (to bake) and *freír* (to fry), to all the nouns of sweets we make with heat. There is a simple solution: we can divide the nouns of sweets made with heat into three groups, each one of them for each verb. Then, from the original semantic label '*dulces y postres'* (sweets) we obtain four different semantic labels: *'dulces horneados'* (baked sweets, such as cake), which combine with *hornear* (bake); *'dulces cocidos'* (cooked sweets, such as custard) which combines with *cocer* (to cook); *'dulces fritos'* (fried sweets, such as churros), which combines with *freír* (to fry); and *dulces congelados'* (frozen sweets, such as ice-cream) which combines with *congelar* (to freeze) and *derretir* (to melt). Note that some of our paraphrases sound quite unusual in Spanish, such as *dulces cocidos*, however they are explicative enough and useful for the inheritance.
We can conclude that our concept of semantic label presents some differences to Milićević's concept. The semantic label we work with demands three conditions: a) it is useful for a group of words (such as 'blue fish' for *salmon, sardine, tuna*, and 'baked sweet' for *pie, cake, cookies*); b) the label is a meaning (such as 'fish' or 'sweet') or a restricted meaning (such as 'blue fish', or 'cooked sweet'), which implies that it is not necessarily part of the definition (note that 'fish' is part of the meaning of *salmon* but 'blue fish' is not); c) the label should reflect at least one co-occurrence (*such as the salmon is fatty, to bake the cookies, to cook the custard)*.

4 Results

In *Diretes*, presently, we have made 1,614 predictions, 819 were predicted by introspection and 795 were inherited from some other predictions, which means that almost half predictions were automatically obtained from some manually added predictions. A total of 233 semantic labels were involved in these predictions. All of these semantic labels were used when labelling 7,774 words and multi-words, which is the number of entries of *Diretes* labelled up until the present (12,069 words are labelled temporally as *sin asignar*, not allocated yet, most of them verbs, adjectives and adverbs). A total of 101,988 lexical and semantic relations were described by means of Lexical Functions, and 32,948 of them were automatically obtained and formalized.
After applying automatic inheritance, we manually add the rest of the lexical-semantic relations. Figure 4 shows some of the 139 lexical relations we formalized around the word *pan* (bread). We use mostly standard LFs, some of them adjectival (see the examples underlined in blue). The first one is AntiBon, which means "bad", and is applied to relations such as *pan sobado*/*resobado* (rubbing bread), *pan de ayer* (lit. bread from yesterday, which is a not fresh bread in our culture). A second adjectival LF is A_0Degrad, which means "damaged", and is applied to relations such as *pan correoso* (lit. flexible bread), *pan enmohecido* (moldy bread), *pan seco* (old dry bread), *pan duro* (hard bread).
There are some verbal LFs (underlined in red), such as $CausFunc_0$, which means "create", applied to *elaborar pan* (to produce the bread), *hacer pan* (to make the bread), etc. A second verbal LF is Degrad, which means "degenerate", applied to *fermentarse el pan* (to ferment the bread) and to *revenirse el pan* (to go off the bread). The third one is $IncepReal_1$ that means "to start doing what is expected to be done with this object": *probar el pan* (to try the bread), *catar el pan* (to taste the bread).
The set of LFs underlined in green, however, does not correspond to standard LFs. As mentioned in the first section, the set of standard LFs is a powerful tool for the description of lexical-semantic relations, however this set is not complete. We can create non-standard LFs if it is necessary; consider that these LFs can be empirically found (Mel'cuk 1996: 45), and that they are a sort of candidate for new standard LFs.
In our dictionary there is a problem relating to the richness of semantic and ontological relations and the lack of standard LFs: the actual set of standard LFs does not allow reflection on the relation between words such as *pan* (bread) and *panadería* (bakery), *empanada* (patty), *panificadora* (bread maker), *empanar* (to bread), *panadero* (baker), *barra de pan* (baguette), etc. It is necessary therefore to create non-standard Lexical Functions for them.

In order to propose a new non-standard LF we should consider two particularities of the standard LFs: there is a significant diversity between the values of any standard LF and there is necessarily a large number of cases for any standard LF (Polguère 2007: 52-53). The author claims that there are LFs which have been proven to satisfy the preliminary conditions, called breadth and diversity, respectively, but only for one language. This set of LFs are then called *local standard LFs*, and we should write them in the local language in which they exist. Polguère summarizes some proposals of Èrastov in 1968 based on the lexicographic task, which saw the LFs Cap, Culm and Prox recognized as standard LFs and added to the MTT model. Polguère (2007) proposes De nouveau as a new non-Standard LF meaning "again", and claims that we need to develop dictionaries in many other languages before proposing it as a universal and standard LF.
In *Diretes* we work with some non-standard LFs that are candidates to be labelled as local standard LFs. Figure 4 shows some of them, ARTIFEX, FACERE CUM, FACTUS CUM and LOCAL (underlined in green):



Relaciones Semánticas

Id-RS	Id-FA	Id-Argumento	Id-Valor	Here	Rec	ELE
369684	A0Degrad	pan (s. m. sg.) 1 - Producto para comer	correoso (adj. c. c.) 1 - Sin asignar	No	No	B
369685	A0Degrad	pan (s. m. sg.) 1 - Producto para comer	enmohecido (adj. c. c.) 1 - Sin asignar	No	No	B
369683	A0Degrad	pan (s. m. sg.) 1 - Producto para comer	seco (adj. c. c.) 1 - Propiedad física	No	No	A
369682	A0Degrad	pan (s. m. sg.) 1 - Producto para comer	duro (adj. c. c.) 1 - Rasgo físico	No	No	A
369686	AntiBon	pan (s. m. sg.) 1 - Producto para comer	sobado (adj. c. c.) 1 - Sin asignar	No	No	B
369687	AntiBon	pan (s. m. sg.) 1 - Producto para comer	resobado (adj. c. c.) 1 - Sin asignar	No	No	B
369792	AntiBon	pan (s. m. sg.) 1 - Producto para comer	de ayer (loc. adv. - -) 1 - Sin asignar	No	No	A
532567	ARTIFEX	pan (s. m. sg.) 1 - Producto para comer	panadero (s. m. sg.) 1 - Tendero	No	No	S
369689	Bon	pan (s. m. sg.) 1 - Producto para comer	crujiente (adj. c. c.) 1 - Sin asignar	No	No	S
369789	Bon	pan (s. m. sg.) 1 - Producto para comer	reciente (adj. c. c.) 1 - Sin asignar	No	No	B
18956	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	hacer (v. - -) 1 - Sin asignar	Sí	No	S
18957	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	hornear (v. - -) 1 - Sin asignar	No	No	S
18955	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	elaborar (v. - -) 1 - Sin asignar	Sí	No	S
18954	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	cocinar (v. - -) 1 - Sin asignar	Sí	No	S
18953	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	cocer (v. - -) 1 - Sin asignar	No	No	S
18952	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	amasar (v. - -) 1 - Acción	No	No	S
18958	CausFunc0	pan (s. m. sg.) 1 - Producto para comer	preparar (v. - -) 1 - Sin asignar	Sí	No	S
21288	Degrad	pan (s. m. sg.) 1 - Producto para comer	fermentar (v. - -) 1 - Sin asignar	No	No	S
21289	Degrad	pan (s. m. sg.) 1 - Producto para comer	revenir (v. - -) 1 - Sin asignar	No	No	S
532557	FACERE CUM_	pan rallado (p. c. m. sg.) 1 - Alimento	empanar (v. - -) 1 - Acción	No	No	S
22270	Fact1	pan (s. m. sg.) 1 - Producto para comer	gustar (v. - -) 1 - Sensación	Sí	No	S
15911	Fact1	pan (s. m. sg.) 1 - Producto para comer	alimentar (v. - -) 1 - Acción	No	No	S
532562	FACTUS CUM_	pan (s. m. sg.) 1 - Producto para comer	tosta (s. f. sg.) 1 - Sin asignar	No	No	S
532561	FACTUS CUM_	pan (s. m. sg.) 1 - Producto para comer	sándwich (s. f. sg.) 1 - Refrigerio	No	No	S
532560	FACTUS CUM_	pan (s. m. sg.) 1 - Producto para comer	medianoche (s. f. sg.) 1 - Sin asignar	No	No	S
532559	FACTUS CUM_	pan (s. m. sg.) 1 - Producto para comer	tostada (s. f. sg.) 1 - Sin asignar	No	No	S
532556	FACTUS CUM_	pan (s. m. sg.) 1 - Producto para comer	bocadillo (s. m. sg.) 1 - Refrigerio	No	No	S
27052	IncepReal1	pan (s. m. sg.) 1 - Producto para comer	probar (v. - -) 1 - Sin asignar	Sí	No	S
27051	IncepReal1	pan (s. m. sg.) 1 - Producto para comer	catar (v. - -) 1 - Sin asignar	Sí	No	S
532558	LOCUS	pan (s. m. sg.) 1 - Producto para comer	panificadora (s. f. sg.) 2 - Local	No	No	S
532554	LOCUS	pan (s. m. sg.) 1 - Producto para comer	panadería (s. f. sg.) 1 - Local comercial	No	No	S
369709	Parte de	pan (s. m. sg.) 1 - Producto para comer	regojo (s. m. sg.) 1 - Sin asignar	No	No	C
369701	Parte de	pan (s. m. sg.) 1 - Producto para comer	migaja (s. f. sg.) 1 - Sin asignar	No	No	C
369702	Parte de	pan (s. m. sg.) 1 - Producto para comer	meaja (s. f. sg.) 1 - Sin asignar	No	No	C
369703	Parte de	pan (s. m. sg.) 1 - Producto para comer	migajón (s. m. sg.) 1 - Sin asignar	No	No	C
369705	Parte de	pan (s. m. sg.) 1 - Producto para comer	canterito (s. m. sg.) 1 - Sin asignar	No	No	C

Registro: 38 de 139 Filtrado Buscar

Figure 4: Extract from the table of Lexical-Semantic Relations: 38 from a total of 139 lexical-semantic *bread's* relations

We write these non-standard LFs in Latin and in capitals (we do not write them in Spanish) because this formal convention helps us in our daily task. Some of them are attached to productive Spanish morphological rules. That is the case of ARTIFEX: it is the name of the person who works professionally with something; see the case of *pan* (bread) and *panadero* (baker) (first example underlined in green in Figure 4).
The set of nouns of professionals and workers in Spanish is quite broad (we have 291 nouns in our database) and there are some productive suffixes attached to this field, such as -ero (*banquero*, banker; *barbero*, barber), -ista (*periodista*, journalist; *dentista*, dentist), -or (*conductor*, driver; *constructor*, builder), etc.
The non-standard LF ARTIFEX is useful for the nouns we labelled as 'professionals'. Frequently we find this lexical relation between two words (*pan, panadero; periódico, periodista; barba, barbero; diente, dentista*) attached to a morphological link between these two words, naming one of them a professional and the other one an object this person works with.
What we call FACERE CUM and FACTUS CUM (see the second and third set of examples underlined in green) means respectively "to do with" and "made with". The first one is useful for verbs, such as *empanar con pan rallado* (bread with breadcrumbs); the second one is useful for relations between nouns, such as *sandwich*, *bocadillo* (a type of sandwich), *tostada* (toast), *medianoche* (bread roll), etc.
Finally, what we call LOCUS (see the fourth set of examples underlined in green) is a noun place related, in this case, to *bread*: *panadería* (bakery) and *panificadora* (a sort of semi-industrial bakery where the daily bread is made, the fresh bread consumed in Spain).

5 Conclusions

We have summarized some of the problems we face when applying the automatic inheritance in the Spanish e-dictionary *Diretes*. The point from which we start in our methodology is the map of semantic labels we work with. It was designed originally for *BADELE.3000*, which is the data base our e-dictionary is based on, and is now growing by new semantic labels.

The semantic label we are working with is a word or an expression that: a) is useful for a group of words (such as 'baked sweet' for *pie, cake, cookies*); b) reflects at least one co-occurrence (*such as to bake the cookies)*; c) is a meaning (such as 'sweet') or a restricted meaning (such as 'baked sweet'), but it is not necessarily part of the definition.

Based on this concept of semantic labels, we are able to predict the relations that can be inherited, and automatically obtain collocations formalized by means of Lexical Functions, such as *to bake*, the LF $CausFunc_0$ (which means "create") and all the nouns labelled as 'baked sweet', *cookies*, *cake*, *pie*, etc. After the inheritance of all the lexical-semantic relations predicted, we manually add the non-predicted relations (which are taken from the data of some corpora we use) and the corresponding LFs. The predictions demand a strong knowledge of the MTT model, which implies the lexicographer is expert not only in standard LFs but also in non-standard LFs.

Summarizing the novelty of this project is the implementation of the automatic inheritance (both lexical and domain) and the particular concept of semantic label on which it is based: the condition upon which a new semantic label is created is that there is at least one particular collocation that distinguishes this set of words from some other sets of words.

6 References

Apresjan, J. (2000). *Systematic Lexicography*. Oxford: Oxford University Press.

Apresjan, J., Boguslavsky, I., Iomdin, L., Tsinman L. (2002). Lexical Functions in NLP: Possible Uses. Computational Linguistics for the New Millennium: Divergence or Synergy? In M. Klenner, H. Visser (eds.). *Festschrift in Honour of Peter Hellwig on the occasion of his 60th Birthday*. Frankfurt: Peter Lang, pp. 55–72.

Atkins, S.B.T., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Barrios Rodríguez, M.A. (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 [30/05/2020].

Barrios Rodríguez, M.A. (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 [30/05/2020].

Barrios Rodríguez, M.A. (2019). ¿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.), *Lo que hablan las palabras. Estudios de lexicología, lexicografía y gramática en honor de Manuel Alvar Ezquerra*. Lugo: Axac.

Barrios, M. A., Bernardos, S. (2007). BaDELE3000: An implementation of the lexical inheritance principle. In Reuther, Tillman; Wanner, Leo. (eds.) *Wienner Slawistischer Almanach. Sonderband* 69, pp. 68-77.

Bosque, I. (2004). *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: S. M.

Bosque, I. (2006). *Diccionario combinatorio práctico del español contemporáneo*. Madrid: S. M.

Hanks, Patrick (2013). *Lexical analysis. Norms and Exploitations*. Cambridge, Massachussetts: the MIT Press.

Fontenelle, T. (2008). "Using a Bilingual Dictionary to Create Semantic Networks". In Fontenelle (ed.), *Practical Lexicography. A reader*. Oxford: Oxford University Press, 169-189.

Gader, N., Lux-Pogodalla, V. and A. Polguére (2012). Hand-Crafting a Lexical Network with a Knowledge-based graph editor. *Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon (CogALex-III)*, pp. 109–126.

Granger, S. (2012). "Introduction. Electronic Lexicography: from challenge to opportunity". In Granger, Sylviane & Paquot, Magali (eds.). *Electronic Lexicography*. Oxford: Oxford University Press, 1-11.

Hanks, P. (2009). The impact of corpora on dictionaries. In Baker, Paul (ed.). *Contemporary Corpus Linguistics*, London: Continuum, 114-236.

Hanks, P. (2013). *Lexical Analysis*, Cambridge: The MIT Press.

L'Homme, M. C. (2016). Terminologie de l'environnement et Semantique des cadres. *Congrès Mondial de Linguistique Française - CMLF 2016,* SHS Web of Conferences 2 05010 (2016)

Mel'čuk, I. (1996). "Lexical functions: A tool for the description of lexical relations in a lexicon". In Wanner, L. *Lexical functions in lexicography and natural language processing*. Amsterdam/ Philadelphia: John Benjamin, 209-278.

Mel'čuk, Igor (2014): *Semantics. from meaning to text*. Vol 3. Amsterdam/Philadelphia: John Benjamins.

Mel'čuk, I., Clas. A., Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-le-neuve: Duculot.

Mel′čuk, I., Wanner, L. (1996). Lexical Functions and Lexical Inheritance. In L. Wanner (ed.), *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam: John Benjamins Publishing Company, pp. 209-278.

Mel'čuk, I., Polguère, A. (2008). *Lexique actif du français - l'apprentissage du vocabulaire fondé sur 20.000 dérivations sémantiques et collocations du français.* Bruxelles : De Boeck & Larcier.

Milićević, J. (1997). Étiquettes sémantiques dans un dictionnaire formalisé du type Dictionnaire Explicatif et Combinatoire. Montreal : Université de Montreal.

Polguère, A. (1997). Etiquetage semantique des lexies dans la base de donnees dico. *Traitement automatique des langues*, 44(2), pp. 39–68.

Polguère, A. (1997). Meaning-Text Semantic Networks as Formal Language. In L. Wanner (ed.). *Recent Trends in Meaning-Text Theory*. Amsterdam: John Benjamins Publishing Company, pp. 1–24.

Polguère, A. (2007). Lexical Function Standardness. In L. Wanner, *Selected Lexical and Grammatical Issues in the*

Meaning-Text Theory. Amsterdam/Philadelphia: John Benjamins, pp. 43-96.
Polguère, A. (2011). Classification sémantique des lexies fondée sur le paraphrasage. *Cahiers de Lexicologie*, 98, pp.197–211.
Polguère, A. (2014). From writing dictionaries to weaving lexical networks. *International Journal of Lexicography*, Vol. 27 No. 4, pp. 396–418.
Reuther, T. (1996) On Dictionary Entries for Support Verbs: The Cases of Russian VESTI, PROVODIT' and PROIZVODIT'. In Wanner, L. (ed), *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins, pp. 181-208.

Acknowledgements

The project "Diretes: Diccionario Reticular Español. Diccionario analógico y relacional con acceso en red desde el sentido y desde la forma" has been founded by the grant FFI2017-83293 of the Spanish Ministry of Science and Economy. I would like to express my gratitude to Deborah Paton for all her suggestions on the English revision of the paper and to the anonymous reviewers for their helpful comments. Any remaining errors are our own.