

**Perspectives on Phraseology from Lexicography, Terminology, and  
Translation: from Language for General Purposes to Language for Specific  
Purposes**

NHH



José Luis Rojas Díaz

Supervisors:

Prof. Dr. Gisle Andersen

Assoc. Prof. Dr. Beate Sandvei

Dissertation for the degree of Philosophiae Doctor (PhD)

NHH Norwegian School of Economics

Department of Professional and Intercultural Communication

Bergen, Norway

November 2021

*a Jorge*

*Quien hace 30 años, sin siquiera imaginarlo, hizo que este viaje iniciara.*

## Acknowledgements

Although a personal and autonomous one, the completion of this PhD journey is thanks in large part to the people who challenged and supported me along the way.

First of all, I would like to thank the public education systems of both Norway and Colombia. To the Norwegian public education system, for giving me the opportunity to be part of the PhD program at the *Norwegian School of Economics* where I was provided with the financial support needed to carry out my research and the *Norwegian Graduate Researcher School in Linguistics and Philology* (LingPhil) of which I was one of the student representatives. I would like also to thank the Colombian public education system, for giving me the basis to make this happen. I would not have had the opportunity of writing these lines without the support of public education.

I would like to express my utmost gratitude to my supervisors Professor Gisle Andersen and Associate Professor Beate Sandvei. You both brought a depth of knowledge that few others could match. Your trust in me as a scholar, and in this project since day one, encouraged me to complete this PhD dissertation. In the last four years you have taught me, by example, what to be a PhD really is. You showed me not only a new way to explore linguistics, terminology, and translation but also a new way to understand science and how to respectfully argue the point of view of others. Gisle, your critical eye and your discerning comments shaped the scope of this PhD dissertation without stifling my creativity. The times I shared with you at different conferences and networks, where you invited me to participate, allowed me to meet top scholars in disciplines related to my project. This created an arena for me to share and test the methods and findings of my PhD dissertation. Beate, in your capacity as head of *Department of Professional and Intercultural Communication*, you are without a doubt one of the best leaders I have ever encountered. Your devotion to your students and everyone in the department is something that I will take with me wherever I go. Thank you also for your encouraging words during the supervision and writing sessions of the last year. Thank you both for all your patience and commitment.

Thank you to all the professors who helped me and inspired me along the way:

- To Professor Gabriel Ángel Quiroz Herrera who wrote one of the recommendation letters when I applied for this position.
- To the Professors of the PhD courses at NHH: Professor Sigurd Villads Troye, Professor Lars Jacob Tynes Pedersen, Professor Erik Øiolf Sørensen. I would like to extend special thanks to Professor Trine Dahl who was truly an inspiration during this PhD dissertation and who taught the invaluable PhD course on academic writing.
- To Professor Marie-Claude L’Homme, Professor Magali Paquot, and Professor Stefan Th. Gries who delivered the lectures in the PhD course on phraseology held at NHH in 2019.
- To all the professors that contributed with their lectures to my PhD education through *LingPhil*: Professor Steve Mann, Professor Michael Oakes, Professor Ray Jackendoff, Professor Penny Eckert, Professor Anders Björkvall, and Professor Marjo van Koppen
- For exchanging ideas with me and being approachable: Professor Miriam Buendía Castro, Professor Pilar León Araúz, Professor Jorge Leiva Rojo, Professor Carmen Mellado Blanco, Professor María Ángeles Recio Ariza, Professor José Agustín Torijano Pérez, Professor Jaime Hunt, Professor Elizabeth Peterson, Professor Ivana Bozděchová, and Professor Aleš Klégr. Special thanks to Professor Gloria Corpas and Professor Pamela Faber your insightful comments and ideas helped me to shape the scope of this PhD dissertation.

Thank you to all my colleagues at the Department of Professional and Intercultural Communication at the Norwegian School of Economics:

- To Professor Emerita Ingrid Simonnæs. There are not enough words to thank you for all the help received. You took care of me as if I were one of your scholars. You read and comment all my drafts until the very end of this PhD dissertation.

- To Professor Anne Kari Bjørge, former head of department, I will always remember the kind words you welcomed me with on the first day at the office.
- To Associate Professor Kristin Rygg for being there when I needed a person to talk to.
- To Professor Emeritus Jan Roald, for reading and commenting on the third article of my dissertation.
- To Maja Dame, Rikke Linn Salbu, Silje Grønner Stang, and Fitia Andrianaivoravelona. Nothing in the department would be possible without your help and assistance.
- To my former fellow PhD scholars at the department: Associate Professor Claudia Förster Hegrenæs, Associate Professor Annelise Ly, and Marte Reenskaug Fjørtoft. You all made me feel welcome since day one. Thanks for so many encouraging words at the right moments.
- To my fellow PhD scholars: Ina Celise Sortland, Victoria Susanne Nydegger Schrøder, Ziyuan Zhang for all the time and conversations shared. Here special thanks to Agnes Bamford, more than colleague you will always have a friend in me. Thanks for our conversations and your energy in the office you really know how to make a bad day become a brighter one.

Thank you to Alejandro Arroyave Tobón, Juan Guillermo Ramírez Giraldo, and Juan Manuel Pérez Sánchez: “we few, we happy, we band of brothers.” Thanks for always being there ready to answer my questions and to provide valuable opinions and criticisms of my project thanks for all your support. There are not enough words to express my gratitude to Juan Manuel who read all my article drafts. You inspired me to set sail into the uncharted waters of semantics. You were the unofficial third co-supervisor of this PhD dissertation, I never heard a NO as an answer from you when I needed someone to discuss conceptual and theoretical problems. Thank you for hours of unconditional help and support. I will be always there if you ever need me. To my dear friend Paula Montoya whose valuable advice drove me towards the publication of Article 1. To my very good friend Pedro Patiño who not only wrote a

recommendation letter for me to be here, but also prepared the ground for another Colombian to be part of the department and NHH.

To my friends here in Norway and around the world for all the support: Dale, Jessica, Giovanni, Daniel, Marte, Lena, Stefania, Averkios, Paris, Ionanis. Special thanks to Jimmy and Amal for taking care of me when things got rough; Cecilia who helped me on the last leg of the journey with good vibes, joy, and suggestions for my drafts. Also, thanks to Rebecca, your company and help were invaluable in the last year. Thank you to Lasse Bjørkhaug and Solveig Aase for opening the doors of your home and introducing me to the Norwegian way of life.

To my family. *A mi Papá. a ti me debo (que estas palabras queden impresas para que cuando tu memoria falle, sepas que no hubo un día en estos cuatro años en que te dejara de pensar). Gracias por inculcarme el amor por el conocimiento, por permitirme ser, por la libertad que me diste, por todo el apoyo siempre. Alguna vez me pediste no olvidar mi origen, hoy te digo no hay orgullo más grande que saberme ser hijo de un humilde obrero, por encima del grado de doctor que esta disertación me otorga. Espero que te sientas orgulloso de lo que has logrado porque mis logros son los tuyos.* To my brothers, thanks for all the values you taught me while growing up. This project would have been even harder without your support throughout these four years. Here I want to thank you Jorge for actively participating in the creation process of the databases and the data visualization tools, I continue learning from you. To my nieces, and nephews thanks for your love and kind words when you had the time to be there. Miyei, thanks for always taking care of me. Last but not least, thank you Lu. Not only you took the time to design many of the graphs of this PhD dissertation but were there next to me and my Dad even before this journey started. You gave me the final push to be here and you were always there when I needed you regardless of the distance. Thanks for all the unconditional love, care, and support. No matter what happens in life, you will always be part of my family.

## Table of Contents

Acknowledgements.....	ii
Table of Contents.....	vi
List of Tables.....	vii
List of Figures.....	viii
List of Abbreviations.....	ix
1. Introduction and rationale.....	1
1.1. Objectives.....	3
1.1.1. General objective.....	3
1.1.2. Specific objectives.....	3
1.2. Research propositions and questions.....	4
1.3. Structure of the dissertation.....	4
2. Outline of the articles.....	5
2.1. Article 1: ‘From Head to Toe’: a Lexical, Semantic, and Morphosyntactic Study of Idioms in Phraseological dictionaries in English and Spanish.....	5
2.2. Article 2: ‘Arm’s length’ phraseology? Building Bridges from general language to Specialized Language Phraseology – a Study Based on a Specialized Dictionary of International Commerce and Economics.....	6
2.3. Article 3: ‘Worlds apart’? Phraseological Equivalence through the Lenses of Translation, Terminology, and Lexicography.....	7
3. Theoretical and conceptual framework.....	8
3.1. Key notions within phraseology.....	8
3.1.1. Philosophy and the object of study of phraseology.....	8
3.1.2. Phraseology in language for general purposes and its object of study.....	13
3.1.3. Phraseology in language for specific purposes and its object of study.....	18
3.2. Key notions within lexicography.....	19
3.3. Key notions within terminology.....	21
3.4. Key notions within translation.....	22
4. Methods, data, and tools.....	24
4.1. The methodology.....	24
4.2. The dictionaries.....	26
4.3. The databases.....	28
4.4. The corpus.....	30
4.5. The tools.....	30
5. Dissertation contributions.....	35
5.1. Theoretical contributions.....	35
5.2. Descriptive data contributions.....	38
5.3. Applied and methodological contributions.....	40
5.4. Limitations.....	40
5.5. Future research.....	41
6. References.....	43
7. Articles.....	50
Appendices.....	I
Appendix A.....	I

## List of Tables

<b>Table 1.</b> Definitions of idiomaticity levels (Baldwin & Kim, 2010, pp. 269-271) .....	13
<b>Table 2.</b> Forefather denominations of PUs before the works of Bally (1909) .....	14
<b>Table 3.</b> Denominations of PUs proposed by the most representative authors related to general phraseology in Spanish .....	15
<b>Table 4.</b> Denominations of phraseology by different authors according to Cowie (1998, p. 7) .....	17
<b>Table 5.</b> Definitions of ‘method,’ ‘strategy,’ and ‘technique’ by Molina and Hurtado Albir (2002, pp. 507-509) .....	23
<b>Table 6.</b> General lexicographic information of the DFDEA and the CCDOI .....	27
<b>Table 7.</b> General lexicographic information of the DCI.....	27
<b>Table 1A.</b> Definitions and examples of PUs (Mel’čuk, 1998, pp. 6-8), and classification of idioms and collocations (Mel’čuk, 2012, pp. 37-40).....	I
<b>Table 2A.</b> Denominations and definitions of LSP phraseology gathered in the work by (Rojas Díaz, forthcoming-a) .....	III
<b>Table 3A.</b> Some authors and definitions of equivalence in translation studies addressed by Halverson (1997), Leonardi (2000), and Panou (2013) .....	V
<b>Table 4A.</b> Some authors and definitions of equivalence in Lexicography.....	VII
<b>Table 5A.</b> Some authors and definitions of equivalence in Terminology .....	VIII
<b>Table 6A.</b> Some authors and definitions of phraseological equivalence.....	IX



## List of Figures

<b>Figure 1.</b> Outline and details of the articles of the PhD dissertation.....	8
<b>Figure 2.</b> General methodology chart of the PhD dissertation.....	25
<b>Figure 3.</b> No. of word forms of the PUs in the DFDEA and the CCDOI (Rojas Díaz, 2020, p. 302) .....	29
<b>Figure 4.</b> Data visualization example through a Microsoft Power BI dashboard (1) slicers and matrices (2) slicers and graphs .....	32
<b>Figure 5.</b> Example of POS-Tagging of SpI equivalents from (1) Spanish-English and from (2) English-Spanish.....	33
<b>Figure 6.</b> Example of POS-tagging of word forms (1) and morphosyntactic pattern ensemble process through regular expression in plain text (2) and (3).....	34
<b>Figure 7.</b> McArthur's (1981) discourse fields (Archer et al., 2002, p. 2). .....	34
<b>Figure 8.</b> New taxonomy for SPU classification (Rojas Díaz, forthcoming-a, p. 27) based on the works by Mel'čuk (1998, pp. 6-8; 2012, pp. 37-40; 2013).....	37

## List of Abbreviations

- CCDOI:** Collins COBLUID Dictionary of idioms
- DCI:** *Diccionario de Comercio Internacional – Importación y Exportación, Inglés-Español/Spanish-English*
- DFDEA:** *Diccionario fraseológico documentado del español actual*
- LGP:** Language for general purposes
- LSP:** Language for specific purposes
- POS:** Part-of-Speech
- PU:** Phraseological unit
- SL:** Source language
- SpC:** Specialized collocation
- SpI:** Specialized idiom
- SpP:** Specialized pragmateme
- SPU:** Specialized phraseological unit
- ST:** Source text
- TL:** Target language
- TT:** Target text
- USAS:** UCREL's Semantic Analysis System

*“At the present time, phraseologists remain weird and wonderful animals,  
but will probably come into their own in the near future as clients become  
more and more demanding when it comes to using  
the ‘right’ phraseology”*

Daniel Gouadec

## 1. Introduction and rationale

Every day, globalization, integration, and economic opening processes, as well as interdisciplinary communication and high technological progress, generate knowledge in diverse disciplines and science, thus increasing the vocabulary of their discourses. A significant number of professional and specialized users including translators, subject field experts, proofreaders, and publishing house editors use such vocabulary (including terms, multiword expressions [MWEs], phraseological units [PUs], among others) allowing them to act as intermediaries in different oral and written communicative acts. Hence, this requires those scientific and technical discourses to have access to consultation resources that could meet their needs.

Lorente Casafont (2002) states that there is a significant problem when users look for MWEs or PUs in a dictionary: “the users of specialized vocabularies need information about recurrent terminological combinations, [...] information which cannot be found in most of the existing resources”<sup>1</sup> (p. 159 [my translation]). Moreover, the author highlights two main problems framed in the study of phraseology within general and specialized lexicographic resources (e.g., dictionaries and databases): the difficulties regarding the indexation of PUs and the lack of descriptive studies on phraseology. These problems have been of interest to several scholars proceeding from diverse traditions and languages during the last two decades (Alonso Ramos, 2006; Bevilacqua, 2004; Buendía Castro & Faber, 2015; Heid, 2008; Leroyer, 2006; Mel’čuk, 2012; Mellado Blanco, 2008; Moon, 2008; Nuccorini, 2020; Paquot, 2015; Rojas Díaz & Pérez Sánchez, 2019; Siepmann, 2008; Sosiński, 2006; Tschichold, 2008; Veisbergs, 2020). Nevertheless, there is still the need for more research on specialized phraseological units (SPUs), and the parameters that allow differentiating PUs in language for general purposes (LGP), e.g., *abogado del diablo* ‘devil’s advocate’, *respirar por la herida* ‘breath by the wound’ | *an early bird, shoot yourself in the foot*<sup>2</sup>, from SPUs in language for specific purposes (LSP), e.g., *letra de cambio mutilada* ‘mutilated

---

<sup>1</sup> Original in Spanish: *Los usuarios de vocabularios especializados precisan información sobre combinaciones terminológicas recurrentes [...] que no queda resueltas con la consulta de la mayoría de recursos existentes*

<sup>2</sup> A pipe is used to separate the examples from Spanish and English.

bill', *hacerse a la mar* 'lift anchor'| *heads of agreement*, *smurf money*, and their indexation in lexicographic resources. Previous literature reviews, as presented in Articles 2 and 3, have explicitly shown that SPUs contain a term within their word forms. However, the descriptive data and the analyses performed in this PhD dissertation show that some SPUs do not include a monolexical term among their word forms. This PhD dissertation aims to fill the gap by offering a new definition and a taxonomy of SPUs based on data extracted from PU and SPU entries and equivalents from dictionaries.

For my Master in Linguistics, I made a statistical description of the data retrieved (16,456 entries) from the *Diccionario de Comercio Internacional – Importación y Exportación, Inglés-Español/Spanish-English* (DCI) (Alcaraz & Castro Calvín, 2007) including lexicographic features (number of entries, subentries, marking (particularly those marks that identify SPUs) and linguistic features (part-of-speech, number of forms and morphosyntactic patterns of the entries of the dictionary). The interpretation of the statistical results from the analysis of the database offered a list of morphosyntactic patterns of SPUs extracted from the dictionary and some methodological inconsistencies found in the marking, lemmatization, and indexation of the SPU entries and sub-entries in the DCI. I decided to broaden this study with the current PhD dissertation and the Norwegian School of Economics provided the perfect research environment in which to carry out this project.

The motivation for this PhD dissertation is two-fold. First, it aims to address the lack of studies on what linguistic features PUs and SPUs share, as well as how PUs and SPUs differ from one another. Second, the aim is to explore how SPUs are indexed in Spanish-English a specialized dictionary, including their lexicographic equivalents.

As regards the language pair included in the present study, Spanish and English are the official languages of several institutions and organizations in the Americas and Europe related to international commerce and economics (e.g., the Council of Europe [CE], the Economic Commission for Latin America and the Caribbean [ECLAC], the Inter-American Development Bank [IADB], the International Monetary Fund [IMF], the International Telecommunications Union [ITU], the Organization of American States [OAS], the United Nations [UN], the World Trade Organization [WTO], among others). Therefore,

Spanish and English constitute a suitable pair for choosing the resources for the extraction of the data and carrying out the present PhD dissertation.

## **1.1. Objectives**

This PhD dissertation has a general objective and several descriptive, methodological, and theoretical objectives.

### **1.1.1. General objective**

- The main goal of this PhD dissertation is to study the behavior and characterization of SPUs related to commerce and economics by means of a series of linguistic analyses (lexical, i.e., word classes and frequency of words; morphosyntactic, i.e., part-of-speech patterns; and semantic, i.e., regarding semantic annotation, classification in discourse and semantic fields and identification of metaphors and metonymies) of entries and equivalents selected from specialized Spanish/English lexicographic resources and corpora.

### **1.1.2. Specific objectives**

- To identify the marking of PUs and SPUs in selected general and specialized dictionaries (Article 1 and Article 2).
- To describe the characteristics (morphosyntactic and semantic patterns, figurative language usage, and constituent elements) of PU entries in Spanish and English (Article 1).
- To describe the characteristics (morphosyntactic and semantic patterns, figurative language usage, and constituent elements) of SPU entries and equivalents in Spanish and English related to commerce and economics (Article 2).
- To propose an alternative indexation method for PUs and SPUs in lexicographic resources (Article 1 and Article 2).

- To analyze SPU entries and equivalents from the selected lexicographic resources in terms of their idiomaticity and their frequency in the EUR-Lex corpus (Article 3).
- To identify the differences between SPU equivalents offered in the selected lexicographic resources and the EUR-Lex corpus, as well as analyze decisions made by translators when dealing with SPUs (Article 3).

## **1.2. Research propositions and questions**

The following questions are addressed in this PhD dissertation:

- Q1: What are the linguistic (i.e., lexical, morphosyntactic, and semantic) features of PUs in Spanish and English LGP dictionaries?
- Q2: How are SPUs classified and indexed in a Spanish-English / English-Spanish LSP dictionary?
- Q3: What linguistic characteristics do PUs and SPUs have in common? Or, conversely, how are PUs different from SPUs according to the data retrieved from lexicographic resources and corpora?
- Q4: Do SPU lexicographic equivalents preserve their idiomatized characteristics (e.g., occurrence of metaphors and metonymies) in translations from Spanish to English and English to Spanish?
- Q5: How do translators deal with SPU translation and how are these equivalents registered in corpora?

## **1.3. Structure of the dissertation**

This PhD dissertation consists of the present cover article that presents background and overview of the topic along with the theoretical and methodological frameworks as well as the main contributions, limitations, and future work of the dissertation, and the three articles that answer the research questions.

Q1 is answered in Article 1, where I make a description of phraseological units within monolingual general language dictionaries. Q2 and Q3 are tackled in Article 2, where I make a description of

phraseological units in English-Spanish specialized dictionaries and offer a new definition of SPU based on previous studies of PUs. Finally, Q4 and Q5 are addressed in Article 3, where I compare a sample of the entries and equivalencies of SPUs from a bilingual (Spanish-English / English-Spanish) dictionary and the EUR-Lex corpus.

This cover article is organized as follows: section 2 provides an outline of the dissertation articles. In section 3, key notions from phraseology, lexicography, terminology, and translation used in this PhD dissertation are presented. Section 4 deals with the methods, data, and tools employed in this study. Section 5 describes the contributions and limitations of this dissertation and suggestions for future research. In section 6, the references of this work are given, Finally, section 7 includes the three articles that compose this PhD dissertation.

## **2. Outline of the articles**

### **2.1. Article 1: ‘From Head to Toe’: a Lexical, Semantic, and Morphosyntactic Study of Idioms in Phraseological dictionaries in English and Spanish<sup>3</sup>**

Article 1 offers an exhaustive characterization of idioms (as a subcategory of PUs) in LGP extracted from two phraseological dictionaries in Spanish and English. My initial intention was to focus on morphosyntactic patterns and Part-of-Speech (POS)-tagging as the focus of Article 1. However, after carrying out the semantic annotation, I decided to explore aspects related to lexical semantics (e.g., semantic annotation) and idiomaticity. This article offered a series of linguistic features (e.g., POS, semantic annotation, and the occurrence of metaphors and metonymies) that encouraged me to do a similar analysis of entries and sub-entries from a specialized dictionary on commerce and economics, i.e., from an LSP lexicographic resource. A revision of the dictionary, prior to the study, offered clues about

---

<sup>3</sup> This paper was published as a journal article in January 2021: Rojas Díaz, J. L. (2020). “From Head to Toe”: a Lexical, Semantic, and Morphosyntactic Study of Idioms in Phraseological Dictionaries in English and Spanish. *MonTI* (Special Issue 6), 287-326. <https://doi.org/http://dx.doi.org/10.6035/MonTI.2020.ne6.9>. I have been granted permission to use this article for my dissertation by the publisher *MonTI: Monographs in Translation and Interpreting* in an email dated 9.11.2021



the occurrences of specialized phraseological units that acquired a specialized meaning (e.g., *llave en mano* ‘key in hand’ | *at arm’s length*), although they did not include any monolexical term among its word forms. Therefore, an answer to the question “what is an SPU?” was needed. Also, it was important to explore whether SPUs shared any characteristics with PUs in order to be studied by phraseology.

## **2.2. Article 2: ‘Arm’s length’ phraseology? Building Bridges from general language to Specialized Language Phraseology – a Study Based on a Specialized Dictionary of International Commerce and Economics<sup>4</sup>**

This article creates a bridge between LGP and LSP phraseology by revisiting the notions and taxonomies from LGP phraseology. The works by Mel’čuk (1998, 2012, 2013) were used to identify classification criteria for SPUs to answer the question “what is an SPU?” Similarly to Article 1, the SPUs used to build the databases for this Article 2 were POS-tagged, and semantically annotated and their morphosyntactic and semantic patterns were extracted. Besides the linguistic information extracted from the SPUs, the new taxonomy proposed in this article offered the possibility to classify the expressions in three main subcategories: specialized collocations (SpCs, e.g., *inundar el mercado* ‘overload the market,’ *tipo de cambio verde* ‘green exchange rate’ | *enter good for free circulation, cancellation of a debt*), specialized idioms (SpIs, e.g., *los cinco tigres* ‘the five tigers,’ *poner en clave* ‘code’ | *sweat in hold, break the deadlock*), and specialized pragmatemes (SpPs, e.g., *este lado hacia arriba* ‘this side up,’ *abrir por este lado* ‘open this end’ | *this way up, do not tilt*). SpIs were chosen to finish the intended studies for this PhD dissertation since the starting of the dissertation (Article 1) was the study of idioms in LGP. As mentioned in section 2.1., Articles 2 and 3 show the shift of focus of the PhD dissertation from the morphosyntactic analyses of PUs and SPUs towards idiomaticity aspects of the SPUs including the analyses of metaphors and metonymies that underlie SpI entries and their equivalents. To achieve the dissertation’s objectives, a cross-linguistic analysis, involving translation and corpora, was carried out in Article 3.

---

<sup>4</sup> This manuscript has been submitted to *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication* and it is currently under revision.

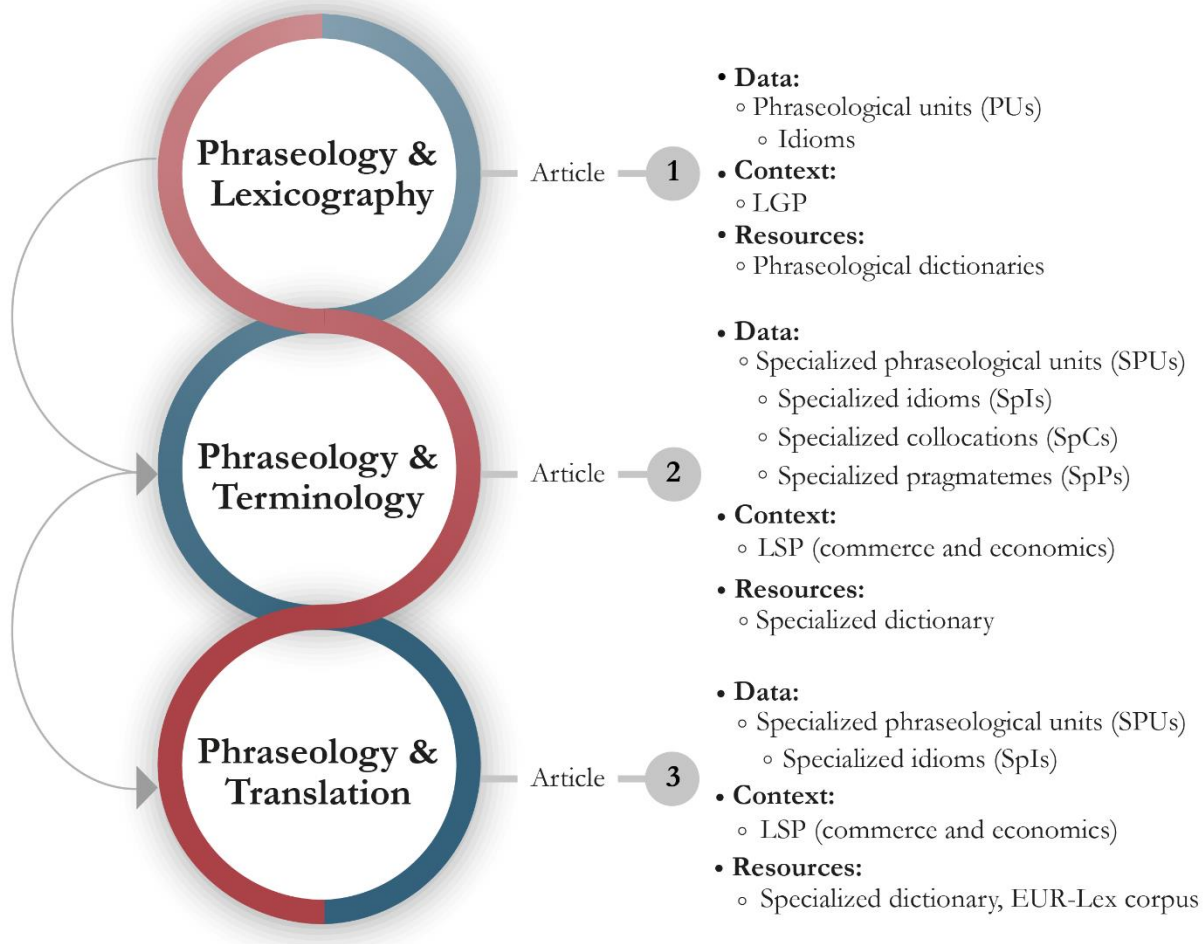
### **2.3. Article 3: ‘Worlds apart’? Phraseological Equivalence through the Lenses of Translation, Terminology, and Lexicography<sup>5</sup>**

Finally, Article 3 offers a cross-linguistic study of the SpI entries found in Article 2 and their equivalents. Article 3 analyses the idiomaticity of both SpI lexicographic entries and equivalents, as well as some other equivalents found in the EUR-Lex corpus. The analysis of the idiomaticity allowed for identification of different levels of idiomaticity (e.g., semantic idiomaticity and syntactic idiomaticity). The absence or occurrence of certain levels of idiomaticity of the lexicographic equivalents shows structural divergences between languages affecting the composition of morphosyntactic and semantic patterns (see section 5.2.). Furthermore, Article 3 offers the frequency of occurrence of SpI lexicographic entries and equivalents in the EUR-Lex corpus. The findings (an average of 24.55 times per million words in Spanish and 3.98 times per million words in English) suggest that idioms (SpIs in this case) are more frequent than reported in the previous LGP phraseology literature. The findings from Article 3 suggest that LGP and LSP phraseology have more characteristics in common than previous definitions and taxonomies of SPUs propose.

A graphical scheme of this dissertation’s outline is offered in Figure 1. The arrows on the left side represent the way in which the findings and conclusions of the previous article help to develop the next article. The linked circles present the object of study of this dissertation (phraseology), along with the three disciplines –lexicography, terminology, and translation– that constitute the different perspectives on the object of study of the present work. The right side of the graph presents an overview of the data, contexts, and resources employed in each article. Red and blue are used again to highlight the use of data in Spanish (red) and in English (blue).

---

<sup>5</sup> This manuscript is been prepared to be submitted to *Meta: Journal des traducteurs*.



*Figure 1. Outline and details of the articles of the PhD dissertation*

### 3. Theoretical and conceptual framework

This section provides the key notions and theoretical concepts related to phraseology (section 3.1.), lexicography (section 3.2.), terminology (section 3.3.), and translation (section 3.4.), that form the basis of this PhD dissertation.

#### 3.1. Key notions within phraseology

##### 3.1.1. Philosophy and the object of study of phraseology

The three articles that constitute this PhD dissertation offer extensive literature reviews regarding (among others) theoretical problems on denominations (e.g., the object of study of phraseology in LGP and LSP),

shared characteristics between notions (e.g., similarities and differences between PUs and SPUs), the notion of compositionality, and phraseological equivalence. These topics are addressed in the three articles progressively: Article 1 deals with the denominations used in LGP phraseology. Article 2, besides tackling the problem of denominations in LSP phraseology, explores idiomaticity and compositionality in LSP phraseology and employs them as criteria for the identification and classification of SPUs. Article 3 focuses on the notion of equivalence. This sub-section will offer a concise overview of some philosophical notions and paradigms that were used as inspiration and theoretical basis of the three articles.

### ***3.1.1.1. On denomination: The Cratylus***

*Cratylus* (Plato, Cambridge: Harvard University Press, 1921) is the name of a dialogue by Plato in which Cratylus and Hermogenes, with Socrates acting as a middle man, intend to find an answer of whether names are ‘natural’ or ‘conventional’. On the one hand, Hermogenes is described as an extreme linguistic conventionalist when he claims that “nothing but local or national convention determines which words are used to designate which objects” (Sedley, 2020, para. 1). On the other hand, Cratylus is depicted as an extreme linguistic naturalist when he argues that “names cannot be arbitrarily chosen in the way that conventionalism describes or advocates, because names belong *naturally* to their specific objects” (Sedley, 2020, para. 1).

Some may argue that Plato’s *Cratylus* is more related to the origin of denominations (etymology) than to denominations themselves. For instance, Ademollo (2011) argues that:

The fact is this: throughout the dialogue all characters express themselves as if there were no difference between being a correct name of something and being just a name of that thing. They continuously speak as if the phrases ‘correct name of X’ and ‘name of X’ were perfectly interchangeable and equivalent to each other. (p. 2)

However, a recent work by Jørgensen (2019, p. 2) presents contextual and textual evidence to counterargue Ademollo (2011, p. 2):

The interlocutors in the dialogue do not share and take for granted the view that there is no difference between being a correct name of something and being a name of something. Rather,

the interlocutors – Socrates, certainly, but I include Hermogenes and Cratylus – make a distinction between being a name and being a correct name. (Jørgensen, 2019 para. 31)

The dialogue provides insights into the importance of the ‘correctness of words’ more than merely offering the resolution offered by Plato through the participation of Socrates. Surprisingly, Plato considers the idea of variants or synonymic denominations:

Two names have the same ‘power’ provided that both succeed in marking off the same object, even if they do so by means of different descriptions, i.e. without being simple synonyms. (Sedley, 2020, para. 19)

Furthermore, ‘conventionalism’ could be observed in the fixation (reiterative use through time) of PUs in language as well as in notions applied to terminological work (see sections 3.1.2. and 3.1.3.) while the ‘naturalism’ (reflected on the etymological work) is key for the identification of metaphors and metonymies. Both notions of ‘conventionalism’ and ‘naturalism’ are needed for a complete analysis in studies of LGP and LSP phraseology. This PhD dissertation offers an explanation for the denomination problem regarding the object of study of phraseology (as explained in sections 3.1.2. and 3.1.3.) in Articles 1 and 2, and Plato’s *Cratylus* was used as a source of inspiration for the analysis of this problem.

### ***3.1.1.2. On similarities and differences between LGP and LSP phraseology: Plutarch’s Ship of***

#### ***Theseus***

The complexity of answering the question what is the object of study of LGP and LSP phraseology? goes beyond offering a certain denomination (see sections 3.1.1.1., 3.1.2. and 3.1.3.). After an extensive literature review, and as explained in Article 2, LSP phraseology scholars have tried to offer solutions to questions on LSP phraseology ranging from the criteria for the identification of SPUs, e.g., Bevilacqua (2004); Cabré, Lorente, and Estopà (1996); Kjær (2007); L’Homme (2000), to practical studies regarding methodologies for their extraction, e.g., Heid and Weller (2008); Patiño García (2011); Seretan (2011). Nevertheless, detaching the object of study of LSP phraseology from its LGP counterpart is the motivation for using different denominations in LSP phraseology:

However, it has not yet been proven that **specialized lexical combinations** behave like general language collocations. As will be discussed below, some studies have underlined the

discrepancies between the word groups that have attracted the interest of lexicographers and terminologists. (L'Homme & Bertrand, 2000, p. 498 emphasis added)

In the specific case of phraseology, some scholars, e.g., Patiño García (2013, p. 82) , have arrived at the conclusion that SPUs do not share enough characteristics with their LGP counterpart to be considered part of phraseology. This change over time (and eventually from discipline to discipline as in this case) is addressed in philosophy by the paradox of the ship of Theseus. The paradox is offered in Plutarch's *Theseus* (Cambridge: Harvard University Press, 1914):

The ship on which Theseus sailed with the youths and returned in safety, the thirty-oared galley, was preserved by the Athenians down to the time of Demetrius Phalereus. They took away the old timbers from time to time, and put new and sound ones in their places, so that the vessel became a standing illustration for the philosophers in the mooted question of growth, some declaring that it remained the same, others that it was not the same vessel. (Plutarch, Cambridge: Harvard University Press, 1914, pp. ch. 23, section 21)

However, this paradox illustrates not only how the loss of characteristics derives in other denominations, but also the concept of identity and identity as an equivalence relation.

The most commonly agreed on distinguishing feature of identity is that it conforms to the Indiscernibility of Identicals, what was earlier called Leibniz's Law. Taking '∀F' to be a quantifier ranging over properties, here is one way to formulate Leibniz's Law:

$$LL: \forall x \forall y [x = y \rightarrow \forall F (Fx \rightarrow Fy)]$$

LL (Leibniz's Law), understood to range over identity properties, if any, such as being identical with a, says that if *x* is identical with *y*, then any property of *x* is a property of *y*. (Gallois, 2016 para. 18-19)

The conceptions of identity and identity as equivalence relation from mathematics were the starting point for the analysis of equivalence in Article 3.

### **3.1.1.3. Idiomaticity and the notion of compositionality**

The notion of compositionality is key for the distinction of PUs among other MWEs. As presented in Article 1, the study of phraseology in American linguistics started with the criticism that Wallace Chafe made of Noam Chomsky's linguistic theory (Norricks, 2007, p. 616). Chomsky argued that the lexicon is a "simple and unordered list of all lexical formatives" (Chomsky, 1965, p. 84) which should include the

idioms. Three years later, Chafe (1968) showed that idioms were not satisfactorily explained by

Chomskyan linguistic theory:

The importance of idioms on language cannot be doubted. Their ubiquity makes them anything but a marginal phenomenon, and surely a linguistic theory has the obligation to explain them in a natural way. I shall suggest that the present paradigm has been unable to do so, and that a different view of language can account for idioms naturally and convincingly. (p. 111)

Chafe's awareness that idiomaticity could happen at different language levels could be seen in the exemplification of 'peculiarities' of idioms:

These four peculiarities of idioms – their anomalous meaning, the transformational deficiencies, the ill-formedness of some of them, and the greater text frequency of well-formed idioms relative to their literal counterparts – must all be explained by a theory of language adequate to cope with idiomaticity. (Chafe, 1968, pp. 111-112)

Chafe (1968) devotes part of this paper to the semantic analysis of idioms and, although he does not mention it explicitly, to the notion of compositionality:

[The] meaning of an idiom, arrived at through the operation of the semantic component on such a deep structure, is not some kind of amalgamation of the meanings of the parts of the structure. Rather, the meaning of an idiom is comparable to the meaning of a single lexical item. For example, the meaning of the idiom frequently used as example – 'kick the bucket' – is not made up of the meanings associated with 'kick', 'bucket', 'definite article', etc., but it is very much like the meaning of 'die'. (Chafe, 1968, p. 111)

In other words, idioms are not compositional. Philosophy, which is the origin of this principle, explains compositionality in the following terms: "the meaning of a complex expression is determined by its structure and the meanings of its constituents" (Szabó, 2020).

Similarly to what happens in phraseological studies, the definition of compositionality, offered by philosophy, is linked only to semantics (the opacity-transparency continuum). Support for this claim is provided by Baldwin and Kim (2010) who argue that "compositionality is often construed as applying exclusively to semantic idiomaticity (hence by 'non-compositional MWE<sup>6</sup>,' researchers tend to mean a semantically idiomatic MWE)" (p. 269).

---

<sup>6</sup> The authors use 'multi-word expression' (MWE) "as a synonym of 'multiword unit', 'multiword lexical item', 'phraseological unit' and 'fixed expression'; there is also variation in the hyphenation of 'multiword', with 'multi-word' in common use." (Baldwin and Kim 2010, 267)

Baldwin and Kim (2010) differentiate idiomaticity from compositionality. On the one hand, the authors define *idiomaticity* as the “markedness or deviation from the basic properties of the component lexemes.” On the other hand, *compositionality* is defined as “the degree to which the features of the parts of a MWE combine to predict the features of the whole.” Furthermore, the authors distinguish five different levels of linguistic idiomaticity, namely lexical, pragmatic, semantic, statistical, and syntactic (pp. 269-271) (see Table 1).

*Table 1. Definitions of idiomaticity levels (Baldwin & Kim, 2010, pp. 269-271)*

<b>Levels of idiomaticity</b>	
Level	Definition
Lexical idiomaticity	Occurs when one or more components of an MWE are not part of the conventional English lexicon. For example, <i>ad hoc</i> is lexically marked in that neither of its components ( <i>ad</i> and <i>hoc</i> ) are standalone English words.
Pragmatic idiomaticity	It is the condition of an MWE being associated with a fixed set of situations or a particular context [...] ‘all aboard’ [is an] example of a pragmatic MWE [...] [it] is a command associated with the specific situation of a train station or dock, and the imminent departure of a train or ship.
Semantic idiomaticity	Semantic idiomaticity is the property of the meaning of an MWE not being explicitly derivable from its parts [...] for example, ‘middle of the road’ usually signifies “non-extremism, especially in political views.”
Statistical idiomaticity	Occurs when a particular combination of words occurs with markedly high frequency, relative to the component words or alternative phrasings of the same expression.
Syntactic idiomaticity	Occurs when the syntax of the MWE is not derived directly from that of its components. [...] For example, ‘by and large’, is syntactically idiomatic in that it is adverbial in nature but made up of the anomalous coordination of a preposition ( <i>by</i> ) and an adjective ( <i>large</i> ).

These levels are the basis of the analyses and discussions of this PhD dissertation (see section 5).

The next two sub-sections present the main definitions offered by phraseology scholars in both LGP (section 3.1.2.) and LSP (section 3.1.3.).

### **3.1.2. Phraseology in language for general purposes and its object of study**

According to García-Page Sánchez (2008), phraseology should be defined in terms of its object of study. Thus, the question to ask would be “what is the object of study of phraseology?” (p. 7). The denomination of phraseology is a matter of discussion among theorists and researchers from several disciplines (e.g., linguistics, terminology, translation, etc.). Some authors suggest that the term was originally coined by



western structuralist linguists (García-Page Sánchez, 2008; Zuluaga, 1980) while some others state that phraseology –as a discipline– appeared in the linguistic theories in the 40s in Soviet linguistic studies (Carneado Moré & Tristá, 1985; Cowie, 1998).

Carneado Moré and Tristá (1985) assert that phraseology is indeed influenced by Bally’s (1909) ‘stable (fixed) combination of words’ but also recognizes that it was Vinogradov (1947) who set the “fundamental concepts of phraseology” (p. 7). A century later, García-Page Sánchez (2008) states the following on Bally:

Most scholars agree on proclaiming Ch. Bally as the founding father of Phraseology based on his brief, yet substantial notes on fixed combinations, which he presented mainly in his *Traité de stylistique française* (1909). Bally is also the first author who coined the term phraseology. Thus, we believe that he might rightfully be named founding father of this discipline since his was the first study of phraseological units in a systematical and scientific manner<sup>7</sup>. (p. 39 [my translation])

Nevertheless, in the work by Zuluaga (1980, pp. 31-37), and as presented in Article 1, three other authors, who have used other denominations of PUs earlier than the works of Bally (1909) and Vinogradov (1947), are presented (see Table 2).

*Table 2. Forefather denominations of PUs before the works of Bally (1909)*

<b>Author</b>	<b>Denomination</b>
Paul (1880)	<i>isolierte Formel</i>
	<i>stehende Formel</i>
	<i>stehende Verbindung</i>
	<i>feste Verbindung</i>
	<i>Erstarrung einer Flexionsform</i>
	<i>freier Akk</i>
	<i>gebundener Akk</i>
Bréal (1897)	<i>Formeln</i>
	<i>schöpferische Gruppen</i>
	<i>Groupes articulés</i>
	<i>Locution</i>
von der Gabelentz (1901)	<i>Formule</i>
	<i>stehende Redensarten</i>
	<i>gebundene Rede</i>

<sup>7</sup> Original in Spanish: *La mayoría de los estudiosos está de acuerdo en proclamar a Ch. Bally como el fundador de la Fraseología en virtud de los breves, pero enjundiosos apuntes sobre las combinaciones fijas expuestos sobre todo en su Traité de stylistique française (1909); él es también el primer autor en acuñar el término fraseología, creemos que con justicia puede atribuírsele tal distinción al ser la primera vez que lo que hoy conocemos por unidades fraseológicas recibe un estudio mínimamente científico y sistemático.*

As seen in the work by Zuluaga (1980, pp. 31-37), there was an extensive number of denominations and definitions used to determine the object of study of phraseology, even before it became a standalone discipline. The question regarding the object of study of phraseology entails a problem that is still a matter of controversy in phraseological studies. Article 1 offers an overview of some authors and the denominations that influenced some works written in Spanish and English (see Tables 3 and 4).

In the Spanish tradition<sup>8</sup>, the types of units that are the object of study of phraseology have been roughly bound to two phraseological notions: the narrow one and the wide one. According to López (2012, p. 57), the wide notion is based on the work by Corpas Pastor (1996) while the narrow notion is based on the work by Ruiz Gurillo (1997). The narrow notion sets the most restrictive and semantically opaque expressions, i.e., ‘idioms,’ to be the object of study of phraseology. The wide notion includes both ‘idioms’ and ‘collocations’ as the objects of study of phraseology.

*Table 3. Denominations of PUs proposed by the most representative authors related to general phraseology in Spanish*

<b>Author</b>	<b>Denomination</b>	<b>Definition</b>
Casares (1950)	<i>Locuciones</i> <i>Frases hechas</i> <i>Refranes</i> <i>Modismos</i>	
Zuluaga (1980)	<i>Locuciones</i> <i>Enunciados</i>	
Carneado Moré & Tristá (1985)	<i>Unidad fraseológica (fraseologismo)</i> <i>(verbal, reflexivo, propositivo, participial,</i> <i>conjuntivo, pronominal, nominal, adjetival,</i> <i>adverbial)</i>	Wide
Corpas Pastor (1996)	<i>Unidad fraseológica</i> <i>(Colocación, Locución, Enunciado</i> <i>Fraseológico)</i>	
Ruiz Gurillo (2001)	<i>Locuciones</i> <i>(nominal, adjetival, verbal, adverbial,</i> <i>marcadora, propositiva, clausal)</i>	Narrow
García-Page Sánchez (2008)	<i>Locuciones</i>	

<sup>8</sup> This includes all works and authors contributing to phraseological studies regarding the Spanish language.

In contrast, the English tradition<sup>9</sup>, according to Norrick (2007), could be divided into two different sub-traditions: the American and the British traditions. Norrick (2007, p. 615) points out that Anglo-American linguists have been focused on studying phraseology and its relationship with the lexicon, syntax, and semantics rather than claiming that phraseology is an independent field within linguistics. This fact could be one of the reasons for the lack of consensus among scholars around the denomination of the object of study of phraseology, as well as the interchangeable use of ‘idiom’ and ‘collocation.’

In the American tradition, three milestones could be identified as the first peripheral studies related to phraseology, namely: (i) the “list of the irregularities in a language” written by Bloomfield (1933), (ii) the grouping of phraseological units in a category denominated ‘idioms’ by Hockett (1958), and (iii) the grammar written for those units by Householder (1959) (Norrick, 2007, p. 615). However, Chafe’s (1968) criticism to the compositionality criterion of Chomskyan linguistic theory establishes the principles for phraseological research (Norrick, 2007, pp. 615-316).

Furthermore, Norrick (2007) states that two notions started the first peripheral studies on phraseology in the British tradition: the notion of ‘collocation,’ put forward by Firth (1957) and the notion of lexis as a linguistic level argued by Halliday (1966). However, it was the development of corpus linguistics and the works by Svartvik and Quirk (1980) and Sinclair (1991) that gave impulse to the growth of phraseology studies in the UK.

Moreover, Norrick (2007, pp. 616-617) highlights the contribution made by several other Anglo-American authors to the study of phraseology from many linguistic subfields including pragmatics and speech acts (Morgan, 1978; Sadock, 1972), cognitive linguistics (Lakoff, 1987; Lakoff & Johnson, 1980),

---

<sup>9</sup> This includes all works and authors contributing to phraseological studies regarding the English language.

and language acquisition (Wray, 2002). Another important milestone in the study of phraseology in the English tradition is given to the work by Makkai (1972) which was derived from one of the first monographs completely devoted to phraseology. Undoubtedly, as asserted by Cowie (1998, p. 7), both the Spanish and the English traditions have been influenced by many other authors from several other linguistic traditions when referring to the object of study of phraseology. Table 4 presents a list of denominations used by several other authors from different language traditions.

*Table 4. Denominations of phraseology by different authors according to Cowie (1998, p. 7)*

<b>Author</b>	<b>Denomination</b>
Vinogradov (1947)	Phraseological unit (phraseological fusion, phraseological unity, phraseological combination)
Amosova (1963)	Phraseological unit (idiom, phraseme/phraseoloid)
Cowie (1981)	Composite (pure idiom, figurative idiom, restricted collocation)
Mel'čuk (1988)	Semantic phraseme (idiom, collocation)
Gläser (1986)	Nomination (idiom, restricted collocation)
Howarth (1996)	Composite unit (pure idiom, figurative idiom, restricted collocation)

Among the previously mentioned authors, the work of Mel'čuk plays a leading role in the present PhD dissertation (see Article 2). Mel'čuk (1998, 2012, 2013) offers an extensive study on phraseological units including a detailed taxonomy for the classification of PUs in LGP. Mel'čuk's initial definition of a 'non-free phrase' sets the basis for the subsequent definition of its subcategories.

A phrase is non-free (= phraseologized) if at least one of its lexical components  $L_1$  is selected by the speaker in a linguistically constrained way – that is, as a function of the lexical identity of other component(s). (Mel'čuk, 2012, p. 33)

A synthesis of Mel'čuk's taxonomy and definitions is offered in Article 2, and illustrated in Appendix A, Table 1A.

As seen in Table 1A, Mel'čuk's taxonomies (1998, pp. 6-8; 2012, pp. 37-40) offered a groundbreaking characterization of PUs. However, in Articles 2 and 3, all these subcategories were regrouped into three main subcategories taken from the work by Mel'čuk (1998), namely: 'idioms,' 'collocations,'

and ‘pragmatemes.’ The purpose of denominative reduction is two-fold: (i) it intends to offer a more intuitive set of marks for PU and SPU identification in lexicographic resources and (ii) it can be employed as a starting point to find a common ground for the analysis of differences and similarities between LGP and LSP phraseology as presented in Article 2.

### **3.1.3. Phraseology in language for specific purposes and its object of study**

As asserted by Kjær, phraseology is an “independent academic discipline within linguistics” (2007, p. 507). Nevertheless, LSP phraseology research has not been widely explored or institutionalized to the point of becoming a non-coherent research field (Kjær, 2007, p. 507). Indeed, phraseology has not been studied for as long as its LGP counterpart, it is necessary to highlight that scholars have been working exhaustively on this matter during the last two decades, e.g., Aguado de Cea (2007); Bevilacqua (2004); Buendía Castro and Faber (2015); Fraile Vicente (2007); Hourani-Martín and Tabares-Plasencia (2020); Kübler and Pecman (2012); L’Homme and Bertrand (2009); Leroyer (2006); Lorente Casafont (2002); Montero Martínez (2008) (Rojas Díaz, forthcoming-a, p. 5).

Nevertheless, the delimitation of the object of study of LSP phraseology is even more problematic than in its LGP counterpart. This is due to the lack of agreement as regards its denominations, definitions, and identification criteria. This issue is addressed in Article 2:

Regarding the object of study of LSP phraseology, one might say that the different denominations used (e.g., LSP phrase, phraseologism, LSP collocation, specialized lexical combinations, legal phraseological unit) intend to distinguish it from the object of study of LGP phraseology (especially within lexicography). Therefore, the question arises whether LSP phraseology should be denominated as such or whether another denomination should be used instead to name the study of phraseological units specifically in the context of specialized languages. (Rojas Díaz, forthcoming-a, p. 6)

Appendix A, Table 2A presents some denominations and definitions, consulted in this PhD dissertation gathered as a basis for the research in Article 2.

As pointed out in Article 2, three identification criteria could be extracted from these definitions: “(i) they refer to phrases consisting of two or more elements, (ii) these phrases include a term as part of their lexical components (i.e., they are plurilexical), and (iii) they are used in LSP.”

However, Article 2 offers a comprehensive SPU taxonomy in which the identification criteria were extended by the inclusion of different levels of idiomaticity and the absence of monolexical terminological constituents among its word forms. On idiomaticity (especially semantic idiomaticity), Article 2 claims that it is an aspect that has been understudied in LSP phraseology:

[...] the lack of explicit details regarding the semantic opacity of SPUs, which could be considered one of the main characteristics of PUs, brings up the question of whether SPUs lack this characteristic or whether lexicographic resources are not registering —or overlooking— those units. (Rojas Díaz, forthcoming-a, p. 9)

Concerning the absence of monolexical terminological constituents, 109 SPUs, that did not include terminological constituents among its word forms, were identified in Article 2:

[...] previous definitions in LSP phraseology tend to focus on the presence of a terminological unit in the phrase. Nevertheless, several SPUs (e.g., ‘at arm’s length’, ‘*los cinco dragones*’ [the five dragons]) were metaphorical in nature, meaning that they entail a terminological *tenor*, [...] by means of non-terminological *vehicles* [...]. (Rojas Díaz, forthcoming-a, p. 25)

Article 2 tackles these two issues about idiomaticity and terminology by proposing an alternative SPU definition with its correspondent taxonomy based on the work by Mel’čuk (1998, pp. 6-8; 2012, pp. 37-40) (see section 5.1.) However, as this PhD dissertation uses dictionaries as part of the data sources, some notions related to lexicography must also be considered and the next section is devoted to them.

### **3.2. Key notions within lexicography**

For the dictionary characterization, seven concepts are defined in this section, namely: (i) megastructure, (ii) macrostructure, (iii) microstructure, (iv) entry, (v) sub-entry, (vi) onomasiology, and (vii) semasiology.

According to Hartmann and James (1998), the lexicographic megastructure “includes the macrostructure and the outside matter” (p. 93); the macrostructure is “the overall list structure which allows the compiler and the user to locate information” (p. 91); finally, the microstructure is defined as “the internal design of a reference unit” (p. 94). Among those ‘reference units’ mentioned by Hartmann and James (1998, p. 94), we can find the entry, which is defined as “the basic reference unit in a dictionary or

other reference system such as a library catalogue” (Hartmann & James, 1998, p. 50). Moreover, the authors offer a list of the component parts of the entry:

In the dictionary, depending on its content and purpose, these component parts are common: the lemma (which allows the compiler to locate and the user to find the entry within the overall word-list); the formal comment on the ‘topic’ introduced by the lemma (spelling, pronunciation, grammar); and the semantic ‘comment’ (definition, usage, etymology). In case of multiple meanings of the lemma, the entry is subdivided into (usually numbered or otherwise marked) sections called ‘sub-entries’ or ‘sub-senses’, each of which provide the same basic information categories. (Hartmann & James, 1998, p. 50)

The definition of entry offered by Hartmann and James (1998), does not include any explicit information about phraseology. The authors explain it through the notion of ‘phrasal entry’: “a multi-word expression appearing as a headword or as a sub-lemma” (Hartmann & James, 1998, p. 108). The work by Martínez de Sousa (1995) offers a definition that encompasses the possibility of having PUs and SPUs as entries:

Word, idiom, phrase, sign or group of letters or signs that head a lexicographic article, vocabulary, glossary, terminology, index, index card, etc., and it is an object of definition or explanation and which could be selected for encyclopedic treatment<sup>10</sup>. (p. 180 [my translation])

As presented in Articles 2 and 3, the SPUs could appear in the dictionary as entries or sub-entries. Sub-entries are defined by Martínez de Sousa (1995) as: an “entry embedded in another one (a main one) on which it depends”<sup>11</sup> (p. 315 [my translation]). This PhD dissertation offers a hybrid method for the indexation of PUs and SPUs (see Article 2 and section 5.3.) in which the notions of onomasiology and semasiology need to be looked into. Hartmann and James (1998) define onomasiology as:

An approach in semantics which is concerned with the matching of the most appropriate word or phrase to a given concept. When its principles are applied to the ‘tip-of-the-tongue’ phenomenon as the result can be a reference work (an onomasiological dictionary) which guides the user from relatively well-known concepts to relatively less familiar words [...]. (p. 102)

---

<sup>10</sup> Original in Spanish: *palabra, locución, frase, sintagma, signo o conjunto de letras o signos que encabeza un artículo del diccionario, vocabulario, glosario, terminología, índice, ficha, etc., y es un objeto de definición o explicación y, eventualmente, de tratamiento enciclopédico*

<sup>11</sup> Original in Spanish: *Entrada refundida en otra principal de la cual depende.*

In other words, when applied to lexicographic resources onomasiology would answer the question ‘how can I express Y (concept)?’ Contrary to onomasiology, semasiology is defined by the authors as:

An approach in semantics concerned with the explanation of the meaning of given words or phrases. Traditional monolingual and bilingual dictionaries supply such semasiological information (e.g., in terms of definitions and translations equivalents.) (Hartmann & James, 1998, p. 124)

Therefore, a semasiological lexicographic resource offers ‘the possible meanings of X (word/phrase/idiom/proverb).’

### 3.3. Key notions within terminology

Articles 2 and 3 are mainly focused on the study of LSP phraseology. As presented in Appendix A, Table 2A, most definitions of SPUs include either a definition of ‘LSP’ or ‘term’. Regarding the notion of ‘term,’ L’Homme (2020, p. 55) states that “there is not consensus” about it (p. 55). For instance, Cabré (2000) defines term as:

These units (terminological units/terms) are, at the same time, similar and different from the lexical units of a language, denominated words by lexicology. Their specialized character can be identified through their pragmatic aspects and the mode of constructing their meaning. Their signified is the outcome of negotiation among experts. This negotiation happens within the specialized discourse through the use that determines the meaning of each unit. (Cabré, 2000, p. 14 [my translation])<sup>12</sup>

A different definition is offered by Faber Benítez and López Rodríguez (2012):

Trying to find a distinction between terms and words is no longer fruitful or even viable, and the best way to study specialized-knowledge units is by studying their behavior in texts. (p. 22)

As mentioned in Article 2:

The difference between the two previously presented notions of term, reside in the approach of each of these theoretical postulations regarding the object of study of terminology, rather than on the object of study itself. Furthermore, Faber Benítez and López Rodríguez (2012, pp. 22-23) explicitly mention characteristics that their definition share with Cabré’s (2000) notion of term,

---

<sup>12</sup> Original in French “*Ces unités sont en même temps semblables et différentes des unités lexicales d’une langue, appelées mots par la lexicologie. Leur spécificité se trouve dans leur aspect pragmatique et dans leur mode de signification. Leur signifié est le résultat d’une négociation entre experts. Cette négociation se produit dans le discours spécialisé à travers des prédications qui déterminent le signifié de chaque unité.*”



such as the predominance of nominal units, the relations between the TUs and LSP domains, and the combinatorial value of TUs. (Rojas Díaz, forthcoming-a, p. 7)

L’Homme (2020) addresses a middle-point notion of term which will be used as the working definition in this PhD dissertation as it allows for the classification of linguistic items according to their meaning within a certain subject field. Hence, this characteristic (see section 4.2.) could be used for the classification of word forms into semantic fields related to a certain LSP domain (commerce and economics for this study):

Stating that a linguistic item is a term is considering its meaning from the perspective of a special subject field. There is no such a thing as a term in essence; a linguistic unit becomes a term relative to their subject field in which it is considered. [...]. This also means that even common linguistic items can become terms in specialized domains. [...] Finally, a linguistic item can also be a relevant term in fields of knowledge. (p. 59)

Concerning the definition of LSP, Hoffmann asserts that:

A specialized language (LSP) is the group of all the linguistic resources that are used in a communication field –delimited by the specialized discipline– to ensure the understanding among the people that work in that field. (Hoffmann, 1998, p. 57 [my translation])<sup>13</sup>

After presenting the key notions taken from terminology, the next section is used to present concepts from translation.

### **3.4. Key notions within translation**

The notions of (i) equivalence and (ii) translation technique were explored in Article 3. Several definitions of equivalence were offered from diverse disciplines: translation studies, lexicography, terminology, and phraseology. Detailed tables with some of the definitions offered by these disciplines are presented in Appendix A Table 3A, 4A, 5A, and 6A. Article 3 also shows examples of decisions made by translators in the EUR-Lex corpus when dealing with SPUs. Nevertheless, the classification of these techniques pose a problem since, as stated by Muñoz Martín (2000) “terminological homogeneity is,

---

<sup>13</sup> Translation in Catalan: “*Un llenguatge d’especialitat és el conjunt de tots els recursos lingüístics que s’utilitzen en un àmbit comunicatiu — delimitable pel que fa a l’especialitat— per tal de garantir la comprensió entre les persones que treballen en aquest àmbit*”. Original in German “*Fachsprache – das ist die Gesamtheit aller sprachlichen Mittel, die in einem fachlich begrenzten Kommunikationsbereich verwendet werden, um die Verständigung zwischen den in diesem Bereich tätigen Menschen zu gewährleisten.*“

paradoxically, not one of the characteristics of translation theory” (p. 129). For instance, Gil-Bardaji

(2011) asserts that:

‘Procedures,’ ‘techniques,’ ‘strategies,’ ‘processes,’ ‘strategic processes,’ ‘rules,’ ‘plans,’ etc., are some of the most common terms used to refer to this more general notion of ‘translator operative knowledge’ or part of it. (p. 162)

The work by Molina and Hurtado Albir (2002, p. 507) distinguishes ‘methods,’ ‘strategies,’ and ‘techniques’ as different categories. Their definitions are presented in Table 5

*Table 5. Definitions of ‘method,’ ‘strategy,’ and ‘technique’ by Molina and Hurtado Albir (2002, pp. 507-509)*

Denomination	Definition
Translation method	Translation method refers to the way a particular translation process is carried out in terms of the translator’s objective, i.e., a global option that affects the whole text.
Translation strategy	Strategies are the procedures (conscious or unconscious, verbal or nonverbal) used by the translator to solve problems that emerge when carrying out the translation process with a particular objective in mind.
Translation technique	procedures to analyse and classify how translation equivalence works. They have five basic characteristics: 1) They affect the result of the translation 2) They are classified by comparison with the original 3) They affect micro-units of text 4) They are by nature discursive and contextual 5) They are functional

The denominations, definitions and classification of translation techniques offered by Molina and Hurtado Albir (2002, pp. 507, 509-511) were adopted for this PhD dissertation in Article 3 and employed in the analysis of translation equivalents of SPUs extracted from the EUR-Lex corpus.

## **4. Methods, data, and tools**

### **4.1. The methodology**

In this PhD dissertation, tasks developed within corpus linguistics were employed in the data collected for the three articles. These tasks included the empirical description of language through data tagging (Leech, 1992; Sinclair, 1992), and the instantiation by corpus interrogation (Halliday, 1992) of PU and SPU entries and equivalents. A summarized methodology of the research process is offered in Figure 2.

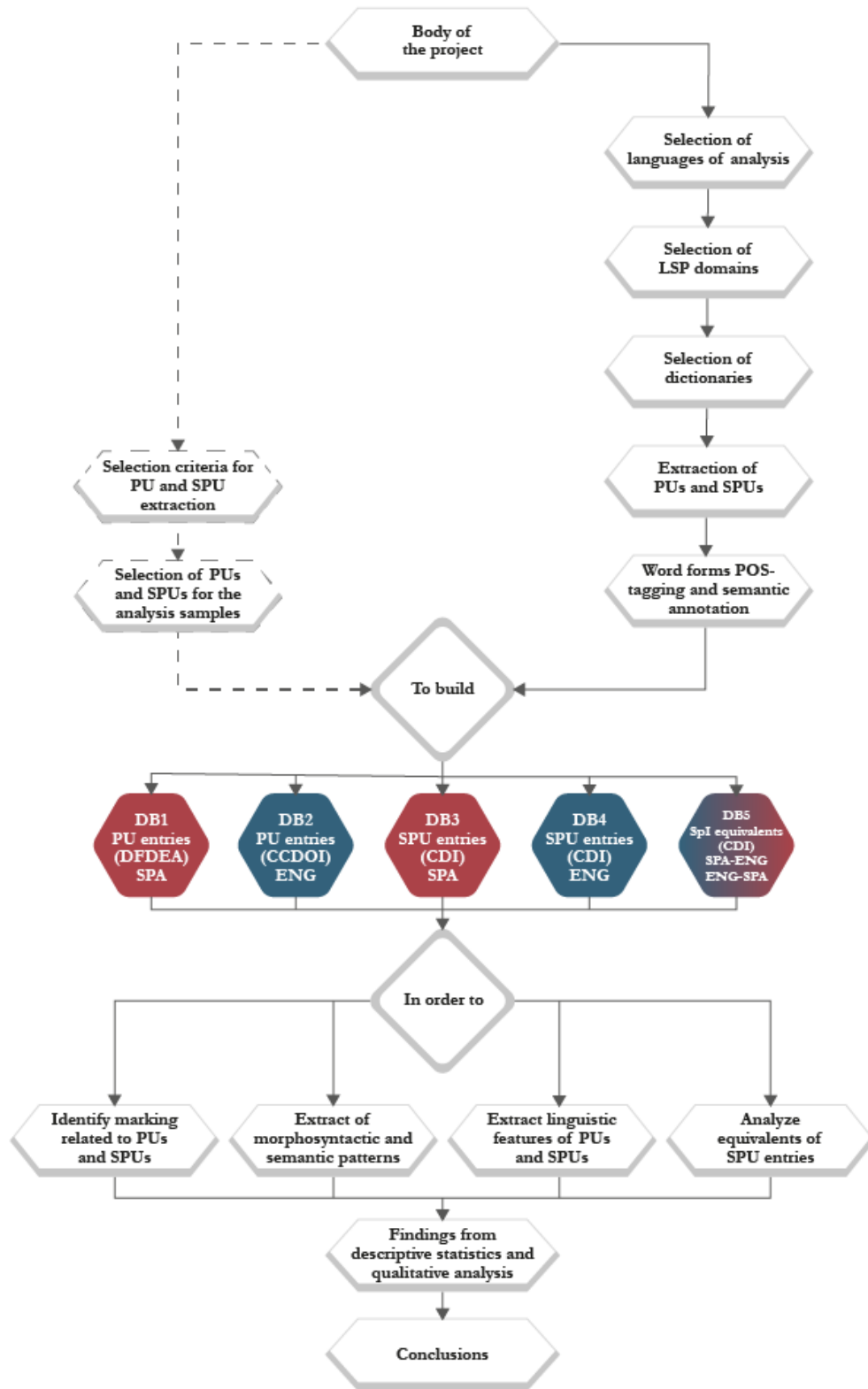


Figure 2. General methodology chart of the PhD dissertation

Since the motivation for this PhD dissertation lies in taking different perspectives on LGP and LSP phraseology, three sets of data were chosen for the analyses throughout the PhD dissertation: lexicographic PU entries, lexicographic SPU entries, and lexicographic SPU equivalents (see Figure 2).

The datasets have three main objectives: (i) to offer a series of linguistic (i.e., lexical, lexicographical, morphosyntactic, and semantic) characteristics for the identification of PUs and SPUs extracted from dictionaries (Articles 1 and 2), (ii) to retrieve empirical data to support the theoretical reflection contributions (Articles 1 and 2), and (iii) to carry out a cross-linguistic analysis of SPUs in corpora (Article 3).

#### **4.2. The dictionaries**

This sub-section offers the reasons and the criteria for choosing the selected dictionaries as well as the corpus for the contrastive analyses.

As stated in Article 2:

As noted by Kübler and Pecman (2012, p. 187), globalization and standardization processes have had an impact on the need for LSP lexicographic resources that can both standardize and describe specific domains by offering definitions. Undoubtedly, commerce and economics have played a leading role in those globalization processes. Moreover, these LSP domains offer an interesting case of terminological overlap, since, as stated by Simonnæs and Kristiansen, “business, finance and economics are in many cases intertwined with law” (2018, p. 157). (Rojas Díaz, forthcoming-a, p. 10)

Two reasons are highlighted for choosing lexicographic resources to build the analysis datasets:

(i) dictionaries and terminological databases ensure the presence of TUs from several levels of abstraction as proposed by Hoffmann (1998, pp. 72-73) and (ii) they are aimed at different users’ needs (Kübler & Pecman, 2012, p. 187). Hence, the following criteria were considered in the choice of the lexicographic resources:

- Article 1.
  - That it be an LGP phraseological dictionary
  - That it be monolingual (one in Spanish and one in English)
  - That its publishing house be recognized as a lexicographic authority

- Article 2.
  - That it be related to commerce and economics
  - That it be bilingual (Spanish-English/English-Spanish)
  - That its publishing house be recognized as a lexicographic authority

Based on these criteria, two dictionaries were chosen to extract the units of Article 1: the *Diccionario fraseológico documentado del español actual* (DFDEA) (Seco, Andrés, & Ramos, 2004) and The Collins COLBUILD dictionary of Idioms (CCDOI) (Sinclair & Moon, 1997). Some generalities of the DFDEA and the CCDOI are presented in Table 6.

*Table 6. General lexicographic information of the DFDEA and the CCDOI*

<b>Lexicographic information</b>	
Spanish	
Editors	Seco, Manuel; Andrés, Olimpia; and Ramos Gabino
Entries	Approximately 16,000
Format	Physical printing (16 x 24 cm)
Pages	1,120
Publishing house	Aguilar
English	
Editors	Sinclair, John and Moon, Rosamund
Entries	Approximately 4,400
Format	Physical printing (13 x 20 cm)
Pages	493
Publishing house	HarperCollins Publishing

In Article 2, the dictionary chosen was the DCI. On an editorial note, the DCI is presented as a dictionary aimed at various users including “field experts and scholars from diverse areas of Economics, International Commerce, and linguistic mediators” (Alcaraz & Castro Calvín, 2007). The DCI could be categorized as a descriptive, semasiological, and synchronic dictionary. Some other generalities of the DCI are presented in Table 7.

*Table 7. General lexicographic information of the DCI*

<b>Lexicographic information</b>	
Field	Information
Editors	Alcaraz varó, Enrique and Castro Calvín, José
Entries	Approximately 17,000
Format	Physical printing (16 x 25 cm)

Languages	Spanish-English/English-Spanish
Pages	1,168
Publishing house	Ariel

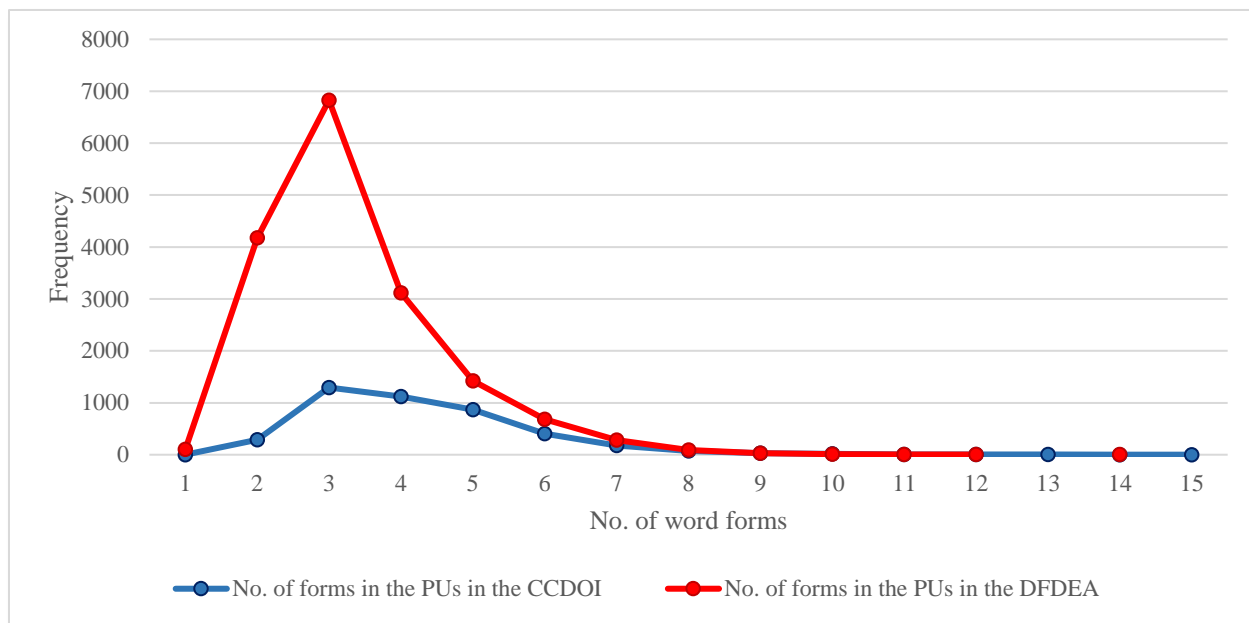
### 4.3. The databases<sup>14</sup>

After choosing the dictionaries, the boundaries for the analyses were set, i.e., the criteria for the selection of the datasets to build the databases were defined. Since LSP phraseology has not been studied as much as its LGP counterpart (Kjær, 2007, p. 507), it was a logical choice that Article 1 was devoted to gathering the core of linguistic features (lexical, morphosyntactic, and semantic) from where to take a *tertium comparationis* to make the later comparative analyses, e.g., LGP phraseology-LSP phraseology and SPU entries-SPU equivalents (Spanish-English/English-Spanish).

The extracted data for Article 1 included 16,760 PUs (composed by 55,831 word forms) in Spanish and 4,285 PUs (composed by 18,123 word forms) in English (Rojas Díaz, 2020, p. 302). Figure 3 shows the length of the PUs consisting of three, four and five word forms covered 67% of the PUs from the DFDEA and 76% of the PUs from the CCDOI (Rojas Díaz, 2020, pp. 303-304).

---

<sup>14</sup> Concerning data availability of the datasets and databases built for this work, I am committed to finding a solution to make it available for future research (e.g., in the CLARINO portal or other data repositories available).



*Figure 3. No. of word forms of the PUs in the DFDEA and the CCDOI (Rojas Díaz, 2020, p. 302)*

For feasibility reasons, it was decided that all the expressions containing three, four and five word forms were extracted to build two databases (one in Spanish with 4,932 PUs from the DFDEA and one in English with 2,387 PUs from the CCDOI).

Article 2 followed this same criterion of word form length. However, unlike the DFDEA and the CCDOI, the DCI did not offer any explicit marking to allow the differentiation of the extracted SPUs. It was not possible to determine if a certain SPU was a ‘specialized idiom’ (SpI), a ‘specialized collocation’ (SpC), or a ‘specialized pragmateme’ (SpP). By following the word form length criterion, 6,230 SPUs (composed by 22,818 word forms) in Spanish and 4,856 SPUs (composed of 16,525 word forms) in English were used to build two more databases to analyze several of the linguistic features (including the type of phrase, morphosyntactic pattern, semantic pattern, among others) that were already identified in Article 1. As seen in section 2, Article 2 offered a definition and a taxonomy for the classification of SPUs including the SpIs. After classifying all the entries, a subset of SpIs (composed by 69 SpIs in Spanish and 40 SpIs in English) was identified. These SpIs were isolated (along with their equivalents) to create a bilingual (Spanish-English/English-Spanish) database for the analysis in Article 3.



#### **4.4. The corpus**

Article 3 was devoted to the analysis of SpI equivalents from the DCI and a corpus. The queries were carried out on the EUR-Lex corpus. As Simonnæs and Kristiansen (2018) assert, “it may be difficult to clearly delineate what constitutes ‘economic’ language” (p. 152). Moreover, they argue that this difficulty originates in how the term ‘economic’ has been “frequently used with reference to a number of domains and subdomains” (Simonnæs & Kristiansen, 2018, p. 152). Further discussion on this issue is presented in Article 2 (see sections 2 and 3.1.) and Article 3 (see section 4).

Therefore, the chosen corpus should meet similar criteria to those used in the dictionary selected for Article 2. From the available corpora in the Sketch Engine<sup>15</sup>, the EUR-Lex corpus was the most suitable option since it is composed from texts that range from regulations, decisions, and preparatory acts to treaties, international agreements, and EFTA documents. Furthermore, it is one of the largest multilingual corpora available in the Sketch Engine (Kilgarriff et al., 2014) with approximately 633.4 million words translated from Spanish into English and 594.1 million words translated from English into Spanish (Baisa, Michelfeit, Medved', & Jakubiček, 2016, pp. 2800-2802).

#### **4.5. The tools**

The data, extracted from three dictionaries and a corpus, needed to be processed and analyzed (see Figure 2). For that reason, three databases (two in LGP [one in Spanish and one in English] and one in LSP [in Spanish and English]) were built. The software used for building the databases, as well as to carry out the descriptive statistics, was Microsoft Excel for Microsoft 365. The visualization of the databases was done by linking the Microsoft Excel books, containing the databases, with Microsoft Power BI as seen in Figure 4.

---

<sup>15</sup>Available online at: <https://www.sketchengine.eu/>

(1)

Language	Morphosyntactic pattern	Count	Semantic pattern	Count	Expression
<input type="checkbox"/> English	N Prep N	2540	I Z I	252	a bordo de un buque
<input type="checkbox"/> Spanish	Adj N N	605	A Z I	215	a corto plazo
<b>POS</b>	N Prep Det N	489	M Z M	188	a costa de
<input type="checkbox"/> noun phrase	N N N	459	Q Z I	145	a cuatro bandas
<input type="checkbox"/> other	N Prep N Adj	453	S Z I	143	a dos bandas
<input type="checkbox"/> verb phrase	V Det N	438	I Z M	115	a efectos aduaneros
<b>Foreign words</b>	N Prep N Prep N	313	M Z I	110	a efectos de
<input type="checkbox"/> (Blank)	N Adj Prep N	260	I Z A	105	a efectos fiscales
<input type="checkbox"/> Foreign word (English)	N Adj Adj	251	A Z M	102	a fondo perdido
<input type="checkbox"/> Foreign word (French)	N Prep Adj N	162	A Z A	96	a forfait finance
<input type="checkbox"/> Foreign word (Latin)	V Prep N	149	Q Z M	89	a la entrada
<b>Hyphen</b>	N Prep N N	139	Q Z A	81	a la espera de instrucciones
<input type="checkbox"/> (Blank)	N Contr N	137	A Z S	77	a la llegada
<input type="checkbox"/> Hyphen	Adj N Prep N	120	S Z M	70	a la orden
<b>Abbreviation / Acronym</b>	Adj Adj N	118	I Z S	69	a la orden de
<input type="checkbox"/> (Blank)	Adj Prep N	118	A Z Q	67	a la par
<input type="checkbox"/> Abbreviation / Acronym	N Prep N PP	101	M Z A	64	a la presentación
<b>Semantic relations</b>	N Prep Det N Adj	96	N Z I	64	a la presentación de
<input type="checkbox"/> metaphor	N Adj Prep Det N	95	S Z S	61	a la vista
<input type="checkbox"/> metonymy	Adj - N N	94	I Z T	58	a largo plazo
<input type="checkbox"/> metonymy / metaphor	N N Prep N	94	I Z Z I	58	a lo largo de
<input type="checkbox"/> none (compositional)	V Prep Det N	80	A Z Z I	57	a más tardar
<b>SPU type</b>	Adj N N N	59	Q Z S	54	a medio plazo
<input type="checkbox"/> hybrid specialized collocation	N Prep Adj	53	T Z I	48	a menos que
<input type="checkbox"/> specialized collocation	N Adj N	52	X Z I	47	a modo de anticipo
<input type="checkbox"/> specialized idiom	Prep Adj N	52	A Z O	42	a partir de la fecha
<input type="checkbox"/> specialized pragmateme	Prep N Prep N	52	M Z N	42	a portes debidos
<b>No. of word forms</b>	Prep N Adj	47	Q Z Q	41	a portes pagados
<input type="checkbox"/> 3	V Det N Prep N	46	N Z N	40	a reserva de
<input type="checkbox"/> 4	N Adj Prep N Adj	45	T Z M	40	a son de mar
<input type="checkbox"/> 5	N Adv Adj	45	M Z O	39	a su debido momento
	PP N N	44	S Z A	38	a su valor nominal
	N Prep N Contr N	43	I Z N	36	a tanto alzado
	Adj PrP N	41	M Z S	35	
	<b>Total</b>	<b>11086</b>	<b>Total</b>	<b>11086</b>	

(2)

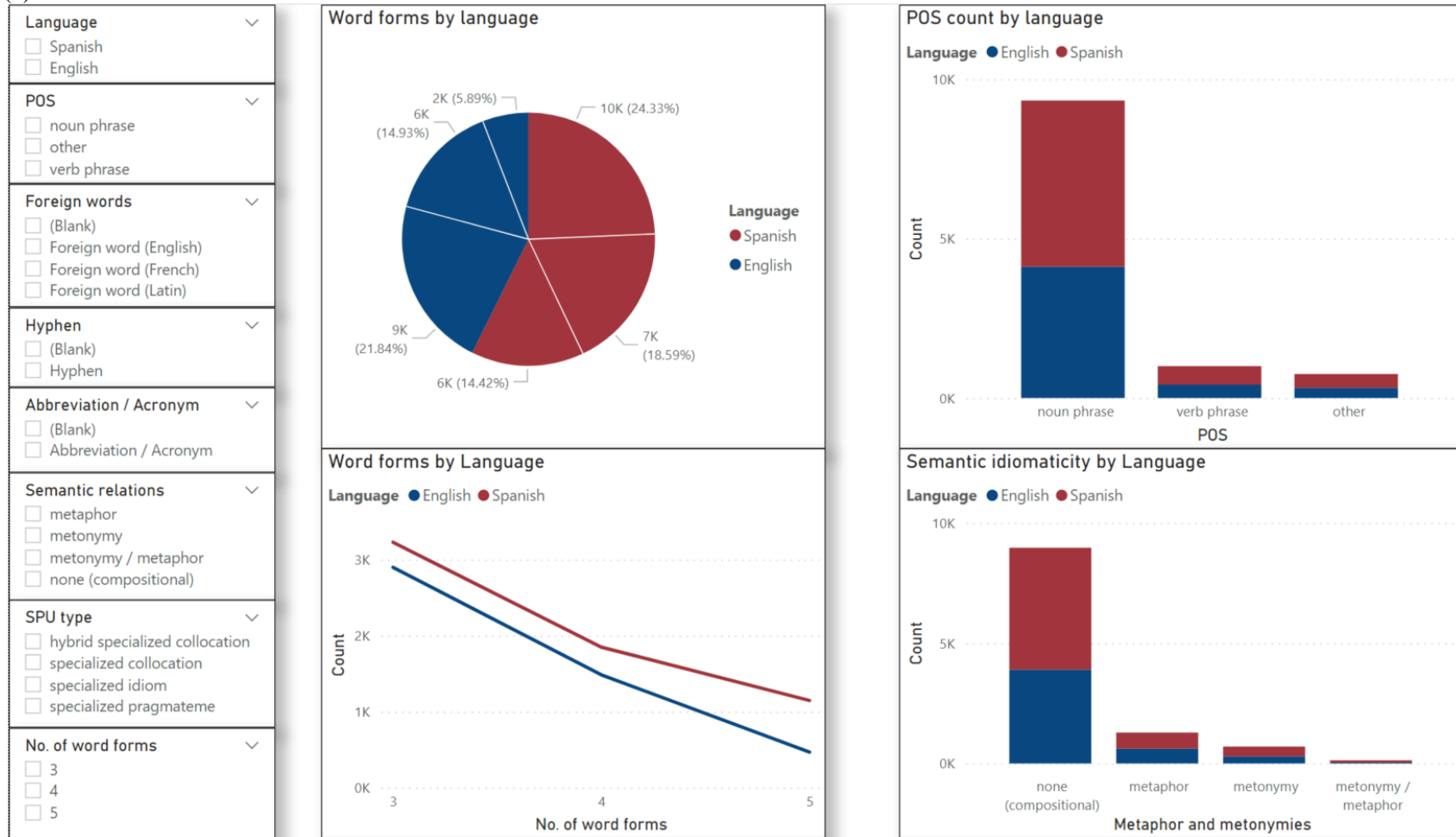


Figure 4. Data visualization example through a Microsoft Power BI dashboard (1) slicers and matrices (2) slicers and graphs

The databases contained the PU and the SPU entries extracted from the selected dictionaries as well as information about their individual word forms. Annotating each entry with this information allowed for their lexical and semantic analyses and the analysis of the morphosyntactic and semantic patterns of entries and their equivalents.

The morphosyntactic patterns were extracted based on the POS of each of the word forms. The POS-tagging was done by means of two different taggers: TreeTagger (Schmid, 1994) and RNNTagger (Schmid, 2019) (see Figure 5).

(1)

```

(base) rojasj@ubuntu: ~/RNNTagger
files/art_iii_eq_spa-eng.txt
at5 IN at
the DT the
expense NN expense
of IN of
# # #
cuadrilateral JJ cuadrilateral
# # #
bilateral JJ bilateral
# # #
for IN for
purposes NNS purpose
# # #
non FW non
- : -
refundable JJ refundable
# # #
non FW non
- : -
repayable JJ repayable
# # #
to TO to
the DT the
order NN order
# # #
order NN order
of IN of
# # #
o UH o

```

(2)

```

(base) rojasj@ubuntu: ~/RNNTagger
files/art_iii_eq_eng-spa.txt
ingresos NOUN.Masc.Plur ingreso
y CCONJ y
gastos NOUN.Masc.Plur gasto
corrientes ADJ.Plur corriente
# PUNCT #
partida NOUN.Fem.Sing partida
ordinaria ADJ.Fem.Sing ordinario
# SYM.Fem.Sing #
de ADP de
aeropuerto NOUN.Masc.Sing aeropuerto
a ADP a
aeropuerto NOUN.Masc.Sing aeropuerto
# SYM #
a ADP a
raya NOUN.Fem.Sing raya
# SYM #
distanciado VERB.Masc.Sing.Part distanciado
# SYM.Fem.Sing #
de ADP de
igual NOUN igual
a ADP a
igual NOUN igual
# SYM #
manteniendo VERB.Ger mantener
las DET.Def.Fem.Plur.Art el
distancias NOUN.Fem.Plur distancia
# SYM #
sin ADP sin

```

Figure 5. Example of POS-Tagging of SpI equivalents from (1) Spanish-English and from (2) English-Spanish

As seen in Figure 6, the extraction of morphosyntactic patterns was a process in three steps: (1) the individual POS-tags of each word form of the extracted PUs and SPUs were manually checked and homogenized. Next, a wild card (#) was used to mark the delimitation of each expression (2). Finally, regular expressions were used to ensemble the individual tags and extract the morphosyntactic patterns (3) of the PUs and SPUs.

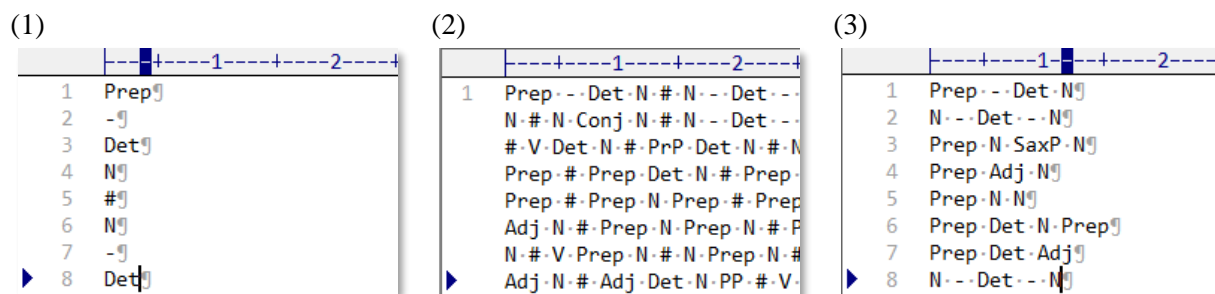


Figure 6. Example of POS-tagging of word forms (1) and morphosyntactic pattern ensemble process through regular expression in plain text (2) and (3).

Regarding the semantic patterns, the UCREL's Semantic Analysis System<sup>16</sup> (USAS) was employed for the semantic annotation of all the word forms in the database. USAS is a POS and semantic tagger with tags classified in 453 semantic categories based on 21 discourse fields (See Figure 7) identified by McArthur (1981) and Archer, Wilson, and Rayson (2002, p. 2).

<b>A</b> general and abstract terms	<b>B</b> the body and the individual	<b>C</b> arts and crafts	<b>E</b> emotion
<b>F</b> food and farming	<b>G</b> government and public	<b>H</b> architecture, housing and the home	<b>I</b> money and commerce in industry
<b>K</b> entertainment, sports and games	<b>L</b> life and living things	<b>M</b> movement, location, travel and transport	<b>N</b> numbers and measurement
<b>O</b> substances, materials, objects and equipment	<b>P</b> education	<b>Q</b> language and communication	<b>S</b> social actions, states and processes
<b>T</b> Time	<b>W</b> world and environment	<b>X</b> psychological actions, states and processes	<b>Y</b> science and technology
<b>Z</b> names and grammar			

Figure 7. McArthur's (1981) discourse fields (Archer et al., 2002, p. 2).

Since the online version of the USAS tagger for Spanish is not currently available, the semantic tagging of PU and SPU entries and equivalents had to be done manually using the single word semantic

<sup>16</sup> Available online at: <http://ucrel.lancs.ac.uk/usas/>

lexicon<sup>17</sup> used for earlier versions of the Spanish semantic tagger. The semantic annotation processes are explained in section 3.2. in Article 1 and section 4.2. in Article 2.

## **5. Dissertation contributions**

This section gives an overview of the main contributions of the three articles and of their limitations and suggestions for future research. The first three sub-sections will be the theoretical (section 5.1.), methodological (section 5.2.), and applied contributions (section 5.3.) of the dissertation. The last two sub-sections are devoted to present the limitations (section 5.4.) and future research lines (section 5.5.) of this project.

### **5.1. Theoretical contributions**

The starting point of this PhD dissertation was the study of LGP phraseology. The study carried out in Article 1 was aimed at retrieving the core set of linguistic features (lexical, lexicographic, morphosyntactic, and semantic) of PU entries extracted from phraseological dictionaries (DFDEA and CCDOI). The same linguistic features were used in Article 2 to identify and classify SPUs from a specialized dictionary (DCI) by offering a new definition of SPU and a taxonomy.

Before starting the extraction and analysis processes, it was necessary to understand and delimitate the objects of study of LGP and LSP phraseology. Articles 1 and 2 offer an exhaustive literature review of (what is considered for this PhD dissertation) the most significant postulations offered by authors that have influenced both the Spanish and the English traditions. This summary includes the denominations and definitions used by these authors and it is complemented in the present cover article by a list of denominations of PUs used before phraseology became a standalone discipline (see section 3.1.2.). Furthermore, Article 2, offers theoretical reflections upon the discipline by comparing the definitions and characterization of PUs and SPUs that went beyond the descriptive analysis of data. Given

---

<sup>17</sup> Available online at: [https://raw.githubusercontent.com/UCREL/Multilingual-USAS/master/Spanish/semantic\\_lexicon\\_es.usas](https://raw.githubusercontent.com/UCREL/Multilingual-USAS/master/Spanish/semantic_lexicon_es.usas)

the evidence of the occurrence of SpIs among the dictionary entries and that the existing definitions did not explain, or describe them, an opportunity arose to offer a more comprehensive SPU definition that could explain the occurrence of SpIs:

A combination of words (including, but not necessarily, monolexical terms) that evidences idiomaticity at least at one of the possible levels (lexical, pragmatic, semantic, statistical, or syntactic) and that, when used in a certain LSP domain, acquires a specialized meaning. (Rojas Díaz, forthcoming-a, p. 26)

Furthermore, this definition, along with the findings resulting from the database analyses, allowed for the creation of a taxonomy for SPU classification based on the works by Mel'čuk (1998, pp. 6-8; 2012, pp. 37-40). This new taxonomy uses the denomination of 'specialized phraseological unit' (SPU) as a hypernym of three other subcategories, namely: (i) 'specialized idiom' (SpI), (ii) 'specialized collocation' (SpC), and (iii) 'specialized pragmateme' (SpP) (see Figure 8).

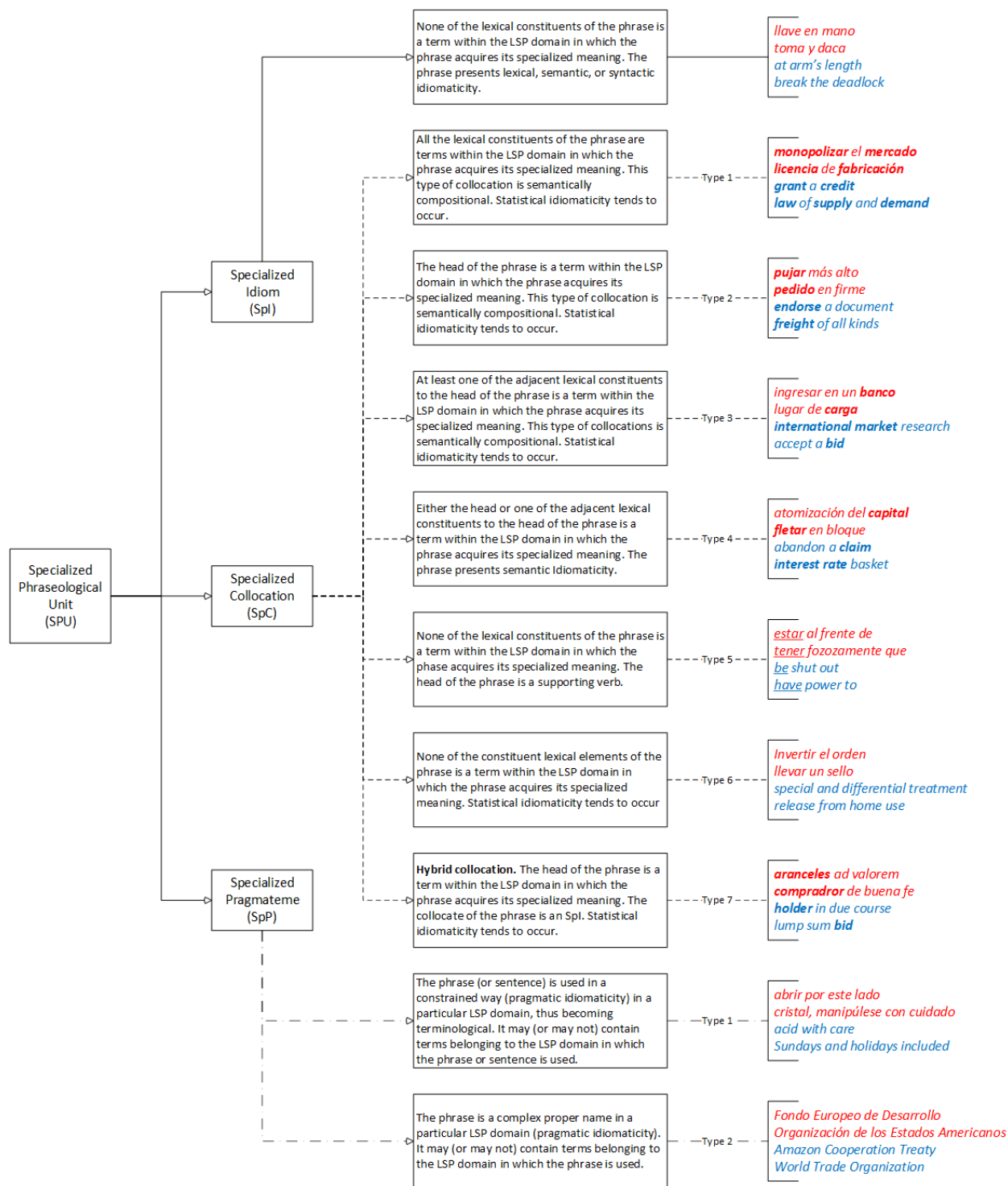


Figure 8. New taxonomy for SPU classification (Rojas Díaz, forthcoming-a, p. 27) based on the works by Mel'čuk (1998, pp. 6-8; 2012, pp. 37-40; 2013)



Article 3 offers an interdisciplinary literature review of the notion of equivalence through the lenses of translation, lexicography, terminology, and phraseology. The theoretical reflection after this extensive literature review highlights the debated and still debatable status of the notion of equivalence. Furthermore, the findings resulting from the analyses carried out in Article 3 demonstrate that:

SpIs are characterized by their lack of a monolexical term within their word forms (in this case, related to commerce and economics, e.g., *in bad faith*, *door-to-door*, *de mala fe*, *puerta a puerta*) that undergo a certain level of idiomaticity (lexical, semantic, or syntactic). (Rojas Díaz, forthcoming-b, p. 2)

The theoretical reflections of the present PhD dissertation intend to contribute to the solution of the continuing problem of the delimitation of the object of study of LSP phraseology. However, this new definition and the taxonomy linked to it must be tested to corroborate whether it can be useful for the enhancement of lexicographic and terminographic resources.

## 5.2. Descriptive data contributions

The three articles that comprise this PhD dissertation offered a large amount of data resulting from the analyses carried out. In Article 1, 21,045 idioms (16,760 in Spanish and 4,285 in English) were analyzed. To extract the different patterns, 73,954 word forms were tokenized and classified according to their POS, discourse field, and semantic category (Rojas Díaz, 2020, p. 302). The resulting information from the tagging and annotation tasks allowed for the extraction of morphosyntactic patterns (e.g., *la mínima expresión* ‘the minimum expression’ – Det Adj N | *a bean counter* – Det N N) and semantic patterns<sup>18</sup> (e.g., *lengua de trapo* ‘tongue of cloth’ – B Z O | *know the ropes* – X Z O). 816 morphosyntactic patterns (420 in Spanish and 396 in English) and 2,655 semantic patterns (1,365 in Spanish and 1,290 in English) were extracted following a ratio of three semantic patterns per each morphosyntactic pattern (3:1) in both the DFDEA and the CCDOI.

---

<sup>18</sup> chains of semantic tags (see Figure 7) –that I called semantic patterns– taken from the works of McArthur (1981) and Archer et al. (2002, p. 2)

For Article 2, 11,086 entries and sub-entries (6,230 in Spanish and 4,856 in English) were extracted from the DCI. These units posed a challenge since their phraseological typology (i.e., idioms, collocations, etc.) was not provided in the linguistic information of the phraseological entries of the dictionary. Thus, the analyses had to be aimed at their classification. In Article 2, 39,832 word forms (22,839 in Spanish and 16,993 in English) were tokenized, POS-tagged, and semantically annotated in order to identify morphosyntactic (e.g., *incluir en la lista negra* ‘include in the black list’ – V Prep Det N Adj | *break into a foreign market* – V Prep Det Adj N) and semantic patterns (e.g., *penetrar en el mercado* ‘penetrate the market’ – M Z Z I | *make out a receipt* – A Z Z Q). In total, in 1,054 morphosyntactic patterns (457 in Spanish and 597 in English) and 4,369 semantic patterns (2,131 in Spanish and 2,238 in English) were extracted. The semantic annotation was used to identify word forms primarily related to the LSP domain of commerce and economics:

From the 21 major fields used by the USAS, three of them, ‘government and public’ (G), ‘money commerce and industry’ (I), and ‘language and communication’ (Q) offered many of the word forms related to commerce and economics. (Rojas Díaz, forthcoming-a, p. 24)

However, to classify these expressions according to phraseological categories, the type of idiomaticity (see section 3.1. and Table 1) needed to be identified. In Article 1, semantic idiomaticity is identified as one of the characteristics that PUs could present. This type of idiomaticity was recognized through the occurrence of metaphors and metonymies in the analyzed PUs (Rojas Díaz, 2020, p. 313). In Article 2, metaphors (e.g., *los cinco dragones* ‘the five dragons’ | *sweat in hold*), metonymies (e.g., *de ida y vuelta* ‘of going and returning’ | *quay to quay*), or a combination of both (e.g., *fuga de cerebros* ‘leak of brains’ | *at arm’s length*) are identified in some DCI SPU entries: 2,118 entries (19% of the sample) presented semantic idiomaticity of which 1,285 were metaphors (667 in Spanish and 618 in English), 702 were metonymies (406 in Spanish and 296 in English), and 132 a combination of both (92 in Spanish and 40 in English).

The semantic annotation along with the idiomaticity identification allowed for the classification of the entries in the different SPU sub-categories (see Figure 8) of which 10,345 are SpCs (5,943 in

Spanish and 4,402 in English), 561 are SpPs (196 in Spanish and 365 in English), 109 are SpIs (69 in Spanish and 40 in English), and 71 Hybrid SpCs (22 in Spanish and 49 in English).

Finally, Article 3 presents a detailed description of the behavior of the 109 SpIs identified in Article 2, including the frequency of occurrence in the EUR-Lex corpus of the SpI entries and equivalents extracted from the DCI. Moreover, this description provides information about the (i) the type of idiomaticity and (ii) the monolexical/pluriverbal aspect of the SpI equivalents extracted from the DCI.

### **5.3. Applied and methodological contributions**

The three articles present a methodology that could be used to derive important linguistic information for the analysis of PUs, SPUs, and their word forms from lexicographic entries through corpus-based methods as presented in section 4.1. Moreover, this PhD dissertation shows how the semantic information of word forms (semantic annotation by semantic/discourse fields [see Figure 7]) could be employed in the identification of different types of idiomaticity.

However, the hybrid indexation method is the most important applied contribution of this dissertation. This method offers an alternative to index SPUs according to their subcategory using monolexical terminological entries (headwords) as semantic anchors for their indexation (Rojas Díaz, forthcoming-a, p. 31). The indexation method that was first proposed in Article 1, explored in detail in the work by Rojas Díaz and Pérez Sánchez (2020), and adapted it to LSP and terminography in Article 2.

### **5.4. Limitations**

This dissertation has shown the important role that semantic annotation has in identifying SPUs. However, automatic (or semiautomatic) semantic annotation tools are scarce and not available in some languages (e.g., Spanish). The unavailability of such tools makes this a challenging and time-consuming task within the time limits of a PhD project. Furthermore, although 31,131 expressions were analyzed, they only reflect the behavior of PUs and SPUs between three- and five-word forms from three dictionaries (the DFDEA, the CCDOI, and the DCI). Consequently, it is difficult to generalize from the

findings of this PhD dissertation. The lack of semantically annotated corpora makes it difficult to test whether pattern nesting (Rojas Díaz, 2020, pp. 315-316; forthcoming-a, pp. 24-25) enhances the extraction of PUs and SPUs. Such annotated corpora can also be used to explore further lexical and semantic changes by contrasting the semantic annotation of PU and SPU entries or candidates and their equivalents, using the methods applied in Article 3.

To classify the SPUs into subcategories, it was necessary to (i) identify metaphors and metonymies, (ii) differentiate words from terms, and (iii) disambiguate the word classes and semantic information of the individual word forms. More accurate results could have been achieved if inter-annotator agreement measurements were performed. However, carrying out such a task was beyond the scope and the time limitations of the current PhD dissertation and it was deliberately omitted in this study.

## **5.5. Future research**

The different approaches and the interdisciplinarity that comprises this PhD dissertation opens a series of possibilities for future research.

First, the identification criteria of SPUs, that is considered to be subjective, should be verified by experts and evaluated by means of inter-annotator agreement<sup>19</sup>. This would allow this data to be used as a Gold Standard for enhancing Natural Language Processing methods for the identification and extraction of SPUs.

Second, the morphosyntactic and semantic patterns extracted from PUs and SPUs could be contrasted to analyze their structural divergences. Although this information has already been extracted in Articles 2 and 3, a more in-depth analysis of SPUs could be done in which morphosyntactic and semantic patterns of metaphors and metonymies could be compared to look for coincidences and differences among them. Moreover, other LSP domains and languages should be added to the analyses made in this PhD dissertation to increase the repertoire of linguistic features of word forms; and morphosyntactic and

---

<sup>19</sup> E.g., Cohen's kappa, Fleiss's kappa, Krippendorff's alpha.

semantic patterns of PUs and SPUs. Furthermore, the identification criteria and the definitions provided in this PhD dissertation could be tested in a custom-made LSP corpus that includes both POS-tagging and semantic annotation.

Third, a diachronic study including LGP and LSP dictionaries and corpora from different time periods would allow one to determine whether they underwent migration from LGP into LSP at some point and to assess the way they acquired their specialized meaning. Moreover, a more in-depth analysis of the occurrence of metaphors and metonymies in SPUs should be done.

Fourth, it seems relevant to explore the usefulness of the hybrid indexation method in an LSP phraseological database and study the reception that it could have among linguistic mediators (e.g., translators, interpreters, proofreaders, etc.). Regarding linguistic mediators, further cross-linguistic studies about how translators and interpreters deal with PUs and SPUs should be carried out to improve our understanding of LGP and LSP phraseology.

By doing an extensive study of the lexical, morphosyntactic, and semantic characteristics of PUs and SPUs, this PhD dissertation provides a new definition of specialized phraseological unit. Indeed, this dissertation and the methods here developed could be used as a starting point for the further study of both LGP and LSP phraseology.

## 6. References

- Ademollo, F. (Ed.) (2011). *The Cratylus of Plato: a Commentary*. New York: Cambridge University Press.
- Aguado de Cea, G. (2007). A Multiperspective Approach to Specialized Phraseology: Internet as a Reference Corpus for Phraseology. In *The Texture of Internet: Netlinguistics in Progress* (pp. 182-207): Ontology Engineering Group – OEG.
- Alcaraz, V. E., & Castro Calvín, J. (2007). *Diccionario de Comercio Internacional – Importación y Exportación, Inglés-Español/Spanish-English*. Barcelona: Ariel.
- Alonso Ramos, M. (Ed.) (2006). *Diccionarios y fraseología*.
- Amosova, N. N. (1963). *Osnovy Anglijskoj Frazelogii*. Leningrad: University Press.
- Archer, D., Wilson, A., & Rayson, P. (2002). *Introduction to the USAS Category System*. Retrieved from
- Atkins, B. T. S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- Baisa, V., Michelfeit, J., Medved', M., & Jakubíček, M. (2016). European Union Language Resources in Sketch Engine. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *The Proceedings of tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 2799-2803). Portorož: European Language Resources Association (ELRA).
- Baker, M. (1992). *In Other Words: a Coursebook on Translation*. New York: Routledge.
- Baldwin, T., & Kim, S. N. (2010). Multiword Expressions. In N. Indurkha & F. J. Damerau (Eds.), *Handbook of Natural Language Processing*, (Second Edition ed., pp. 267-292). Boca Raton: CRC Press.
- Bally, C. (1909). *Traité de Stylistique Française* (Second ed. Vol. 1). Paris: Librairie C. Klincksieck
- Bevilacqua, C. (2004). *Unidades fraseológicas especializadas eventiva: descripción y reglas de formación en el ámbito de la energía solar*. (PhD Thesis). Universitat Pompeu Fabra, Barcelona.
- Blais, E. (1993). Le phraséologisme. Une hypothèse de travail. *Terminologies nouvelles. Terminologie et diversité culturelle*, 10, 50-56.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston.
- Bréal, M. (1897). *Essai de sémantique*. Paris: Hachette.
- Buendía Castro, M., & Faber, P. (2015). Phraseological Units in English-Spanish Legal Dictionaries: a Comparative Study. *Fachsprache*(3-4), 161-175.
- Cabré, M. T. (2000). Terminologie et linguistique: la théorie des portes. *Terminologies nouvelles. Terminologie et diversité culturelle*, 21, 10-15.
- Cabré, M. T., Lorente, M., & Estopà, R. (1996). Terminología y fraseología. In *Actas del V Simposio de Terminología Iberoamericana* (pp. 67-81). Mexico City: Colegio de México.
- Carneado Moré, Z., & Tristá, A. M. (1985). *Estudios de fraseología*: Academia de Ciencias de Cuba - Instituto de Literatura y Lingüística.
- Casares, J. (1950). *Introducción a la lexicografía moderna*. Madrid: CSIC.
- Catford, J. C. (1965). *A Linguistic Theory of Translation*. Oxford: Oxford University Press.
- Chafe, W. (1968). Idiomaticity as an Anomaly in the Chomskyan Paradigm. *Foundations of Language*, 4, 109-127.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

- Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid: Gredos.
- Corpas Pastor, G. (2003). *Diez años de investigaciones en fraseología: análisis sintáctico-semanticos, contrastivos y traductológicos*. Madrid: Iberoamericana.
- Cowie, A. P. (1981). The Treatment of Collocations and Idioms in Learners' Dictionaries. *Applied Linguistics*, 2(3), 223-235.
- Cowie, A. P. (1998). Introduction. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 1-20). New York: Oxford University Press.
- Dobrovolskij, D., & Piirainen, E. (2005). *Figurative Language: Cross-Cultural and Cross-Linguistic Perspectives*. Amsterdam: Elsevier.
- Faber Benítez, P., & López Rodríguez, C. I. (2012). Terminology and specialized language. In P. Faber (Ed.), *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin: De Gruyter Mouton.
- Firth, J. R. (1957). *Papers in linguistics, 1934-1951*. New York: Oxford University Press.
- Fraile Vicente, E. (2007). The Relationship between the Typical Errors in the Translation of Business Idioms and their Lexicographical Treatment. *Revista Alicantina de Estudios Ingleses*(20), 65-93.
- Gabelentz vor der, G. (1901). *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Leipzig: C. H. Tauchnitz.
- Gallois, A. (2016, Winter 2016 Edition). Identity over Time. Retrieved from <https://plato.stanford.edu/archives/win2016/entries/identity-time/>
- García-Page Sánchez, M. (2008). *Introducción a la fraseología española: estudio de las locuciones* (Vol. 6). Barcelona: Anthropos.
- Gil-Bardaji, A. (2011). Procedures, techniques, strategies: translation process operators. *Perspectives: Studies in Translatology*, 17(3), 161-173. doi:10.1080/09076760903249372
- Gläser, R. (1986). The Grading of Idiomaticity as a Presupposition for a Taxonomy of Idioms. In R. Schulze & W. Hüllen Llen (Eds.), *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics* (pp. 264-279). Tübingen: Max Niemeyer.
- Gouadec, D. (1993). Terminologie & terminotique: outils, modèles & méthodes. In D. Gouadec (Ed.), *Actes de la première Université d'automne en Terminologie Rennes 2 - 21 au 26 septembre 1992*. Paris: La Maison du Dictionnaire.
- Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J.R. Firth* (pp. 148– 162). London: Longmans.
- Halliday, M. A. K. (1992). Language as system and language as instance: The corpus as a theoretical construct. In J. Svartvik (Ed.), *Directions in Corpus Linguistics* (pp. 61-77). Berlin: De Gruyter Mouton.
- Halverson, S. (1997). The concept of equivalence in translation Studies: Much Ado About Something. *Target*, 9(2), 207-233. doi:10.1075/target.9.2.02hal
- Hartmann, R. R. K., & James, G. (1998). *Dictionary of Lexicography*. New York: Routledge.
- Heid, U. (2008). Computational Phraseology: an Overview. In S. Granger & F. Meunier (Eds.), *Phraseology: an interdisciplinary perspective* (pp. 337–360). Amsterdam: John Benjamins Publishing Company.
- Heid, U., & Weller, M. (2008). Tools for Collocation Extraction: Preferences for Active vs. Passive. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, & D. Tapias (Eds.), *Proceedings of the Sixth International Conference on Language Resources*

- and Evaluation (LREC'08)* (pp. 1266-1272). Marrakech: European Language Resources Association (ELRA).
- Hockett, C. (1958). *A Course in Modern Linguistics*. New York: Macmillan.
- Hoffmann, L. (1998). *Llenguatges d'especialitat: selecció de textos* (J. Brumme Ed.). Barcelona: Institut Universitari de Lingüística Aplicada.
- Hourani-Martín, D., & Tabares-Plasencia, E. (2020). Morphosyntactic and Semantic Behavior of Legal Phraseological Units: a Case Study in Spanish Verb-Noun Constructions about Money Laundering. *Terminology*, 26(1), 108-131.
- House, J. (1977). A model for Assessing Translation Quality. *Meta*, 22(2), 103–109.  
doi:<https://doi.org/10.7202/003140ar>
- Householder, F. W. (1959). On Linguistic Primes. *Word*, 15(2), 231-239.  
doi:10.1080/00437956.1959.11659696
- Howarth, P. (1996). *Phraseology in English Academic Writing: some Implications for Language Learning and Dictionary Making*. Tübingen: Max Niemeyer.
- ISO. (2019). Terminology work and terminology science — Vocabulary. In *Terms and definitions*: ISO.
- Jakobson, R. (1959). On Linguistic Aspects of Translation. In R. A. Brower (Ed.), *On Translation* (pp. 232-239). Cambridge: Harvard University Press.
- Jørgensen, S. L. (2019). The notion of a correctness of names in Plato's Cratylus. Arguments for a basic distinction. *Methodos: savoirs et textes*, 19.  
doi:<https://doi.org/10.4000/methodos.6133>
- Kade, O. (1968). *Zufall und Gesetzmässigkeit in der Übersetzung* (Vol. I). Leipzig: Verlag Enzyklopädie.
- Karpinska, L. (2019). Specialized bilingual dictionaries as a translation tool: a case study of Latvian-English-Latvian dictionaries of legal terms. In D. Osuchowska & L. Harmon (Eds.), *Contents, Use, Usability: Dictionaries from the Perspective of a Translator and a Language Teacher* (pp. 35-48). Berlin: Peter Lang.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., . . . Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1, 7-36.  
doi:<https://doi.org/10.1007/s40607-014-0009-9>
- Kjær, A.-L. (2007). Phrasemes in legal texts. In H. Burger, D. Dobrovolskij, P. Kühn, & N. Norrick (Eds.), *Phraseology* (Vol. I, pp. 506-516). New York: Walter de Gruyter.
- Koller, W. (1989). Equivalence in translation theory. In A. Chesterman (Ed.), *Reading in Translation Theory* (pp. 99-104). Helsinki: Oy Finn Lectura Ab.
- Kromann, H.-P. (1991). Principles of Bilingual Lexicography. In F. J. Hausmann, O. Reichmann, H. E. Wiegand, & L. Zgusta (Eds.), *Dictionaries* (Vol. III, pp. 2711-2728). Berlin: Walter de Gruyter.
- Kübler, N., & Pecman, M. (2012). The ARTES Bilingual LSP Dictionary: from Collocation to Higher Order Phraseology. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography* (pp. 187-209). Oxford: Oxford University Press.
- L'Homme, M.-C. (1998). Caractérisation des combinaisons lexicales spécialisées par rapport aux collocations de langue générale. In T. Fontenelle, P. Hilgsmann, & A. Michiels (Eds.), *EURALEX 98 Proceedings, VIII International Congress of the European Association for Lexicography* (Vol. II, pp. 513-522). Liège: Éd. Université de Liège.



- L'Homme, M.-C. (2000). Understanding specialized lexical combinations. *Terminology: international journal of theoretical and applied issues in specialized communication*, 6(1), 89-109. doi:<https://doi.org/10.1075/term.6.1.06hom>
- L'Homme, M.-C. (2020). *Lexical Semantics for Terminology: an Introduction*. Amsterdam: John Benjamins Publishing Company.
- L'Homme, M.-C., & Bertrand, C. (2000). Specialized Lexical Combinations: Should they be Described as Collocations or in Terms of Selectional Restrictions? In U. Heid, S. Evert, E. Lehmann, & C. Rohrer (Eds.), *Proceedings of the IX EURALEX International Congress* (pp. 497-506). Stuttgart: Institut für Maschinelle Sprachverarbeitung.
- L'Homme, M.-C., & Bertrand, C. (2009). Specialized Lexical Combinations: Should they be Described as Collocations or in Terms of Selectional Restrictions? In U. Heid, S. Evert, E. Lehmann, & C. Rohrer (Eds.), *Proceedings of the IX EURALEX International Congress* (pp. 497-506). Stuttgart: Institut für Maschinelle Sprachverarbeitung.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: Chicago University Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in Corpus Linguistics* (pp. 105-122). Berlin: De Gruyter Mouton.
- Leonardi, V. (2000). Equivalence in translation: Between myth and reality. *Translation Journal*, 4(4). Retrieved from <https://translationjournal.net/journal/14equiv.htm>
- Leroyer, P. (2006). Dealing with Phraseology in Business Dictionaries: Focus on Dictionary Functions - not Phrases. *Linguistik online*, 27(2), 183-194.
- López, X. P. (2012). *Fraseología española de origen latino y motivo grecorromano*. (PhD Thesis). Universitat de Lleida, Lleida.
- Lorente Casafont, M. (2002). Terminología y fraseología especializada: del léxico a la sintaxis. In M. F. Pérez Lagos & G. Guerrero Ramos (Eds.), *Panorama actual de la terminología* (pp. 159-180). Granada: Comares.
- Makkai, A. (1972). *Idiom Structure in English*. The Hague: De Gruyter Mouton.
- Martínez de Sousa, J. (1995). *Diccionario de lexicografía práctica*. Barcelona: Bibliograf.
- McArthur, T. (1981). *Longman Lexicon of Contemporary English*. Harlow: Longman.
- Mel'čuk, I. (1988). Semantic Description of Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria. *International Journal of Lexicography*, 1(3), 165-188.
- Mel'čuk, I. (1998). Collocations and Lexical Functions. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp. 23-53). Oxford: Oxford University Press.
- Mel'čuk, I. (2012). Phraseology in the Language, in the Dictionary, and in the Computer. In K. Kuiper (Ed.), *Yearbook of Phraseology* (Vol. 3, pp. 31-56). New York: De Gruyter Mouton.
- Mel'čuk, I. (2013). Tout ce que nous voulions savoir sur les phrasèmes, mais.... *Cahiers de lexicologie*, 102, 129-149. doi:10.15122/isbn.978-2-8124-1259-2.p.0129
- Mellado Blanco, C. (2015). Parámetros específicos de equivalencia en las unidades fraseológicas (con ejemplos del español y el alemán). *Revista de Filología*, 33, 153-174.
- Mellado Blanco, C. (Ed.) (2008). *Colocaciones y fraseología en los diccionarios*. Frankfurt am Main: Peter Lang.

- Molina, L., & Hurtado Albir, A. (2002). Translation Techniques Revisited: A Dynamic and Functionalist Approach. *Meta : journal des traducteurs*, 47(4), 498-512.
- Montero Martínez, S. (2008). A Constructional Approach to Terminological Phrasemes. In E. Bernal & J. DeCesaris (Eds.), *Proceedings of the XIII EURALEX International Congress* (pp. 1015-1022). Barcelona: Institut Universitari de Lingüística Aplicada.
- Moon, R. (2008). Dictionaries and Collocation. In S. Granger & F. Meunier (Eds.), *Phraseology: an interdisciplinary perspective* (pp. 313-336). Amsterdam: John Benjamins.
- Morgan, J. L. (1978). Two Types of Convention in Indirect Speech Acts. In P. Cole (Ed.), *Pragmatics* (pp. 261-280): Brill.
- Muñoz Martín, R. (2000). Translation strategies: Somewhere over the rainbow. . In A. Beeby, D. Ensinger, & M. Presas (Eds.), *Investigating translation* (pp. 129-138). Amsterdam: John Benjamins.
- Newmark, P. (1981). *Approaches to translation*. New York: Pergamon Press.
- Nida, E. A., & Taber, C., Russell. (1982). *The Theory and Practice of Translation* (Vol. VIII). Leiden: E. J. Brill.
- Norrick, N. (2007). English Phraseology. In H. Burger, D. Dobrovolskij, P. Kühn, & N. Norrick (Eds.), *Phraseology* (Vol. II, pp. 615-619). New York: Walter de Gruyter.
- Nuccorini, S. (2020). Digging into a Thesaurus for Treasures: Phraseological Conceptual Maps in the Longman Language Activator. In P. Cotta Ramusino & F. Mollica (Eds.), *Contrastive Phraseology: Languages and Cultures in Comparison* (pp. 349-375). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Panou, D. (2013). Equivalence in Translation Theories: A Critical Evaluation. *Theory and Practice in Language Studies*, 3(1), 1-6. doi:<http://10.4304/tpls.3.1.1-6>
- Paquot, M. (2015). Lexicography and Phraseology. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics* (pp. 460-477). Cambridge: Cambridge University press.
- Patiño García, P. (2011). A specialized parallel corpus of English and Spanish Free Trade Agreements for the study of specialized collocations. *SYNAPS - A Journal of Professional Communication*, 26, 85-89. Retrieved from <http://hdl.handle.net/11250/2393998>
- Patiño García, P. (2013). FTA Corpus: a parallel corpus of English and Spanish Free Trade Agreements for the study of specialized collocations. *Bergen Language and Linguistics Studies*, 3(1), 81-91. doi:<https://doi.org/10.15845/bells.v3i1.363>
- Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Halle: Max Niemeyer.
- Pavel, S. (1993). Neology and Phraseology as Terminology-in-the-Making. In H. Sonneveld & K. Loening (Eds.), *Terminology : applications in interdisciplinary communication* (pp. 21-34). Amsterdam: John Benjamins.
- Picht, H. (1987). Terms and their LSP Environment - LSP Phraseology. *Meta: Journal des traducteurs*, 32(2), 149-155. doi:<https://doi.org/10.7202/003836ar>
- Picht, H., & Draskau, J. (1985). *Terminology: an introduction*. Guildford: University of Surrey.
- Plato. (Cambridge: Harvard University Press, 1921). *Plato in Twelve Volumes* (H. N. Fowler, Trans. Vol. 12): Perseus Digital Library.
- Plutarch. (Cambridge: Harvard University Press, 1914). *Theseus, Plutarch's Lives, Volume I* (B. Perrin, Trans.): Perseus Digital Library.
- Pym, A. (2014). *Exploring translation theories* (Second Edition ed.). New York: Routledge.

- Rojas Díaz, J. L. (2020). “From Head to Toe”: a Lexical, Semantic, and Morphosyntactic Study of Idioms in Phraseological Dictionaries in English and Spanish. *MonTI* (Special issue 6), 287-326. doi:<http://dx.doi.org/10.6035/MonTI.2020.ne6.9>
- Rojas Díaz, J. L. (forthcoming-a). ‘Arm’s length’ phraseology? *Building bridges from general language to specialized language phraseology – a study based on a specialized dictionary of International Commerce and Economics in Spanish and English*. Manuscript submitted for publication.
- Rojas Díaz, J. L. (forthcoming-b). ‘Worlds apart?’: *phraseological equivalence through the lenses of translation, terminology, and lexicography*. Manuscript submitted for publication.
- Rojas Díaz, J. L., & Pérez Sánchez, J. M. (2019). “You Took the Word Out of My Mouth”: a Morphosyntactic and Semantic Analysis of a Phraseological Lexicon of Colombian Spanish. In G. M. Corpas Pastor, Ruslan (Ed.), *Computational and Corpus-Based Phraseology - Third International Conference, Europhras 2019 Malaga, Spain, September 25–27, 2019 Proceedings* (pp. 375-390): Springer.
- Rojas Díaz, J. L., & Pérez Sánchez, J. M. (2020). “A Baptism of Fire”: towards a Practical Hybrid Approach for the Lexicographic Indexation of Phraseological Units with Religious Lexical Components in English and Spanish. *CLINA: An Interdisciplinary Journal of Translation, Interpreting and Intercultural Communication*, 6(2), 95-117. doi:<https://doi.org/10.14201/clina20206295117>
- Ruiz Gurillo, L. (1997). *Aspectos de fraseología teórica española*. Valencia: Servicio de Publicaciones de la Universidad de Valencia.
- Ruiz Gurillo, L. (2001). *Las locuciones en el español actual*. Madrid: Arco Libros.
- Sadock, J. M. (1972). Speech act idioms. In P. M. Peranteau, J. N. Levi, & G. C. Phares (Eds.), *Papers from the eighth regional meeting of the Chicago Linguistic Society* (pp. 329 – 339). Chicago: Chicago Linguistic Society.
- Schmid, H. (1994). *Probabilistic Part-of-Speech Tagging Using Decision Trees*. Paper presented at the Proceedings of the International Conference on New Methods in Language Processing, Manchester.
- Schmid, H. (2019). Deep Learning-Based Morphological Taggers and Lemmatizers for Annotating Historical Texts. In *DATeCH2019: Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage* (pp. 133–137). Brussels: ACM.
- Seco, M., Andrés, O., & Ramos, G. (Eds.). (2004). *Diccionario fraseológico documentado del español actual*. Madrid: Aguilar.
- Sedley, D. (2020, Winter 2020 Edition). Plato’s Cratylus. Retrieved from <https://plato.stanford.edu/archives/win2020/entries/plato-cratylus/>
- Seretan, V. (2011). *Syntax-Based Collocation Extraction* (Vol. 44). New York: Springer.
- Siepmann, D. (2008). Phraseology in Learners’ Dictionaries: What, Where and How? In F. G. Meunier, Sylviane (Ed.), *Phraseology in Foreign Language Learning and Teaching* (pp. 185-202). Amsterdam: John Benjamins Publishing Company.
- Simonnæs, I., & Kristiansen, M. (2018). Economic Language. In J. Humbley, G. Budin, & C. Laurén (Eds.), *Languages for special puposes: an international handbook* (pp. 151-185). Berlin: De Gruyter Mouton.
- Sinclair, J. (1992). The automatic analysis of corpora. In J. Svartvik (Ed.), *Directions in Corpus Linguistics* (pp. 379-397). Berlin: De Gruyter Mouton.

- Sinclair, J. (Ed.) (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J., & Moon, R. (Eds.). (1997). *Collins COBUILD Dictionary of Idioms*. Glasgow: HarperCollins Publishers.
- Sosiński, M. (2006). *Frasesología comparada del polaco y del español: su tratamiento en los diccionarios bilingües*. Retrieved from Granada:
- Svartvik, J., & Quirk, R. (Eds.). (1980). *A corpus of English conversation* (Vol. 82). Lund: Lund University Press.
- Szabó, Z. G. (2020, Fall 2020 Edition). Compositionality. Retrieved from <https://plato.stanford.edu/archives/fall2020/entries/compositionality/>
- Toury, G. (1981). Translated Literature: System, Norm, Performance: Toward a TT-Oriented Approach to Literary Translation. *Poetics Today*, 2(4), 9-27.  
doi:<https://doi.org/10.2307/1772482>
- Tschichold, C. (2008). A Computational Lexicography Approach to Phraseologisms. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 361–376). Amsterdam: John Benjamins Publishing Company.
- Veisbergs, A. (2020). Phraseology in General Bilingual Dictionaries: Idioms as Equivalents of Single Words. In P. Cotta Ramusino & F. Mollica (Eds.), *Contrastive Phraseology: Languages and Culture in Comparison* (pp. 331-343). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Vinay, J.-P., & Darbelnet, J. (1995). *Comparative Stylistics of French and English: A methodology for translation* (J. C. Sager & M.-J. Hamel, Trans.). Amsterdam: John Benjamins Publishing Company.
- Vinogradov, V. V. (1947). Ob osnovnuikh tipakh frazeologicheskikh edinit v russkom yazyu. In A. A. Shakhmatov (Ed.), *1864-1920. Sbornik statey i materialov* (pp. 339-364). Moscow: Nauka.
- Wilss, W. (1982). *The science of translation: problems and methods* (Vol. 180). Tübingen: G. Narr.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Zgusta, L. (1971). *Manual of lexicography*. Prague: Academia.
- Zuluaga, A. (1980). *Introducción al estudio de las expresiones fijas*. Frankfurt: Peter Lang.

## 7. Articles

Rojas Díaz, J. L. (2020). “From Head to Toe”: a Lexical, Semantic, and Morphosyntactic Study of Idioms in Phraseological Dictionaries in English and Spanish. *MonTI* (Special Issue 6), 287-326. <https://doi.org/http://dx.doi.org/10.6035/MonTI.2020.ne6.9>.

Recibido / Received: 10/11/2019  
Aceptado / Accepted: 28/01/2020

Para enlazar con este artículo / To link to this article:  
<http://dx.doi.org/10.6035/MonTI.2020.ne6.9>

Para citar este artículo / To cite this article:

Rojas Díaz, José Luis. (2020) "From head to toe: A lexical, semantic, and morphosyntactic study of idioms in phraseological dictionaries in English and Spanish." In: Mogorrón Huerta, Pedro (ed.) 2020. *Análisis multidisciplinar del fenómeno de la variación fraseológica en traducción e interpretación / Multidisciplinary Analysis of the Phenomenon of Phraseological Variation in Translation and Interpreting*. MonTI Special Issue 6, pp. 287-326.

## “FROM HEAD TO TOE”: A LEXICAL, SEMANTIC, AND MORPHOSYNTACTIC STUDY OF IDIOMS IN PHRASEOLOGICAL DICTIONARIES IN ENGLISH AND SPANISH

JOSÉ LUIS ROJAS DÍAZ

jose.diaz@nhh.no

NHH Norwegian School of Economics

### Abstract

In recent times, interest in the study of phraseology in general and specialized lexicographic resources has increased (Castillo Carballo 2006, Aguado de Cea 2007, Mellado 2008, Buendía Castro & Faber 2015). However, to date, a lack of knowledge related to the characterization and indexation of phraseological units (PUs) in lexicographic resources remains. That issue is addressed here through an analysis of phraseological units in the entries of two phraseological dictionaries, one in Spanish, and one in English: the *Diccionario fraseológico documentado del español actual* (Seco, Andrés & Ramos 2004) and the *Collins COBUILD Dictionary of Idioms* (Sinclair & Moon 1997). To perform this analysis, two databases containing 21,045 entries extracted from the two dictionaries mentioned above were compiled. The databases were tagged syntactically and semantically in order to extract 816 morphosyntactic patterns, 2,655 combinations of semantic categories (Semantic patterns) and a series of lexical and lexicographic information about indexation of PUs in dictionaries.

**Keywords:** Phraseology; Dictionaries; Lexicography; Semantic patterns; Morphosyntactic patterns.



Esta obra está bajo una licencia de Creative Commons Reconocimiento 4.0 Internacional.

## Resumen

En los últimos años ha aumentado el interés en el estudio de la fraseología en lengua general y de especialidad en recursos lexicográficos (Castillo Carballo 2006, Aguado de Cea 2007, Mellado 2008, Buendía Castro & Faber 2015). Sin embargo y a la fecha, aún existen algunos vacíos en cuanto a la caracterización e indización de unidades fraseológicas (UFs) en recursos lexicográficos. Esta problemática se aborda en el presente artículo por medio del análisis de unidades fraseológicas en las entradas de dos diccionarios (uno en español y otro en inglés): el *Diccionario fraseológico documentado del español actual* (Seco, Andrés & Ramos 2004) y el *Collins COBUILD Dictionary of Idioms* (Sinclair & Moon 1997). Para llevar a cabo este análisis, se compilieron dos bases de datos con 21 045 entradas extraídas de los diccionarios antes mencionados. Las bases de datos fueron etiquetadas sintáctica y semánticamente para extraer 816 patrones morfosintácticos, 2 655 combinaciones semánticas (patrones semánticos) y una serie de datos léxicos y lexicográficos sobre la indización de UFs en diccionarios.

**Palabras clave:** Fraseología; Diccionarios; Lexicografía; Patrones semánticos; Patrones morfosintácticos.

## 1. Introduction

In recent times, the study of phraseology in general language and specialized language lexicographic resources has gained particular interest (e.g. dictionaries and databases) (Castillo Carballo 2006: 8, Aguado de Cea 2007: 184-185, Mellado 2008, Buendía Castro & Faber 2015: 161). However, more in-depth knowledge is needed about the characterization and indexation of phraseological units.

This article will shed light on how PUs are indexed in dictionaries as well as the lexicographic, lexical, semantic, and morphosyntactic characteristics of the selected PUs. The analyses and the article are structured as follows: firstly, a summary of the different concepts regarding phraseology in the Spanish and English traditions is presented. Secondly, the lexicographic description of the dictionaries used for the compilation of the database is introduced. Thirdly, the results of the lexical, semantic and morphosyntactic analyses are presented. Lastly, the final section is devoted to the most salient conclusions reached, and to a practical lexicographic proposal for the indexation of PUs in lexicographic resources.



### 1.1. *Phraseology: denomination and definition*

According to García-Page (2008), phraseology should be defined in terms of its object of study. Thus, the question to ask would be “what is the object of study of phraseology?” (2008: 7). However, answering this question entails a problem, namely: the extensive number of denominations and definitions used to determine the object of study of phraseology. Bushnaq (2015: 173) states that the terms ‘phraseologism’, ‘phraseme’, ‘phraseological expression’, ‘phraseological unit’, ‘idiomatic expression’, and ‘idiom’ are used in English to describe an expression the meaning of which cannot be deduced from the individual meaning of its constituent words. Although the definition given by Bushnaq is correct, it is still vague and corresponds to the classical definition of phraseology. Among theoreticians, it is possible to find the most diverse taxonomies to categorize expressions according to their compositionality/idiomaticity, functional categories, and fixation in language, among other features. Ruiz Gurrillo (2001: 44) and Cowie (2001: 7) present a summary of those categories for some phraseology research in both Spanish and English. However, three categories are common to almost all taxonomies. Those categories include: (i) expressions that behave as sentences (proverbs/sayings), (ii) other expressions in which one of the constituents is not idiomatic (collocations), and (iii) other that are fully idiomatized (idioms).

The approach to the study of phraseological units in Spanish and English is considered to have major differences. On the one hand, the Spanish tradition tends to be taxonomic in nature, having two fundamental notions on the study of phraseology: a narrow one—in which only idioms are considered to be PUs—, and a wide one—where not only idioms, but also sayings, proverbs, collocations, among others are considered to be PUs—. On the other hand, the English tradition is more flexible—similar to the wide notion of Spanish phraseology—, and it includes many subsets of phrases that would not be considered as *phraseological units* by some theoreticians in Spanish. Particularly, the most restricted subset of units in both languages will be used in this paper, i.e. *locuciones* in Spanish and *idioms* in English, and they will be referred to as *phraseological units* in an attempt to use a denomination that encompasses the characteristics of both subsets in both languages. When looking up the definitions of *locución* and *idiom* in general dictionaries in

Spanish and in English, it becomes evident that those definitions evidently differ. Thus, on the one hand, the *Diccionario de la lengua española* (DLE [online]) defines *locución* in its third sense as a “group of words that function as a single lexical unit with a unitary meaning and certain degree of formal fixation” (*locución*, n.d., author’s translation). On the other hand, the *Cambridge Dictionary* (online) defines *idiom* as “a group of words in a fixed order that have a particular meaning that is different from the meanings of each word on its own” (*idiom*, n.d.).

In Spanish, when resorting to the literature on phraseology studies, one of the most accepted definitions of *locución* is given by Casares (1950), who states that a *locución* is: “a stable combination of two or more terms that function as an element in a sentence and whose unitary meaning cannot be simply justified as the sum of the usual meanings of its components” (1950: 170, author’s translation). As a further elaboration to the conception of Casares, Ruiz Gurillo (1997) says that a ‘phraseological unit’ is “a fixed combination of words that presents a certain level of fixation, and eventually, idiomaticity” (1997: 14). Likewise, in English, Moon (1998) states that the definition of *idiom* is ambiguous due to its different uses. Nonetheless, the author also asserts that the most restrictive definition of *idioms* could be “a particular kind of unit: one that is fixed and semantically opaque or metaphorical, or, traditionally, not the sum of its parts” (1998: 4). Similarly, Mel’čuk (2012) proposes a definition of ‘pure idiom’ in the following terms: “an idiom AB is a full idiom if its meaning does not include the meaning of any of its lexical components: ‘AB’  $\not\supset$  ‘A’ and ‘AB’  $\not\supset$  ‘B’” (2012: 37). This last definition put forward by Mel’čuk, will be the one applied to ‘phraseological units’ in this paper. In the next section, the Spanish and English phraseology traditions will be discussed in more detail.

### 1.2. Spanish and English theoretical traditions on phraseology

Norrick (2007) states that there are two different traditions related to the study of phraseology in English: the British tradition, and the American one. He also suggests that both traditions were originally driven by either anthropological or literary approaches (2007: 615). For the British tradition, Norrick proposes three stages in the study of phraseology. The first one is

based on the “list of the irregularities in a language” written by Bloomfield (1933). The second one is the study conducted by Hockett (1958), where he grouped phraseological units in a category called idioms. The third one was the grammar written for those units by Householder (1959). The distinction between idioms and collocations was made, among others, by Firth (1957), and later by Sinclair & Moon (1989, 1997).

According to Norrick (2007: 616), the American tradition started with the criticism Wallace Chafe made of Noam Chomsky’s compositionality concept. Chomsky argued that the lexicon is a “simple and unordered list of all lexical formatives” (Chomsky 1965: 84) which should include the idioms. Three years later, Chafe (1968) showed that the concept of idiomaticity, one of the characteristics of phraseological units, is totally opposed to the compositionality criterion of Chomskyan linguistic theory. Table 1 presents a synthesis of phraseological denominations in the English language according to Cowie (2001).

**Table 1.** Denominations of phraseology used in English by different authors according to Cowie (2001: 7)

Author	General category	Opaque, invariable unit	Partially motivated unit	Phraseological bound unit
Vinogradov	Phraseological unit	Phraseological fusion	Phraseological unity	Phraseological combination
Amosova	Phraseological unit	Idiom	Idiom (not differentiated)	Phraseme / Phraseoloid
Cowie	Composite	Pure idiom	Figurative Idiom	Restricted collocation
Mel’cuk	Semantic Phraseme	Idiom	Idiom (not differentiated)	Collocation
Gläser	Nomination	Idiom	Idiom (not differentiated)	Restricted collocation
Howarth	Composite unit	Pure Idiom	Figurative Idiom	Restricted collocation

In Spanish, authors such as Casares (1950), Zuluaga (1980), Carneado & Trista (1985), Corpas Pastor (1996), Ruiz Gurillo (2001), and García-Page

(2008) are among the most quoted ones in phraseology studies. Nevertheless, the denominations of idiomatized units proposed by those authors differ greatly, as shown in Table 2.

**Table 2.** Denominations of PUs proposed by the most representative authors related to general phraseology in Spanish

Author	Denomination	Definition	
Julio Casares (1950)	<i>Locuciones</i>	Wide	
	<i>Frases hechas</i>		
	<i>Refranes</i>		
	<i>Modismos</i>		
Alberto Zuluaga (1980)	<i>Locuciones</i>		
	<i>Enunciados</i>		
Zoila Carneado & Antonia Tristá (1985)	<i>Unidad fraseológica (fraseologismo)</i> [verbal, reflexivo, propositivo, participial, conjuntivo, pronominal, nominal, adjetival, adverbial]		
Gloria Corpas Pastor (1996)	<i>Unidad fraseológica</i> [Colocación, Locución, Enunciado Fraseológico]		
Leonor Ruiz Gurillo (2001)	<i>Locuciones [nominal, adjetival, verbal, adverbial, marcadora, propositiva, clausal]</i>		Narrow
Mario García-Page (2008)	<i>Locuciones [nominales, adjetivales, adverbiales, propositivas, conjuntivas, verbales, oracionales]</i>		

The Spanish tradition of the study of phraseology includes two basic conceptions: the wide one and the narrow one. The wide conception could include everything from proverbs or collocations (depending on the author) to idioms. The narrow conception focuses only on *locuciones* (idioms), as evidenced by works such as those by Carneado & Trista (1985: 68), Ruiz Gurillo (1998: 12), Rakotojoelima (2004: 25), Sosiński (2006: 23), Školníková (2010: 7), and López (2012: 57).

Although both the English and the Spanish traditions have denominations for each kind of PU, and authors have undoubtedly developed complex

taxonomies to classify them, there are some aspects related to semantics and pragmatics that have not yet been addressed. For instance, literature on phraseology lacks information related to PUs' semantic patterns, or to the way in which their two semantic macro-components—the figurative and the mental image (Molina Plaza 2005: 176)—change from one language to another. This limitation is due to the lack of linguistic information—related to the composition of the PUs, their meaning, and how they are used in a communicative context—which can only be obtained through descriptive studies.

### *1.3. Phraseology and lexicography: a shared-ground proposal*

In order to deal with phraseological units in dictionaries, it is necessary to talk about lexicography in general, and lexicographic resources (i.e. dictionaries, glossaries, databases) in particular. Lexicography is considered as an applied discipline related to linguistics. According to Sinclair (1984):

“It is clearly an applied science or craft, rather than a pure one. That is to say, it relies for a theoretical framework on external disciplines. I know this is a contentious point and that this paper is not the proper forum for its debate, but the shape proposed for lexicography as an academic subject depends on the attitude taken to this issue. There is, for example, no subject heading ‘Lexicography theory’ in my syllabus because I have nothing to put there; on the other hand there is substantial input from IT and LINGUISTICS because I believe that the relevant theory is to be found in these areas or via these areas”. (1984: 6-7).

According to Moon (2009), Sinclair showed that lexicography does not have a theoretical background due to its applied nature, but at the same time, she recognizes that the methodology Sinclair developed for the COBUILD project was based on principles that could be applied to lexicography in general, one of them being the use of corpus linguistics for the creation of the dictionary.

On lexicographers and lexicography, Atkins & Rundell (2008) state that “by the nature of the work they do, lexicographers are applied linguists”, and although these authors think “a grounding in linguistic theory is not a prerequisite”, they also believe that “there are certain basic linguistic concepts which are invaluable in preparing people to analyze data and to produce concise, accurate dictionary entries” (2008: 130). In turn, regarding phraseology and lexicography, Leroyer (2006) states that the relationship between these

two disciplines should be considered a “scientific marriage” since they have been related for a long time. According to him, more than 1,700 reference entries can be found in the EURALEX site concerning both phraseology and lexicography (2006: 183). Leroyer also suggests that there are two ways to look at the relationship between these two disciplines: firstly, the treatment of phraseology by lexicographers and, secondly, the phraseological studies of linguists drawing recommendations on how to deal with phraseology in dictionaries (2006: 183). Furthermore, Paquot (2015) draws attention to several problems related to the phraseological information (related to collocations) that dictionaries provide. Among her findings, Paquot found a systematic lack of consistency in dictionary entries (2015: 5-6). This problem is also tackled by Moon (2008). She explains that lexicographic resources struggle with providing the description of phraseological units that meet the requirements of phraseological theories, and with the evidences of occurrence of those units in real texts. She further states that dictionaries must provide information about how idioms behave in context (2008: 314).

The study of the inclusion of phraseology in dictionaries has not only been of interest to linguists in English. It is also possible to find a number of articles related to the study of phraseology and dictionaries in Spanish. For instance, the papers in two books edited by Alonso (*Diccionarios y fraseología*, 2006) and Mellado Blanco (*Colocaciones y fraseología en los diccionarios*, 2008). On the one hand, included in the book edited by Alonso (2006), the study by González (2006) addresses how collocations and idioms are registered in the DRAE (Spanish Royal Academy’s Dictionary of the Spanish Language). This study made by González arrives at the conclusion that the selection criteria for the inclusion of collocations follow the classification system developed by Corpas Pastor (1996), while idioms are categorized using the taxonomy proposed by Casares (1950). On the other hand, also included in the book edited by Alonso, the work by Penadés (2006: 252-253) discusses issues related to the marking of phraseological units in dictionaries.

In the book edited by Mellado Blanco (2008), Ortega Ojeda & González Aguilar present the marks used in two general language dictionaries in Spanish, and they conclude that the marking in both dictionaries is inaccurate. The same holds true for the criteria that lexicographers used to classify and mark PUs in the dictionaries studied (González Aguilar 2008: 244).

All these studies show the tendency for marking and indexation in dictionaries to be incomplete, inconsistent, or inaccurate to some extent. In addition, Buendía Castro & Faber (2015) state that phraseological units have begun to be indexed more frequently in dictionaries in recent years (2015: 161). However, this does not mean that a systematic methodology is followed for the indexation or lemmatization of phraseological entries—this includes, for example, the criteria for choosing a certain word form as the headword of the PU—. One possible explanation for this problem is that although many studies and theoretical-methodological reflections have been proposed on how to deal with phraseological units in dictionaries, the conclusions provided by such works do not seem to be taken into account in the lexicographic practice. For instance, the introduction or guidelines of dictionaries should include information regarding the marking and indexation of phraseological units (Santamaría Pérez 2003: 1045), but that is not always the case.

As shown above (section 1.1), it is possible to find concepts in Spanish and English that are applicable to all the PUs suitable to be indexed in a monolingual or a bilingual dictionary. However, these criteria must be synthesized and shared among experts and publishing houses in an attempt to reach a consensus in aspects such as taxonomy, selection criteria, and marking, as it has been done before in lexicographic manuals regarding monolexical entries. As for the representation and indexation of PUs, Heid (2008) states that many current projects and initiatives involving Natural Language Processing are taking place in relation to the development of standards for PUs. Nevertheless, problems regarding the automatic identification, extraction and productivity of PUs “are far from being solved” (2008: 349-350).

The “quantification of the phenomenon” and the succeeding recording of PUs (Heid 2008: 349-350) is one of the several challenges that lexicography faces regarding phraseology. On this matter, Jackendoff (1997) observes that “there are vast numbers of such memorized fixed expressions; these extremely crude estimates suggest that their number is of about the same order of magnitude as the single words of the vocabulary” (1997: 156). Jackendoff’s claim is in turn quoted by Tschichold (2008) to add that the recording process of such amount of PUs in a language will always be incomplete (2008: 366). Heid (2008) identifies the need for more morphosyntactic and semantic annotated resources as well as research on this aspect of phraseology (2008: 354).

Nevertheless, he also points out that a possible solution for the identification of PUs could be reached by means of “distributional semantics” meaning that “items with similar contexts share meaning components” (Heid 2008: 353). A similar approach is used in this article (see sections 3.2 & 3.3) through the use of semantic annotation for the extraction of semantic patterns that could be used as criteria for the identification and extraction of PUs.

## 2. Data, Tools, and Methods

For the analysis intended here, two dictionaries were used: the *Diccionario fraseológico documentado del español actual* (henceforth DFDEA) (Seco, Andrés & Ramos 2004) and *The Collins COLBUILD dictionary of idioms* (henceforth CCDOI) (Sinclair & Moon 1997). This selection was based on the following criteria: (i) the dictionary is a phraseological or phraseology-related dictionary, (ii) it is a dictionary based on corpora, (iii) it is a reputable dictionary in terms of its publishing house, its editors and the lexicographers involved in its creation. However, before presenting the data and its related statistics, two questions need to be answered: What kinds of units are indexed in each dictionary? What lexicographic information is presented in the megastructure, macrostructure, and microstructure of each dictionary? In order to start answering those questions, the next section will offer a definition of megastructure, macrostructure, and microstructure.

### 2.1. Lexicographic information: megastructure, macrostructure, and microstructure

The present analysis is partly concerned with the ways in which PUs are represented in these two dictionaries. Thus, it is necessary to distinguish the characteristics of the sources from which data have been extracted. In order to do so, three parts of the dictionary had to be analyzed, namely: (i) the dictionary’s megastructure, (ii) its macrostructure, and (iii) its microstructure. The definitions given by Hartmann & James (1998) for these three terms will be the ones adopted in this paper. According to these authors, the megastructure “includes the macrostructure and the outside matter” (1998: 93); the macrostructure is “the overall list structure which allows the compiler



and the user to locate information” (1998: 91); finally, the microstructure is defined as “the internal design of a reference unit” (1998: 94).

The DFDEA’s megastructure encompasses seven sections: (i) the motivation of the dictionary, (ii) the guidelines of use, (iii) a list of abbreviations used in the dictionary, (iv) a glossary of linguistic terms, (v) an alphabetical consultation guide, (vi) the body of the dictionary, and (vii) a list of cited texts. All this information comprises 1,084 pages.

The first section of the DFDEA, related to the motivation of this lexicographic work, explains the choosing of three words included in the title of the dictionary: *fraseológico* (phraseological), *documentado* (documented), and *actual* (current). According to its editors, the dictionary is *fraseológico* because it contains several types of PUs, including idioms, collocations, formulaic expressions, foreign-language idioms, and sayings (Seco, Andrés & Ramos 2004: xvi-xviii) as exemplified in Table 3.

Table 3. PU examples taken from the DFDEA

Type of PU	Example
Idiom	<i>callejón sin salida</i>
Collocation	<i>prestar atención</i>
Formulaic expression	<i>calladito estás mejor</i>
Idioms in other languages	<i>sine qua non</i>
Sayings	<i>a lo hecho, pecho</i>

In the DFDEA two types of sources were used in order to retrieve the phraseological entries: corpora and the press. The corpora used included two resources from the Real Academia Española (CORDE and CREA), one that was compiled for the *Diccionario del español actual* (Seco, Andrés & Ramos 1999), and one *ad hoc* corpus for this specific project. The authors do not add any further information about how newspapers were used for the extraction of PUs; however, the last part of the dictionary has an appendix that contains all the texts cited, including the press references that were used (Seco, Andrés & Ramos 2004: xiii-xiv).

Finally, its temporal aspect turns this lexicographic work into a synchronic dictionary. It was developed by using sources from a period spanning almost 50 years (1955 to 2004), thus offering a picture of phraseology up to that time.

It is worth mentioning, that there is one aspect that was not explained in depth in the first section of the DFDEA regarding how PUs were indexed in the dictionary. The authors only explain that PUs are listed under certain headwords. Those headwords are emphasized in the consultation guide through the use of bold letters (see Table 4).

**Table 4.** Examples of headwords in the DFDEA

Phraseological Unit	Type of PU	Headword
<i>hombre de la calle</i>	Noun idiom	<b>hombre</b> (man)
<i>como un solo hombre</i>	Adverbial idiom	
<i>hacer un hombre</i>	Verbal idiom	
<i>vamos, hombre</i>	Interjectional idiom	

Therefore, at first sight, it looks like the expressions are listed under the noun (when present.) However, after a further analysis of other examples (see Table 5) there is no evidence of any practical or theoretical motivation for choosing a word in particular.

**Table 5.** Incongruence of headword choosing in the DFDEA

Phraseological Unit	Type of PU	headword
<i>clamar al cielo</i>	Verbal idiom	<i>cielo</i> (heaven)
<i>clamar en el desierto</i>	Verbal idiom	<i>desierto</i> (desert)
<i>clamar justicia</i>	Verbal idiom	<b>clamar</b> (to cry out)
<i>clamar venganza</i>	Verbal idiom	

The CCDOI consists of four main sections: (i) the introduction, (ii) the guidelines of use, (iii) the body of the dictionary, and (iv) an alphabetic consultation index of the PUs. The dictionary length is 493 pages.

The first section of the CCDOI includes a detailed explanation of the motivation behind this lexicographic work, the sources used for the extraction of PUs, as well as the definition of *idiom*. Among the PUs that the authors extracted for their inclusion in the CCDOI one can find not only idioms but also a wide range of expressions (see Table 6.) However, it is stated that phrasal verbs such as “give up” or “put off” are not included in this work (Sinclair & Moon 1997: v).

**Table 6.** Examples of PUs included in the CCDOI

Type of PU	Example
Idiom	spill the beans
Multiword metaphors	the acid test
Metaphorical proverbs	in for a penny, in for a pound
Expressions with pragmatic meaning	famous last words

The main source for the extraction of PUs was the Bank of English, a subset of the Collins Corpus, containing approximately 650 million words (HarperCollins n.d.). One feature that this dictionary presents is the frequency-of-occurrence mark based on the corpora from where they were extracted. In the dictionary’s introductory section, the editors explain that idioms have an infrequent level of occurrence in texts. The dictionary offers a scale in which the least frequent idioms occur less often than once per 10 million words and the most frequent at least once per two million words (Sinclair & Moon 1997: v). This scale is included in front of each idiom (see Table 7).

**Table 7.** Frequency bands in the CCDOI

PU	Frequency indicator	Range
prepare the ground	◀◀◀	Once every two million words
fire on all cylinders	◀◀	Not specified in the dictionary
all system go	◀	Between 1 and 3 times every 10 million words
come down in the world	No indicator	Not specified in the dictionary

The consultation index in the CCDOI is organized alphabetically, based on the headwords of the PUs. Another characteristic of this index is that it does not take the determiners ‘the’, ‘a’, or ‘an’ into consideration for the alphabetical-order distribution of the PUs (see Table 8).

**Table 8.** PU examples taken from the CCDOI

Headword	PU order
Light	a leading <b>light</b>
	<b>light</b> a fire under someone
	<b>light</b> as a feather
	the <b>light</b> at the end of the tunnel

At this point, it is noteworthy that the macrostructure of both the DFDEA and the CCDOI share the same lemmatization and indexation of entries as it can be seen in their own consultation guidelines. That means that PUs are listed under certain headwords, and, subsequently, those headwords are listed alphabetically. In contrast, the microstructures of both dictionaries differ in the information they include, as illustrated in Table 9.

**Table 9.** Lexicographic article in the DFDEA and the CCDOI (microstructure)

DFDEA	CCDOI
Headword	Headword
Entry (in alphabetical order)	Entry (in alphabetical order excluding determiners)
Grammatical/Functional marking	<b>Not included</b>
Diasystematical marking (colloquial, jargon, etc.)	<b>Not included</b>
<b>Not included</b>	Frequency band (according to the Bank of English)
Definition (direct or by context of use)	Definition (sentence like definition)
Example (Concordance from corpora or the press)	Example (Concordance from the Bank of English)
Source of the example	<b>Not included</b>

Since one of the objectives of this study is to analyze the morphosyntactic patterns related to idioms, the absence of markings (word class e.g.: nominal, adjectival, verbal, etc.) in the CCDOI was an obstacle for achieving such goal. Therefore, it was necessary to assign a grammatical/functional mark to each entry of the CCDOI. However, this procedure will be explained in detail in the next section of this paper.

As shown in Table 8, the two dictionaries of interest present definitions in different ways. In the DFDEA definitions are presented in a manner in which they can substitute the entry in certain contexts. In cases in which it is not possible to offer a “direct” definition, the dictionary includes an explanation of the use of the expression. The CCDOI include full-sentence definitions that explain the different contexts of use for each expression (see Table 10).

**Table 10.** Examples of definitions in the DFDEA and the CCDOI

Expression	Definition
<i>pedir peras al olmo</i>	<i>Esperar o pretender imposibles.</i> (Seco, Andrés & Ramos 2004: 774)
like getting blood out of a stone	If you have difficulty persuading someone to give you money or information, you can say that it is like getting blood out of a stone. (Sinclair & Moon 1997: 36)
<i>Dios los cría y ellos se juntan</i>	<i>Se usa para comentar la unión de personas de caracteres o intereses similares.</i> (Seco, Andrés & Ramos 2004: 398)
birds of a feather flock together	If you describe two or more people as birds of a feather, you mean that they are very similar in many ways. (Sinclair & Moon 1997: 34)

Determining if one way of defining the PUs is better than the other depends on each reader’s —e.g. a linguist, translator, or enthusiast of phraseology— interests. What becomes apparent is that neither of those definitions could provide a quick solution for a user who does not know exactly which PU he/she is looking for. Nonetheless, a solution to this problem will be humbly proposed in the course of this paper.

## 2.2. Data selection and database compilation

In order to carry out the analyses proposed in this study, two databases were compiled (one containing the entries of the DFDEA, and the other containing the entries of the CCDOI). The database in Spanish includes 16,760 PUs composed by 55,831 word forms, while the database in English contains 4,285 PUs composed by 18,123 tokens. The tokens in both databases include grammar and lexical words as well as punctuation marks.

One limitation in phraseological studies aiming at characterizing sets of units has to do with the selection of the analytical sample. The amount of PUs included in the two databases for this study goes beyond what could be informed about in a single paper. Therefore, it was necessary to reduce the number of units for the analysis while maintaining a representative group of them. It was decided then to single out a limited number of selection criteria from the data starting with the number of forms (see Fig. 1).

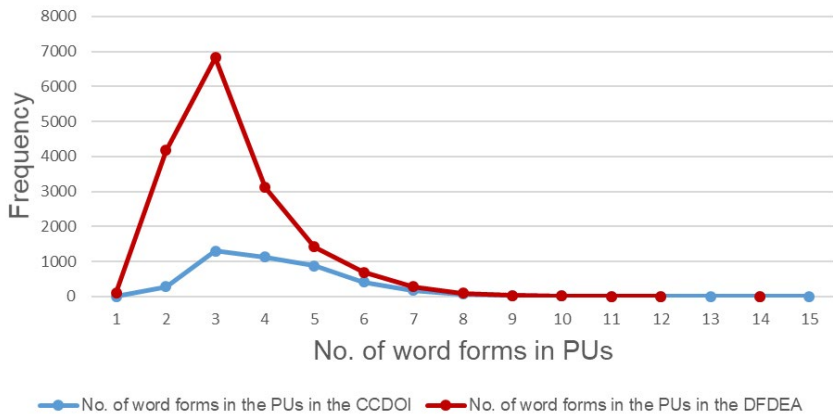


Figure 1. No. of word forms of the PUs in the DFDEA and the CCDOI.

It becomes apparent that Figure 1 shows a certain inconsistency with the criterion of plurilexicality of PUs. For example, a closer look to the data

shows that the DFDEA indexed 105 PUs with only one word form, while the CCDOI indexed only one PU with one word form. The Spanish PUs were enclitic units, and, therefore, they were presented as one-form PUs (which is common use in lexicographic entries). Moreover, the only PU word form in English is an initialism: OTT (over the top) that was indexed in the CCDOI under the headword top (see Table 11).

Additionally, Figure 1 shows an uneven distribution of the frequency of PUs' number of forms. The only group that alters that frequency distribution is the one comprising phraseological bi-grams. One possible explanation for this might be related to an editorial decision to avoid these kinds of units due to the difficulty in drawing a boundary between compounds and PUs. This tendency has already been observed in other lexicographic resources in Spanish (Rojas Díaz & Pérez Sanchez 2019). However, reaching a satisfactory conclusion on the reasons behind this frequency distribution would only be possible through a deeper analysis of this group of units.

**Table 11.** Contexts of use of one-form PUs indexed in the DFDEA and in the CCDOI

Expression	Context of use
<i>componérselas</i>	<i>Tiene mucha familia y parece buen prójimo. Mal se las va a componer el hombre.</i> (Seco, Andrés, & Ramos 2004: 303).
OTT	Each design is very different in style. Some are subtle, some gloriously OTT (Sinclair & Moon 1997: 397).

With the information presented above, the criteria needed in order to carry out the sample selection was finally available. Thus, the first selection criterion was the number of forms. Then, the PUs consisting of three, four, and five forms were chosen. That selection, in turn, allowed for the study of more than 50% of the entries in both dictionaries (see Table 12).

**Table 12.** Distribution of PUs consisting of three - five forms, indexed in the DFDEA and in the CCDOI

No. of forms	Frequency in the DFDEA (%)	Frequency in the CCDOI (%)
Three-form PUs	6828 (40.7%) ( <i>abogado del diablo</i> )	1294 (30.2%) (bite your tongue)
Four-form PUs	3117 (18.6%) ( <i>cara de pocos amigos</i> )	1122 (26.2 %) (dig your own grave)
Five-form PUs	1422 (8.5%) ( <i>el malo de la película</i> )	870 (20.3 %) (get your brain into gear)

Since this study has been conceived as the starting point of a larger project involving specialized dictionaries, the second criterion for sample selection was the functional marking of each PU. The DFDEA offered an extensive set of marks for this purpose; however, that was not the case for the CCDOI, as shown in Table 9, above. Given that this criterion was central both for this study, and, as said before, for further in-depth, specialized-lexicography studies, all the entries of the CCDOI were marked manually by using the marking set provided by the DFDEA (Seco, Andrés, & Ramos 2004: xxvii-xxviii) and the functional/grammatical information available in another dictionary related to the Cobuild project (Sinclair 2006). Once all the entries in both dictionaries were marked, 33 different marks were identified in the DFDEA, and 17 in the CCDOI.

The analysis of functional marking in both dictionaries allowed for the identification and selection of PUs' grammatical functions of interests. Thus, on the one hand, verb PUs were chosen for in-depth analysis because it was the most frequent mark in the DFDEA and in the CCDOI. On the other hand, given that authors such as Sager (1990: 58) and L'Homme (2004) assert that nouns are predominant in concept representation in specialized dictionaries, noun PUs were selected as the second category to be analyzed in depth. Once that selection was made, the analysis databases were finally set for carrying out the study on 4,932 PUs chosen from the DFDEA and 2,387 PUs from the CCDOI (See Fig. 2).



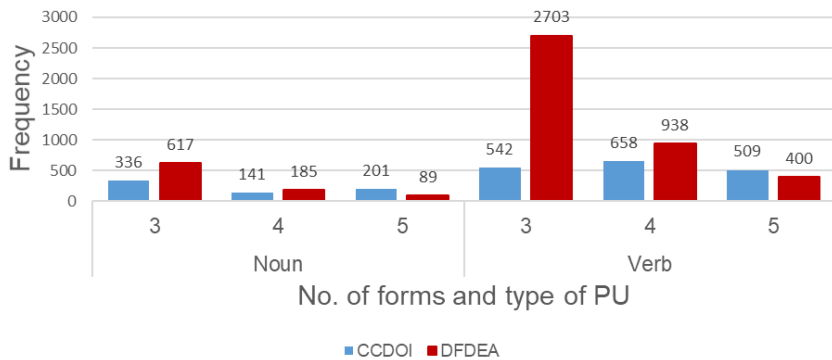


Figure 2. No. of nominal and verbal PUs in the analysis databases and their corresponding number of forms.

In summary, based on criteria such as frequency and importance for concept representation, verb and noun PUs were selected from both dictionaries in order to create the analysis databases for this study. Once the databases were set up, the units included in them were analyzed in depth as explained in the following section.

### 3. Analysis and Results

In view of the fact that the objective of this study is to extract as much linguistic information from the PUs as possible three different analyses were performed: (i) lexical, (ii) semantic, and (iii) morphosyntactic. Those analyses will be explained in detail here.

#### 3.1. Lexical analysis

The first step in order to perform the lexical and the morphosyntactic analyses was to implement a Part-of-Speech (henceforth POS) tagging on the databases. For that task, the TreeTagger (Schmid 1994) was employed, followed by a homogenization of the tags in order for them to be readable. 26,277 forms were tagged including Saxon possessive morphemes (e.g. a baker’s dozen) and hyphens (-) in English, constituted as categories (see Table 12).

**Table 13.** Distribution of component words by POS in the databases

POS	Frequency in the DFDEA (%)	Frequency in the CCDOI (%)
Noun	5,430 (32.14%)	3,354 (35.76%)
Verb	4,349 (25.74%)	1,731 (18.45%)
Determiner	2,876 (17.02%)	1,447 (15.43%)
Preposition	2,166 (12.82%)	1,233 (13.14%)
Adjective	667 (3.95%)	552 (5.88%)
Pronoun	198 (1.17%)	514 (5.48%)
Adverb	526 (3.11%)	139 (1.48%)
Contraction	318 (1.88%)	0 (0.00%)
Conjunction	213 (1.26%)	90 (0.96%)
Past Participle	132 (0.78%)	78 (0.83%)
Saxon possessive	0 (0.00%)	126 (1.34%)
Present Participle	7 (0.78%)	72 (0.77%)
Hyphen	0 (0.00%)	42 (0.45%)
Demonstrative	11 (0.07%)	0 (0.00%)
Interjection	4 (0.02%)	2 (0.02%)

The ‘noun’ category is the most frequent among the component words followed by the ‘verb’ category (see Table 13). This goes in contrast with the predominance of verbal PUs shown in Fig. 2. However, the reason for having more nouns than verbs in the word class counting is that a number of nouns co-occur with verbs in verb PUs.

Once POS frequency was determined, a word cloud was plotted in order to identify the most frequent nouns and verbs among the component words of the PUs (see Fig. 3).



All the words in the database were lemmatized. Thus, it becomes apparent that although the most frequent word form in the word cloud (see Fig. 3) is *cabeza*, once the database was lemmatized, this information could vary (see Table 14). The table below illustrates two different scores. On the one hand, the CORDE (Real Academia Española, n.d.), and the frequency lists by Kilgarriff (2006) show the ranks of such words in a corpus. On the other hand, the *Corpus del español* (2002) and the frequency lists by Leech, Rayson & Wilson (2001) provide ranks based on each word's POS. It is evident that all the words in the databases are ranked within the top 1000 most frequent words in Spanish and in English respectively, according to the corpora consulted. The words in Table 13 are also within the top 200 most frequent nouns and verbs according to the POS frequency list of the *Corpus del español* (2002) and the frequency lists by Leech, Rayson & Wilson (2001).

**Table 14.** Top 5 ranking for nouns and verbs of component lemmas of PUs in corpora

Spanish (DFDEA)				English (CCDOI)			
POS	Lemma	CORDE	<i>Corpus del español</i>	POS	Lemma	Kilgarriff	Leech, Rayson, & Wilson
Noun	<i>ojo</i>	184	14	Noun	hand	176	26
	<i>mano</i>	158	26		eye	240	43
	<i>cabeza</i>	324	47		head	241	38
	<i>vida</i>	99	5		foot	484	163
	<i>cara</i>	522	112		line	278	73
Verb	<i>dar</i>	136	40	Verb	get	44	8
	<i>hacer</i>	140	17		have	8	2
	<i>tener</i>	192	36		go	40	10
	<i>ser</i>	51	6		put	125	26
	<i>poner</i>	387	119		take	54	13

Table 14 presents the top 5 nouns and verbs extracted from the databases. This gives indications about the POS and semantic-category distribution of the words in the databases.

### 3.2. *Semantic analysis*

The second type of analysis performed in this study is a semantic one. In recent years, scholars have progressively explored some semantic aspects of phraseology, more especially in studies related to terminology and languages for specific purposes (Grčić Simeunović & de Santiago 2016; Patiño 2017). For this study, however, a different semantic approach was taken. The UCREI's Semantic Analysis System (henceforth USAS) was employed. USAS is a POS and semantic tagger, containing semantic tags divided into 232 semantic categories based, in turn, on 21 discourse fields identified by McArthur (1981) (Archer, Wilson & Rayson 2002: 2).

All the word forms of the database were tagged with this semantic tagset and revised and corrected manually in both languages, thus creating four analysis layers, namely: lexical, grammatical, discourse filed, and semantic category. These four layers made it possible to observe how certain morpho-syntactic patterns interacted with different sequences of semantic categories (hereinafter semantic patterns) (see Table 15).

**Table 15.** Database tagging sample in Spanish and English

		Spanish (DFDEA)			
Word form	<i>el</i>	<i>ombligo</i>	<i>del</i>	<i>mundo</i>	
POS	Prep	N	Contr	N	
Discourse field	Z	B	Z	W	
(Descriptive)	Names and grammatical words	Body and the individual	Names and grammatical words	World and the environment	
Semantic level	Z5	B1	Z5	W1	
(Descriptive)	Grammatical bin	Anatomy and physiology	Grammatical bin	The universe	
		English (CCDOI)			
Word form	throw	in	the	towel	
POS	V	Prep	Det	N	
Discourse field	M	Z	Z	B	
(Descriptive)	Movement, location, travel, and transportation	Names and grammatical words	Names and grammatical words	Body and the individual	
Semantic level	M2	Z5	Z5	B5	
(Descriptive)	Putting, pulling, pushing, transporting	Grammatical bin	Grammatical bin	Clothes and personal belongings	

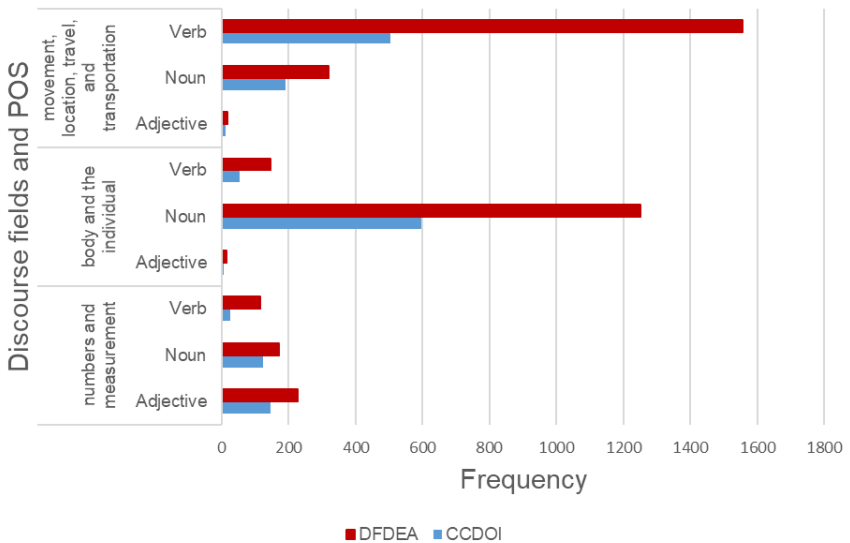
Some information —such as the distribution of word forms in semantic fields— could only be obtained when the databases were tagged by using the USAS tagset. When comparing the information from Fig. 3 with that presented in Table 13, it is possible to state that there is a strong tendency for parts of the body to occur as a word form in the database. Nevertheless, when comparing the whole distribution of word forms in the databases, according to the discourse fields provided by McArthur (1981) and tagged through USAS, it is possible to observe that the category “the body and the individual” ranks fourth (see Table 16). As observed elsewhere, in a lexicographic resource in Spanish, the occurrence of these words used in everyday language is an indicator of embodiment in the creation and fixation of PUs in general language

(Rojas Díaz & Pérez Sanchez 2019: 9). Embodiment is a concept developed in cognitive linguistics, and it is based on the statement that “our concepts, our ideas are influenced and composed by the structure of our bodies, by our own experience of the world that surrounds us” (Ibarretxe-Antuñano & Valenzuela 2016: 44, author’s translation)

**Table 16.** Distribution of discourse fields tags in the databases

Discourse field	Frequency in DFDEA (%)	Frequency in CCDOI (%)
Names and grammar (Z)	6,249 (36.98%)	3,958 (42.39%)
General and abstract terms (A)	2,343 (13.87%)	945 (10.12%)
Movement, location, travel, and transportation (M)	1,979 (11.71%)	800 (8.57%)
Body and the individual (B)	1,431 (8.47%)	665 (7.12%)
Substances, materials, objects, and equipment (O)	763 (4.52%)	657 (7.04%)
Social actions, states, and processes (S)	641 (3.79%)	204 (2.18%)
Numbers and measurement (N)	527 (3.12%)	343 (3.67%)
Psychological actions, states, and processes (X)	466 (2.76%)	239 (2.56%)
Life and living things (L)	455 (2.69%)	300 (3.21%)
Language and communication (Q)	356 (2.11%)	169 (1.81%)
Food and farming (F)	320 (1.89%)	158 (1.69%)
Emotion (E)	294 (1.74%)	120 (1.29%)
World and environment (W)	221 (1.31%)	109 (1.17%)
Government and public (G)	182 (1.08%)	97 (1.04%)
Entertainment, sports, and games (K)	169 (1.00%)	146 (1.56%)
Architecture, housing and home (H)	165 (0.98%)	112 (1.20%)
Time (T)	161 (0.95%)	98 (1.05%)
Money and commerce in industry (I)	105 (0.62%)	184 (1.97%)
Science and technology (Y)	30 (0.18%)	5 (0.05%)
Arts and crafts (C)	27 (0.16%)	19 (0.20%)
Education (P)	13 (0.08%)	10 (0.11%)

Although this study is not intended to do a contrastive analysis between languages, but to present the information in parallel, some contrastive insights could be obtained when taking a closer look at the databases. There is a tendency in both dictionaries for certain types of words to occur within specific discourse fields, as is the case for verbs indicating movement (7.85% in the databases), nouns related to body parts (7.04% in the databases), and adjectives describing measurements (1.42% in the databases) (see Fig. 4).



**Figure 4.** Distribution of main lexical categories (verbs, nouns, and adjectives) grouped by discourse fields.

At first sight, most semantic relationships between PUs' word forms and their meanings seem to be metaphorical. Nevertheless, a closer look shows that several cases are also metonymical (see Table 17), but an in-depth analysis of such semantic relationships is beyond the scope of the present work.



**Table 17.** Examples of metaphorical and metonymical relationships between PUs and their meaning

DFDEA		CCDOI	
Phraseological Unit	Meaning	Phraseological Unit	Meaning
<i>arma de doble filo</i> (Metaphorical)	<i>Cosa, argumento o procedimiento, que puede ocasionar un resultado opuesto al que se pretende.</i>	the salt of the earth (Metaphorical)	If you describe someone as the salt of the earth, you are showing admiration for their honesty
<i>cargar la barriga</i> (Metonymical)	<i>Quedarse embarazada</i>	give someone a black eye (Metonymical)	If you give someone a black eye, you punish them severely for something they have done, but without causing them permanent harm

Finally, it is necessary to state that semantic analysis is crucial not only for the type of work intended here, but also for phraseological studies in general, given that the study of meaning in phraseology will shed light on both the understanding of PUs and on their proper representation in lexicographical resources.

### 3.3. Morphosyntactic analysis

The third analysis in this study was morphosyntactic. 816 morphosyntactic patterns were extracted from the DFDEA and CCDOI. 388 (47.5%) of those patterns had two or more occurrences in the database. When combining both variables, it was possible to make a query about morphosyntactic and semantic patterns among the PUs. Table 18 includes the top-five most frequent patterns for noun and verb PUs in the dictionaries.

**Table 18.** Top five morphosyntactic patterns of noun and verb PUs in the DFDEA and the CCDOI

Morphosyntactic pattern	Frequency	Example	Type of PU	Dictionary
V Det N	1567	<i>abandonar el barco</i>	Verb	DFDEA
V Prep N	422	<i>bailar de alegría</i>		
V Prep Det N	398	<i>caber en la cabeza</i>		
V Contr N	150	<i>dar del vientre</i>		
V Adj N	94	<i>echar buen pelo</i>		
N Prep N	393	<i>faena de alino</i>	Noun	
N Prep Det N	63	<i>gatos en la barriga</i>		
N Contr N	59	<i>hombre del saco</i>		
Det Adj N	37	<i>la mínima expresión</i>		
Det N Prep Det N	37	<i>un cero a la izquierda</i>		
V Det N	233	<i>jump the gun</i>	Verb	CCDOI
V Pron N	134	<i>keep your cool</i>		
V Prep Det N	102	<i>lay down the law</i>		
V Det N Prep N	90	<i>make a meal of something</i>		
V N Prep Det N	69	<i>play things by the book</i>		
Det N N	109	<i>a bean counter</i>	Noun	
Det Adj N	108	<i>the acid test</i>		
Det N Prep Det N	71	<i>a skeleton in the cupboard</i>		
Det N Prep N	48	<i>the kiss of death</i>		
Det PrP N	26	<i>a sitting duck</i>		

It is possible to nest the morphosyntactic patterns with the semantic patterns, which makes it possible to identify nouns related to certain discourse fields. In Table 19, some examples of the semantic patterns linked to the most frequent morphosyntactic patterns are shown. Each of the letters composing the

semantic patterns corresponds to one of the discourse field labels previously presented in Table 16. From another point of view, it is also possible to look for semantic patterns and to observe morphosyntactic patterns that follow a specific semantic combination.

**Table 19.** Example of pattern nesting of semantic tags in morphosyntactic patterns

Morphosyntactic pattern	Semantic pattern	Example	Type of PU	Dictionary
V Det N	M Z B	<i>alzar la hombro</i>	Verb	DFDEA
	B Z B	<i>cagarse los calzones</i>		
N Prep N	B Z O	<i>lengua de trapo</i>	Noun	
	Q Z S	<i>cuento de hadas</i>		
V Det N	E Z H	hit the roof	Verb	CCDOI
	X H O	know the ropes		
Det Adj N	Z O O	a bright spark	Noun	
	Z O Q	a dirty word		

Similarly, it is also possible to nest the PUs by taking semantic patterns as a starting point, and then looking at what morphosyntactic patterns can be derived from those semantic patterns. This means that semantic tags could be used as a variable for either the extraction or classification of PUs in corpora, or for the indexation in lexicographic resources (see Table 20).

**Table 20.** Example of pattern nesting of morphosyntactic tags in semantic patterns

Semantic pattern	Morphosyntactic pattern	Example	Type of PU	Dictionary	
M Z B	V Det N	<i>correr la sangre</i>	Verb	DFDEA	
	V Prep N	<i>caerse de culo</i>			
	V Contr N	<i>salir del corazón</i>			
	V Prep V	<i>echarse a dormir</i>			
	V Prep PP	<i>ir de dormida</i>			
	V Prep Adj	<i>pasar a limpio</i>			
B Z O	N Prep N	<i>lengua de trapo</i>	Noun		
	N Contr N	<i>cana al aire</i>			
	N Adv N	<i>ojos como platos</i>			
	N Conj N	<i>pelos y señales</i>			
A Z O	V Det N	<i>fan the flames</i>	Verb		CCDOI
	V Prep N	<i>be in overdrive</i>			
	V Pron N	<i>blow your stack</i>			
	V N N	<i>give someone stick</i>			
	V Conj V	<i>crash and burn</i>			
	V Adj N	<i>spread like wildfire</i>			
	V N Adj	<i>catch someone cold</i>			
Z O O	Det Adj N	<i>a black mark</i>	Noun		
	Der N N	<i>the brass ring</i>			

The count of morphosyntactic and semantic patterns in the databases is presented in the following table. The information includes the frequency and relative percentages of each of the patterns that were extracted (see Table 21).

**Table 21.** Summary of the frequencies and percentages of the patterns in the databases

Dictionary	Type of PU	Morphosyntactic pattern		Semantic patterns	
DFDEA	Noun	124	(29.5 %)	461	(33.8 %)
	Verb	296	(70.5 %)	904	(66.2 %)
	<b>Total DFDEA</b>	<b>420</b>		<b>1,365</b>	
CCDOI	Noun	126	(31.8 %)	455	(35.3 %)
	Verb	270	(68.2 %)	835	(64.7 %)
	<b>Total CCDOI</b>	<b>396</b>		<b>1,290</b>	
<i>Total dictionaries</i>		<b>816</b>	<b>(23.5%)</b>	<b>2,655</b>	<b>(76.5 %)</b>
<i>Grand total</i>		<b>3,471</b>			

Traditionally, the morphosyntactic patterns resulting from the analyses (that have been carried out in the phraseology studies) have been used for the recognition and the extraction of candidates of PU in corpora. Nevertheless, the results regarding the nesting of morphosyntactic patterns and semantic patterns offered in this article will shed light on how to enhance the recognition method through semantic annotation but also the analysis of metaphorical and metonymical patterns.

#### 4. Discussion, Conclusions, and a Practical Proposal

This study offers 3,471 patterns. They are divided as follows: 816 (23.5%) are morphosyntactic patterns (420 from the DFDEA and 396 from the CCDOI) and 2,655 (76.5%) are semantic patterns (1,365 from the DFDEA and 1,290 from the CCDOI). The distribution of these patterns follow a ratio of almost 1:3, meaning that for each morphosyntactic pattern extracted three semantic patterns were identified. This ratio (1:3) is consistent within the dictionaries and the different types of idioms. The most frequent morphosyntactic patterns and semantic patterns of nominal and verbal units consisting of three, four, and five forms were exemplified and presented. This information can be used as a gold standard in order to make a comparison between the linguistic

features found in PUs in general language dictionaries and PUs in specialized dictionaries.

Semantic tagging of databases or corpora offers the opportunity for testing different parameters for the extraction of PUs aiming at lexicographic or terminographic work. The use of morphosyntactic and semantic annotation for PUs opens the discussion on how PUs should be indexed in dictionaries nowadays. Although semasiology and onomasiology are two very well-known concepts in lexicography and terminography, it is evident that most phraseological dictionaries follow a semasiological approach for the indexation of entries, i.e., those dictionaries answer the question of what does X (word/phrase/idiom/proverb) mean? However, such approach requires the user to know the form or the expression he/she is looking for (Kocjančič 2004). Additionally, the results of an analysis like the one presented here can also provide empirical data useful for the study of the semantic composition of metaphorical and metonymical constructions.

The frequency analysis along with the semantic information extracted from the component words of the PUs of the DFDEA and the CCDOI shows the use of common words of our daily experiences to describe more complex conceptions through rhetorical devices such as similes, metonymies and metaphors as it has already been observed in several studies in corpora and lexicographic resources. (Ellis 2008; Sharma 2018; Torijano & Recio 2019; Rojas Díaz & Pérez Sanchez 2019). The results of the previously mentioned analyses support some of the views of Cognitive Semantics regarding the embodiment hypothesis (Ibarretxe-Antuñano & Valenzuela 2016: 37).

In many cases, dictionary users (e.g. translators) do not know the exact form or expression they are looking for. As a result, looking for similar or equivalent expressions in semasiological dictionaries becomes a time-consuming task. One solution that could be offered to users so that they can do better and more efficient searches for PUs in dictionaries would be to transform the way in which dictionaries present phraseological entries. That could be done by grouping PUs' entries semantically, following a hybrid indexation that uses semasiological and onomasiological approaches. Therefore, suggestion derived from the present study entails the display of information in phraseological entries somehow as it is exemplified in Figure 5

**(1) HAPPINESS [FELICITY, HAPPY, JOY]**

◆ Adjectival (5)

- If someone is very happy, someone is...
  - happy as a clam
  - happy as a lark
  - happy as a pig in muck
  - happy as a sandboy
  - happy as Larry

Figure 5. Example of an entry with an onomasiological/semasiological hybrid approach related to happiness.

The five idioms presented in Figure 4 have the same meaning in the CCDOI “If you are X, you are very happy”. A representation like the one in Figure 5 not only allows the user to look for the expression needed, but it also provides the user with similar expressions. The lexicographic article could be expanded in order to provide the user with more lexicographic information such as diatopic marking (related to the place), diaphasic marking (related to language register), and contexts, among others (see Fig. 6).

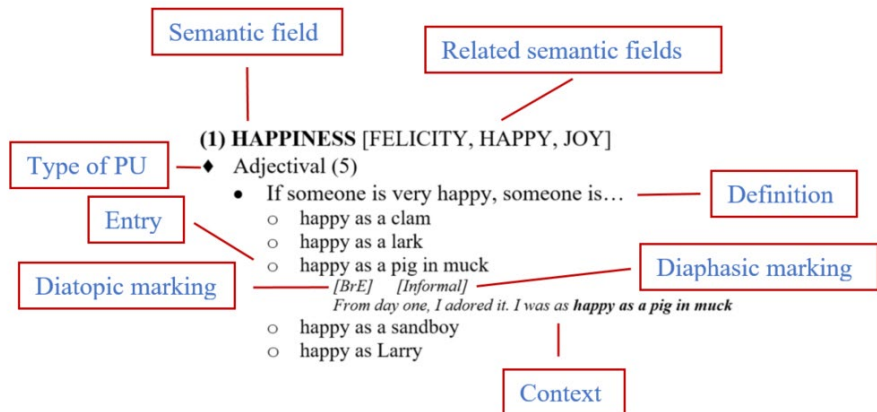


Figure 6. Example of an entry with an onomasiological/semasiological hybrid approach.

Information could also be presented in an electronic format (see Fig. 7) allowing the user to make different kind of queries if the entries are annotated morphosyntactically and syntactically.

The screenshot shows a web application window titled "Onofrasikon - NHH". The main area is titled "QUERY BY SEMANTIC FIELD" and contains several input fields and text boxes. The "Semantic field" dropdown is set to "HAPPINESS". The "Type of idiom" dropdown is set to "Adjectival". The "Entry" dropdown is set to "happy as a pig in muck". Below these are two text boxes: "Diatopic:" with "[BrE]" and "Diaphasic:" with "[informal]". The "Definition:" text box contains the text "If you are happy as a pig in muck, you are very happy". The "Context:" text box contains the text "From day one I adored it. I was as happy as a pig in muck". At the bottom of the window is an "OK" button and a footer area containing logos for NHH, EFMD EQUIS ACCREDITED, and CEMS.

Figure 7. Example of an entry with an onomasiological/semasiological hybrid approach in an electronic format.

Evidently, this reflection on the lexicographic techniques used to compile dictionaries needs to be broadened and verified with users and lexicographers to test its suitability as a possible approach for the enhancement of the compilation of dictionaries.

Finally, although it is true that it is impossible to offer the whole picture of the paradigm of phraseology for a language on the basis of the analysis of dictionaries, the information, statistics, and findings presented here can be used as a starting point for a transformation in the description and indexation of PUs in future studies and projects related to phraseology and lexicography.



## References

- AGUADO DE CEA, Guadalupe. (2007) "A multiperspective approach to specialized phraseology: Internet as a reference corpus for phraseology." In: Posteguillo, Santiago; María José Esteve & María Lluïsa Gea Valor (eds.) 2007. *The Texture of Internet: Netlinguistics in Progress*. Newcastle: Cambridge Scholars Publishing. 182-207.
- ALONSO, Margarita (ed.) (2006) *Diccionarios y fraseología*. La Coruña: Universidade da Coruña.
- ARCHER, Dawn; Adrew Wilson & Paul Rayson. (2002) *Introduction to the USAS category system*. Benedict project report.
- ATKINS, Sue & Michael Rundell. (2008) *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- BLOOMFIELD, Leonard. (1933) *Language*. New York: Henry Holt.
- BUENDÍA CASTRO, Miriam & Pamela Faber. (2015) "Phraseological units in English-Spanish legal dictionaries: a comparative study." *Fachsprache* 3:4, pp. 161-175.
- BUSHNAQ, Tatiana. (2015) "A Retrospective Analysis of the Term Phraseological Unit." In: Boldea, Iulian (ed.) 2015. *Debates on Globalization. Approaching National Identity through Intercultural Dialogue*. Tirgu-Mureş: Arhipelag XXI Press, pp. 167-176.
- CARNEADO, Zoila & Antonia Trista. (1985) *Estudios de Fraseología*. La Habana: Academia de Ciencias de Cuba, Instituto de Literatura y Lingüística.
- CASARES, Julio. (1950) *Introducción a la lexicografía moderna*. Madrid: CSIC.
- CASTILLO CARBALLO, María Auxiliadora. (2006) *El lema: tipos de entradas*. Electronic version: <<https://www.liceus.com/producto/lema-tipos-entradas/>>
- CHAFE, Wallace. (1968) "Idiomaticity as an Anomaly in the Chomskyan Paradigm". *Foundations of Language* 4, pp. 109-127.
- CHOMSKY, Noam. (1965) *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- CORPAS PASTOR, Gloria. (1996) *Manual de fraseología española*. Madrid: Gredos.
- COWIE, Anthony. (2001) "Introduction" In: Cowie, Anthony (ed.) 2001. *Phraseology: Theory, Analysis, and Application*. Oxford: Oxford University Press, pp. 1-22.
- DAVIES, Mark. (2002) *Corpus del español: 100 million word, 1200s-1900s*. (Corpus) Electronic version: <<http://corpusdelespanol.org/hist-gen>>

- ELLIS, Nick. (2008) "Phraseology: The periphery and the heart of language." In: Meunier, Fanny & Sylviane Granger (eds.) 2008. *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins, pp. 1-13.
- FIRTH, John. (1957) *Papers in Linguistics, 1934-1951*. London: Oxford University Press.
- GARCÍA-PAGE, Mario. (2008) *Introducción a la fraseología española*. Barcelona: Anthropos.
- GONZÁLEZ, María Isabel. (2006) "La definición lexicográfica de las unidades fraseológicas: la aplicación de modelos formales." In: Alonso, Maragarita (ed.) 2006. *Diccionarios y fraseología*. La Coruña: Universidade da Coruña, pp. 221-233.
- GRČIĆ SIMEUNOVIĆ, Larisa & Paula de Santiago. (2016) "Semantic approach to Phraseological Patterns in Karstology." In: Margalitatze, Tinatin & George Meladze (eds.) 2008. *Proceedings of the XVII Euralex International Congress*. Tbilisi: Ivane Javakhishvili Tbilisi State University, pp. 685-693.
- HARPERCOLLINS. (n.d.) *The Collins Corpus*. (Corpus) Electronic version: <<https://collins.co.uk/pages/elt-cobuild-reference-the-collins-corpus>>
- HARTMANN, Reinhard & Gregory James. (1998) *Dictionary of Lexicography*. New York: Routledge.
- HEID, Ulrich. (2008) "Computational phraseology: An overview." In: Granger, Sylviane & Fanny Meunier (eds.) 2008. *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins Publishing Company, pp. 337-360.
- HOCKETT, Charles. (1958) *A Course in Modern Linguistics*. New York: Macmillan.
- HOUSEHOLDER, Fred. (1959) "On Linguistic Primes." *Word* 15, pp. 231-239.
- IBARRETXE-ANTUÑANO, Irialde & Javier Valenzuela (eds.) (2016) *Lingüística Cognitiva*. Barcelona: Anthropos.
- IDIOM. (n.d.) *Cambridge dictionary*. Electronic version: <<https://dictionary.cambridge.org/dictionary/english/idiom>>
- JACKENDOFF, Ray. (1997) *The Architecture of Language Faculty*. Cambridge: The MIT Press.
- KILGARRIFF, Adam. (2006) *BNC database and word frequency lists*. (Word list) Electronic version: <<http://www.kilgarriff.co.uk/bnc-readme.html>>
- KOČJANČIČ, Polonca. (2004) "Acerca de la macroestructura y la microestructura en el diccionario bilingüe." *Verba Hispanica* 12:1, pp. 171-185.

- LEECH, Geoffrey; Paul Rayson & Andrew Wilson. (2001) *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Longman.
- LEROYER, Patrick. (2006) "Dealing with phraseology in business dictionaries: focus on dictionary functions - not phrases." *Linguistik online* 27:2, pp. 183-194 Electronic version: <<https://bop.unibe.ch/linguistik-online/article/view/750/1279>>
- L'HOMME, Marie-Claude. (2004) *La terminologie: principes et techniques*. Montreal: Presses de l'Université de Montréal.
- LOCUCIÓN. (n.d.) *Diccionario de la lengua española*. Electronic version: <<https://dle.rae.es/?id=NYSj8PH>>
- LÓPEZ, Xavier Pascual. (2012) *Fraseología española de origen latino y motivo grecorromano*. Lleida: Universitat de Lleida. PhD Thesis.
- MCARTHUR, Tom. (1981) *Longman Lexicon of Contemporary English*. London: Longman.
- MEL'ČUK, Igor. (2012) "Phraseology in the language, in the dictionary, and in the computer." In: Kuiper, Koenraad (ed.) 2012. *Yearbook of Phraseology* Vol. 3. New York: De Gruyter Mouton, pp. 31-56.
- MELLADO BLANCO, Carmen. (2008) *Colocaciones y fraseología en los diccionarios*. Frankfurt am Main: Peter Lang.
- MOLINA PLAZA, Silvia. (2005) "English and Spanish phraseology in contrast." *RAEL: revista electrónica de lingüística aplicada*, pp. 174-189.
- MOON, Rosamund. (1998) *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. New York: Oxford University Press.
- MOON, Rosamund. (2008) "Dictionaries and collocation." In: Granger, Sylviane & Fanny Meunier (eds.) 2008. *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins, pp. 313-336.
- MOON, Rosamund. (2009) *Words, Grammar, Text: Revisiting the Work of John Sinclair*. Amsterdam: John Benjamins.
- NORRICK, Neal. (2007) "English Phraseology". In: Burger, Harald; Dmitriy Dobrovolskij & Peter Kühn (eds.) 2007. *Phraseology* Vol. II. New York: Walter de Gruyter, pp. 615-619.
- ORTEGA OJEDA, Gonzalo & María Isabel González Aguilar. (2008) "La técnica fraseográfica: el DRAE-2001 frente al DEA-1999." In: Mellado, Carmen (ed.) 2008. *Colocaciones y fraseología en los diccionarios*. Frankfurt am Main: Peter Lang, pp. 232-245.

- PAQUOT, Magali. (2015) "Lexicography and Phraseology." In: Biber, Douglas & Randi Reppen (eds.) 2015. *The Cambridge Handbook of Corpus Linguistics*. Electronic version: <[https://dial.uclouvain.be/pr/boreal/en/object/boreal%3A139795/datastream/PDF\\_01/view](https://dial.uclouvain.be/pr/boreal/en/object/boreal%3A139795/datastream/PDF_01/view)>
- PATIÑO, Pedro. (2017) *Description and representation in language resources of Spanish and English specialized collocations from Free Trade Agreements*. Bergen: NHH Norwegian School of Economics. PhD Thesis.
- PENADÉS, Inmaculada. (2006) "La información gramatical sobre la clasificación de las locuciones en los diccionarios." In: Alonso, Maragita (ed.) 2006. *Diccionarios y fraseología*. La Coruña: Universidade da Coruña. pp. 249-259.
- RAKOTOJOELIMARIA, Agathe. (2004) *Esbozo de un diccionario de locuciones verbales español-malgache*. Alcalá de Henares: Universidad de Alcalá. PhD Thesis.
- REAL ACADEMIA ESPAÑOLA. (n.d.) *Banco de datos CORDE*. (Corpus) Electronic version from Corpus diacrónico del español: <<http://www.rae.es>>
- ROJAS DÍAZ, José Luis & Juan Manuel Pérez Sánchez. (2019) "You Took the Word Out of My Mouth: a Morphosyntactic and Semantic Analysis of a Phraseological Lexicon of Colombian Spanish." In: Corpas Pastor, Gloria & Ruslav Mitkov (eds.) 2019. *Computational and Corpus-Based Phraseology*. Berlin: Springer, pp. 375-390.
- RUIZ GURILLO, Leonor. (1997) *Aspectos de la fraseología teórica española*. Valencia: Universidad de Valencia.
- RUIZ GURILLO, Leonor. (1998) *La fraseología del español coloquial*. Barcelona: Ariel.
- RUIZ GURILLO, Leonor. (2001) *Las locuciones en el español actual*. Madrid: Arco Libros.
- SAGER, Juan Carlos. (1990) *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.
- SANTAMARÍA PÉREZ, Isabel. (2003) "Localización de la información fraseológica en el diccionario: cuándo y cómo situarla en la macroestructura." In: Alemany Bay, Carmen; Beatriz Aracil Varón; Remedios Mataix Azuar; Pedro Mendiola Oñate; Eva Valero Juan & Abel Villaverde Pérez (eds.) 2003. *Con Alonso Zamora Vicente: Actas del Congreso Internacional "La Lengua, la Academia, lo Popular, los Clásicos, los Contemporáneos..."*. Alicante: Universidad de Alicante. pp. 1045-1057.

- SCHMID, Helmut. (1994) "Probabilistic Part-of-Speech Tagging Using Decision Trees". In: *Proceedings of the International Conference on New Methods in Language Processing*. Manchester.
- SECO, Manuel; Andrés Olimpia & Gabino Ramos (eds.) (1999) *Diccionario del español actual*. Madrid: Aguilar.
- SECO, Manuel; Andrés Olimpia & Gabino Ramos (eds.) (2004) *Diccionario fraseológico documentado del español actual*. Madrid: Aguilar.
- SHARMA, Sunil. (2018) "Happiness and metaphors: a perspective from Hindi phraseology." *Yearbook of Phraseology* 8:1, pp. 171-190.
- SINCLAIR, John. (1984) "Lexicography as an Academic Subject." In: Hartmann, Reinhard (ed.) 1984. *LEXeter '83 Proceedings*. Tübingen: Niemeyer. pp. 3-12.
- SINCLAIR, John. (2006) *Collins Cobuild Advanced Learner's English Dictionary*. Glasgow: HarperCollins.
- SINCLAIR, John & Rosamund Moon (eds.) (1989) *Collins COBUILD Dictionary of Phrasal Verbs*. London: Harper Collins.
- SINCLAIR, John & Rosamund Moon (eds.) (1997) *Collins COBUILD Dictionary of Idioms*. Glasgow: HarperCollins publishers.
- SKOLNÍKOVÁ, Pavlína. (2010) *Las colocaciones léxicas en el español actual*. Brno: Masarykova univerzita. PhD Thesis.
- SOSIŃSKI, Marcin. (2006) *Fraseología comparada del polaco y del español: su tratamiento en los diccionarios bilingües*. Granada: Universidad de Granada. PhD Thesis.
- TORIJANO, José Agustín, & María Ángeles Recio. (2019) "Translating Emotional Phraseology: A Case Study." In: Corpas Pastor, Gloria & Ruslav Mitkov (eds.) 2019. *Computational and Corpus-Based Phraseology*. Berlin: Springer, pp. 391-403.
- TSCHICHOLD, Cornelia. (2008) "A computational lexicography approach to phraseologisms." In: Granger, Sylviane & Fanny Meunier (eds.) 2008. *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins. pp. 361-376.
- ZULUAGA, A. (1980) *Introducción al estudio de las expresiones fijas*. Frankfurt: Peter Lang.

**BIONOTE / BIONOTA**

JOSÉ LUIS ROJAS DÍAZ is a PhD Scholar who joined NHH Norwegian School of Economics (Norway) in 2017. He holds a Master in Linguistics from the University of Antioquia (Colombia) (2014). His research focuses on the representation of phraseological units in Spanish and English dictionaries. His study involves dictionaries from general language and specialized fields. His major research interests cover phraseology, lexicography, linguistics, and translation.

JOSÉ LUIS ROJAS DÍAZ es un becario doctoral que en 2017 se unió a la *NHH Norwegian School of Economics* (Noruega). Es Magíster en Lingüística de la Universidad de Antioquia (Colombia) (2014). Su investigación se enfoca en la representación de unidades fraseológicas en diccionarios en inglés y español. Sus estudios incluyen diccionarios de lengua general y lenguajes especializados. Sus áreas de interés en investigación abarcan temas relacionados con la fraseología, la lexicografía, la lingüística y la traducción.

## Article 2

Rojas Díaz, J. L. 'Arm's length' phraseology? Building bridges from general language to specialized language phraseology – a study based on a specialized dictionary of International Commerce and Economics in Spanish and English (*submitted to Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication*).

## **'Arm's length' phraseology? Building bridges from general language to specialized language phraseology – a study based on a specialized dictionary of International Commerce and Economics in Spanish and English**

### **Abstract**

In the last decades, the study of phraseology within general and specialized lexicographic resources has been of interest to scholars. However, phraseology has not been studied in language for specific purposes (LSP) as much as in language for general purposes (LGP). Therefore, this study (i) offers an overview of the definitions regarding LSP phraseology, (ii) provides a series of linguistic analyses of specialized phraseological units (SPUs) extracted from a specialized bilingual dictionary, and (iii) draws a comparative line between LGP and LSP phraseology. To do so, 11,086 entries were extracted to build the analysis database. This study provides 1,054 morphosyntactic and 4,369 semantic patterns, a definition and a taxonomy of SPUs based on the data analysis and revision of LGP phraseology notions, and a hybrid lexicographic indexation method for SPUs. The contributions of this paper answer the question 'what is an SPU?'; while highlighting similarities and differences with LGP phraseology.

**Keywords:** phraseology, terminology, specialized lexicographic resources, commerce, economics, LSP, LGP



# **'Arm's length' phraseology? Building bridges from general language to specialized language phraseology – a study based on a specialized dictionary of International Commerce and Economics in Spanish and English**

## **1. Introduction**

In the last two decades, the study of phraseology within general and specialized lexicographic resources (e.g., dictionaries and databases) has been of interest to several scholars originating from diverse traditions and languages (Bevilacqua 2004; Alonso Ramos 2006; Leroyer 2006; Sosiński 2006; Heid 2008; Mellado Blanco 2008; Moon 2008; Siepmann 2008; Tschichold 2008; Mel'čuk 2012; Buendía Castro and Faber 2015; Paquot 2015; Rojas Díaz and Pérez Sánchez 2019; Nuccorini 2020; Veisbergs 2020). Regarding the definition of phraseology in language for general purposes (LGP), García-Page asserts that instead of defining the discipline itself, one must answer the question “what is the object of study of phraseology?” (2008, 7). In other words, what is a phraseological unit (PU)? Although, as pointed out by Bushnaq (2015, 173), the answers to this question tend to differ among phraseology scholars in terms of denominations (e.g., phraseologism, phraseme, idiom, collocation, etc.), a set of common characteristics of PUs could be found among the different definitions e.g., Corpas Pastor (1996, 6); Mellado Blanco (2004, 17); Gries (2008), namely (i) they are phrasal structures that follow syntax rules, (ii) they tend to be stereotypical (fixed in language by reiterative use), but not fully fixed, (iii) they allow the variation and the insertion of new elements, and (iv) they have figurative meanings.

Similarly to LGP phraseology, the definitions offered by scholars within language for specific purposes (LSP) and terminology e.g., (Picht 1987; Blais 1993; Pavel 1993; L'Homme 1998; Lorente Casafont 2002; Bevilacqua 2004) share (at least the first three of) the previously presented characteristics of PUs. However, in an attempt to detach its object of study from the one offered in LGP phraseology, these definitions clearly set out the occurrence of a terminological unit (i.e., a term) as part of the word forms of specialized phraseological units (SPUs).

Regarding the proximity between SPUs and multiword terms (MWTs) León Araúz and Cabezas-García (2020, 212) state that although some authors e.g., Zuluaga (1975); García-Page

(2008) do not consider multiword terms (MWTs) as phraseological units, they share some of the defining characteristics of PUs, namely the formation by two or more word forms, the frequent co-occurrence, the functioning as a whole, and a certain degree of lexicalization. However, similarly to previous works e.g., (Benson, Benson, and Ilson 1986; Pawley 2001; Ramisch 2015; León Araúz and Cabezas-García 2020), this study considers MWTs to be SPUs.

After revisiting the notions used to study LGP phraseology put forward by Mel'čuk (1998, 2012, 2013), this study answers the question "what is an SPU?", within the LSP domain of commerce and economics, by means of a new definition of SPUs and a taxonomy for their classification based on the occurrence (or absence) of terminological units within their word forms and an understudied aspect in SPU definitions, their level of idiomaticity. Thus, this study (i) presents different points of view and definitions regarding the study of LSP phraseology, (ii) offers a series of linguistic analyses traditionally used in the study of LSP and LGP phraseology (e.g., lexical and morphosyntactic) as well as a semantic analysis of a sample of SPUs extracted from a specialized bilingual (English-Spanish) dictionary, and (iii) draws a comparative line between phraseology in LGP and in LSP based on the analyzed data.

As regards the language pair analyzed in the present study, English and Spanish are the official languages of a significant number of institutions and organizations in Europe and the Americas (e.g., the Economic Commission for Latin America and the Caribbean [ECLA], the Inter-American Development Bank [IADB], the International Monetary Fund [IMF], the International Telecommunications Union [ITU], the World Trade Organization [WTO], the Organization of American States [OAS], the Council of Europe [CE], the United Nations [UN], among others) related to international commerce and economics. Being the official languages of these institutions makes them a suitable pair for descriptive data that could be used for future contrastive studies.

This paper is structured as follows: section 2 provides an overview of the denominations and definitions needed as a framework for this study. In section 3, data, tools, and methods are described. Section 4 presents the characterization of a sample of SPUs extracted from a bilingual

dictionary of international commerce along with an alternative definition and a taxonomy for the classification of SPUs. Section 5 offers a hybrid indexation method for SPUs in specialized lexicographic resources. And finally, in section 6, the conclusions resulting from this study are presented.

## 2. Phraseology within LGP and LSP

*Disentangling the Phraseological Web* is the title of an insightful work by Granger and Paquot (2008). This title is probably one of the most accurate ways of describing what the authors consider to be the two main factors preventing phraseology from establishing itself as a research field in its own right: (i) “the highly variable and wide-ranging scope of the field” and (ii) “the vast and confusing terminology associated with it” (2008, 27). Moreover, this terminological confusion has been pointed out repeatedly in phraseological studies over the last four decades (Zuluaga 1980, 31-37; Cowie 1998, 7; Ruiz Gurillo 2001, 44; Burger et al. 2007, 11; Norrick 2007, 615; Bushnaq 2015, 173; Rojas Díaz 2020, 6). Although there is still no consensus among scholars regarding the denomination, the types of units that are the object of study of phraseology have been roughly bound to two phraseological notions (narrow and wide) in the Spanish tradition. According to López (2012, 57), these two notions are based on the works by Corpas Pastor (1996) (wide notion) and Ruiz Gurillo (1997) (narrow notion). The narrow notion sets the most constrained and semantically opaque expressions, i.e., ‘idioms’, to be the object of study of phraseology. The wide notion includes both ‘idioms’ and ‘collocations’ as the objects of study of phraseology.

In this paper, the object of study of LGP phraseology will be denominated ‘phraseological unit’, while ‘idiom’, ‘collocation’, and ‘pragmateme’ are different types of PUs, and they will be defined according to Mel’čuk’s proposals (1998, 2012, 2013). Mel’čuk’s initial definition of a ‘non-free phrase’ sets the basis for the subsequent definition of its subcategories.

A phrase is non-free (= phraseologized) if at least one of its lexical components  $L_1$  is selected by the speaker in a linguistically constrained way – that is, as a function of the lexical identity of other component(s) (Mel’čuk 2012, 33).

Mel'čuk warns his readers a few lines later that: "(non)-constrain must be understood strictly in the technical sense indicated above as selection of a lexeme regardless of the individual identity of any other lexeme of the same expression" (Mel'čuk 2012, 33). From this, it could be deduced that 'idioms,' 'collocations,' and 'pragmatemes' are non-free phrases. When studying Mel'čuk's phraseological taxonomies, it is possible to notice a change from his 1998 work to his 2012 and 2013 works regarding the strict use of the principle of compositionality: "the meaning of a complex expression is determined by its structure and the meanings of its constituents" (Szabó 2020). In fact, Mel'čuk's later works (2012, 2013) are extensive and they include a ground-breaking theoretical contribution to the discipline as well as very detailed PU typologies (see Figure 1).

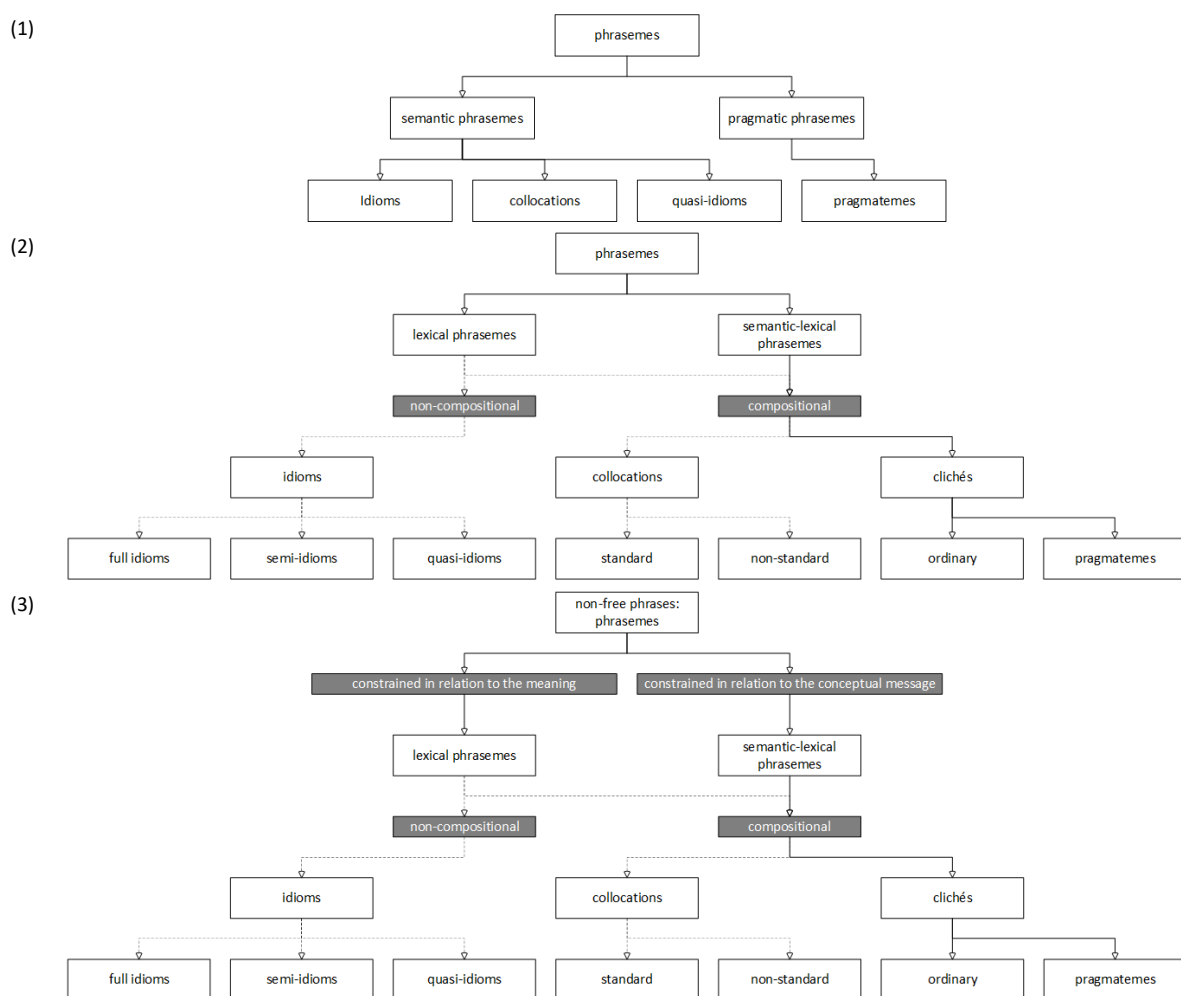


Figure 1. Mel'čuk's phraseme<sup>1</sup> typologies: (1) (1998, 30), (2) (2012, 42), and (3) (2013)

<sup>1</sup> Phraseme is used in the works by Mel'čuk as a synonym of phraseological unit.

However, the denominations used in his 2012 and 2013 taxonomies are extensive and (due to that) not quite suitable for the lexicographic indexation of PUs or SPUs. Therefore, in the present study, only three of the categories used in Mel'čuk's 1998, 2012, and 2013 taxonomies to group SPUs were chosen: (i) specialized idioms (covering only what Mel'čuk calls full idioms), (ii) specialized collocations (including both types of semi-idioms, quasi idioms, standard collocations, and non-standard collocations), and (iii) specialized pragmatemes. This classification aims to find a common ground for LGP phraseology and LSP phraseology by taking as a starting point the works by Mel'čuk (1998, 2012, 2013). Both the identification criteria of each subcategory and the working definition of SPU were adapted for this study (see section 4.4.).

Kjær states that phraseology is, without doubt, an "independent academic discipline within linguistics" (2007, 507). However, she also points out that LSP phraseology is an under-explored and a non-institutionalized line of research to the point of considering it a non-coherent research field (Kjær 2007, 507). Although it is possible to agree with her that LSP phraseology has not been studied for as long as its LGP counterpart, it is necessary to highlight the fact that scholars have been working extensively on this matter during the last twenty years (L'Homme and Bertrand 2000; Lorente Casafont 2002; Bevilacqua 2004; Aguado de Cea 2007; Fraile Vicente 2007; Kübler and Pecman 2012; Montero Martínez 2008; Buendía Castro and Faber 2015; Leroyer 2006; Hourani-Martín and Tabares-Plasencia 2020).

As within LGP phraseology, defining what an SPU is has been a prolific research topic in LSP phraseology. However, if definitions and denominations vary in LGP phraseology, this issue becomes even more problematic in LSP phraseology.

On the one hand, it is possible to find wide definitions regarding LSP phraseology, e.g., "Every entity worthy of interest and bigger than the standard terminological unit is called a phraseological

unit<sup>2</sup>” (Gouadec 1993, 550 [my translation]). On the other hand, some authors offer detailed information regarding the characteristics of their structure and word formants, e.g., Picht (1987, 151), Blais (1993, 550), Pavel (1993, 29), L'Homme and Meynard (1998, 515), Lorente Casafont (2002), Bevilacqua (2004, 28), Hourani-Martín and Tabares-Plasencia (2020, 115). The definitions and denominations of this later group vary across authors, but they have some characteristics in common, namely (i) they refer to phrases consisting of two or more elements, (ii) these phrases include a term as part of their lexical components (i.e., they are plurilexical), and (iii) they are used in LSP or acquire a specialized meaning when used in a certain LSP domain.

Regarding the object of study of LSP phraseology, one might say that the different denominations used (e.g., LSP phrase, phraseologism, LSP collocation, specialized lexical combinations, legal phraseological unit) intend to distinguish it from the object of study of LGP phraseology (especially within lexicography). Therefore, the question arises whether LSP phraseology should be denominated as such or whether another denomination should be used instead to name the study of phraseological units specifically in the context of specialized languages.

As mentioned above, ‘term’ and ‘LSP’ are recurrent when defining SPUs. Regarding the notion of term (terminological unit), L'Homme (2020) states that “there is no consensus about the notion of ‘term’” (55). An example could be drawn from the definitions offered by Cabré (2000) and Faber Benítez and López Rodríguez (2012). On the other hand, Cabré (2000) states that:

These units (terminological units/terms) are, at the same time, similar and different from the lexical units of a language, denominated words by lexicology. Their specialized character can be identified through their pragmatic aspects and the way of constructing their meaning. Their signified is the outcome of negotiation among experts. This negotiation happens within the specialized discourse through the use that determines the meaning of each unit” (Cabré 2000, 14 [my translation])<sup>3</sup>.

On the other hand, Faber Benítez and López Rodríguez (2012) assert that:

---

<sup>2</sup> Original in French “*Toute entité digne d'intérêt et plus grande que l'unité terminologique standard est dite unité phraséologique.*”

<sup>3</sup> Original in French “*Ces unités sont en même temps semblables et différentes des unités lexicales d'une langue, appelées mots par la lexicologie. Leur spécificité se trouve dans leur aspect pragmatique et dans leur mode de signification. Leur signifié est le résultat d'une négociation entre experts. Cette négociation se produit dans le discours spécialisé à travers des prédications qui déterminent le signifié de chaque unité.*”

Trying to find a distinction between terms and words is no longer fruitful or even viable, and the best way to study specialized-knowledge units is by studying their behavior in texts. (22)

The difference between the two previously presented notions of term, reside in the approach of each of these theoretical postulations regarding the object of study of terminology, rather than on the object of study itself. Furthermore, Faber Benítez and López Rodríguez (2012, 22-23) explicitly mention characteristics that their definition share with Cabré's (2000) notion of term, such as the predominance of nominal units, the relations between the TUs and LSP domains, and the combinatorial value of TUs.

A middle point is addressed by L'Homme's (2020) notion of term when she asserts that:

Stating that a linguistic item is a term is considering its meaning from the perspective of a special subject field. There is no such a thing as a term in essence; a linguistic unit becomes a term relative to their subject field in which it is considered. (...) This also means that even common linguistic items can become terms in specialized domains. (...) Finally, a linguistic item can also be a relevant term in fields of knowledge. (59)

This later definition offered in the work by L'Homme (2020, 59) allows for the classification of linguistic items according to their meaning into a certain subject field. Hence, this characteristic (see section 4.2.) could be used for the classification of word forms into semantic fields related to a certain LSP domain (commerce and economics for this study).

Concerning the definition of LSP, Hoffmann asserts that:

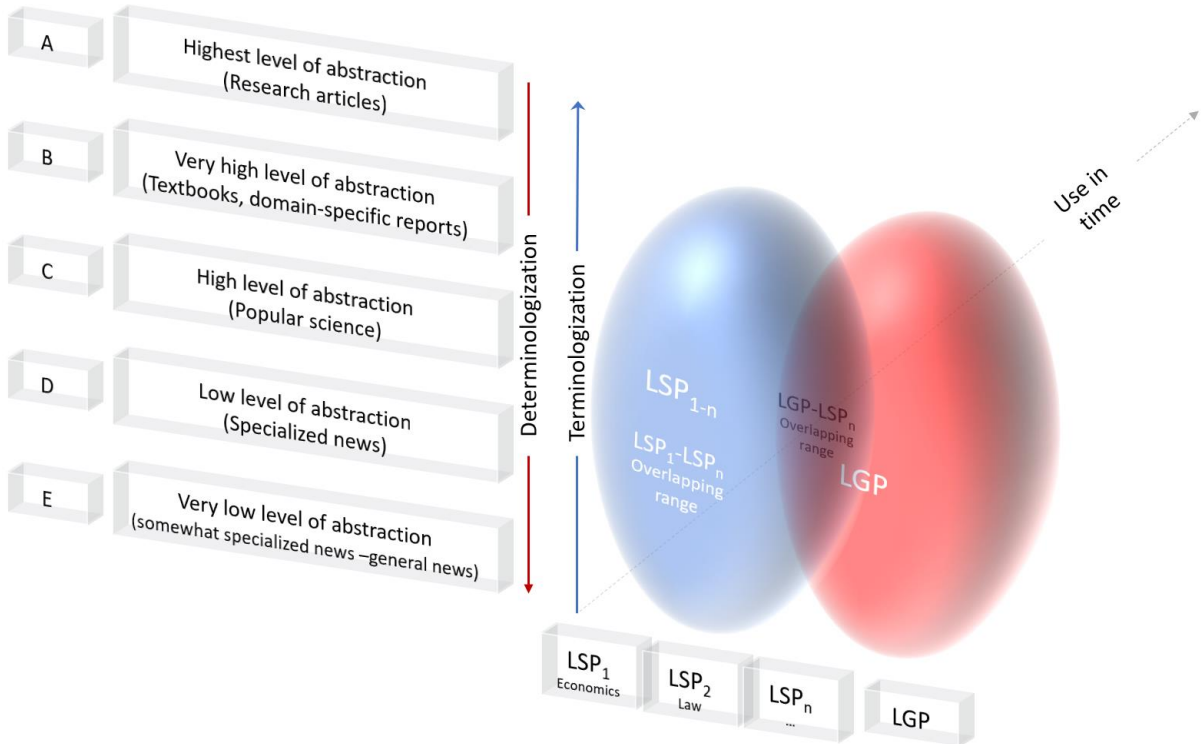
A specialized language (LSP) is the group of all the linguistic resources that are used in a communication field –delimited by the specialized discipline– to ensure the understanding among the people that work in that field (Hoffmann 1998, 57 [my translation])<sup>4</sup>.

Since this study concerns LSP phraseology within commerce and economics (See section 1), Hoffmann's definition of LSP raises the question how can 'language of economics' be defined? Regarding this question, Simonnæs and Kristiansen state that "it may be difficult to clearly delineate what constitutes 'economic' language" (2018, 152). The authors suggest that this difficulty derives from the fact that the term 'economic' has been "frequently used with reference to a number of

---

<sup>4</sup> Translation in Catalan "Un llenguatge d'especialitat és el conjunt de tots els recursos lingüístics que s'utilitzen en un àmbit comunicatiu — delimitable pel que fa a l'especialitat— per tal de garantir la comprensió entre les persones que treballen en aquest àmbit". Original in German "Fachsprache – das ist die Gesamtheit aller sprachlichen Mittel, die in einem fachlich begrenzbaren Kommunikationsbereich verwendet werden, um die Verständigung zwischen den in diesem Bereich tätigen Menschen zu gewährleisten."

domains and subdomains” (Simonnæs and Kristiansen 2018, 152). Furthermore, they offer a series of examples of these domains, which include, among others, ‘individuals,’ ‘businesses,’ ‘markets,’ ‘monetary issues,’ and ‘global political issues’ (Simonnæs and Kristiansen 2018, 152-153). Furthermore, these authors [just like Kristiansen and Andersen (2012, 45)] highlight the well-known overlapping interaction between LGP and LSP. They explain this phenomenon as a continuum of terminologization and determinologization.



**Figure 2.** Illustration of the overlapping range based on the terminological overlapping notion extracted from the works by Hoffmann (1998, 65), Kristiansen and Andersen (2012, 45), and Simonnæs and Kristiansen (2018, 160)

The overlap proposed initially by Kristiansen and Andersen (2012, 45), and then expanded in the work by Simonnæs and Kristiansen (2018, 160) occurs in both axes of Hoffmann’s division of languages, namely the horizontal one (regarding the differentiation of LSP domains), and the vertical one (related to language register or level of specialty). Figure 2 illustrates this double continuum that generates an overlapping range in which a terminological unit may migrate from one LSP to another or from LGP to a specific LSP at a different level of specialty. It is important to point out that this migration could also be part of the behavior of both PUs and SPUs.



An understudied aspect in SPU definitions is the semantic idiomaticity (as well as other levels of idiomaticity explained in section 4.4.). In fact, the lack of explicit details regarding the semantic opacity of SPUs, which could be considered one of the main characteristics of PUs, brings up the question of whether SPUs lack this characteristic or whether lexicographic resources are not registering —or overlooking— those units.

Vangehuchten states that a terminological study would not be complete until both semasiological and onomasiological descriptive lexicographic approaches are adopted (2005, 148). Semasiology and onomasiology are linked to lexicographic methods of indexation. Generally, the repertoire of dictionaries and databases available for consultation has been crafted by using a semasiological approach. Semasiology is defined by Hartmann and James as “an approach in semantics concerned with the explanation of the meaning of given words or phrases” (1998, 124). This approach requires the users to know the form of the expression or lexical unit they are looking for (Kocjančič 2004), i.e., semasiological dictionaries answer the question ‘what are the possible meanings of X (word/phrase/idiom/proverb)?’

Nonetheless, the semasiological approach poses a series of problems depending on the kind of user who consults lexicographic resources. On the one hand, for many novel users (e.g., language learners), the problem has to do with the headwords used to index a PU or an SPU in lexicographic resources. On some occasions, the headwords are not very intuitive, or the dictionary does not include the necessary information in the user’s guidelines as for how to look for those headwords (Pawley 2001, 130). On the other hand, advanced users (e.g., translators and other linguistic mediators) may not find a semasiological dictionary useful if their query is ‘how can I express Y (concept) in a certain language?’ Such query, based on concepts, would require an onomasiological approach, as is common in terminological databases. The definition of onomasiology given by Hartmann and James states that “it is an approach in semantics which is concerned with matching the most appropriate word or phrase to a given concept” (1998, 102).

Vangehuchten's statement (see above) regarding the use of semasiological and onomasiological approaches in studies within terminology, and the semantic analysis presented in this paper (section 4.2) were the key to developing an alternative, hybrid (semasiological/onomasiological) indexation method for SPUs in specialized lexicographic resources in the current study (see section 5).

### **3. Data, tools, and methods**

The motivation for this study derives from previous studies on lexicographic resources in general language (Rojas Díaz 2020; Rojas Díaz and Pérez Sánchez 2020). Those studies offered extensive morphosyntactic and semantic information regarding patterns based on part-of-speech (POS) tags along with detailed information concerning the semantic fields into which each of the word forms of the PUs could be categorized. The semantic and morphosyntactic patterns that emerged from the POS-tagging and the semantic annotation were used to identify metaphors and metonymies, as well as patterns for querying the analysis database. However, these studies did not include any specialized lexicographic resources.

#### **3.1. Dictionary selection criteria and lexicographic information**

As noted by Kübler and Pecman (2012, 187), globalization and standardization processes have had an impact on the need for LSP lexicographic resources that can both standardize and describe specific domains by offering definitions. Undoubtedly, commerce and economics have played a leading role in those globalization processes. Moreover, these LSP domains offer an interesting case of terminological overlap, since, as stated by Simonnæs and Kristiansen, "business, finance and economics are in many cases intertwined with law" (2018, 157). Furthermore, the use of lexicographic resources as a source for the creation of the analysis database ensures the presence of terms from several levels of abstraction as proposed by Hoffmann (1998, 72-73) and aimed at different users' needs (Kübler and Pecman 2012, 187).

In consequence, the lexicographic resource for this study matches the following criteria: (i) that it is related to commerce and economics, (ii) that it is bilingual (in English and Spanish), and (iii) that its publishing house is recognized as a lexicographic authority. The *Diccionario de Comercio Internacional* (DCI) (Alcaraz Varó and Castro Calvín 2007) meets all these criteria (see Tables 1 and 2). On an editorial note, the DCI is presented as a dictionary aimed at various users including “field experts and scholars from diverse areas of Economics, International Commerce, and linguistic mediators” (Alcaraz Varó and Castro Calvín 2007). Besides, the DCI could be categorized as a descriptive, semasiological, and synchronic dictionary. The characterization of the DCI is presented in Table 1.

**Table 1.** *General lexicographic information of the DCI*

<b>Lexicographic information</b>	
<b>Field</b>	<b>Information</b>
Editors	Enrique Alcaraz Varó & José Castro Calvín
Entries	17,000 approximately (according to the dictionary)
Format	Physical printing (16 x 25 cm)
Languages	English-Spanish / Spanish-English
Pages	1,168
Publishing house	Ariel

The DCI, in contrast to other dictionaries of the same publishing house, does not include a user’s guide, and the marking in the dictionary was reduced to a series of labels related to the sub-fields of the LSP domain of commerce and economics with examples of the entries or sub-entries categorized under those labels (see Table 2).

**Table 2.** *Sub-field-of-specialty labels, and examples of entries linked to them in the DCI*

Thematic / Specialty sub-domain label (diatechnical marking)			
Label	Meaning	Entry example (Spanish)	Entry example (English)
CONTA/ACCTG	contabilidad/accounting	<i>moneda de cuenta</i>	<i>accounting currency</i>
DER/LAW	derecho/law	<i>normativa reguladora de las exportaciones</i>	<i>export regulations</i>
DOC/DOCMT	documentación/documentation	<i>conocimiento de embarque</i>	<i>bill of lading</i>
ECON/ECON	economía/economics	<i>sumarse a un bloqueo</i>	<i>adhere to a boycott</i>
FIN/FIN	finanzas/finance	<i>adelanto del vencimiento</i>	<i>acceleration of maturity</i>
GRAL/GEN	general/general	<i>compra sobre muestra</i>	<i>purchase by sample</i>
FISC/TAXN	fiscalidad/taxation	<i>arancel aduanero</i>	<i>customs duty</i>
REC.HUM./HH.RR.	recursos humanos/human resources	<i>formación sobre asuntos conflictivos</i>	<i>sensitivity training</i>
LOGÍST/LOGIST	logística/logistics	<i>hora real de llegada</i>	<i>actual time of arrival</i>
MERCAD/MKTG	mercadotecnia/marketing	<i>compra al por mayor</i>	<i>bulk procurement</i>
ORGAN/ORGAN	organización/organization	<i>Instituto Monetario Europeo</i>	<i>European Monetary Institute</i>
SEGUR/INSCE	seguros/insurance	<i>póliza vigente</i>	<i>unexpired policy</i>

Although it is explicitly stated in the DCI that it is aimed at being used by linguistic mediators, there are no user guidelines that show how to interpret the linguistic marking. Specifically, it was impossible to accurately determine the meaning of eight of these marks: (i) ‘col,’ (ii) ‘coll,’ (iii) ‘coloq,’ (iv) ‘colloq,’ (v) ‘Exp,’ (vi) ‘fr,’ (vii) ‘phr,’ and (viii) ‘phrase’ (see Table 3). When analyzing them in context, the marks ‘Exp’ and ‘fr’ were used to identify a certain type of phraseological marking. However, some marks did not make any sense, such as ‘phrase’ and ‘fr’ used to identify acronyms. Furthermore, it is uncertain whether marks like ‘col,’ ‘coll,’ ‘coloq,’ indicate the register of a certain expression (diaphasic marking such as ‘colloquial’) or if they make reference to a collocation (see Table 5).

**Table 3.** Undetermined linguistic labels in the DCI: (i) col, (ii) coll, (iii) coloq, (iv) colloq, (v) Exp, (vi) fr, (vii) phr, and (viii) phrase (Alcaraz Varó and Castro Calvín 2007)

Undetermined linguistic labels in the DCI	
Label example	
(i) <b>smurf money</b> <i>col fr</i> : FIN atomizar dinero. [Exp: <b>smurfing</b> ( <i>col</i> FIN atomización de dinero; alude a la transferencia –transfer– de muchas	(ii) <b>seven sisters</b> <i>n</i> : GRAL/ECON <i>coll</i> . las siete multinacionales más importantes; también llamadas <i>major</i> <i>multinationals</i> .
(iii) V. Classification Society), <b>Lloyd's Registry, LR</b> (LOGÍST <i>coloq</i> . registro de buques mercantes de Lloyd's; es una sociedad clasifi-	(iv) <i>ation, rules of origin</i> ), <b>Harmonized System, HS</b> (FISC <i>colloq</i> Sistema Armonizado de Codificación y Descripción; se refiere al sistema
(v) <b>ábrase aquí</b> <i>fr</i> : LOGIST open here; S. <i>abrir aquí</i> ; <i>instrucciones de embarque</i> . <b>abreviado</b> <i>a</i> : GEN/LAW/TAXN informal; S. <i>simpli-</i>	(vi) <b> echar</b> <i>v</i> : GEN throw, throw away/out, expel. [Exp: <b> echar al correo</b> (GEN mail, post; S. <i>enviar/remitir por correo</i> ), <b> echar amarras</b>

(vii) **caballo, a** *phr*: GEN piggyback; S. *a cuestras*.  
**cabeza** *n*: GEN head. [Exp: **cabeza, que va a la**  
(GEN leading; S. *destacado, importante, líder*;

(viii) **NCR** *phrase*: V. *no carbon required*.  
**NCV** *fr*: V. *no commercial value, no customs*  
*value*.

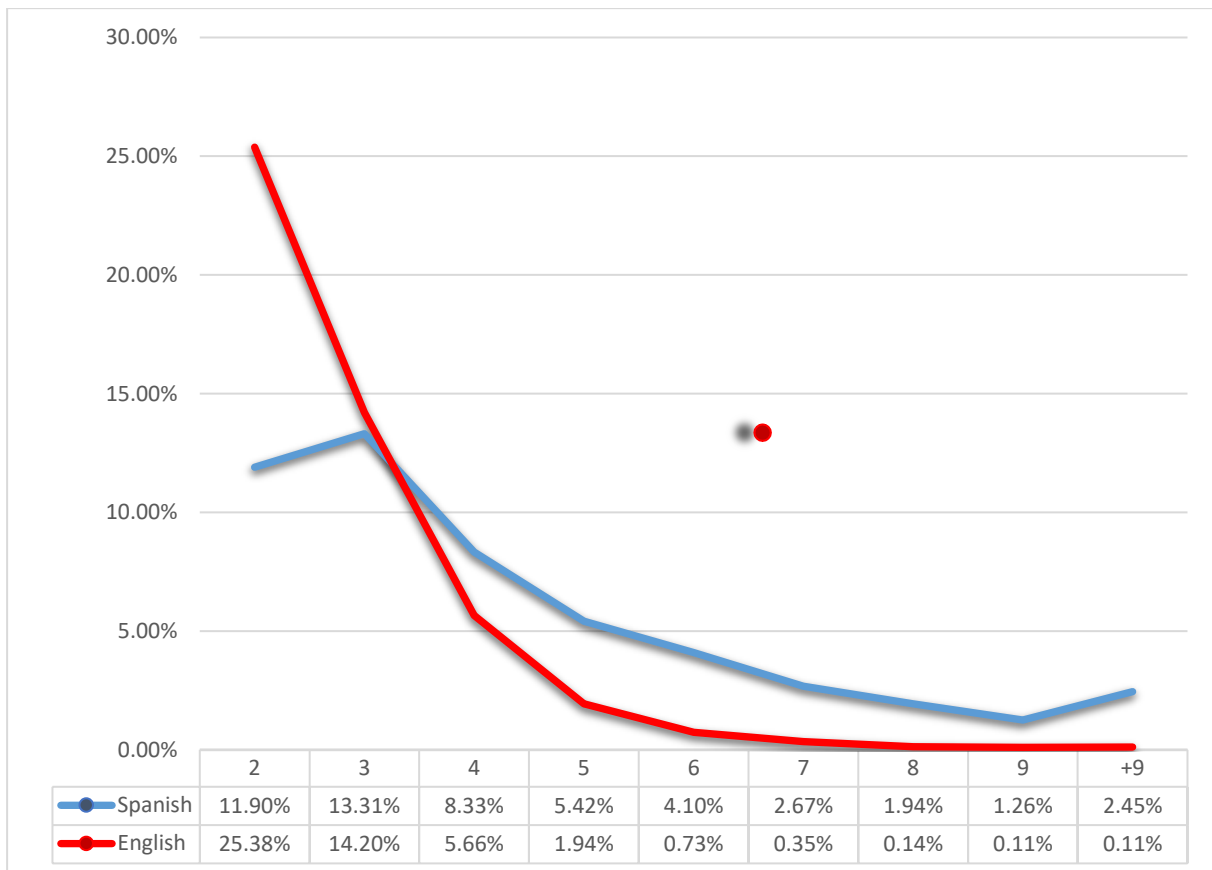
---

Marking could have been used as a criterion for the extraction of SPUs. However, as presented in Table 3, when the marking is not consistent, it does not allow the user to understand a given expression's behavior in context or its typology. Thus, marking was not used as a criterion for the selection of the entries of the analysis sample.

### 3.2. Analysis database creation and sample selection

In the present study, several selection criteria and tagging techniques (e.g., No. of word forms, POS-tagging, semantic annotation, among others) from the work by Rojas Díaz (2020, 301-303) were implemented. Nevertheless, this study intends to identify and describe the full phraseological repertoire of the sample. The first criterion for constructing the sample database was the number of word forms of the dictionary entries.

All the entries and sub-entries from the DCI were digitalized. Since PUs and SPUs are plurilexical units, expressions with more than two-word forms were extracted. The resulting extraction offered 22,773 plurilexical entries and sub-entries (11,702 in Spanish and 11,071 in English) from the DCI (see Figure 3).



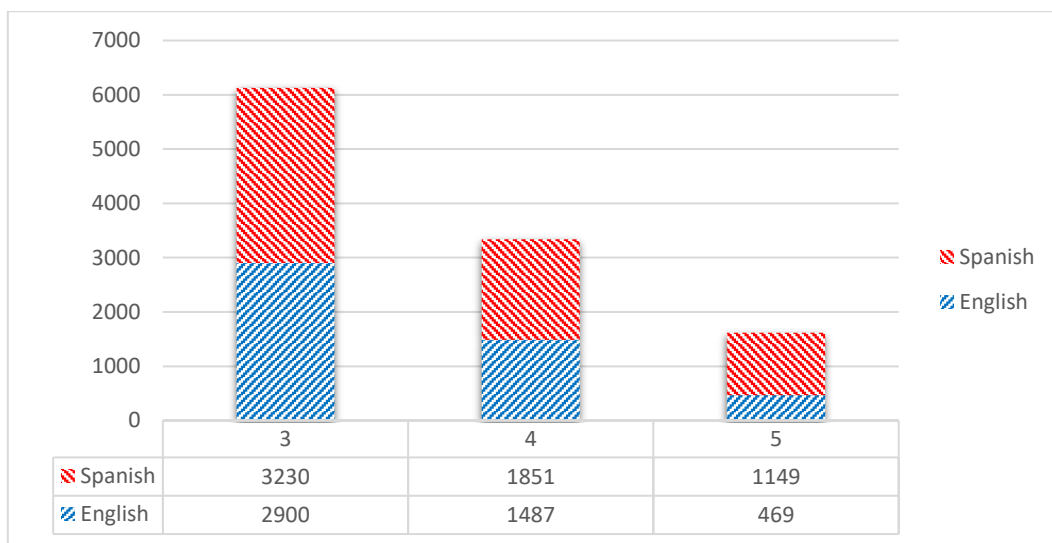
**Figure 3.** Distribution of entries according to their number of word forms in Spanish and English in the DCI

For capacity reasons, it was decided to reduce the number of units selected for the sample analysis considering the analyses that were going to be carried out (including semantic annotation and identification of metaphors and metonymies).

The results, presented in Figure 3, show a difference of more than twice as many two-word-form expressions in English (25.38% of the DCI entries) than in Spanish (11.90% of the DCI entries). A previous study (Rojas Díaz 2020, 302-304), shows not only that the correlation of word forms and occurrence in Spanish in the DCI was similar to those found in the *Diccionario Freseológico Documentado del Español Actual* (DFDEA) (Seco, Andrés, and Ramos 2004) and the Collins COBUILD Dictionary of Idioms (CCDOI) (Sinclair and Moon 1997), but that more than 50% percent of the idioms (67.8% in Spanish and 76.7% in English) in the DFDEA and the CCDOI occurred in the range of three, four, or five-word forms.

The only group that alters that frequency distribution is the one comprising phraseological bi-grams. One possible explanation for this might be related to an editorial decision to avoid these kinds of units due to the difficulty in drawing a boundary between compounds and PUs. [...] However, reaching a satisfactory conclusion on the reasons behind this frequency distribution would only be possible through a deeper analysis of this group of units. (Rojas Díaz 2020, 303)

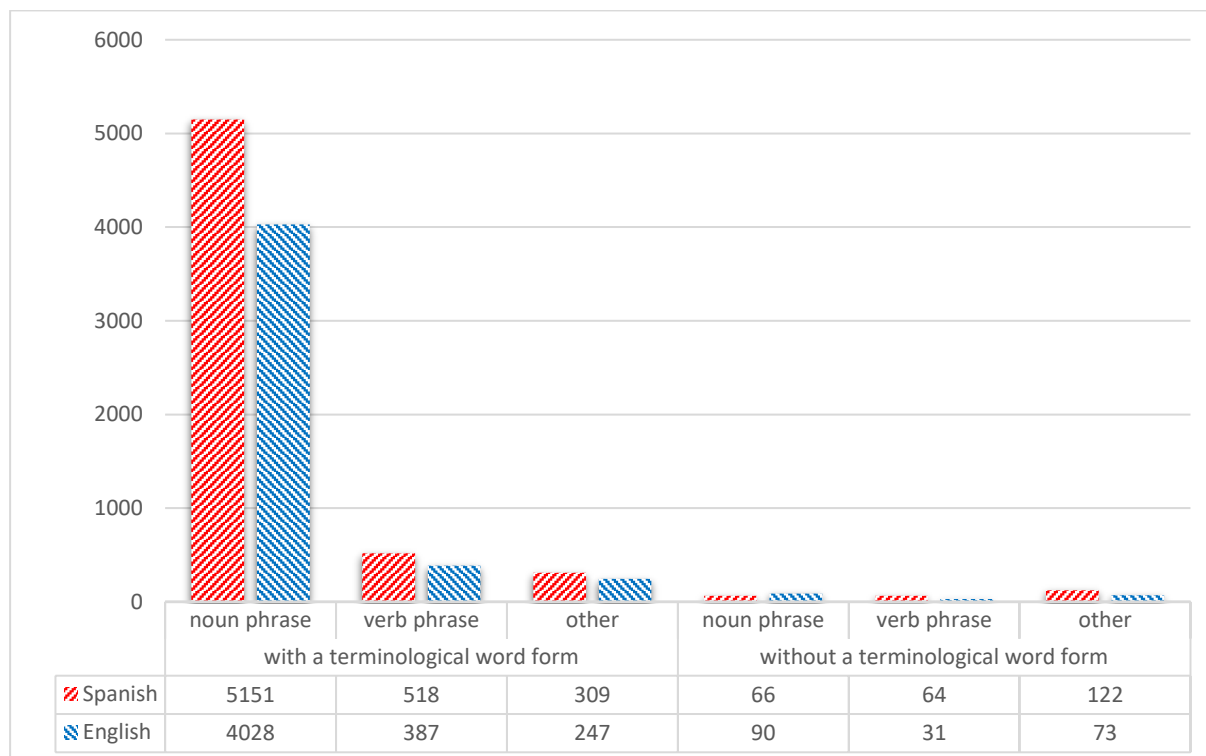
Therefore, and similarly to the work by Rojas Díaz (2020), the database in this study only includes entries and sub-entries extracted from the DCI consisting of three, four, and five-word forms. The distribution of the sample is presented in Figure 4. 11,086 (48.86% of the DCI) entries and sub-entries, constituted by 39,832 word forms (including commas and hyphens) were extracted to build the database.



**Figure 4.** Distribution of three, four, and five-word-form expressions extracted from the DCI to create the database

Authors such as Sager (1990, 58) and L'Homme (2004) assert that nouns are predominant in concept representation in specialized dictionaries. When plotting the frequency distribution of the extracted expressions according to their POS in the database, the results supported these statements (see Figure 5). Nevertheless, a marginal yet interesting finding appears in the data when combining the functional-mark information with the presence or absence of a term in the component word forms. Contrary to what several scholars have asserted (see section 2), not all extracted expressions contained a monolexical term (within their word forms) as the head of the phrase or as an adjacent to the head of the phrase. Therefore, it is possible to say that (i) either PUs migrate from LGP to LSP in what could be called an overlapping range as graphed in Figure 2, or (ii) there are combinations of

words that become terminological (MWTs) when used in a certain LSP domain. Since the DCI does not offer the POS of the expressions, it was necessary to classify them manually.



**Figure 5.** Frequency of phrases according to their POS and occurrence of a monolexical term within their word forms in the database

After completing the tagging of the database according to their POS, the next step in the study was to start the linguistic analyses and queries to the database to see what lexical, semantic, and morphosyntactic information could be extracted to characterize the SPUs of the sample.

#### 4. Analyses and results

Four different analyses were carried out (i) lexical, (ii) semantic, (iii) morphosyntactic, and (iv) phraseological. These analyses will be explained in detail in this section.

##### 4.1. Lexical analysis

After extracting dictionary entries with three, four, and five-word forms, all the lexical and grammatical word components underwent two tagging processes, namely POS and semantic annotation. First, a POS tagging was done by means of *TreeTagger* (Schmid 1994) (see Table 6).

**Table 4.** Distribution of word forms by POS in the database



POS		Frequency in Spanish		Frequency in English	
Noun	<b>(N)</b>	11,296	(49.50%)	9,558	(57.84%)
Preposition	<b>(Prep)</b>	5,559	(24.36%)	2,158	(13.06%)
Adjective	<b>(Adj)</b>	2,673	(11.71%)	2,185	(13.22%)
Determiner	<b>(Det)</b>	1,156	(5.07%)	703	(4.25%)
Verb	<b>(V)</b>	752	(3.30%)	630	(3.81%)
Past participle	<b>(PP)</b>	550	(2.41%)	444	(2.69%)
Adverb	<b>(Adv)</b>	257	(1.13%)	163	(0.99%)
Present participle	<b>(PrP)</b>	5	(0.02%)	413	(2.50%)
Contraction	<b>(Contr)</b>	377	(1.65%)	0	(0.00%)
Conjunction	<b>(Conj)</b>	152	(0.67%)	146	(0.88%)
Prefix	<b>(Pref)</b>	7	(0.03%)	58	(0.35%)
Saxon Possessive	<b>(SaxP)</b>	0	(0.00%)	56	(0.34%)
Pronoun	<b>(Pron)</b>	30	(0.13%)	11	(0.07%)
Demonstrative	<b>(Dem)</b>	2	(0.01%)	0	(0.00%)
Interjection	<b>(Interj)</b>	2	(0.01%)	0	(0.00%)
<b>Grand Total</b>		<b>22,818</b>		<b>16,525</b>	

39,832 word forms in the database were tagged. Some morphemes and punctuation marks were part of the constituent elements of the DCI entries; for instance, English possessive morphemes (e.g., *at arm's length*<sup>5</sup>), commas (e.g., *crystal, manipúlese con cuidado*), and hyphens (e.g., *glass-with care | acuerdo stand-by*). These elements were tokenized and counted as word forms.

Not surprisingly, among the lexical words in the database, nouns are the most frequent (20,854 occurrences [53.01%]) followed by adjectives (4,858 occurrences [12.35%]) and verbs (1,382 occurrences [3.51%]). That information matches the predominance of nominal SPUs shown in Figure 5.

After the POS-tagging was carried out, the data was plotted by means of word clouds offering a graphic overview of the word forms with font size signifying frequency weight. This word cloud generator<sup>6</sup> allows for the inclusion of a complete list of single constituent words of the analysis sample (see Table 7).

<sup>5</sup> Italics are used when mentioning or citing a certain PU or SPU in the running text or a word form belonging to a certain PU or SPU. Single quotes are used in the captions of tables and figures to distinguish PUs from SPUs in the text.

<sup>6</sup> The word cloud generator is available online at <https://www.wordclouds.com/>.



## 4.2. Semantic analysis

This study adopts a semantic approach similar to the one proposed by Rojas Díaz (2020, 108-112) to analyze the word forms of the database. Among the possibilities, the UCREL's Semantic Analysis System (USAS) was chosen for the semantic annotation of the word forms of the database. USAS is a POS and semantic tagger that allows for the classification of word forms into word classes using 453 semantic tags classified in 232 semantic categories based on 21 discourse fields identified by McArthur (1981) and (Archer, Wilson, and Rayson 2002, 2).

All the word forms of the database were revised and corrected manually in both languages. The information obtained from the word clouds (see Table 5) indicated that there is a frequent occurrence of word forms related to commerce and economics. However, the semantic annotation classified the word forms in several semantic and discourse fields. The data showed that the largest category was related to 'money and commerce in industry' (23.59%), followed by 'general and abstract terms' (16.10%); mostly associated with supporting words (e.g., 'be' | 'ser', 'do' | 'hacer', 'have' | 'tener'), and 'movement, location, travel, and transport' (15.57%) (see Table 6).

*Table 6. Distribution of discourse-field tags in the databases*

Discourse field (example)	Tag	Frequency in Spanish	Frequency in English
Money and commerce in industry ( <i>exportación, pagar</i>   <i>trade, pay</i> )	(I)	3,873 (24.93%)	3,008 (22.46%)
General and abstract terms ( <i>acuerdo, ser</i>   <i>agreement, be</i> )	(A)	2,686 (17.29%)	1,976 (14.75%)
Movement, location, travel, and transport ( <i>transporte, poner</i>   <i>freight, enter</i> )	(M)	2,430 (15.64%)	2,078 (15.52%)
Social actions, states, and processes ( <i>condición, aceptar</i>   <i>exchange, comply</i> )	(S)	1,555 (10.01%)	1,546 (11.54%)
Language and communication ( <i>certificado, visado</i>   <i>certificate, request</i> )	(Q)	1,152 (7.42%)	973 (7.26%)
Numbers and measurement ( <i>carga, reducir</i>   <i>rate, spread</i> )	(N)	1,140 (7.34%)	949 (7.09%)
Substances, materials, objects, and equipment ( <i>contenedor, congelar</i>   <i>goods, fade</i> )	(O)	426 (2.74%)	804 (6.00%)
Psychological actions, states, and processes ( <i>conocimiento, cumplir</i>   <i>system, fulfil</i> )	(X)	515 (3.32%)	501 (3.74%)
Time ( <i>plazo, anular</i>   <i>date, terminate</i> )	(T)	543 (3.50%)	405 (3.02%)
Government and public ( <i>derecho, violar</i>   <i>licence, smuggle</i> )	(G)	425 (2.74%)	275 (2.05%)
Names and grammar ( <i>Londres</i>   <i>London</i> )	(Z)	319 (2.05%)	353 (2.64%)
Architecture, housing, and the home ( <i>almacén, inmobiliario</i>   <i>warehouse, build</i> )	(H)	185 (1.19%)	222 (1.66%)
World and environment ( <i>país, mundial</i>   <i>country, global</i> )	(W)	128 (0.82%)	123 (0.92%)
Emotion ( <i>preferencia, satisfacer</i>   <i>relief, satisfy</i> )	(E)	45 (0.29%)	67 (0.50%)
The body and the individual ( <i>mano, sanear</i>   <i>health, sanitary</i> )	(B)	51 (0.33%)	51 (0.38%)
Life and living things ( <i>vida, muerto</i>   <i>life, dead</i> )	(L)	21 (0.14%)	19 (0.14%)
Science and technology ( <i>motor, electrónico</i>   <i>technology, automated</i> )	(S)	13 (0.08%)	19 (0.14%)
Education ( <i>curso, cursar</i>   <i>assignment, test</i> )	(P)	9 (0.06%)	9 (0.07%)
Food and farming ( <i>faena, fitosanitario</i>   <i>meat, phytosanitary</i> )	(F)	9 (0.06%)	6 (0.04%)
Entertainment, sports, and games ( <i>feria, juego</i>   <i>competition, entertain</i> )	(K)	5 (0.03%)	6 (0.04%)
Arts and crafts ( <i>cultura, cultural</i>   <i>culture, cultural</i> )	(C)	3 (0.02%)	3 (0.02%)

The annotation showed the occurrence of certain discourse fields not primarily related to economics e.g., emotions (E), the body and the individual (B), entertainment, sports, and games (K), etc., suggesting idiomaticity through the occurrence of metaphors and metonymies (see Table 7). Therefore, the semantic annotation could be used as one of the classification criteria of the SPUs (see Figure 6).

**Table 7.** Examples of metaphorical and metonymical relationships between SPUs' word forms and their meaning

Spanish		English	
SPU	Definition/Meaning	SPU	Definition/Meaning
<i>venta en firme</i> (Metaphorical)	firm sale, outright sale	<i>advice of shipment</i> (Metaphorical)	<i>es un documento, que normalmente lleva anexo un ejemplar de la factura y otro del conocimiento de embarque, mediante el cual se notifica al comprador que se ha realizado el embarque de las mercancías</i>
<i>estar en libre práctica</i> (Metonymical)	be in free circulation	<i>call a meeting</i> (Metonymical)	<i>Convocar una reunión; equivale a convene a meeting</i>

Although, at first sight, most of the SPUs extracted from the dictionary seemed to be compositional, a further examination of the entries showed the occurrence of semantic idiomaticity among the SPUs' word forms through metaphors (1,284 [11.59%], e.g., *bolsa de fletes* | *basket of rates*) and metonymies (702 [6.33%], e.g., *de domicilio a muelle* | *door to port*). Furthermore, it was possible to identify some SPUs with word forms containing both metaphors and metonymies (132 [1.19%], e.g.,  *echar mano de* | *at arm's length*). The identification of the semantic relationships was done manually, and the results are seen in Table 8.

**Table 8.** Frequency of semantic relationships in each language

Semantic relationship	Frequency		
	Spanish	English	Total
None (compositional or other idiomaticity)	5,065 (81.30%)	3,902 (80.35%)	<b>8,968</b>
Metaphor	667 (10.70%)	618 (12.73%)	<b>1,284</b>
Metonymy	406 (6.52%)	296 (6.10%)	<b>702</b>
Metonymy / Metaphor	92 (1.48%)	40 (0.82%)	<b>132</b>
<b>Grand total</b>	<b>6,230</b>	<b>4,856</b>	<b>11,086</b>

### 4.3. Morphosyntactic analysis

Morphosyntactic information has been widely used for the identification of monolexical terms, and multiword terms (EAGLES 1999; Drouin 2003; Drouin, Morel, and L'Homme 2020). Linguistic approaches by queries of productive patterns (Justeson and Katz 1995, 16-17) and statistical approaches by means of mutual information, log-likelihood coefficient, among other association ratio scores (Dunning 1993; Church and Hanks 1989) use morphosyntactic patterns or POS-tagging information from word forms to extract MWTs (EAGLES 1999, 176-178).

When grouping the POS-tagging information of the word forms, 1,054 morphosyntactic patterns were extracted, of which 538 (51.04%) occurred twice or more in the database. Tables 9 and 10 present the top five most frequent patterns for noun and verb phrases in the dictionary in Spanish and English.

**Table 9.** Top five morphosyntactic patterns of noun and verb SPUs in Spanish in the DCI

Morphosyntactic patterns in Spanish				
Noun / Verb	No. of word forms	Morphosyntactic pattern	Freq.	SPU Example
Noun	3	N Prep N	1,872	<i>ánimo de lucro</i>
		N Adj Adj	251	<i>bien económico básico</i>
		N Contr N	137	<i>compra al contado</i>
		N Adv Adj	44	<i>día no laborable</i>
		N Prep Adj	38	<i>endoso en blanco</i>
	4	N Prep N Adj	452	<i>fianza de avería gruesa</i>
		N Prep Det N	370	<i>gestor de un consorcio</i>
		N Adj Prep N	251	<i>hora real de llegada</i>
		N Prep N PP	98	<i>letra de cambio mutilada</i>
		N Prep Adj N	78	<i>negociaciones a tres bandas</i>
	5	N Prep N Prep N	292	<i>operación de divisas a plazo</i>
		N Prep Det N Adj	96	<i>precio en el mercado global</i>
		N Adj Prep Det N	93	<i>reservas oficiales de un país</i>
		N Adj Prep N Adj	45	<i>título pagadero en divisa extranjera</i>
		N Prep N Contr N	43	<i>daños por agua del mar</i>
Verb	3	V Det N	229	<i>burlar la ley</i>
		V Prep N	72	<i>continuar en vigor</i>
		V N Prep	18	<i>echar mano de</i>
		V Contr N	11	<i>dedicarse al contrabando</i>
		V N Adj	7	<i>instituir barreras comerciales</i>
	4	V Prep Det N	37	<i>llegar a un acuerdo</i>
		V Prep N Adj	26	<i>mantener en un lugar fresco</i>
		V Det N Adj	14	<i>rescatar las acciones propias</i>
		V Prep N Prep	4	<i>actuar en calidad de</i>
		V Prep Adj N	3	<i>llegar a buen puerto</i>
	5	V Det N Prep N	32	<i>protestar una letra de cambio</i>
		V Prep Det N Adj	9	<i>incluir en la lista negra</i>
		V Prep N Prep N	6	<i>poner en tela de juicio</i>
		V N Prep Det N	5	<i>hacer frente a la competencia</i>
		V Det N Prep Adj	4	<i>vincular una moneda a otra</i>

In total, 457 patterns were identified in Spanish (43.8% in the database). 244 of them were noun phrases and 72 verb phrases. In English, 597 (56.2% in the database) patterns were identified, of which 402 correspond to noun phrases and 64 to verb phrases.

**Table 10.** Top five morphosyntactic patterns of noun and verb SPUs in English in the DCI

Morphosyntactic patterns in English				
Noun / Verb	No. of word forms	Morphosyntactic pattern	Freq.	SPU Example
Noun	3	N Prep N	651	<i>airport to entry</i>
		Adj N N	601	<i>Black-market goods</i>
		N N N	455	<i>cash box company</i>
		Adj Adj N	117	<i>direct foreign investment</i>
		N Adj N	42	<i>exchange cost rate</i>
	4	N Prep Det N	116	<i>face of a document</i>
		Adj N Prep N	109	<i>gross amount of invoice</i>
		Adj - N N	94	<i>communist – block countries</i>
		N Prep N N	92	<i>table of denial orders</i>
		N Prep Adj N	84	<i>holder in good faith</i>
	5	N Prep Det N N	21	<i>sale on the installment plan</i>
		N Prep N Prep N	21	<i>referee in case of need</i>
		N N Prep Det N	13	<i>trade effect of a tariff</i>
		N Prep Det Adj N	13	<i>invoice in a third currency</i>
		Adj - N N N	12	<i>bare – hull charter party</i>
Verb	3	V Det N	182	<i>back a bill</i>
		V Prep N	53	<i>come into operation</i>
		V N N	12	<i>eliminate tax frontiers</i>
		V PP Adj	2	<i>be declared void</i>
		V V Prep	2	<i>be borne by</i>
	4	V Prep Det N	34	<i>fit out a ship</i>
		V Det N Prep	9	<i>run the risk of</i>
		V Det N N	8	<i>make a customs entry</i>
		V Prep N N	7	<i>release for home use</i>
		V Prep Adj N	6	<i>sell for ready money</i>
	5	V Det N Prep N	11	<i>open a letter of credit</i>
		V N Prep Det N	5	<i>put work out to contract</i>
		V N Prep Adj N	4	<i>declare goods for free circulation</i>
		V Prep Det Adj N	3	<i>break into a foreign market</i>
		V N Prep Adj PrP	2	<i>enter goods for inward processing</i>

In line with the information offered in Tables 5 and 6, Tables 9 and 10 present entries in which monolexical terms related to commerce and economics occur (e.g., *precio en el mercado global* ‘price in the global market’ and *cash box company*). Drouin (2003) and Drouin, Morel, and L’Homme (2020) explain how these MWTs (most of them specialized collocations) can be extracted by contrasting technical and non-technical corpora.

The entries in Tables 9 and 10 do not offer any monolexical terms that could be associated to commerce and economics (e.g., *echar mano de* ‘take hand of’ and *run the risk of*). Moreover, as previously presented (see section 4.2.), the identification of certain semantic fields could lead to the identification of metaphors and metonymies. Furthermore, the sole use of morphosyntactic patterns does not allow for the identification of idiomaticity (see section 4.4.) of the entries nor their

classification as SPUs (e.g., specialized idioms, specialized collocations, and specialized pragmatemes).

Therefore, as shown in Table 11, the lexical analysis along with the POS tagging and the semantic annotation (see tags in Table 6) make possible the enhancement of the traditional identification patterns.

**Table 11.** Database tagging sample in English and Spanish

<b>Spanish</b>				
<b>Word forms</b>	<i>de</i>	<i>muelle</i>	<i>a</i>	<i>domicilio</i>
<b>POS</b>	Prep	N	Prep	N
<b>Discourse field (1)</b>	Z	M	Z	H
<b>Semantic field (2)</b>	Z5	M7	Z5	H4
<b>(1) Description</b>	Names and grammar	Movement, location, travel, and transport	Names and grammar	Architecture, housing, and the home
<b>(2) Description</b>	Grammatical bin	Place	Grammatical bin	Residence
<b>English</b>				
<b>Word forms</b>	<i>at</i>	<i>arm</i>	<i>'s</i>	<i>length</i>
<b>POS</b>	Prep	N	SaxP	N
<b>Discourse field (1)</b>	Z	B	Z	N
<b>Semantic field (2)</b>	Z5	B1	Z5	N3.7
<b>(1) Description</b>	Names and grammar	The body and the individual	Names and grammar	Number and measurement
<b>(2) Description</b>	Grammatical bin	Anatomy and physiology	Grammatical bin	Measurement: length & height

When combining the lexical information of the entries and the POS and semantic tagging, it is possible to extract semantic patterns that could be nested with morphosyntactic patterns (see Table 11). The semantic (or discourse) fields, like the one designed by McArthur (1981) used by the USAS, could be used to identify word forms of a certain (or related LSP domain). From the 21 major fields used by the USAS, three of them, ‘government and public’ (G), ‘money commerce and industry’ (I), and ‘language and communication’ (Q) offered many of the word forms related to commerce and economics. Table 12 presents the most productive nested patterns from the analysis sample and highlights the word forms identified within the semantic fields (G), (I), and (Q).

**Table 12.** Examples of morphosyntactic and semantic pattern nesting



Morphosyntactic and semantic nesting							
Lang.	Noun / Verb	No. of word forms	Morphosyntactic pattern	Semantic pattern	Freq.	Example	
Spa	Noun	3	N Prep N	I Z I Q Z I	162 99	<i>moneda de pago</i> <i>certificado de comercio</i>	
		4	N Prep N Adj	A Z I A I Z I M	14 13	<i>tipo de cambio fijo</i> <i>fondo de inversión global</i>	
		5	N Prep N Prep N	A Z A Z I S Z I Z I	10 8	<i>riesgo de falta de liquidez</i> <i>intercambio de títulos de renta</i>	
		Verb	3	V Det N	A Z I A Z Q	26 20	<i>abrir una plica</i> <i>romper las negociaciones</i>
			4	V Prep Det N	M Z Z I A Z Z I	3 3	<i>penetrar en el mercado</i> <i>operar en un banco</i>
	5		V Det N Prep N	A Z Q Z I S Z Q Z I	3 2	<i>abrir una letra de crédito</i> <i>avaluar una letra de cambio</i>	
	Eng	Noun	3	N Prep N	I Z I A Z I	35 34	<i>tax on export</i> <i>presentation for payment</i>
			4	N Prep Det N	I Z Z I T Z Z I	11 8	<i>amortization of a debt</i> <i>renewal of a bill</i>
			5	N Prep Det N N	A Z Z Q I S Z Z I G	2 2	<i>issuance of a documentary credit</i> <i>control of the customs authorities</i>
			Verb	3	V Det N	A Z I M Z I	26 15
4				V Prep Det N	A Z Z I A Z Z Q	6 4	<i>issue of an invoice</i> <i>make out a receipt</i>
5		V Det N Prep N		M Z S Z O A Z I Z A	2 2	<i>lay an embargo on goods</i> <i>present a bill for acceptance</i>	

#### 4.4. Phraseological analysis

There remained two problems to be solved regarding phraseological issues: the first one had to do with choosing the appropriate criteria for the classification of SPUs. As presented in section 2, previous definitions in LSP phraseology tend to focus on the presence of a terminological unit in the phrase. Nevertheless, several SPUs (e.g., ‘at arm’s length’, ‘*los cinco dragones*’ [the five dragons]) were metaphorical in nature, meaning that they entail a terminological *tenor*, the term used in metaphor studies for the figurative meaning of a word or expression (Richards 1965, 96) by means of non-terminological *vehicles*, the term used in metaphor studies for the literal meaning of a word or expression (Richards 1965, 96).

The second problem, linked to SPUs’ definitions, was related to how the notions of compositionality and idiomaticity are treated in phraseological studies. In the case of the units extracted for this study, some of them behaved like SPUs, but they did not show evidence of any

semantic idiomaticity (e.g., ‘in account for’, ‘*por cuenta de*’). These two issues did not allow for a complete classification of the extracted units according to previous SPU definitions or taxonomies (see section 2).

Baldwin and Kim (2010, 269-271) tackle the second issue by differentiating these two concepts. The authors define idiomaticity as the “markedness or deviation from the basic properties of the component lexemes,” while compositionality is “the degree to which the features of the parts of a MWE<sup>7</sup> combine to predict the features of the whole” (Baldwin and Kim 2010, 269). Moreover, these authors also assert that, in most cases, researchers have used compositionality to refer only to semantic idiomaticity, while idiomaticity can occur at different linguistic levels (e.g., lexical, pragmatic, semantic, statistical, or syntactic, see Table 13) (2010, 269).

**Table 13.** Idiomaticity levels and their definitions according to Baldwin and Kim (2010, 269-271)

Levels of idiomaticity	
Level	Definition
Lexical idiomaticity	Occurs when one or more components of an MWE are not part of the conventional English lexicon. For example, <i>ad hoc</i> is lexically marked in that neither of its components ( <i>ad</i> and <i>hoc</i> ) are standalone English words.
Pragmatic idiomaticity	It is the condition of a MWE being associated with a fixed set of situations or a particular context [...] ‘all aboard’ [is an] example of a pragmatic MWE [...] [it] is a command associated with the specific situation of a train station or dock, and the imminent departure of a train or ship.
Semantic idiomaticity	Semantic idiomaticity is the property of the meaning of a MWE not being explicitly derivable from its parts [...] for example, ‘middle of the road’ usually signifies “non-extremism, especially in political views.”
Statistical idiomaticity	Occurs when a particular combination of words occurs with markedly high frequency, relative to the component words or alternative phrasings of the same expression.
Syntactic idiomaticity	Occurs when the syntax of the MWE is not derived directly from that of its components. [...] For example, ‘by and large’, is syntactically idiomatic in that it is adverbial in nature but made up of the anomalous coordination of a preposition ( <i>by</i> ) and an adjective ( <i>large</i> ).

Hence, based on all the descriptive data retrieved from the analyses previously presented in this paper, I consider it necessary to offer an alternative working definition of SPU: a combination of words (including, but not necessarily, monolexical terms) that evidences idiomaticity at least at one

<sup>7</sup> The authors use ‘multiword expression’ (MWE) “as a synonym of ‘multiword unit’, ‘multiword lexical item’, ‘phraseological unit’ and ‘fixed expression’; there is also variation in the hyphenation of ‘multiword’, with ‘multi-word’ in common use” (Baldwin and Kim 2010, 267).

of the possible levels (lexical, pragmatic, semantic, statistical, or syntactic) and that, when used in a certain LSP domain, acquires a specialized meaning. This working definition and a taxonomy, based on the works by Mel'čuk (1998, 2012, 2013) (see Figure 1), allows the classification of SPUs into three subcategories; specialized idioms (Spl), specialized collocations (SpC), and specialized pragmatemes (SpP), according to semantic categories of their word forms and the type of idiomaticity they undergo as shown in Figure 6.

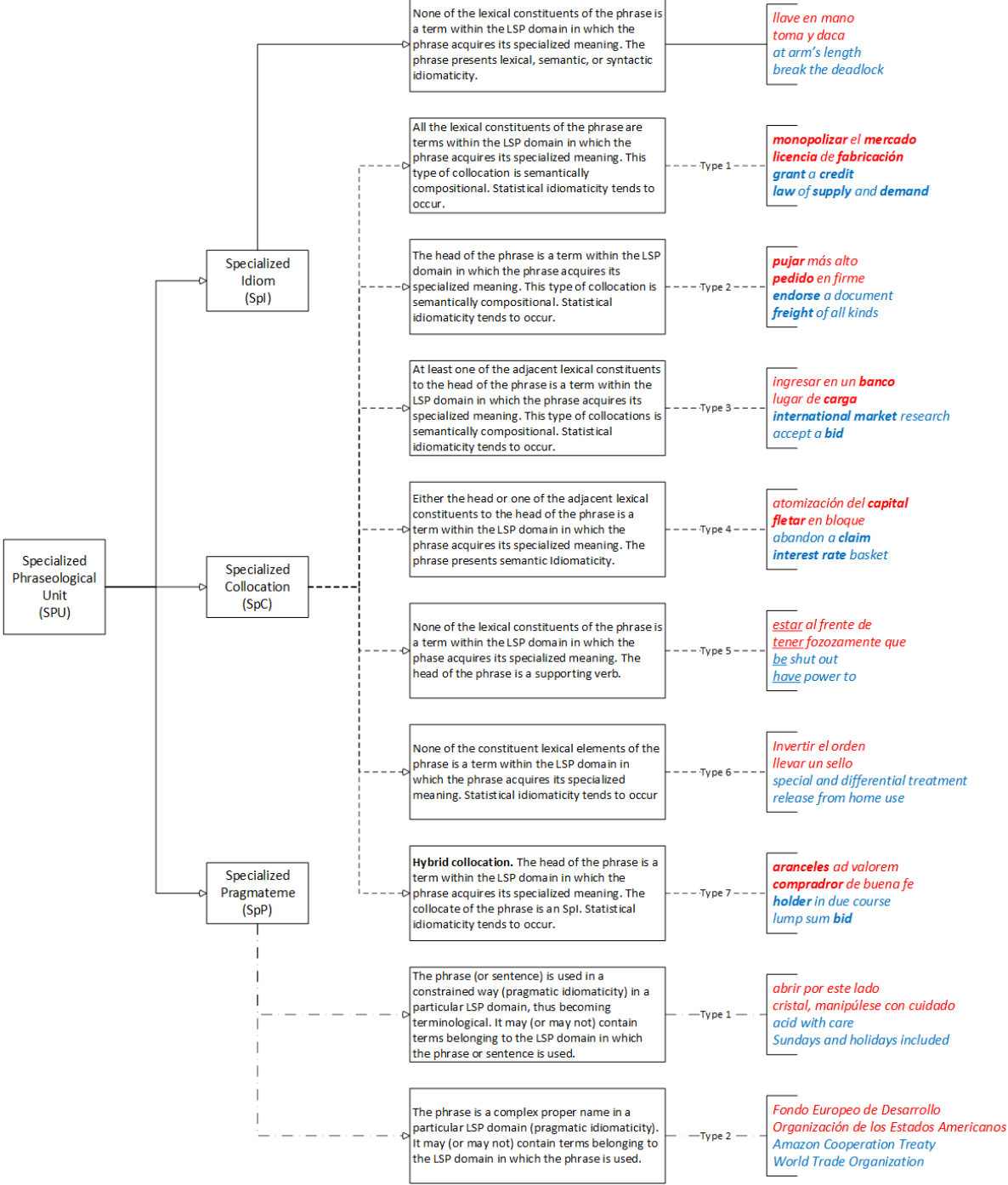


Figure 6. Classification of SPUs based on the works by Mel'čuk

Figure 7 presents the distribution of the SPU subcategories and similarly to what was presented in section 3.2. SpCs were the most prolific SPU in the sample (93.31% in the database [95.39% in Spanish and 90.63% in English]). However, a particularly interesting finding was the identification of SpPs (5.06% in the database [3.15% in Spanish and 7.52% in English]) and Spls (0.99% in the database [1.11% in Spanish and 0.84% in English]) in the sample.

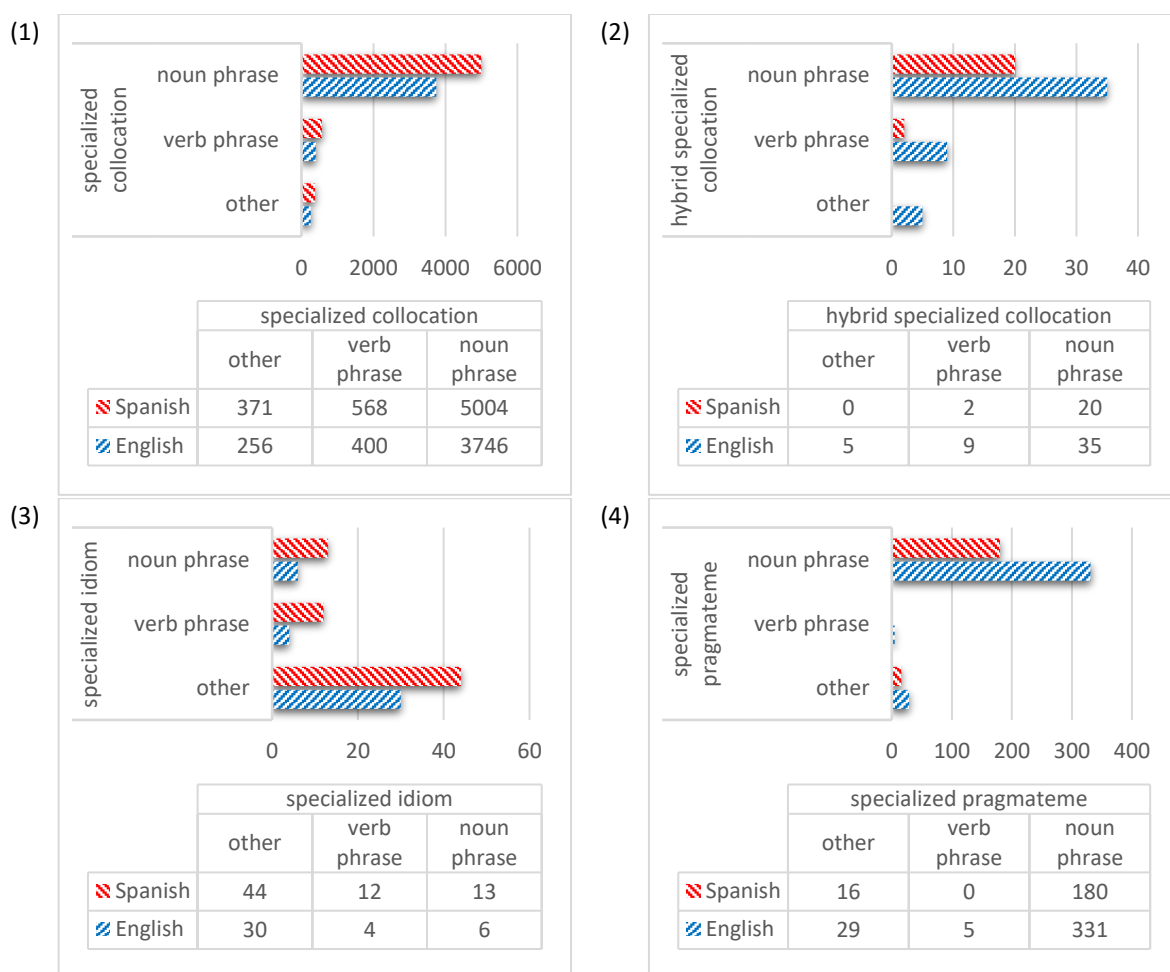


Figure 7. Distribution of SPUs according to the classification offered in Figure 5: SpCs (1), hybrid SpCs (2), Spls (3), and SpPs (4)

When analyzing the SPUs, a specific type of SpC was found: the hybrid SpC. This kind of collocation is an SPU composed of a word form (not necessarily a terminological one) as the head of

the phrase or node and an SpI (as the adjacent or collocate)<sup>8</sup> (see examples in Table 14). 70 (0.63% in the database) hybrid SpCs were identified (49 [1.01%] in English and 21 [0.34%] in Spanish).

**Table 14.** Examples of hybrid collocations in the database

Hybrid specialized collocation examples		
Language	Example	Idiomatycity type
Spanish	<i>Fletamento</i> <sub>(node)</sub> <i>de ida y vuelta</i> <sub>(collocate)</sub>	Semantic
	<i>travesía</i> <sub>(node)</sub> <i>de ida y vuelta</i> <sub>(collocate)</sub>	Semantic
	<i>viaje</i> <sub>(node)</sub> <i>de ida y vuelta</i> <sub>(collocate)</sub>	Semantic
	<i>aranceles</i> <sub>(node)</sub> <i>ad valorem</i> <sub>(collocate)</sub>	Lexical
	<i>derechos de aduana</i> <sub>(node)</sub> <i>ad valorem</i> <sub>(collocate)</sub>	Lexical
	<i>flete</i> <sub>(node)</sub> <i>ad valorem</i> <sub>(collocate)</sub>	Lexical
English	<i>arm's length</i> <sub>(collocate)</sub> <i>agreement</i> <sub>(node)</sub>	Semantic
	<i>arm's length</i> <sub>(collocate)</sub> <i>negotiation</i> <sub>(node)</sub>	Semantic
	<i>arm's length</i> <sub>(collocate)</sub> <i>price</i> <sub>(node)</sub>	Semantic
	<i>del credere</i> <sub>(collocate)</sub> <i>agent</i> <sub>(node)</sub>	Lexical
	<i>del credere</i> <sub>(collocate)</sub> <i>commission</i> <sub>(node)</sub>	Lexical
	<i>del credere</i> <sub>(collocate)</sub> <i>risk</i> <sub>(node)</sub>	Lexical

## 5. A hybrid approach for the indexation of SPUs in specialized lexicographic resources

SPUs pose a lexicographic challenge regarding their indexation according to semasiological approaches. Take the example of an expression that has different meaning in both LGP and in the LSP domain of commerce and economics: ‘at arm’s length.’

**Table 15.** General and specialized definitions of ‘at arm’s length’

SPU	LGP/LSP	Definition	Source
<i>at arm's length</i>	LGP	“As far as one can reach with one's arm; (hence) at a distance, remote, not on familiar or friendly terms”	OED ("arm, n.1," n.d.)
	LSP	“Law. Of two parties: without legal obligations to each other, esp. fiduciary obligations; (also more generally) in an independent or impartial position; conducted by independent or impartial parties”	OED ("arm, n.1," n.d.)
	LSP	“fr: GRAL a raya, distanciado; de igual a igual; manteniendo las distancias, sin concederse favores; con total independencia; en condiciones de plena concurrencia; con escasas muestras de cordialidad”	DCI (Alcaraz & Castro Calvín, 2007, p. 37)

As presented in Table 15, the Oxford English Dictionary online (OED) offers two definitions for the expression *at arm’s length*, one for LGP and another one for Law. Similarly, the DCI also offers both definitions, although they are not presented with context of use or a clearer diasystematic

<sup>8</sup> ‘Node’ and ‘collocate’ were the denominations used by Patiño García (2017) to name the component parts of a ‘specialized collocation’.

marking, which makes it difficult for the user to know what equivalent to use under a given circumstance.

Although both use semasiological indexation approaches, the OED indexes the expression under the headword *arm* while the DCI decided to index the expression as a whole entry using the first noun as the lemmatizing headword *arm's length, at*. See example (2) in Table 16.

*Table 16. Indexation of 'at arm's length' in two dictionaries from Ariel*

<b>at arm's length entry</b>	
(Alcaraz Varó and Hughes 2004, 56)	(Alcaraz Varó and Castro Calvín 2007, 37)
<p>(1) <b>arm<sup>1</sup> n:</b> GRAL brazo; arma. [Exp: ARM<sup>2</sup> (FINAN V. <i>adjustable rate mortgage</i>), <b>arm's length, at</b> (GRAL en condiciones de plena competencia, distanciado, a raya; guardando las distancias, sin concederse favores), <b>arm's length contract/deal/transaction</b> (DER contrato o transacción sin favores o entre partes</p>	<p><b>area n:</b> GRAL/DOC área, campo, esfera, terreno, ámbito; campo de un documento; sector ∅ <i>Do not use this area.</i>  <b>arm's length, at</b> <i>fr:</i> GRAL a raya, distanciado; de igual a igual; manteniendo las distancias, sin concederse favores; con total independencia; en condiciones de plena concurrencia; con escasas muestras de cordialidad ∅ <i>He is a supplier at arm's length.</i> [Exp: <b>arm's length agreement or contract</b> (DER contrato cerrado</p>

The indexation change from the 2004 to 2007 editions of the DCI seems to be justified in the fact that *arm* and *length* might not be considered terms and, therefore, should not be used as headwords for the indexation of the SPU. Moreover, in previous dictionaries of the same publishing house, e.g., Alcaraz Varó and Hughes (2004), *at arm's length* was indexed under *arm*; however, the definition of the headword does not offer any terminological meaning or use in the interest of the LSP domain of commerce and economics. See example (1) in Table 16.

Another solution to this problem is proposed in the work by Rojas Díaz and Pérez Sánchez (2020, 112) by means of a hybrid approach for the indexation PUs (or SPUs in this case). It is defined as hybrid because it uses both semasiological and onomasiological approaches for the indexation of entries.

This hybrid model consists of two parts. The first one is a series of terminological headwords that are lemmatized alphabetically following a semasiological approach (see Figure 8).

<b>indenture</b>	<b>n:</b>	<i>escritura, contrato, instrumento, partida.</i>
	[FIN]	1. (US) A document or agreement devised to prevent forgery, especially one between an apprentice and his master. It was so called because it was duplicated on a single sheet and separated or indented by a zigzag cut so that each party held identical halves. Their authenticity could be proved by matching the jagged edges.
	<b>n:</b>	<i>contrato bilateral, contrato de aprendizaje.</i>
	[FIN]	2. The legal document establishing the terms and conditions of a securities issue and the obligations of the trustee.
	<b>n:</b>	<i>escritura de emisión de bonos u obligaciones</i>
	[FIN]	3. (US) The contract covering a public offering of bonds
<b>independent</b>	<b>adj:</b>	<i>externo, independiente, neutral.</i>
	[DER]	1. Not subject to the control or influence of another.   2. Not associated with another (often larger) entity.   3. Not dependent or contingent on something else.
<b>index</b>	<b>n:</b>	<i>índice.</i>
	[GRAL]	1. An alphabetized listing of the topics or other item included in a single book or document, or in a series of volumes, usu. found at the end of the book, document, or series.
	<b>n:</b>	<i>índice, coeficiente</i>
	[ECON]	2. A number usu. expressed in the form of a percentage or ratio, that indicates or measures a series of observations, esp. those involving a market or the economy

Figure 8. Example of semasiological indexation for monolexical entries

Similarly to monolexical entries, SpCs would be indexed following the semasiological approach when the head or the collocate of the phrase corresponds to a headword due to their compositional or semi-compositional structure (see Figure 6). Differently from SpCs, Spls would apply an onomasiological approach for their indexation by using the headwords as semantic anchors according to their signified (see Figure 9). As seen in Table 15, the definitions offered by the OED and the DCI link the concepts of concurrent, equal, and independent to *at arm's length* when used in LSP. An example of a hybrid entry, using independent as headword is presented in Figure 9.

<b>independent</b>	<b>adj:</b> [DER]	<i>externo, independiente, neutral.</i>
		1. Not subject to the control or influence of another.   2. Not associated with another (often larger) entity.   3. Not dependent or contingent on something else.
		o collocation.
		• ~ taxation <b>n:</b>
		<i>declaración de renta separada.</i>
		[FISC] (UK) 1. A system of personal taxation in which married women are treated as completely separate and independent taxpayers for both income tax and capital gains tax. Prior to April 1990 in the UK, the income of a married woman was added to the income of her husband and taxed accordingly.
		• of ~ means <b>adj:</b>
		<i>de posición acomodada, que vive de rentas, que no depende de un sueldo.</i>
		[FISC] 1. Having enough money to support oneself without help from others.
		o idiom.
		• at arm's length <b>adj:</b>
		<i>de igual a igual, sin concederse favores, con total independencia, en condiciones de plena concurrencia.</i>
[DER] 1. Denoting a transaction entered into by unrelated parties, each acting in their own best interests in paying or charging prices based on fair market values.		

Figure 9. Example of a lexicographic entry using a hybrid indexation method

As observed in Figure 7, *at arm's length* keeps the same syntactic function (adjective) of the conceptual headword (independent). As demonstrated in the works by Rojas Díaz (2020, 320) and Rojas Díaz and Pérez Sánchez (2020), this hybrid model can also be applied to electronic lexicographic resources, allowing for the inclusion of more detailed information in each entry (see Figure 10).

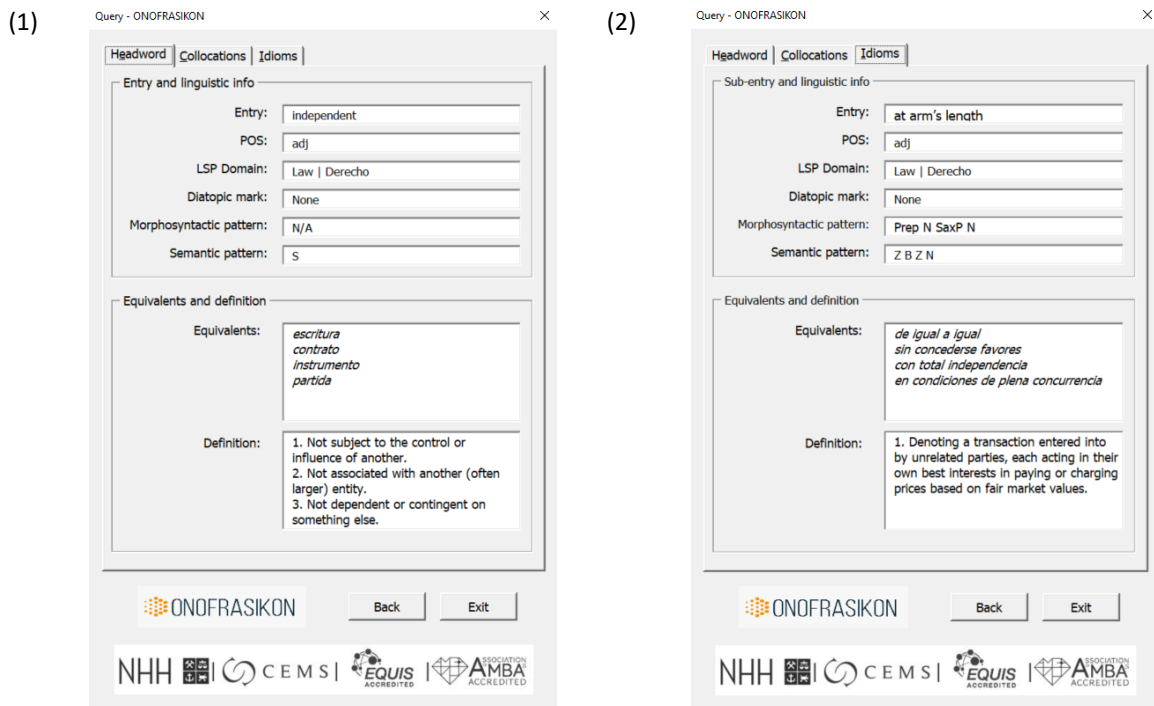


Figure 10. Example of a hybrid entry: (1) the main headword and (2) an *Spl* using the headword as semantic anchor

Example (2) from Figure 10, shows how an electronic entry allows the possibility of including the morphosyntactic and semantic patterns. Although this information could be of interest to linguists, pattern nesting proves to be useful when executing queries combining lemmas, POS tagging, and semantic annotation. Figure 11 presents an example for the consultation of *at arm's length* through an advanced query.







Advanced query - ONOFRASIKON ×


Advanced query

Lemma:

POS tag:

Semantic tag:



**Figure 11.** Example of an advanced query for 'at arm's length' using lemmas and pattern nesting (morphosyntactic/semantic)

## 6. Conclusions

This paper has shed light on the delimitation of the object of study of phraseology in LSP by offering a working definition of SPU along with a classification taxonomy based on a large descriptive analysis of data: 11,086 SPUs, that were classified in specialized idioms (SpIs), specialized collocations (SpCs), and specialized pragmatemes (SpPs) composed of 39,832 word forms (that were POS-tagged and annotated semantically). Moreover, the analyses of the data have provided 1,054 morphosyntactic patterns (457 in Spanish and 597 in English) of which 51% were repeated 2 or more times and 4,369 semantic patterns (2,131 in Spanish and 2,238 in English) of which 35.4% were repeated more than one time. Both the offered taxonomy and the definition should be contrasted with other LSP lexicographic resources and corpora to assess their accuracy to identify SPUs in other LSP domains. Moreover, the concept of overlapping range allows for the analysis of phraseological migration from LGP to SLP and between LSP domains.

The morphosyntactic and lexical analyses of the extracted SPUs have shown how uneven the distribution of phrases is according to their type (e.g., nouns vs. verbs). This finding, although documented before, raises the question of whether the entries and equivalents gathered in lexicographic resources are a result of what occurs in corpora or whether the information

lexicographic resources offer is totally biased towards a certain category (e.g., the noun phrases in this particular study. See Figure 5).

The analysis of the database containing the entries and sub-entries extracted from the DCI, showed that several SPUs did not include monolexical terms within their word forms in the LSP domain of commerce and economics, contradicting what previous definitions and the taxonomies have stated. Since SPU extraction from corpora is based on theoretical criteria, it is possible to assert that SPUs (e.g., *los cinco dragones, llave en mano, de banda a banda* | *break the deadlock, above the line, door to door*) might be under-registered in LSP lexicographic resources. An LSP corpus-driven study in which PUs are analyzed in context would provide information about whether these PUs have acquired a specialized meaning into a certain LSP domain.

Beyond the descriptive statistics, the morphosyntactic and semantic patterns resulting from this study (see section 4.3.) could be used for the identification of SPUs within the field of commerce and economics. Furthermore, the collected linguistic data could potentially be used in Natural Language Processing, namely for phraseological extraction and Machine Learning purposes. However, the semantic annotation should be verified by experts and evaluated by means of inter-annotator agreement (e.g., Cohen's kappa, Fleiss's kappa, Krippendorff's alpha) to use it as a Gold Standard criterion for the identification and extraction of SPUs.

Although a more in-depth analysis of the metaphors and metonymies is outside the scope of this study, 2,118 SPUs in which metaphors (e.g., *hacer frente a* | *take the ground*) and metonymies (e.g., *de ida y vuelta* | *house to house*) or both (e.g., *at arm's length*) were identified, along with the morphosyntactic and semantic patterns linked to them. This finding suggests, as presented in section 4.4., that idiomaticity should be considered as one of the main criteria for the identification of SPUs as is the case with their LGP counterpart.

This paper has presented a parallel study that shows how SPUs are formed in Spanish and English. Nevertheless, it is the task for future work to carry out a crosslinguistic study in which the behavior of SPUs (when translated) is described.

Regarding lexicography and terminography, the study has proposed an indexation method of SPUs in dictionaries and terminological databases. This hybrid (semasiological / onomasiological) proposal benefits from both approaches by indexing the SPUs under a certain headword that better encompasses the meaning of the expression. However, the assessment of this proposal as well as the reception it could receive from the users was not part of the objectives of this study and therefore should be considered for future work.

As a final remark, and as proven in this paper, Mel'čuk's contributions in phraseology are still relevant regardless of whether they are framed within LGP or LSP. The delimitation of the object of study of LSP phraseology has been a recurrent topic in these studies. It might be time to revisit concepts and methods from LGP phraseology and contrast them with LSP phraseology to find a broader shared ground between them.

### References

- Aguado de Cea, Guadalupe. 2007. "A Multiperspective Approach to Specialized Phraseology: Internet as a Reference Corpus for Phraseology." In *The Texture of Internet: Netlinguistics in Progress*, 182-207. Ontology Engineering Group – OEG.
- Alcaraz Varó, Enrique, and José Castro Calvín, eds. 2007. *Diccionario de comercio internacional – importación y exportación, Inglés-Español/Spanish-English*. Barcelona: Ariel.
- Alcaraz Varó, Enrique, and Brian Hughes, eds. 2004. *Diccionario de términos económicos, financieros y comerciales, inglés-español / español-inglés*. 4th ed. Barcelona: Ariel.
- Alonso Ramos, Margarita, ed. 2006. *Diccionarios y fraseología, Anexos de Revista de Lexicografía*.
- Archer, Dawn, Andrew Wilson, and Paul Rayson. 2002. *Introduction to the USAS Category System*. [http://ucrel.lancs.ac.uk/usas/usas\\_guide.pdf](http://ucrel.lancs.ac.uk/usas/usas_guide.pdf).
- Baldwin, Timothy, and Su Nam Kim. 2010. "Multiword Expressions." In *Handbook of Natural Language Processing*, edited by Nitin Indurkha and Fred J Damerau, 267-292. Boca Raton: CRC Press.
- Benson, Morton, Evelyn Benson, and Robert F Ilson. 1986. *Lexicographic description of English*. Vol. 14 *Studies in Language Companion*. Amsterdam: John Benjamins.
- Bevilacqua, Cleci. 2004. "Unidades fraseológicas especializadas eventiva: descripción y reglas de formación en el ámbito de la energía solar." PhD Thesis, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Blais, Esther. 1993. "Le phraséologisme. Une hypothèse de travail." *Terminologies nouvelles. Terminologie et diversité culturelle* 10: 50-56.
- Buendía Castro, Miriam, and Pamela Faber. 2015. "Phraseological Units in English-Spanish Legal Dictionaries: a Comparative Study." *Fachsprache* (3-4): 161-175.
- Burger, Harald, Dmitriy Dobrovolskij, Peter Kühn, and Neal Norrick. 2007. "Phraseology: Subject Area, Terminology and Research Topics." In *Phraseology*, edited by Harald Burger, Dmitriy Dobrovolskij, Peter Kühn and Neal Norrick, In *Handbooks of Linguistics and Communication Science (HSK)*, 10-19. New York: Walter de Gruyter.

- Bushnaq, Tatiana. 2015. "A Retrospective Analysis of the Term Phraseological Unit." In *Debates on Globalization. Approaching National Identity through Intercultural Dialogue*, edited by Iulian Boldea, 167-176. Tîrgu-Mureş: Arhipelag XXI Press.
- Cabré, María Teresa. 2000. "Terminologie et linguistique: la théorie des portes." *Terminologies nouvelles. Terminologie et diversité culturelle* 21: 10-15.
- Church, Kenneth Ward, and Patrick Hanks. 1989. "Word Association Norms, Mutual Information, and Lexicography." In *27th Annual Meeting of the Association for Computational Linguistics*, 76-83. Vancouver: Association for Computational Linguistics.
- Corpas Pastor, Gloria. 1996. *Manual de fraseología española*. Madrid: Gredos.
- Cowie, Anthony Paul. 1998. "Introduction." In *Phraseology: Theory, Analysis, and Applications*, 1-20. Oxford: Oxford University Press.
- Drouin, Patrick. 2003. "Term extraction using non-technical corpora as a point of leverage." *Terminology* 9 (1): 99–115. <https://doi.org/10.1075/term.9.1.06dro>.
- Drouin, Patrick, Jean-Benoît Morel, and Marie-Claude L'Homme. 2020. "Automatic Term Extraction from Newspaper Corpora: Making the Most of Specificity and Common Features." In *6th International Workshop on Computational Terminology (COMPUTERM 2020), Proceedings*, edited by Béatrice Daille, Kyo Kageura and Ayla Rigouts Terryn, 1-7. Marseille: European Language Resources Association (ELRA).
- Dunning, Ted. 1993. "Accurate Methods for the Statistics of Surprise and Coincidence." *Computational Linguistics* 19 (1): 61-74. <https://aclanthology.org/J93-1003>.
- EAGLES. 1999. EAGLES LE3-4244 Preliminary Recommendations on Lexical Semantic Encoding. Accessed September 22nd, 2021.
- Faber Benítez, Pamela, and Clara Inés López Rodríguez. 2012. "Terminology and specialized language." In *A Cognitive Linguistics View of terminology and Specialized Language*, edited by Pamela Faber. Berlin: De Gruyter Mouton.
- Frailé Vicente, Esther. 2007. "The Relationship between the Typical Errors in the Translation of Business Idioms and their Lexicographical Treatment." *Revista Alicantina de Estudios Ingleses* (20): 65-93.
- García-Page, Mario. 2008. *Introducción a la fraseología española*. Barcelona: Anthropos.
- Gouadec, Daniel. 1993. "Terminologie & terminotique: outils, modèles & méthodes." In *Actes de la première Université d'automne en Terminologie Rennes 2 - 21 au 26 septembre 1992*, edited by Daniel Gouadec. Paris: La Maison du Dictionnaire.
- Granger, Sylviane, and Magali Paquot. 2008. "Disentangling the Phraseological Web." In *Phraseology: an interdisciplinary perspective*, edited by Sylviane Granger and Fanny Meunier, 27-49. Amsterdam: John Benjamins Publishing Company.
- Gries, Stefan Th. 2008. "Phraseology and linguistic theory: a brief survey." In *Phraseology: an interdisciplinary perspective (p. 6)*. Amsterdam, edited by Sylviane Granger and Fanny Meunier, 3-25. Amsterdam: John Benjamins Publishing Company.
- Hartmann, Reinhard R.K., and Gregory James. 1998. *Dictionary of Lexicography*. New York: Routledge.
- Heid, Ulrich. 2008. "Computational Phraseology: an Overview." In *Phraseology: an interdisciplinary perspective*, edited by Sylviane Granger and Fanny Meunier, 337–360. Amsterdam: John Benjamins Publishing Company.
- Hoffmann, Lothar. 1998. *Llenguatges d'especialitat: selecció de textos*. edited by Jenny Brumme. Barcelona: Institut Universitari de Lingüística Aplicada.
- Hourani-Martín, Dunia, and Encarnación Tabares-Plasencia. 2020. "Morphosyntactic and Semantic Behavior of Legal Phraseological units: a Case Study in Spanish Verb-Noun Constructions about Money Laundering." *Terminology* 26 (1): 108-131.
- Justeson, John S, and Slava M Katz. 1995. "Technical terminology: some linguistic properties and an algorithm for identification in text." *Natural Language Engineering* 1 (1): 9-27. <https://doi.org/10.1017/S1351324900000048>.

- Kjær, Anne-Lise. 2007. "Phrasemes in Legal Texts." In *Phraseology*, edited by Harald Burger, Dmitrij Dobrovolskij, Peter Kühn and Neal Norrick, In *Handbooks of Linguistics and Communication Science (HSK)*, 506-516. New York: Walter de Gruyter.
- Kocjančič, Polonca. 2004. "Acerca de la macroestructura y la microestructura en el diccionario bilingüe." *Verba Hispanica* 12 (1): 171-185.
- Kristiansen, Marita, and Gisle Andersen. 2012. "Corpus Approaches to Neology and their Relevance for Dynamic Domains." *Neologica* 6: 43-62. <https://doi.org/10.15122/isbn.978-2-8124-4232-2.p.0043>.
- Kübler, Natalie, and Mojca Pecman. 2012. "The ARTES bilingual LSP dictionary: From collocation to higher order phraseology." In *Electronic Lexicography*, edited by Sylviane Granger and Magali Paquot, 187-209. Oxford: Oxford University Press.
- L'Homme, Marie-Claude. 1998. "Caractérisation des combinaisons lexicales spécialisées par rapport aux collocations de langue générale." In *EURALEX 98 Proceedings, VIII International Congress of the European Association for Lexicography*, edited by T Fontenelle, P Hiligsmann and A Michiels, 513-522. Liège: Éd. Université de Liège.
- . 2004. *La terminologie: principes et techniques*. Montreal: Presses de l'Université de Montréal.
- . 2020. *Lexical Semantics for Terminology: an Introduction*. Edited by Marie-Claude L'Homme and Kyo Kageura. Vol. 20 *Terminology and Lexicography Research and Practice*. Amsterdam: John Benjamins Publishing Company.
- L'Homme, Marie-Claude, and Claudine Bertrand. 2000. "Specialized Lexical Combinations: Should they be Described as Collocations or in Terms of Selectional Restrictions?" In *Proceedings of the IX EURALEX International Congress*, edited by Ulrich Heid, Stefan Evert, Egbert Lehmann and Christian Rohrer, 497-506. Stuttgart: Institut für Maschinelle Sprachverarbeitung.
- L'Homme, Marie-Claude, and Isabelle Meynard. 1998. "Le point d'accès aux combinaisons lexicales spécialisées : présentation de deux modèles informatiques." *TTR : traduction, terminologie, rédaction* 11 (1): 199-227. <https://doi.org/https://doi.org/10.7202/037322ar>.
- León Araúz, Pilar, and Melania Cabezas-García. 2020. "Term and translation variation of multiword terms." *MonTI* (Special Issue 6): 210-247. <https://doi.org/doi.org/10.6035/MonTI.2020.ne6.7>.
- Leroyer, Patrick. 2006. "Dealing with Phraseology in Business Dictionaries: Focus on Dictionary Functions - not Phrases." *Linguistik online* 27 (2): 183-194.
- López, Xavier Pascual. 2012. "Fraseología española de origen latino y motivo grecorromano." PhD Thesis, Departament de Filologia Clàssica, Francesa i Hispànica, Universitat de Lleida.
- Lorente Casafont, Mercè. 2002. "Terminología y fraseología especializada: del léxico a la sintaxis." In *Panorama actual de la terminología*, edited by Manuel Fernando Pérez Lagos and Gloria Guerrero Ramos, 159-180. Granada: Comares.
- McArthur, Tom. 1981. *Longman Lexicon of Contemporary English*. Harlow: Longman.
- Mel'čuk, Igor. 1998. "Collocations and Lexical Functions." In *Phraseology: Theory, Analysis and Applications*, edited by Anthony Paul Cowie, 23-53. Oxford: Oxford University Press.
- . 2012. "Phraseology in the language, in the dictionary, and in the computer." In *Yearbook of Phraseology*, edited by Koenraad Kuiper, 31-56. New York: De Gruyter Mouton.
- . 2013. "Tout ce que nous voulions savoir sur les phrasèmes, mais...." *Cahiers de lexicologie* 102: 129-149. <https://doi.org/10.15122/isbn.978-2-8124-1259-2.p.0129>.
- Mellado Blanco, Carmen. 2004. *Fraseologismos somáticos del alemán: un estudio léxico-semántico*. Edited by Gerd Wotjak, José Juan Batista Rodríguez and Dolores García Padrón. Vol. 13 *Studien zur romanischen Sprachwissenschaft und interkulturellen Kommunikation*. Frankfurt am Main: Peter Lang.
- , ed. 2008. *Colocaciones y fraseología en los diccionarios*. Frankfurt am Main: Peter Lang.
- Montero Martínez, Silvia. 2008. "A Constructional Approach to Terminological Phrasemes." In *Proceedings of the XIII EURALEX International Congress*, edited by Elisenda Bernal and Janet DeCesaris, 1015-1022. Barcelona: Institut Universitari de Lingüística Aplicada.

- Moon, Rosamund. 2008. "Dictionaries and Collocation." In *Phraseology: an interdisciplinary perspective*, edited by Sylviane Granger and Fanny Meunier, 313-336. Amsterdam: John Benjamins.
- Norrick, Neal. 2007. "English Phraseology." In *Phraseology*, edited by Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn and Neal Norrick, In Handbooks of Linguistics and Communication Science (HSK), 615-619. New York: Walter de Gruyter.
- Nuccorini, Stefania. 2020. "Digging into a Thesaurus for Treasures: Phraseological Conceptual Maps in the Longman Language Activator." In *Contrastive Phraseology: Languages and Cultures in Comparison*, edited by Paola Cotta Ramusino and Fabio Mollica, 349-375. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Paquot, Magali. 2015. "Lexicography and Phraseology." In *The Cambridge Handbook of English Corpus Linguistics*, edited by Douglas Biber and Randi Reppen, 460-477. Cambridge: Cambridge University press.
- Pavel, Silvia. 1993. "Neology and Phraseology as Terminology-in-the-Making." In *Terminology: applications in interdisciplinary communication*, edited by Helmi Sonneveld and Kurt Loening, 21-34. Amsterdam: John Benjamins.
- Pawley, Andrew. 2001. "Phraseology. Linguistics and the Dictionary." *International Journal of Lexicography* 14 (2): 122-134. <https://doi.org/doi.org/10.1093/ijl/14.2.122>.
- Picht, Heribert. 1987. "Terms and their LSP Environment - LSP Phraseology." *Meta: Journal des traducteurs* 32 (2): 149-155. <https://doi.org/https://doi.org/10.7202/003836ar>.
- Ramisch, Carlos. 2015. *Multiword Expressions Acquisition: a Generic and Open Framework. Theory and Applications of Natural Language Processing*. Cham: Springer International Publishing.
- Richards, Ivor Armstrong. 1965. *The Philosophy of Rhetoric*. New York: Oxford University Press.
- Rojas Díaz, José Luis. 2020. "'From Head to Toe': a Lexical, Semantic, and Morphosyntactic Study of Idioms in Phraseological Dictionaries in English and Spanish." *MonTI* (Special issue 6): 287-326. <https://doi.org/http://dx.doi.org/10.6035/MonTI.2020.ne6.9>.
- Rojas Díaz, José Luis, and Juan Manuel Pérez Sánchez. 2019. "'You Took the Word Out of My Mouth': a Morphosyntactic and Semantic Analysis of a Phraseological Lexicon of Colombian Spanish." In *Computational and Corpus-Based Phraseology - Third International Conference, Europhras 2019 Malaga, Spain, September 25-27, 2019 Proceedings*, edited by Gloria & Mitkov Corpas Pastor, Ruslan, In Lecture Notes in Artificial Intelligence, 375-390. Springer.
- . 2020. "'A Baptism of Fire': towards a Practical Hybrid Approach for the Lexicographic Indexation of Phraseological Units with Religious Lexical Components in English and Spanish." *CLINA: An Interdisciplinary Journal of Translation, Interpreting and Intercultural Communication* 6 (2): 95-117. <https://doi.org/https://doi.org/10.14201/clina20206295117>.
- Ruiz Gurillo, Leonor. 1997. *Aspectos de fraseología teórica española*. Valencia: Servicio de Publicaciones de la Universidad de Valencia.
- . 2001. *Las locuciones en el español actual*. Madrid: Arco Libros.
- Sager, Juan Carlos. 1990. *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.
- Schmid, Helmut. 1994. "Probabilistic Part-of-Speech Tagging Using Decision Trees." Proceedings of the International Conference on New Methods in Language Processing, Manchester.
- Seco, Manuel, Olimpia Andrés, and Gabino Ramos, eds. 2004. *Diccionario fraseológico documentado del español actual*. Madrid: Aguilar.
- Siepmann, Dirk. 2008. "Phraseology in Learners' Dictionaries: What, Where and How?" In *Phraseology in Foreign Language Learning and Teaching*, edited by Fanny & Granger Meunier, Sylviane, 185-202. Amsterdam: John Benjamins Publishing Company.
- Simonnæs, Ingrid, and Marita Kristiansen. 2018. "Economic Language." In *Languages for special puposes: an international handbook*, edited by John Humbley, Gerhard Budin and Christer Laurén, In De Gruyter reference, 151-185. Berlin: De Gruyter Mouton.
- Sinclair, John, and Rosamund Moon, eds. 1997. *Collins COBUILD Dictionary of Idioms*. Glasgow: HarperCollins Publishers.

- Sosiński, Marcin. 2006. *Fraseología comparada del polaco y del español: su tratamiento en los diccionarios bilingües*. (Granada).
- Szabó, Zoltán Gendler. 2020. "Compositionality." The Stanford Encyclopedia of Philosophy Last Modified Fall 2020 Edition. Accessed September 29.  
<https://plato.stanford.edu/archives/fall2020/entries/compositionality/>.
- Tschichold, Cornelia. 2008. "A Computational Lexicography Approach to Phraseologisms." In *Phraseology: an interdisciplinary perspective*, edited by Sylviane Granger and Fanny Meunier, 361–376. Amsterdam: John Benjamins Publishing Company.
- Vangehuchten, Lieve. 2005. *El léxico del discurso económico empresarial: identificación, selección y enseñanza en español como lengua extranjera con fines específicos*. Madrid: Iberoamericana.
- Veisbergs, Andrejs. 2020. "Phraseology in General Bilingual Dictionaries: Idioms as Equivalents of Single Words." In *Contrastive Phraseology: Languages and Culture in Comparison*, edited by Paola Cotta Ramusino and Fabio Mollica, 331-343. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Zuluaga, Alberto. 1975. "La fijación fraseológica." *Thesaurus* XXX (2): 225-248.
- . 1980. *Introducción al estudio de las expresiones fijas*. Frankfurt: Peter Lang.

Article 3

Rojas Díaz, J. L. 'Worlds apart?': phraseological equivalence through the lenses of translation, terminology, and lexicography (*in prep. to be submitted to Meta: Journal des traducteurs*)



# ‘Worlds Apart?’: Phraseological Equivalence through the Lenses of Translation, Terminology, and Lexicography

JOSÉ LUIS ROJAS DIAZ

*NHH Norwegian School of Economics, Bergen, Norway*

jose.diaz@nhh.no

## RÉSUMÉ

Cet article examine une question qui intéresse les spécialistes de la phraséologie, de la lexicographie, de la terminologie et de la traduction, à savoir, la traduction d’expressions idiomatiques, telles que les unités phraséologiques spécialisées (UPS). Le type d’entrée UPS choisi pour cette base de données appartient à une sous-catégorie appelée locutions spécialisées (LocSp), caractérisées par l’absence d’un terme monolexical dans leurs formes verbales (liées à un certain domaine de spécialisation dans lequel elles acquièrent leur sens spécialisé) et subissant un certain niveau d’idiomaticité (lexical, sémantique ou syntaxique). L’objectif de cet article est triple, il présente (i) un aperçu de la notion d’équivalence telle qu’elle est traitée dans la littérature en traduction, lexicographie, terminologie et phraséologie. Il fournit (ii) la caractérisation des entrées et des équivalents des LocSp à travers une série d’analyses lexicographiques et sémantiques. Et il offre (iii) l’analyse des techniques de traduction utilisées par les traducteurs à partir des résultats des requêtes effectuées dans le corpus EUR-Lex. Les analyses ont été réalisées dans une base de données composée de 109 LocSp et de 174 équivalents en espagnol et en anglais qui a été construite pour cette étude. Le présent travail a montré que le fait de revisiter des notions issues de disciplines liées à la phraséologie permet de créer un terrain d’entente pour l’étude des similitudes et des différences entre la phraséologie générale et celle des LSP.

## ABSTRACT

This article examines a question of the interest of scholars dealing with phraseology, lexicography, terminology, and translation, namely the task of translating idiomatic expressions, such as specialized phraseological units (SPUs). The type of SPU chosen for this database belongs to a subcategory denominated as specialized idioms (SpIs) which are characterized by their lack of a monolexical term within their word forms (related to the language for specific purposes [LSP] domain in which they acquire their specialized meaning) that undergo a certain level of idiomaticity (lexical, semantic, or syntactic). The aim of this article is three-fold, it presents (i) an overview of the notion of equivalence as treated in the previous literature in translation, lexicography, terminology, and phraseology. It provides (ii) the characterization of SpI entries and equivalents through a series of lexicographic and semantic analyses. And it offers (iii) the analysis of translation techniques used by translators based on results from queries made in the EUR-Lex corpus. The analyses were carried out in a database consisting of 109 SpI entries and 174 equivalents in Spanish and English that was built for this study. The present work has shown that revisiting notions from disciplines related to phraseology allows for creating a shared ground for the study of similarities and differences between general and LSP phraseology.

## RESUMEN

Este artículo examina una cuestión de interés para los estudiosos de la fraseología, la lexicografía, la terminología y la traducción: la tarea de traducir expresiones idiomáticas, como las unidades fraseológicas especializadas (UFEs). El tipo de UFE elegida para esta base de datos pertenece a una subcategoría denominada locución especializada (LE) que se

caracteriza por carecer de un término monoléxico dentro de sus formas verbales (relacionadas con un determinado dominio de especialidad en el que adquiere su significado especializado) y en la que subyace algún tipo de idiomatidad (ya sea léxica, semántica o sintáctica). Los objetivos de este artículo son tres: (i) presenta una visión general de la noción de equivalencia en traducción, lexicografía, terminología y fraseología. (ii) Ofrece la caracterización de las entradas y equivalencias del LEs mediante una serie de análisis lexicográficos y semánticos. Y (iii) brinda el análisis de las técnicas de traducción utilizadas por traductores a partir de los resultados de las consultas realizadas en el corpus EUR-Lex. Los análisis se han realizado en una base de datos compuesta por 109 entradas de LE y 174 equivalentes en español e inglés construida para este estudio. El presente trabajo ha demostrado que la revisión de nociones procedentes de disciplinas relacionadas con la fraseología permite crear un punto de encuentro para el estudio de las similitudes y diferencias entre la fraseología general y especializada.

### **MOTS-CLÉS/KEYWORDS/PALABRAS CLAVE**

phraséologie, équivalence, lexicographie, terminologie, traduction  
phraseology, equivalence, lexicography, terminology, translation  
fraseología, equivalencia, lexicografía, terminología, traducción

## **1. Introduction**

The Oxford English Dictionary online<sup>1</sup> defines *equivalence* as: “the condition of being equivalent; equality of value, force, importance, significance, etc.” (“equivalence, n.1” n.d.). This “non-specific definition” of the concept of equivalence is explained by Halverson (1997) in terms of “a relationship existing between two (or more) entities, and the relationship is described as one of likeness/sameness/similarity/equality in terms of any of a number of potential qualities.” (209). The notion of equivalence has been a matter of controversy in translation studies for decades (Halverson 1997; Leonardi 2000; Panou 2013). Several studies in other disciplines related to translation (e.g., terminology and lexicography) have shown similar interest in the concept of equivalence. However, equivalence tends to be a conflicting notion when linguistic mediators face the challenge of translating a text containing language for specific purposes (LSP) phraseology. This study aims to shed light on the characteristics of specialized phraseological entries and equivalents in lexicographic resources related to commerce and economics, and corpora.

To carry out this cross-linguistic study, I built a database consisting of SPU entries extracted from a dictionary of commerce and economics. The SPU entries chosen for this database belong to a subcategory denominated as specialized idioms (SpIs) identified in the work by Rojas Díaz (forthcoming). SpIs are characterized by their lack of a monolexical term within their word forms (in this case, related to commerce and economics, e.g., *in bad faith*, *door-to-door*, *de mala fe*, *puerta a puerta*) that undergo a certain level of idiomatity (lexical, semantic, or syntactic). 109 SpI entries and 174 equivalents in Spanish and English were selected for the database. This study offers (i) an overview of the notion of equivalence as treated in the previous literature in translation, lexicography, terminology, and phraseology; (ii) a characterization of SpI entries and equivalents through a series of lexicographic and semantic analyses which include the identification of metaphors and metonymies; and (iii) an analysis of translation techniques used in the generation of SpI equivalents to describe the decisions made by translators based on findings from queries made in the EUR-Lex corpus<sup>2</sup>.

---

<sup>1</sup> Available online at: <https://www.oed.com/>

<sup>2</sup> Available online at: <https://www.sketchengine.eu/eurlex-corpus/>

This paper is organized as follows: section 2 provides a literature review of the notions of equivalence within translation studies, lexicography, terminology, and phraseology. In section 3, concise definitions of the notions of term and SPU are offered. Section 4 deals with the data, tools, and methods used in this study. Section 5 presents examples from the analyses carried out to the SpI entries and their equivalents and descriptive statistical results from the queries made to the corpus. Finally, in section 6, the conclusions and suggestions for future research resulting from this study are presented.

## 2. The notion of equivalence: a general overview

### 2.1. Equivalence in Translation Studies

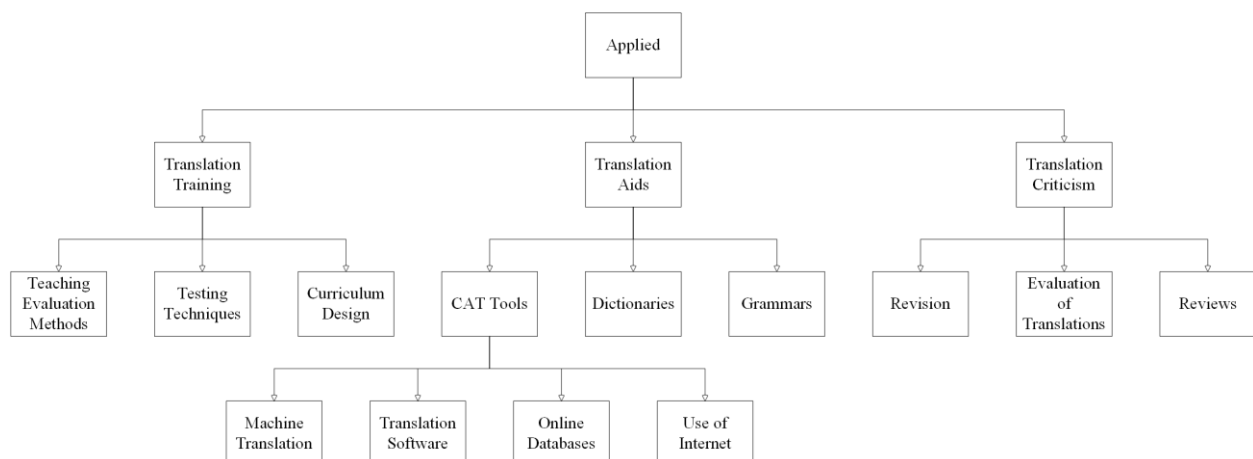
Several authors have offered their definitions regarding equivalence in translation studies (Jakobson 1959: 233; Catford 1965: 27; Kade 1968: 72-90; House 1977: 103; Toury 1981: 13; Nida and Taber 1982: 200-01; Wilss 1982: 145; Koller 1989: 100-01; Baker 1992: 11-12, 82-83, 217-18; Vinay and Darbelnet 1995: 342; Pym 2014: 67, 24-5) as reviewed by Halverson (1997); Leonardi (2000); and Panou (2013). Other authors have stated their concerns regarding the notion of equivalence on topics related to the symmetry between languages (Snell-Hornby 1995: 22), the superiority of source text (Vermeer 2012: 191-92), and equivalence vs. similarity (Chesterman 1996: 74). Pym's (2014) response to these debates was the notion of 'directional equivalence':

*Directional equivalence* is an asymmetric relation where the creation of an equivalent by translating one way does not imply that the same equivalence will also be created when translating the other way. (39, emphasis added)

Before Pym, Toury (1985: 36-37) tried to give equivalence a new home by moving equivalence to the "domain of applied extension of translation studies." He considered equivalence a "single target-source relationship" that "has little importance in itself" for descriptive translation studies (see Figure 1).

FIGURE 1

The map of the Applied Translation Studies schematized by Munday (2008: 12)



However, placing equivalence under 'applied translation studies' did not protect it from the criticism:

Translation involves so much more than the simple engagement of an individual with a printed page and a bilingual dictionary; indeed, the bilingual dictionary itself is an object lesson in the inadequacy of any concept of equivalence as linguistic sameness. (Lefevere and Bassnett 1995: 11)

Nonetheless, dictionaries and terminological databases are still being used worldwide as part of the translator's toolkit, resting on the implicit assumption that equivalence exists between listed forms. Moreover, generalized assumptions on bilingual dictionaries (Lefevere and Bassnett 1995), show the lack of connection between theoretical translation studies and the disciplines involved in applied translation studies, such as lexicography.

Dictionaries share the same epitome of equivalence in terms of being old-fashioned, prescriptive, and normative. However, it is undeniable that lexicography has changed through time (Trap-Jensen 2018: 26-34) offering a broader series of options depending on dictionaries' target users and intended functions.

## *2.2. Equivalence in lexicography*

Regardless of all the discussions around the notion of equivalence, it is a key concept for lexicographic work, especially in bilingual lexicography (Zgusta 1971: 312; Martínez de Sousa 1995: 193; Hartmann and James 1998: 51; Atkins and Rundell 2008: 468; Adamska-Sałaciak 2010: 397-99; Karpinska 2019: 38-39). In consequence, the position assumed by some translation theoreticians, about getting rid of the notion of equivalence, has been criticized by lexicography scholars:

Clearly, whatever the merits, or otherwise, of such an anti-equivalence stance in the study of translation, in lexicography a similar rejection would amount to throwing the baby out with the bathwater. (Adamska-Sałaciak 2010: 399)

Adamska-Sałaciak states that “it has always been taken for granted in our discipline [lexicography] that equivalence need not be symmetrical” (2010: 392). Similarly, Hartmann (1985) explains that “equivalence is a relative, fluid and relational concept: it does not exist until it has been established as a result of a bilingual conscious act” (123) and that “the convergences and divergences depend on the directionality in which one switches between the source and target languages” (128).

The double role of translation and translators as suppliers and consumers of lexicographic equivalents has been highlighted by Hartmann (2007); he asserts that this relationship has not been equally recognized:

Both of these give-and-take operations presuppose that active channels of awareness and collaboration exist. Unfortunately, and similar to our diagnosis of the neglect of interlingual aspects in dictionary research, we can observe relatively scant attention paid to lexicographic topics within translation. (Hartmann 2007: 211)

Despite the differences between lexicography and translation studies, Fuertes-Olivera (2011) points out that the notion of equivalence could be explored by looking outside lexicography. For instance, Adamska-Sałaciak (2010) adopts the ‘prototype concept,’ put forward by Halverson (1999), to explain lexicographic equivalence:

It rests on the assumption that a culturally embedded category (concept), containing central (prototypical) as well as less central members, may be related to corresponding categories in other languages (Adamska-Sałaciak 2010: 403).

For Adamska-Sałaciak (2010), Halverson's (1999) notion allows to define equivalence “as a broad spectrum of relations, from similarity to identity” (403).

### **2.3. Equivalence in terminology**

L'Homme (2020: 229) states that equivalence has not been as debated in terminology as it has been in other disciplines (e.g., lexicography and translation). Most of the time, the terminological work happens in multilingual settings in which terms designate 'specialized realities' for which establishing an equivalent is not as difficult as it could be for other cultural-bonded units. L'Homme's (2020: 230-34) notion of terminological equivalence describes a continuum between equivalence and non-equivalence as possible results of terminological work.

However, terminology has been described as the scenario where one-to-one equivalence tends to happen (Snell-Hornby 1995: 17). This idea was widely spread due to the terminological notion of monosemic reference that was put forward, among others, by Wüster (*Einführung in die Allgemeine Terminologielehre und terminologische Lexikographie: 1979*) (1998: 138)<sup>3</sup> and Felber (1984: 179-83). However, monosemic reference, has been widely criticized by terminology scholars, e.g., (Boulanger 1995; Cabré 1999; Faber Benítez 2009; Gaudin 2003; L'Homme 2005; Temmerman 2000, 1997), who highlight the frequent occurrence of terminological variation in LSP:

The same concept can often be designated by more than one term, and the same linguistic form can be used to refer to more than one concept. Furthermore, terms have distinctive syntactic projections, and can behave differently in texts, depending on their conceptual focus. This is something that happens in texts of all languages, and is a problem that translators inevitably have to deal with. (Faber Benítez 2009: 112)

Similarly to Faber Benítez, Wilss (1982: 131) considers that reducing LSP translation to "the dimension of simple terminological substitutions" is a naïve approach to the study of language communities. Moreover, Sager (1994) asserts that the different levels of coincidence of different terminological systems, can only be obtained after "detailed analyses of knowledge structures in two cultures with their linguistic representations" (55). Therefore, in terminology, the notion of equivalence (as in lexicography) involves more than a one-to-one unit transfer in which it is possible to find some cultural-bonded units.

Culture-bound terminology is defined by Faber Benítez and León Araúz (2014) as "a new approach towards specialized conceptualizations, which underlines that each community parcels reality in a different way. This generates culture-specific concepts and terms" (141). This approach is supported by several studies, e.g., Diki-Kidiri (2014); Kristiansen (2014), with statements such as: "concepts and conceptual structures are created and understood within a cultural, social and situational context" (Fernández-Silva, Freixa, and Cabré 2014: 2).

### **2.4. Equivalence in phraseology**

Jaskot (2016a) highlights three main reasons why translating PUs is a problem: (i) "cross-linguistic definition being still unclear," (ii) "powerful culture-anchored meaning," and (iii) "stylistic and connotative functions" (417). However, this issue was previously identified by Tytler (1907):

The chief difficulty he [a translator] has to encounter will be found in the translation of idioms, [...] those particular idiomatic phrases of which every language has its own collection; phrases which are generally of a familiar nature, and which occur most commonly in conversation, or in that species of writing which approaches to the ease of conversation. (135-37)

---

<sup>3</sup> Translation into Spanish by María Teresa Cabré Castellví

The idea of phraseology as a translation problem derived in some authors considering some PUs as untranslatable, e.g., Veisbergs (1994). This once-generalized idea has been demystified through time in the works by several authors (Morvay 1996: 728; Corpas Pastor 2003: 275). Zuluaga (2019) offers a solution to this problem:

Jakobson formulates the principle that all cognitive experience and its classification can be expressed in each and every one of the existing languages. Gaps can be filled by loans, calques, neologisms, paraphrases, circumlocutions, shifts of meaning, etc. (64)

Another solution is offered by Buendía Castro and Faber (2016: 392) through the use of “a semantically-based approach and the frequency-oriented approach” demonstrating how a semantic (conceptual) annotated database could solve many of the problems usually found in semasiological approaches when looking for SPUs.

Regarding equivalence grading, Corpas Pastor (2003: 208-09, 81-82) and Dobrovol'skij and Piirainen (2005: 78, 154) offer a continuum that goes from total equivalence to non-equivalence. However, similarly to Zuluaga (1999: 64), Mellado Blanco (2015) states that there is always a way to reproduce the meaning and the intention of a certain PU from L1 to L2 by employing different translation techniques. This means that either there is an equivalent or the equivalent does not exist because the “translation is inadequate.” Moreover, Mellado Blanco (2015: 155-56) explains that the lexicographic equivalence of PUs should distinguish (and record) the most representative uses of the PUs.

## ***2.5. Shared ground for phraseological equivalence***

Based on this literature review, a series of common characteristics that underlie an SPU equivalence were identified:

- From lexicography (Adamska-Sałaciak 2010: 397-99) and translation (Pym 2014: 24-25). The relation between the SPU and its equivalent is asymmetric, meaning that the creation of an equivalent by means of translation does not imply that the same equivalent will be created when translating the opposite way.
- From terminology (L'Homme 2020: 230-31). The SPU and its equivalent must denote the same concept in the same LSP domain.
- From LGP phraseology (Mellado Blanco 2015: 155). The SPU equivalent should be described along with its context of use including the possibility of equivalence through a single lexeme.

## **3. A concise definition of term and SPU**

L'Homme (2020) states that “there is no consensus about the notion of ‘term’” (55). This could be exemplified when checking the notions of Cabré (2000) and Faber Benítez and López Rodríguez (2012):

These units [TUs/terms] are, at the same time, similar and different from the lexical units of a language, denominated words by lexicology. Their specialized character can be identified through their pragmatic aspects and the mode of their meaning. Their signified is the outcome of negotiation among experts. This negotiation happens within the specialized discourse through the use that determines the meaning of each unit. (Cabré 2000: 14, my translation)<sup>4</sup>

---

<sup>4</sup> Original in French “*Ces unités sont en même temps semblables et différentes des unités lexicales d'une langue, appelées mots par la lexicologie. Leur spécificité se trouve dans leur aspect pragmatique et dans leur mode de signification. Leur signifié est le résultat d'une négociation entre experts. Cette négociation se produit dans le discours spécialisé à travers des prédications qui déterminent le signifié de chaque unité*”.

Trying to find a distinction between terms and words is no longer fruitful or even viable, and the best way to study specialized-knowledge units is by studying their behavior in texts. (Faber Benítez and López Rodríguez 2012: 22)

However, Faber Benítez and López Rodríguez (2012: 22-23) explicitly mention characteristics that their definition shares with Cabré's (2000) notion of term, such as the preponderance of nominal units, the intrinsic relation between the TUs and LSP domains, and the combinatorial aspect of TUs. Thus, the main difference of these two notions rests in the approach that each of these theoretical postulations have regarding the object of study of terminology.

Regarding terminological work, L'Homme (2020) states that it could be carried out aiming at objectives related to tasks such as “dictionary compilation, translation, knowledge modeling, document indexing, and standardization” (59). This idea suggests that the different terminological theories are complementary rather than exclusive, depending on the task they are framed into. For instance, Cabré's (2000) notion has been proved to be suitable for semasiological work (e.g., Terminus 2.0<sup>5</sup>) while Faber Benítez and López Rodríguez's (2012) approach has enriched the work of the terminological knowledge database EcoLexicon<sup>6</sup>.

The notion of term used in this study offers a middle point between the two previously presented notions and it is offered by L'Homme (2020):

Stating that a linguistic item is a term is considering its meaning from the perspective of a special subject field. There is no such a thing as a term in essence; a linguistic unit becomes a term relative to their subject field in which it is considered. [...] This also means that even common linguistic items can become terms in specialized domains. [...] Finally, a linguistic item can also be a relevant term in fields of knowledge. (59)

The definition of L'Homme goes in line with the definition of SPU chosen for this study:

[SPU] is a combination of lexical units (including, but not necessarily, terminological word forms) that evidences idiomaticity at least at one of the possible levels (lexical, pragmatic, semantic, statistical, or syntactic) and that when used in a certain LSP domain acquires a specialized meaning. (Rojas Díaz forthcoming)

The study by Rojas Díaz (forthcoming) included a taxonomy based on the works by Mel'čuk (1998, 2012, 2013) in which SPU is considered the hypernym of three categories (specialized idioms, specialized collocations, and specialized pragmatemes). In this article, SpIs were chosen due to the lack of studies and data that could shed light on how they behave in a certain LSP domain, in comparison to the exhaustive existing studies on specialized collocations.

SpIs are characterized by the absence of a monolexical term among its word forms evidencing the occurrence of lexical, semantic, or syntactic idiomaticity. Baldwin and Kim (2010: 269-71) assert that idiomaticity can occur at different linguistic levels (see Table 1).

TABLE 1

**Idiomaticity levels and their definitions (Baldwin and Kim 2010: 269-71)**

Level	Definition
Lexical idiomaticity	Occurs when one or more components of an MWE are not part of the conventional English lexicon.
Semantic idiomaticity	Is the property of the meaning of a MWE not being explicitly derivable from its parts.

<sup>5</sup> Available online at: <http://terminus.iula.upf.edu/cgi-bin/terminus2.0/terminus.pl>

<sup>6</sup> Available online at: <http://ecolexicon.ugr.es/en/index.htm>

Statistical idiomaticity	Occurs when a particular combination of words occurs with markedly high frequency, relative to the component words or alternative phrasings of the same expression.
Syntactic idiomaticity	Occurs when the syntax of the MWE is not derived directly from that of its components.

Semantic idiomaticity happens through figuration that can be identified in the occurrence of metaphors or metonymies. Metaphor could be defined as “the use of language to refer to something other than what it was originally applied to, or what it ‘literally’ means, in order to suggest some resemblance or make a connection between the two things” (Knowles and Moon 2006: 3). Moreover, metonymy “broadly [...] involves part-and-whole relations and associations. The word for a part of something is used to refer to the whole, or else the whole is referred to in terms of something associated with it” (Knowles and Moon 2006: 6).

#### 4. Data, tools, and methods

This study employs the same database used by Rojas Díaz (forthcoming) which contains 11,086 SPUs (39,832 word forms) extracted from the entries and sub-entries of the *Diccionario de Comercio Internacional* (Alcaraz and Castro Calvín 2007) (DCI).

The DCI was chosen based on three criteria: “(i) that it be related to Commerce and Economics, (ii) that it be bilingual (in English and Spanish), and (iii) that its publishing house be recognized as a lexicographic authority” (Rojas Díaz forthcoming). Moreover, using a dictionary to build the analysis database, ensures the occurrence of terms from several levels of abstraction (Hoffmann 1998: 72-73) and different users’ needs (Kübler and Pecman 2012: 187). Furthermore, Kübler and Pecman (2012) explain that English and Spanish offer an ideal pair for a cross-linguistic analysis, since both are the official languages of several international organizations related to commerce and economics (e.g., the Economic Commission for Latin America and the Caribbean [ECLAC] and the World Trade Organization [WTO]).

The corpus selected for the queries was the EUR-Lex corpus (Baisa et al. 2016: 2800-02) which was queried through Sketch Engine<sup>7</sup>. The EUR-Lex corpus was the most suitable option since it is one of the largest multilingual corpora available in Sketch Engine (with approximately 633.4 million words translated from Spanish into English and 594.1 million words translated from English into Spanish). The corpus contains texts that range from regulations, decisions, and preparatory acts to treaties, international agreements and EFTA documents (Baisa et al. 2016: 2800).

##### 4.1. Selection of the analysis subset

The SPUs of the database used by Rojas Díaz (forthcoming) contain between three and five word forms (see Figure 2). Some fixed elements (e.g., commas, hyphens, and slashes) were also classified as word forms in this study.

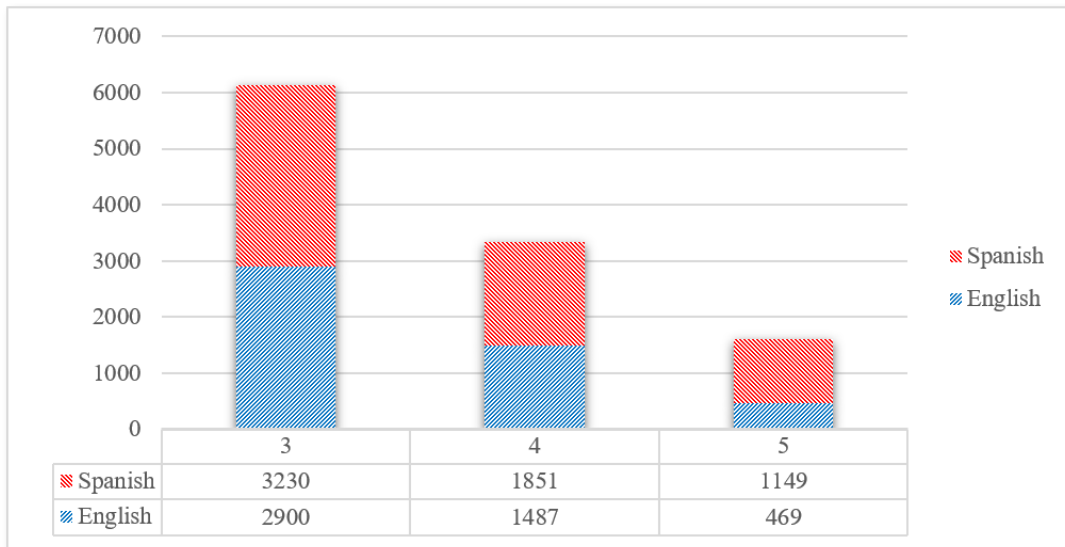
FIGURE 2

**Distribution of SPUs according to their number of word forms (Rojas Díaz, forthcoming)<sup>8</sup>**

<sup>7</sup> Available online at: <https://www.sketchengine.eu/>

<sup>8</sup> A code of colors will be used to identify each language results: red for Spanish and blue for English. Literal translations between single quotation will be offered for the examples in Spanish when the equivalent is not presented next to them.

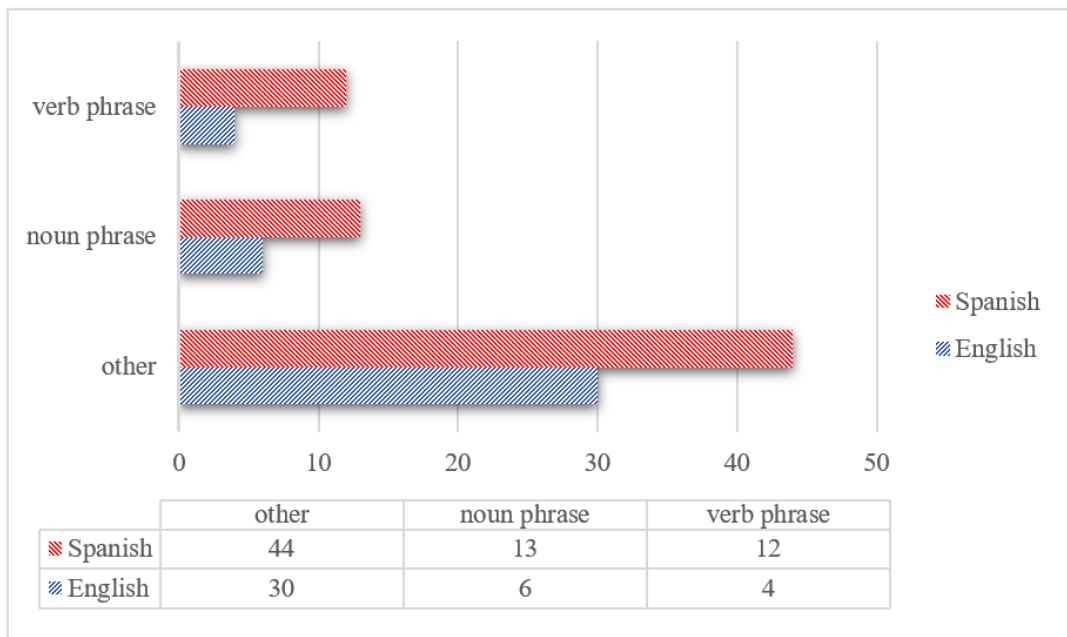




From the analyses carried out by Rojas Díaz (forthcoming: 16-29), 109 SpIs were identified (69 in Spanish and 40 in English) corresponding to 0.99% of the units in the database (see Figure 3) which are the units that conform the analysis subset of the present study.

FIGURE 3

**Phrase types corresponding to the SpIs in the analysis subset**



## 5. Results and analysis

### 5.1. Entries

During the extraction process, the LSP domains, in which the SpIs are classified, were identified. Examples taken from the database in each language are given for each LSP domain in Table 2. Only one occurrence in the LSP domains of economics, marketing, accounting, insurance, and human resources was found in the database (see Table 2).

TABLE 2

**LSP domains of the DCI**

LSP domain (Dictionary LSP marking) Example	Freq. in English	Freq. in Spanish
Law (DER/LAW) <i>con la salvedad de</i> ‘with the exception of’ <i>in bad faith</i>	13	19
Logistics (LOGÍST/LOGIST) <i>a son de mar</i> ‘at the pace of the sea’ <i>gate to gate</i>	17	11
General (GRAL/GEN) <i>a dos bandas</i> ‘at two bands’ <i>at the latest</i>	2	24
Finance (FIN/FIN) <i>dorso con dorso</i> ‘back with back’ <i>at short notice</i>	7	10
Economics (ECON/ECON) <i>los cinco dragones</i> ‘the five dragons’	0	4
Documentation (DOC/DOCMT) <i>dar cuenta de</i> ‘give account of’ <i>at sight draft</i>	2	2
Taxation (FISC/TAXN) <i>perfect the sight</i>	2	0
Marketing (MERCAD/MKTG) <i>al por mayor</i> ‘in bulk’	0	2
Accounting (CONTA/ACCTG) <i>above-the line</i>	1	0
Insurance (SEGUR/INSCE) <i>sweat in hold</i>	1	0
Human resources (REC. HUM/HH.RR.) <i>fuga de cerebros</i> ‘leak of brains’	0	1

As observed in Table 2, the total count of the LSP domains (118) is higher than the number of units of analysis (109) because several entries are classified in more than one LSP domain (see Figure 4). Regarding the marking of LSP domains suggested by the dictionary, 22.03% of the SpIs are used in general contexts, while 50.84% are related to law and logistics (see Table 2).

FIGURE 4

**Examples of multiple LSP domain marking in the DCI in (1) Spanish and (2) English**

(1)

**poner en clave** (LAW/GEN  
code; S. *codificar, cifrar*),

(2)

**sight draft, at**  
(DOC/FIN letra a la vista; también llamada *at sight draft, demand bill/draft, sight bill/draft, sight bill of exchange, presentation draft* en EE.UU.; V. *documents against acceptance sight draft, acceptance, date draft, time draft*),

*5.1.1. Entries: idiomaticity and semantics*

As stated in section 3, a characteristic of SpIs is the occurrence of lexical, semantic, or syntactic idiomaticity. Lexical idiomaticity was not found in the SpI entries. Table 3 presents the idiomaticity levels found and their frequency in the subset.

TABLE 3

**Idiomaticity of the SpIs**

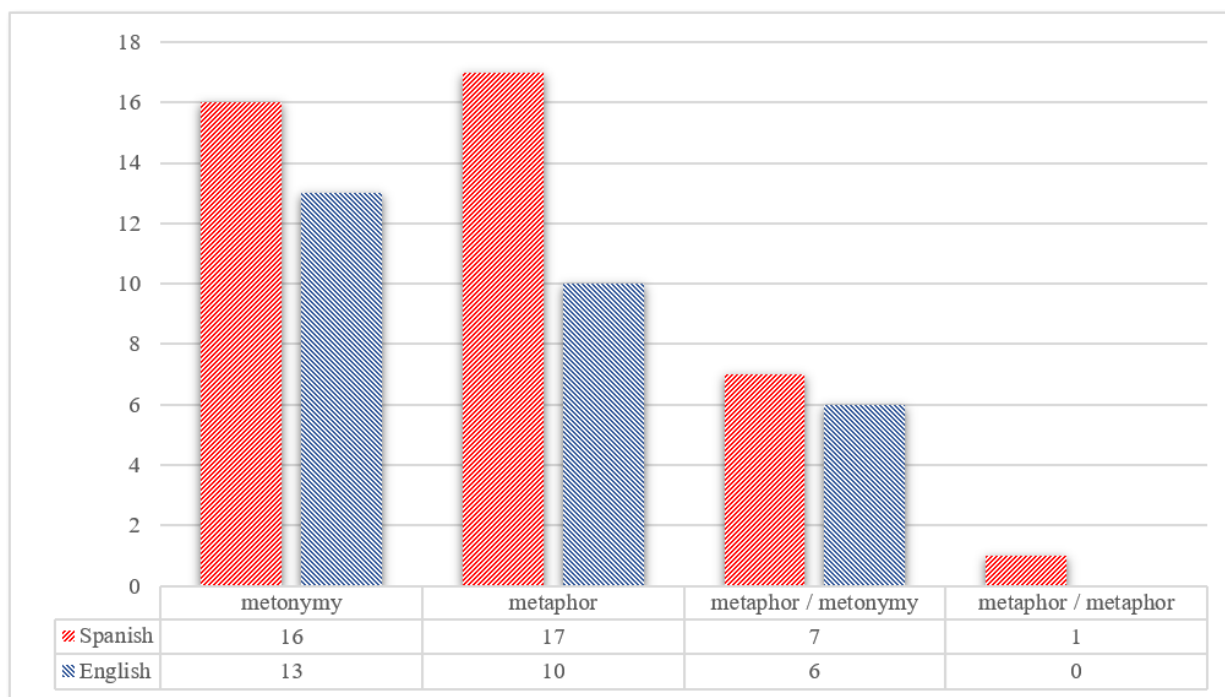
Language	Semantic (freq.)	Syntactic (freq.)
Spanish	<i>dar de alta</i> 'to give the discharge' (metaphor + metonymy) (part of entity for whole entity) (41)	<i>a efectos de</i> 'to the effects of' (28)
English	<i>door-to-door</i> (metonymy) (part of event for whole event) (29)	<i>in due time</i> (40)

Syntactic idiomaticity can be identified in the examples presented in Table 3. Both SpIs (*a efectos de* and *in due time*) are adverbial in nature but they consist of a series of lexical elements from which it is not possible to derive an adverbial expression directly. The examples of semantic idiomaticity in Table 3 show the occurrence of metaphors, metonymies or a combination.

For instance, *dar de alta* 'join/register/discharge oneself from hospital'/'return to work after sickness'/'be off work through sickness' contains a metaphor; in the abstraction of the verb *dar* 'to give', and a metonymy; in the use of the word *alta* originally defined in Spanish as a document issued in the military to discharge a person from duty. Figure 5 presents the frequency of occurrence of metaphors and metonymies in the analysis subset.

FIGURE 5

**Frequency of metaphors and metonymies in the subset**



Knowles and Moon (2006) state that PUs often are conventional metaphors, i.e., “kinds of metaphor [that] are institutionalized as part of the language. Much of the time, we hardly notice them at all, and do not think of them as metaphorical when we use or encounter them” (6). Other times, PUs are metonymic rather than metaphoric in nature, but those metonymies

also tend to be conventionalized. The subset presents similar behavior in terms of frequency of occurrence of metaphors and metonymies (see Figure 5). These results might indicate that the semantic behavior of SPUs is similar to that of PUs.

5.1.2. Entries in corpus

Two queries were carried out in the corpus to find out (i) whether the SpIs from the sample could be found in corpora or not, and (ii) their frequency of occurrence. The results of the first query are presented in Figure 6.

FIGURE 6

No. of SpIs 'found' and 'not found' in the corpus

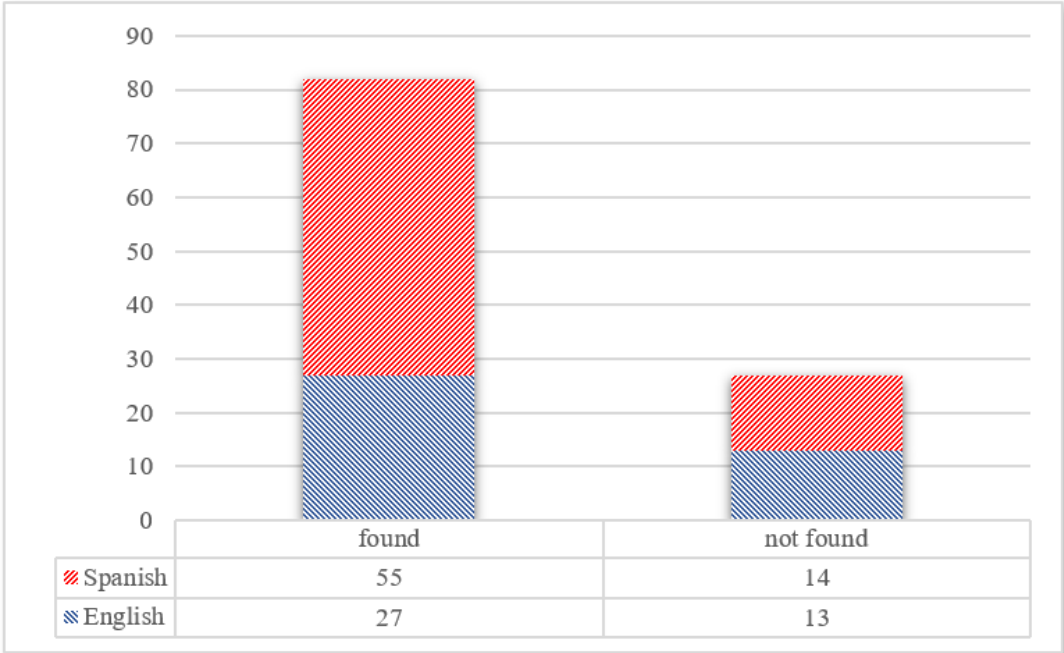
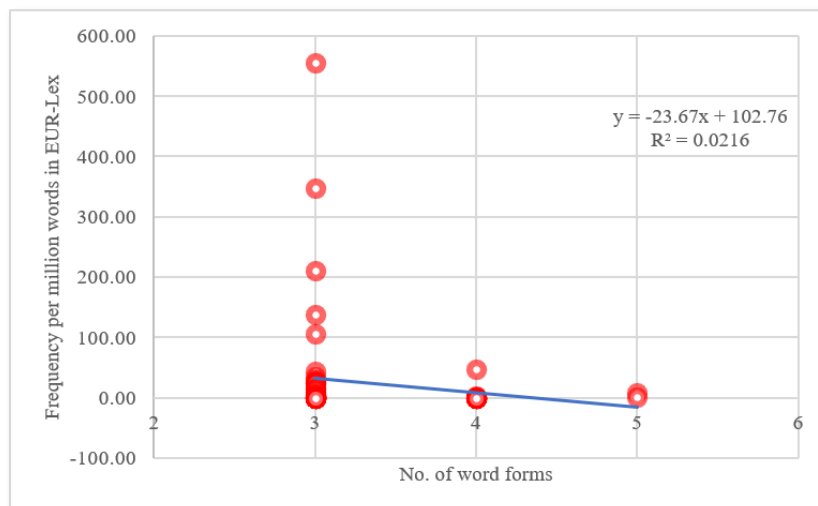


Figure 7 shows the frequency of occurrence in corpora expressed in terms of frequency 'per million words' (pmw).

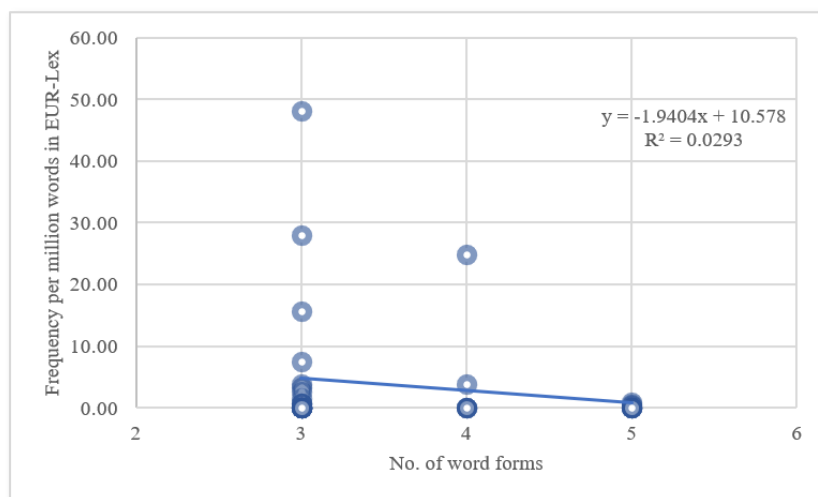
FIGURE 7

Frequency of SpIs in (1) Spanish (red) and (2) English (blue) in the corpus

(1)



(2)



As expected, the frequency of occurrence in the corpus shows a negative correlation between the number of word forms and the frequency of the queried SpIs. The frequencies of occurrence of SpIs in the corpus range from 0.01 to 556.05 pmw in Spanish, and from 0.01 to 48.11 pmw in English. However, the analysis of the coefficient of determination (2.16% in Spanish and 2.93% in English [See  $R^2$  values in Figure 7]) demonstrates the unevenness of SpIs' frequency-of-occurrence distribution according to their number of word forms.

According to Sinclair and Moon (1997: xvii), an idiom is considered to be frequent if it occurs at least one time every two million words in a corpus (0.5 pmw), thus asserting that LGP idioms do not occur frequently in corpora. However, the statistical mean of the SpI entries in the corpus (24.55 times pmw in Spanish and 3.98 times pmw in English [see Figure 10]) shows a high rate of occurrence. Based on Sinclair and Moon's (xvii) measuring scale, the subset offered, in Spanish, 36 SpIs that occurred 0.5 times or more pmw, 19 that occurred less than 0.5 times pmw, and 14 that did not occur in the corpus. In English, 13 SpIs occurred 0.5 times or more pmw, 14 SpIs that occurred less than 0.5 times pmw, and 13 with no occurrences in the corpus. However, a closer analysis of the outliers of the sample (see Figure 7) shows how some SpIs, (particularly in Spanish [see Table 4]), have a high occurrence that causes the sample to be uneven (including the groups with the same word form length). Table 4 presents the top three most frequent SpIs of the sample in each language.

TABLE 4

**Most frequent SpIs in the corpus**

Language	Three word forms (freq. pmw)	Four word forms (freq. pmw)	Five word forms (freq. pmw)
Spanish	<i>de conformidad con</i> 'of acceptance with' (556.05)	<i>a lo largo de</i> 'at the length of' (46.56)	<i>con efectos a partir de</i> 'with effects from' (7.26)
	<i>en virtud de</i> 'in virtue of' (347.84)	<i>de ida y vuelta</i> 'of going and returning' (2.52)	<i>por cuenta y riesgo de</i> 'for the account and risk of' (0.25)
	<i>a efectos de</i> 'to the effects of' (211.47)	<i>con la salvedad de</i> 'with the exception of' (1.58)	
English	<i>at the latest</i> (48.11)	<i>on the ground that</i> (24.80)	<i>door - to - door</i> (0.87)
	<i>not later than</i> (27.99)	<i>at the expense of</i> (3.92)	<i>back - to - back</i> (0.47)
	<i>on account of</i> (15.63)	<i>at arm's length</i> (0.08)	<i>in good and due form</i> (0.38)

## 5.2. Equivalentents

The DCI offers 174 equivalentents (112 in Spanish and 62 in English) for the 109 SpI entries of the subset. Regarding the equivalentents offered by the dictionary, two main characteristics can be drawn from the analyzed data. First, the dictionary registered cases of terminological variation of SpIs (see Table 5) as well as polysemy in both languages (see Table 6).

TABLE 5

### Examples of SpI terminological variation in the DCI

Language pair direction	Entry	Equivalent
Spanish - English	<i>dar cuenta de</i>	<i>report</i>
	<i>dar parte de</i>	
English - Spanish	<i>door-to-door</i>	<i>puerta a puerta</i>
	<i>gate-to-gate</i>	
	<i>warehouse-to-warehouse</i>	

SpI polysemy could reflect the different uses of the expressions in the LSP domain in which they are used (see Table 6). However, as argued in section 4 (above), defining LSP domains or sub-domains related to economics could be troublesome. For instance, the expression *at arm's length* was tagged as 'general.'

TABLE 6.

### Examples of SpI polysemy in the DCI

Language pair direction	Entry	Equivalent
Spanish - English	<i>hacer frente a</i>	<i>counter</i>
		<i>face</i>
		<i>honor</i>
		<i>honour</i>
		<i>meet</i>
English - Spanish	<i>at arm's length</i>	<i>a raya</i>
		<i>distanciado</i>
		<i>de igual a igual</i>
		<i>manteniendo las distancias</i>
		<i>sin concederse favores</i>
		<i>con total independencia</i>
		<i>en condiciones de plena concurrencia</i>

		<i>con escasas muestras de cordialidad</i>
--	--	--

Nevertheless, some of the equivalents offered by the DCI (see Table 6) suggest that *at arm's length* has a specialized meaning which is explicitly mentioned in the OED entry (see Table 7).

TABLE 7

**General and specialized definitions of *at arm's length***

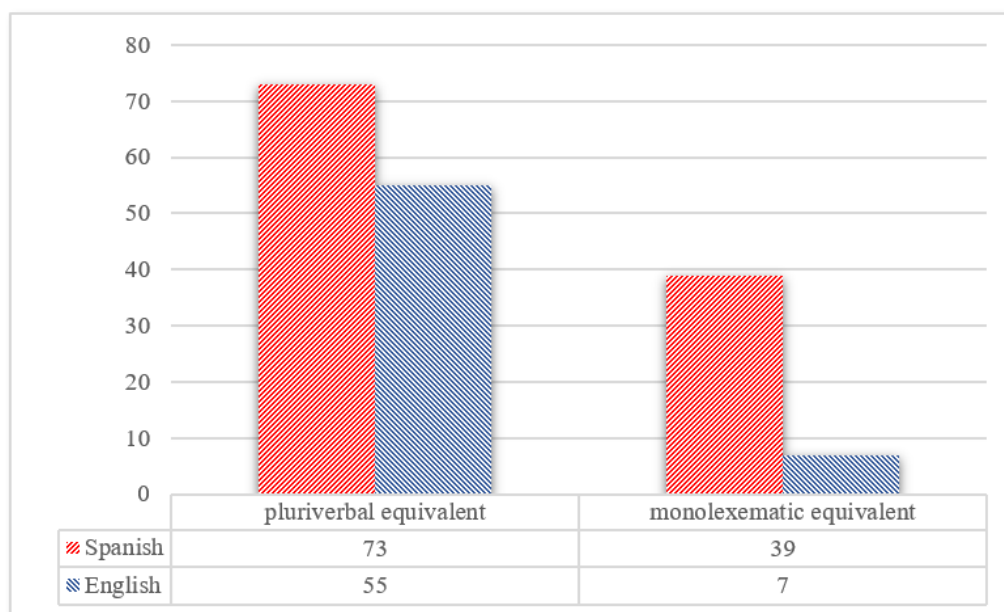
SPU	LGP/LSP	Definition	Source
<i>at arm's length</i>	LGP	“As far as one can reach with one’s arm; (hence) at a distance, remote, not on familiar or friendly terms.”	OED (“arm, n.1” n.d.)
	LSP	“ <i>Law.</i> Of two parties: without legal obligations to each other, esp. fiduciary obligations; (also more generally) in an independent or impartial position; conducted by independent or impartial parties.”	OED (“arm, n.1” n.d.)
	LSP	“ <i>fr. GRAL</i> a raya, distanciado; de igual a igual; manteniendo las distancias, sin concederse favores; con total independencia; en condiciones de plena concurrencia; con escasas muestras de cordialidad.”	DCI (Alcaraz and Castro Calvín 2007: 37)

Second, as observed in Tables 5-8 and in Figure 8, the dictionary presents monolexematic units as possible SpI equivalents. These findings are in line with what Jaskot (2016b) stated regarding LGP phraseological equivalents:

The form is of secondary importance, as equivalence can be achieved by using a simple lexeme and not necessarily a PU” (contrary to what we aim to preserve at the systemic level). (Jaskot 2016b: 63)

FIGURE 8

**Frequency of monolexematic and pluriverbal equivalents in the sample**



Both Dobrovol'skij (2000: 364), and Jaskot (2016b: 60) agree on stating that the main difficulty to set a PU as the equivalent of another PU lies in the meaning grading of both expressions in certain contexts just as it could be observed in the SpIs analyzed in the present study. Moreover, terminological variation and polysemy show the asymmetric characteristic of equivalence put forward by Adamska-Sałaciak (2010: 397-99) and Pym (2014: 24-25).



TABLE 8

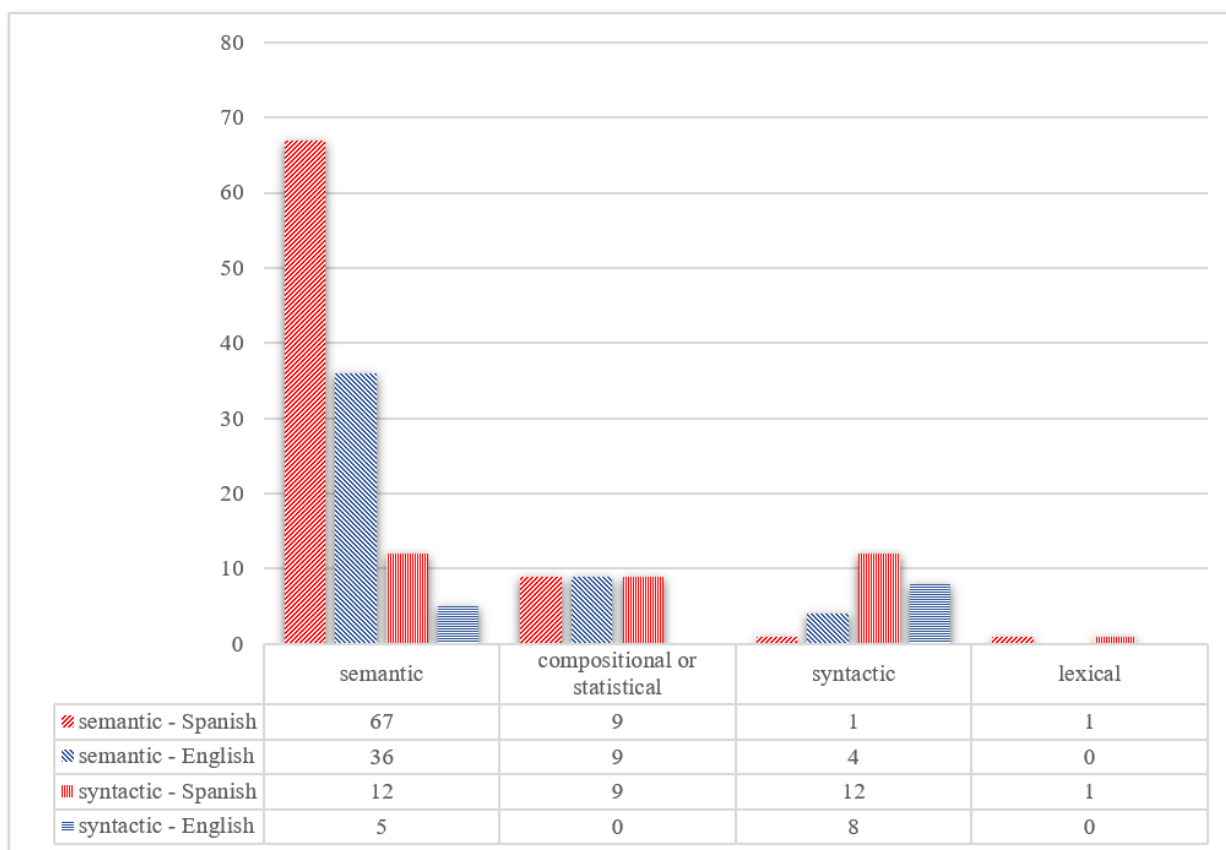
**Examples of monolexical and pruriverbal equivalents in the DCI**

Language pair direction	Entry	Equivalent
Spanish - English	<i>a tres bandas</i>	<i>tripartite</i>
		<i>three way</i>
	<i>dar de baja</i>	<i>cancel</i>
		<i>strike off the rolls</i>
English - Spanish	<i>back-to-back</i>	<i>subsidiario</i>
		<i>dorso con dorso</i>
	<i>under cover of</i>	<i>amparado</i>
		<i>cubierto por</i>

*5.2.1. Equivalents: idiomaticity and semantics*

Equivalents were analyzed to identify the occurrence of any sort of idiomaticity (see Figure 9). This analysis focused on the interaction of idiomatic relations between entries and the equivalents being offered, i.e., if idiomaticity was preserved in the proposed equivalents or not (see Tables 9-13).

FIGURE 9

**Equivalents idiomaticity of the SpI subset**

A trend of semantic idiomaticity occurrence among SpI equivalents is apparent. Nevertheless, some SpI entries with semantic idiomaticity have equivalents that show other sorts of idiomaticity (or even the lack of it) becoming compositional. Some examples of entries containing metaphors and their equivalents are presented in Table 9.

TABLE 9



### Examples of idiomatic interaction between SpI entries with metaphors and their equivalents

Language pair direction	Entry	Equivalent (type of idiomaticity)
Spanish – English	<i>los cinco dragones</i>	<i>five dragons</i> (semantic - metaphor)
	<i>hacer frente a</i>	<i>honor</i> (semantic - metonymy) (member of category for whole event)
	<i>pasar por alto</i>	<i>waive</i> (compositional or statistical)
English - Spanish	<i>break the deadlock</i>	<i>desbloquear las negociaciones</i> (semantic - metaphor)
	<i>take the ground</i>	<i>embarrancar</i> (semantic - metonymy) (part of event for whole event)
	<i>under the date of</i>	<i>fechado a</i> (compositional or statistical)

Metonymies were also identified in the SpI entries. Some examples of their equivalents are presented in Table 10.

TABLE 10

### Examples of idiomatic interaction between SpI entries with metonymies and their equivalents

Language pair direction	Entry (metonymical relation)	Equivalent (type of idiomaticity)
Spanish – English	<i>llave en mano</i> (part of entity for whole entity) (part of event for whole event)	<i>turnkey</i> (semantic - metaphor)
	<i>de ida y vuelta</i> (feature for whole event)	<i>out and home</i> (part of event for whole event)
	<i>de banda a banda</i> (part of entity for whole entity)	<i>athwartships</i> (compositional or statistical)
English - Spanish	<i>above the line</i> (feature for whole entity)	<i>ingresos y gastos corrientes</i> (semantic - metaphor)
	<i>house-to-house</i> (part of event for whole event)	<i>a domicilio</i> (semantic - metonymy) (part of event for whole event)
	<i>door to door</i> (part of event for whole event)	<i>transporte desde las instalaciones del consignador hasta el puerto de importación</i> (compositional or statistical)

Moreover, some equivalents presented semantic idiomaticity while their corresponding SpI entries presented syntactic idiomaticity (see Table 11).

TABLE 11

### Examples of idiomatic interaction between SpI entries with syntactic idiomaticity and their equivalents

Language pair direction	Entry	Equivalent (type of idiomaticity)
Spanish - English	<i>con la salvedad de</i>	<i>subject to</i> (semantic - metaphor)
	<i>al por mayor</i>	<i>wholesale</i> (semantic-metonymic) (part of event for whole event)
	<i>con inclusion de</i>	<i>inclusive of</i>
English - Spanish	<i>in regular turn</i>	<i>cláusula de atraque</i> (semantic-metonymy) (part of event for whole event)

	<i>at the latest</i>	<i>a mas tardar</i> (syntactic)
--	----------------------	------------------------------------

Two of the SpIs analyzed (in Spanish) offered equivalents with lexical idiomaticity (see Table 12); however, they could not be analyzed in the corpus. The equivalent of *a tanto alzado* and the SpI entry *en buen fin* did not appear in the corpus.

TABLE 12

**Examples of idiomatic interaction between SpI entries and equivalents with lexical idiomaticity**

Language pair direction	Entry	Equivalent (type of idiomaticity)
Spanish – English	<i>a tanto alzado</i>	<i>à forfait</i> (lexical from French)
	<i>en buen fin</i>	<i>bona fide</i> (lexical from Latin)

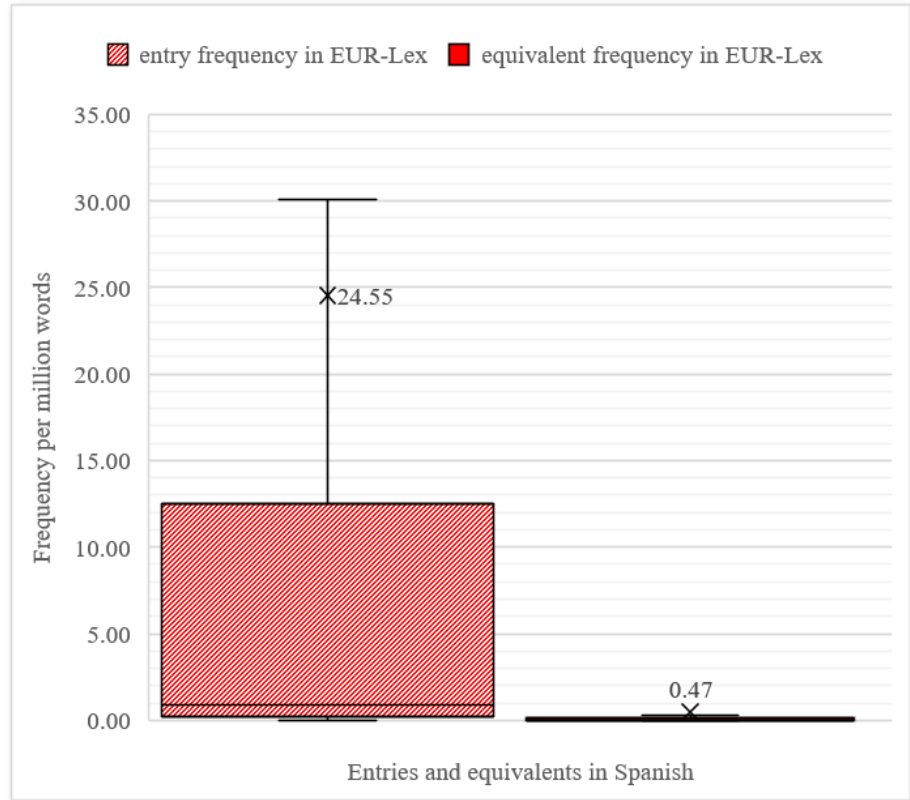
*5.2.2. Equivalents in corpus and translation techniques*

L'Homme (2020: 238) states that besides morphosyntactic patterns and their correspondence, contrasting bilingual texts and the terms they contain offer valuable information regarding divergences. One of these divergences appears when contrasting the frequency of SpI entries and their equivalents in the corpus (see Figure 10). The SpI entries that did not occur in the corpus were excluded from the analysis.

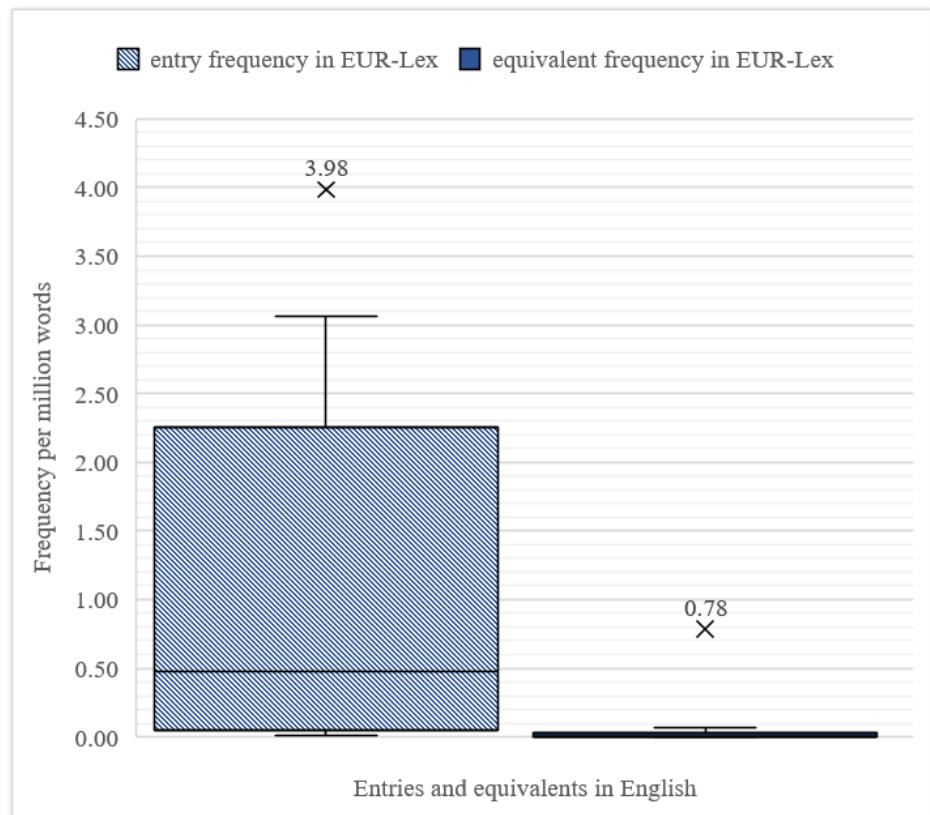
FIGURE 10

**Frequency of SpIs and their equivalents. X is the mathematical mean of occurrences of entries and equivalents in the corpus for both languages (1) Spanish and (2) English**

(1)



(2)



As seen in Figure 10, the difference in frequency (of occurrences) of SpIs in Spanish and English causes the scale to be different for the two languages. This difference shows how the Spanish SpIs occurred at a higher ratio than the English SpIs (see Table 13).

Furthermore, the size difference of entry and equivalent boxes and whiskers in (1) and (2) (see Figure 10) suggest that, on the one hand, the DCI does not register all the possible equivalents found in the corpus. This goes in line with Mellado Blanco's (2015: 155-56) statement (see Section 2.4.) regarding the boundaries of lexicographic equivalence of PUs. On the other hand, the graphs also suggest the lack of indexation (in both languages) of the most common uses of the equivalents of SpI entries in the DCI. The evident differences between the SpIs indexed in the DCI and what was found in the corpus raises the question of whether the former is corpus-based or not.

The numerical findings were followed by the analysis of those equivalents that were not registered in the dictionary but occurred in the corpus. However, one limitation of this analysis is the amount of data that could be presented. Consider for instance, the occurrences of the two most frequent SpIs in Spanish and English (see Table 13).

TABLE 13

**Top two outliers based on the frequency data retrieved from the corpus**

Spanish entry	(freq.) (whole corpus)	(freq.) (parallel SPA-ENG corpus)	English equivalent	(freq.) (parallel SPA-ENG corpus)
<i>de conformidad con</i> 'of acceptance with'	451,333	357,629	<i>under</i>	1,155
<i>en virtud de</i> 'in virtue of'	282,330	230,179	<i>under</i>	3,429
English entry	(freq.) (whole corpus)	(freq.) (parallel ENG-SPA corpus)	Spanish equivalent	(freq.) (parallel ENG-SPA corpus)
<i>at the latest</i>	40,659	30,504	<i>a más tardar</i> 'at more late'	23,953
<i>not later than</i>	23,650	17,500	<i>en un plazo no superior a</i> 'in a period not higher to'	43

Since the EUR-Lex is a translation corpus, not all the occurrences in the whole corpus are registered in the parallel (Spanish-English / English-Spanish) sub-corpus. Therefore, the differences between the frequencies offered by *Sketch Engine* on a particular SpI are expected. However, a problem arises when checking the correspondence between the occurrence of a given expression and that of its equivalent. For instance, there were 357,629 occurrences of *de conformidad con* 'of acceptance with' in the EUR-Lex Spanish-English parallel corpus and 1,155 occurrences of its equivalent *under*. In other words, a large proportion of the corpus data that is unaccounted for should be checked in order to have 'the big picture' of the cross-linguistic analysis of *de conformidad con* (and its equivalents into English) in the corpus.

Therefore, it was decided to show some of the corpus concordance findings of four SpIs from the subset, two in Spanish (*dar de baja* 'to give the discharge' and *de mala fe* 'in bad faith') and two in English (*above the line* and *in due time*). The following concordances will exemplify decisions made by translators when dealing with SpIs not registered as equivalents in the DCI. It was decided to adopt the classification of translation techniques offered by Molina and Hurtado Albir (2002: 509-11) for this task.

The first SpI to be analyzed was *dar de baja* 'to give of discharge' (see Figure 11).

FIGURE 11

**Lexicographic entry for *dar de baja* in the DCI**

**dar** v: GEN give. [Exp: **dar como garantía** (LAW charge; S. *afectar*), **dar cuenta de** (DOCMT account, report; S. *informar, comunicar*), **dar de baja** (LAW cancel, charge off; strike off the list/rolls; remove from membership/subscription; S. *cancelar, anular, rescindir, invalidar, suspender*), **dar entrada** (GEN/ACCTG admit;

There were 1,386 occurrences of *dar de baja* in the corpus. However, only one of the six equivalents offered by the DCI (*cancel*), appeared in the corpus (15 occurrences). Some examples of the concordances from the the corpus are presented in Table 14.

TABLE 14

**Concordances of *dar de baja* from the corpus**

Translation technique	SL (Spanish) segment	TL (English) segment
(1) -Modulation -Linguistic compression	<i>así como el importe <b>dado de baja</b> contablemente de esas partidas y el importe abonado por recuperación de saldos, previamente <b>dados de baja</b>, que han sido recuperados con posterioridad;</i>	the amount charged in the period for loans and advances <b>written off</b> and the amount credited in the period for loans and advances previously <b>written off</b> that have been recovered;
(2) -Linguistic compression -Modulation -Transposition	<i>Las entidades incluirán en este campo el efectivo recibido [...] y operaciones de préstamo con reposición del margen que mantengan en el balance (es decir, respecto de los cuales no se cumplan los criterios para <b>darlos de baja</b>).</i>	Institutions shall include in this field cash received [...] and margin lending transactions and that is retained on the balance sheet (i.e. the accounting criteria for <b>derecognition</b> are not met).
(3) -Modulation -Linguistic compression	<i>Cualquier persona podrá ser <b>dada de baja</b> de la lista de representantes autorizados en las condiciones establecidas con arreglo al artículo 93 bis</i>	A person may <b>be removed from the list</b> of professional representatives under the conditions established in accordance with Article 93a
(4) -Modulation	<i>En tercer lugar, a pesar de que tras el acuerdo de fusión alcanzado el 29 de septiembre de 1998 las partes ofrecieron a sus miembros la posibilidad de <b>darse de baja</b> en un plazo de dos semanas, mediante notificación inmediata, ninguno de los ganaderos se adhirió a una cooperativa extranjera ni inició actividades de exportación.</i>	Thirdly, no pig farmer joined a foreign cooperative or started export operations despite the opportunity given to the members of the parties to <b>terminate their membership</b> with immediate notice during a period of two weeks after the agreement of the parties to merge reached on 29 September 1998.

*Dar de baja* presents semantic idiomatization characterized by the interaction between a metaphor (in *dar* ‘to give’) and a metonymy (in *baja*, ‘discharge’ originally defined as a document issued as a proof of discharge from the military service, [‘part of entity for whole entity’ relation]). In example (1), the equivalent found in the corpus is another verbal SpI (*write off*) in which the semantic idiomatization is evident in a metonymy (in ‘write,’ [‘part of event for whole event’ relation]).

Both, examples (1) and (2) evidence modulation, “to change the point of view, focus or cognitive category in relation to the ST” (Molina and Hurtado Albir 2002: 510), and linguistic compression, “to synthesize linguistic elements in the TT” (Molina and Hurtado Albir 2002: 510). However, example (2) also shows the use of transposition, “to change a grammatical category.” (Molina and Hurtado Albir 2002: 510), and the equivalent offered is a nominalized monolexeme, a possibility that was already stated by Mellado Blanco (2015: 155) (see section 2.4.).

On some occasions the DCI does not offer enough contextual information. On the one hand, examples (1) and (2) are related to figures and accounting. On the other hand, examples (3) and (4) deal with memberships. The equivalents offered by the DCI do not include context of use that could allow the DCI user a better understanding of these SpI equivalents. The lack of this information goes against what Mellado Blanco's (2015: 155) suggests (see section 2.4.).

Example (3) (*to be removed from the list*) presents semantic idiomaticity, characterized by the occurrence of a metonymy (*list* ['part of entity for whole entity' relation]). Finally, in example (4) there is no identifiable evidence of semantic, syntactic, or lexical idiomaticity configuring a specialized collocation, as defined by (Rojas Díaz forthcoming).

The second SpI to be analyzed was *de mala fe* 'in bad faith' (see Figure 12).

FIGURE 12

**Lexicographic entry of *de mala fe* in the DCI**

**mal** *adv/n*: GEN badly; wrongly, mistakenly; poorly; ill, wrong; harm. [Exp: **mal definido** (GEN gray US, grey; S. *gris, ambiguo*), **mala fe** (LAW bad faith, mala fide; S. *dolo; buena fe*), **mala fe, de** (LAW in bad faith; S. *con dolo*), **malo** (GEN bad; S. *deficiente, defectuoso; bueno*)].

There were 502 occurrences of *de mala fe* in the corpus and 225 occurrences of its equivalent (*in bad faith*). Some concordance examples are presented in Table 15.

TABLE 15

**Concordances of *de mala fe* from the corpus**

Translation technique	SL (Spanish) segment	TL (English) segment
(5) -Borrowing (from Latin) -Linguistic compression	<i>Si no hubiésemos hecho nuestra esta opinión habríamos admitido que cuando se ha satisfecho un derecho de aduana a partir de un precio superior al precio real como consecuencia de una declaración hecha <b>de mala fe</b>, el declarante pierde el derecho a obtener la devolución</i>	If I had not come to that view then I would accept that where duty is paid on a price which is higher than the true price, as a result of a <b>mala fide</b> declaration, the person making the declaration is disentitled, or estopped [sic], from recovering the money.
(6) -Linguistic compression	<i>los operadores consideran que el marco jurídico de los Estados miembros no disuade de forma análoga a los deudores <b>de mala fe</b> e incluso puede, en numerosos Estados miembros.</i>	operators feel the legal framework in the Member States is uneven in its deterrent effect on <b>bad</b> debtors and may even work to their financial benefit in many Member States.

*De mala fe* evidences semantic idiomatization characterized by the occurrence of a metonymy (in *fe* 'faith', ['feature for whole entity' relation]). Not surprisingly, the examples (5) and (6) share a characteristic that was also recurrent in the equivalents of *dar de baja*: their linguistic compression in comparison to the Spanish SpI entry.

A possible explanation for linguistic compression are the structural divergences (L'Homme 2020: 237-39) caused by the differences in the syntactic structures in both languages (Spanish and English).

Example (5) offers a SpI for SpI equivalence. However, there is a change in the idiomaticity of the SpI, from semantic, in *de mala fe*, to lexical, in *mala fide*, since it is a borrowing from Latin. As observed in Figure 12, example (5) is offered under *mala fe* but excluded from *de mala fe*. These findings are in accordance with both the asymmetry of the



equivalents (Adamska-Sałaciak 2010: 397-99; Pym 2014: 24-25) and the same concept denotation characteristic in a certain LSP domain (L'Homme 2020: 230-31) (see section 2.4.). In example (6), the expression was reduced to a monolexeme encompassing the basic meaning of the original SpI.

The third SpI to be analyzed was *above-the-line* (see Figure 13).

FIGURE 13

**Lexicographic entry for *above-the line* in the DCI**

**above-the line** *a*: CONTA ingresos y gastos corrientes, partida ordinaria. [Exp: **above-the-line items**<sup>1</sup> (ECON operaciones financieras independientes de la balanza de pagos ◊ *Above-the-line items generate income and improve economic welfare*; se refiere a las transacciones que tienen lugar independientemente

There were 44 occurrences of *above-the-line* in the corpus. However, none of its equivalents (*ingresos y gastos corrientes* ‘current income and expenses’ and *partida ordinaria* ‘ordinary line item’) were found in the corpus. Two examples of the concordances of the corpus are presented in Table 16.

TABLE 16

**Concordances of *above-the-line* from the corpus**

Translation technique	SL (English) segment	TL (Spanish) segment
(7) -Borrowing -Amplification	With regard to competition policy, the main parameters are what is called the " <b>above the line</b> " parameters: overall positioning of the brand (including pricing), general marketing policy (advertising concept, national advertising, promotions) as opposed to "below the line" which is more "point of sale"-related marketing.	<i>Por lo que respecta a la política de competencia, los parámetros principales son los genéricos <b>above the line</b>: posicionamiento global de la marca (precios incluidos), política general de marketing (concepto publicitario, publicidad a nivel nacional, promociones) en oposición a los específicos below the line, que es un tipo de marketing orientado al "punto de venta".</i>
(8) -Omission	Table 5 in the Annex presents details of ' <b>above the line</b> ' advertising expenditure by the major brands (and therefore by manufacturer) in the Great Britain market	<i>El cuadro 5 del Anexo ofrece los datos relativos a los gastos de publicidad de las principales marcas (y por lo tanto, de los principales fabricantes) de Gran Bretaña</i>

*Above-the-line* shows semantic idiomatization characterized by the occurrence of a metonymy (*above the line* was originally used literally to describe the score registered above a line in a scorecard. Its origin suggests the occurrence of a metonymy [‘feature for whole entity’ relation]).

Example (7) is a borrowing taken directly from the ST. Moreover, amplification, “to introduce details that are not formulated in the ST” (Molina and Hurtado Albir 2002: 510), was used as the translation technique in this case by means of the insertion of the word *genéricos* ‘generic’, which makes explicit the basic meaning of the SpI.

In example (8), the translation of *above the line* was omitted suggesting that the translator may have thought the meaning carried by *above-the-line* was already imbued in the translated fragment.

Finally, the fourth SpI to be analyzed was *in due time* (see Figure 14).

FIGURE 14

**Lexicographic entry for *in due time* in the DCI**

*patch*), **due form, in** (DER en buena y debida forma; equivale a *in good and due form*), **due time, in** (DER a su debido momento), **due to arrive at ... o'clock** (LOGÍST con llegada prevista a las ... horas ◊ *The cellular vessel is due to arrive at 17.00 hours*)].

There were 2,588 occurrences of *in due time* and 6 occurrences of its equivalent *a su debido momento*. Three concordance examples are offered in Table 17.

TABLE 17

**Concordances of *in due time* from the corpus**

Translation technique	SL (English) segment	TL (Spanish) segment
(9) -Linguistic amplification	it cannot be concluded that the defendant took all due care to ensure that its financial contributions were entered in the applicant's account <b>in due time</b> .	<i>no cabe admitir que la demandada haya empleado la necesaria diligencia para velar por la inscripción <b>a su debido momento</b> de su contribución financiera en la cuenta de la demandante.</i>
(10) -Linguistic amplification	adjustments to the manning nationality provisions laid down in paragraphs 2 and 3 so that the definitive system shall be approved by the Council <b>in due time</b> and before 1 January 1999.	<i>adaptaciones de las disposiciones sobre la nacionalidad de la tripulación a que se refieren los apartados 2 y 3, de forma que el sistema definitivo sea aprobado por el Consejo <b>en su debido momento</b> y antes del 1 de enero de 1999.</i>
(11) -Linguistic amplification	As far as the former is concerned, clearly Greece failed to challenge <b>in due time</b> the validity of the Regulations dealing with 1981/82, 1982/83	<i>En lo que atañe al primer principio, está claro que Grecia no impugnó <b>en su debido tiempo</b> la validez de los Reglamentos relativos a las campañas 1981/1982 y 1982/1983</i>

The SpI *in due time* evidences syntactic idiomatization due to its adverbial nature that is not suggested by the coordination of its word forms (a preposition [*in*], an adjective [*due*], and a noun [*time*]). The equivalent offered by the DCI and the other three examples found in the corpus illustrate an interesting case of phraseological variation. Among the equivalents found in the corpus, three were chosen for this analysis: (9) *a su debido tiempo* (1,075 occurrences), (10) *en su debido momento* (15 occurrences) and (11) *en su debido tiempo* (2 occurrences). It is evident that the coordination of the morphosyntactic pattern is the same (Prep Adj PP N), although the lexical units may vary. However, the dictionary only indexed one of the equivalents, ranking third in the corpus.

Regarding the translation techniques employed, as expected, ‘linguistic compression’ was recurrent from Spanish to English (see Tables 14 and 15), while English to Spanish (see Tables 16 and 17), shows a tendency for the use of ‘linguistic amplification,’ “to add linguistic elements” (Molina and Hurtado Albir 2002: 510), as part of the translation techniques. However, the occurrence of other translation techniques (e.g., borrowing, modulation, and transposition) in the concordances as well as the occurrence of terminological variation and polysemy (in the dictionary and in the corpus) suggest the asymmetry of SpI equivalence which goes in line with the postulation by Mellado Blanco (2015: 156) and Pym (2014: 24-25).

**6. Conclusions**



This paper has shed light on a set of characteristics (derived from an extensive literature review within translation studies, terminology, lexicography, and linguistics) common to the study of the equivalence of SPUs in an interdisciplinary setting. Despite (the still current) heated debate around the notion of equivalence, the academic exercise of identifying an interdisciplinary shared ground between these disciplines should not be labeled as an attempt to make ‘U-turns’ or reinvent the ‘translational wheel.’

The lexicographic and corpus-driven analyses of both the entries and the equivalents allowed for the identification of a wide range of equivalent options that go from terminological monolexemes (e.g., *al por mayor - wholesale*), to specialized collocations (e.g., *above the line - ingresos y gastos corrientes*), and specialized idioms (e.g., *dado de baja - written off*). In line with Sinclair and Moon (1997), the analysis of SpI entries and equivalents in the corpus indicates sub-registering of SpI equivalents in the DCI. The lack of indexation of the most common SpI equivalents could be caused by the use of a monolexical term as an anchor for the extraction of specialized collocations from corpora that cannot be applied to SpIs since they lack monolexical terms related to the LSP domain they are related to. A more in-depth analysis of SpIs involving other lexicographic resources and corpora must be carried out to provide more accurate extraction methods.

The idiomatic analysis carried out here provided empirical data that could be used as a starting point for a detailed analysis of SpIs in lexicographic resources and corpora. The combination of semantic, morphosyntactic, and translation-technique-related analyses of SpIs, and their equivalents, could lead to a better understanding of the notion of phraseological equivalence (and the behavior of SpIs) by comparing the extracted features in different LSP domains. Regarding SpI idiomaticity and how it was preserved or changed through the translation process, it was observed that some translation techniques (e.g., linguistic compression and linguistic amplification) did not change the semantic features of the SpIs. However, these two techniques deal more with structural divergences, while techniques such as borrowing lead to an obvious change in idiomaticity by using words from the SL. In those cases in which modulation (a change in the cognitive focus e.g., *darse de baja – terminate their membership*) was used, there was a change of lexical items that were used that could alter the metaphorical or metonymical constructions, as presented in section 5.2.2.

As shown throughout this study, SpIs exist and their characterization of this SPU sub-category was done by analyzing some of their lexical, semantic, and translation features in a lexicographic sample. Beyond the theoretical reflection, and contrarily to what LGP literature suggests, this article presents evidence that some SpIs occurred at a high rate in the EUR-Lex (24.55 times pmw in Spanish and 3.98 times pmw in English). Furthermore, the discrepancies observed between the indexed SpI entries and their equivalents in the DCI need to be explored further in future studies which should include other LSP domains and corpora.

## REFERENCES

- ADAMSKA-SALACIAK, Arleta (2010): 'Examining equivalence', *Journal of lexicography*, 23: 387-409.
- ALCARAZ, Varó Enrique and CASTRO CALVÍN José (2007): *Diccionario de Comercio Internacional – Importación y Exportación, Inglés-Español/Spanish-English*. Barcelona: Ariel.
- ""arm, n.1" [Oxford English Dictionary] (n.d.): Oxford University Press. Consulted on 18 March 2021, <<https://www.oed.com/view/Entry/10808?redirectedFrom=at+arm%27s+length>>
- ATKINS, Beryl T Sue and RUNDELL, Michael (2008): *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press
- BAISA, Vít, MICHELFEIT, Jan, MEDVEĎ, Marek and JAKUBÍČEK, Miloš (2016): European Union Language Resources in Sketch Engine. In CALZOLARI, Nicoletta, CHOUKRI, Khalid, DECLERCK, Thierry, GOGGI, Sara, GROBELNIK, Marko, MAEGAARD, Bente, MARIANI, Joseph, MAZO, Helene, MORENO, Asuncion, ODIJK, Jan and PIPERIDIS, Stelios, eds. *The Proceedings of tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož: European Language Resources Association (ELRA)

- BAKER, Mona (1992): *In Other Words: A Coursebook on Translation* New York: Routledge.
- BALDWIN, Timothy and KIM, Su Nam (2010): Multiword Expressions. In INDURKHYA, Nitin and DAMERAU, Fred J eds. *Handbook of Natural Language Processing*. Boca Raton: CRC Press.
- BOULANGER, Jean-Claude. (1995): Présentation : images et parcours de la socioterminologie. *Meta : Journal des traducteurs*. 40: 194-205.
- BUENDÍA CASTRO, Míriam and FABER, Pamela (2016): Phraseological Correspondence in English and Spanish Specialized Texts. In CORPAS PASTOR, Gloria, ed. *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. Geneva: Tradulex.
- CABRÉ, María Teresa (1999): *La terminología: Representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada.
- CABRÉ, María Teresa (2000): Terminologie et linguistique: la théorie des portes. *Terminologies nouvelles. Terminologie et diversité culturelle*. 21: 10-15.
- CATFORD, John Cunnison (1965): *A Linguistic Theory of Translation*. Oxford: Oxford University Press.
- CHESTERMAN, Andrew (1996): On Similarity. *Target*. 8: 159-64.
- CORPAS PASTOR, Gloria (2003): *Diez años de investigaciones en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. Madrid: Iberoamericana.
- DIKI-KIDIRI, Marcel (2014): Le vocabulaire juridique en sängö: une application de la terminologie culturelle. In Rita Temmerman and Marc van Campenhoudt, eds. *Dynamics and Terminology: an interdisciplinary perspective on monolingual and multilingual culture-bound communication*. Amsterdam: John Benjamins Publishing.
- DOBROVOL'SKIJ, Dmitrij (2000): Idioms in contact in contrast: a functional view. In CORPAS PASTOR, Gloria, ed. *Las lenguas de Europa: Estudios de fraseología, fraseografía y traducción* Granada: Comares.
- DOBROVOL'SKIJ, Dmitrij and PIIRAINEN, Elisabeth (2005): *Figurative Language: Cross-Cultural and Cross-Linguistic Perspectives*. Amsterdam: Elsevier.
- "equivalence, n.1." [Oxford English Dictionary] (n.d.): Oxford University Press. Consulted on 30 January 2021, <<https://www.oed.com/view/Entry/63840?rskey=w2S8y8&result=1&isAdvanced=false#eid>>.
- FABER BENÍTEZ, Pamela (2009): The cognitive shift in terminology and specialized translation. *MonTI. Monografías de Traducción e Interpretación*. 1: 107-34.
- FABER BENÍTEZ, Pamela and LEÓN ARAÚZ, Pilar (2014): Specialized knowledge dynamics: from cognition to culture-bound terminology. In TEMMERMAN, Rita and VAN CAMPENHOUDT, Marc, eds. *Dynamics and Terminology: an interdisciplinary perspective on monolingual and multilingual culture-bound communication*. Amsterdam: John Benjamins Publishing.
- FABER BENÍTEZ, Pamela and LÓPEZ RODRÍGUEZ, Clara Inés (2012): Terminology and specialized language. In FABER, Pamela, ed. *A Cognitive Linguistics View of terminology and Specialized Language* Berlin: De Gruyter Mouton.
- FELBER, Helmut (1984): *Terminology Manual*. Paris: UNESCO and Infoterm).
- FERNÁNDEZ-SILVA, Sabela, FREIXA, Judit and CABRÉ, María Teresa (2014): A method for analysing the dynamics of naming from a monolingual and multilingual perspective. In Rita Temmerman and Marc van Campenhoudt, eds. TEMMERMAN, Rita and VAN CAMPENHOUDT, Marc, eds. *Dynamics and Terminology: an interdisciplinary perspective on monolingual and multilingual culture-bound communication*. Amsterdam: John Benjamins Publishing..
- FUERTES-OLIVERA, Pedro A (2011): Equivalent Selection in Specialized e-Lexicography: A Case Study with Spanish Accounting Terms. *Lexikos*. 21: 95-119.
- GAUDIN, François (2003): *Socioterminologie: une approche sociolinguistique de la terminologie.*: Louvain-la-Neuve: De Boeck Supérieur.
- HALVERSON, Sandra (1997): The concept of equivalence in translation Studies: Much Ado About Something. *Target*, 9: 207-33.
- HALVERSON, Sandra (1999): Conceptual Work and the "Translation" Concept. *Target*. 11: 1-31.
- HARTMANN, Reinhard Rudolf Karl (1985): Contrastive Text Analysis and the Search for Equivalence in the Bilingual Dictionary. In HYLDGAARD-JENSEN, Karl and Zettersten, Arne, eds. *Symposium on Lexicography II*. Tübingen: Max Niemeyer Verlag.
- HARTMANN, Reinhard Rudolf Karl (2007): Interlingual References: On the Mutual Relations between Lexicography and Translation. In HARTMANN, Reinhard Rudolf Karl, ed. *Interlingual Lexicography: Selected Essays on Translation Equivalence, Contrastive Linguistics and the Bilingual Dictionary*. Tübingen: Max Miemeyer Verlag.
- HARTMANN, Reinhard Rudolf Karl and JAMES, Gregory (1998): *Dictionary of Lexicography*. New York: Routledge.
- HOFFMANN, Lothar (1998): *Llenguatges d'especialitat: selecció de textos*. Barcelona: Institut Universitari de Lingüística Aplicada.
- HOUSE, Juliane (1977): A model for Assessing Translation Quality. *Meta : Journal des traducteurs*, 22: 103-09.

- JAKOBSON, Roman (1959): On Linguistic Aspects of Translation. In BROWER, Reuben Arthur, ed. *On Translation*. Cambridge: Harvard University Press.
- JASKOT, Maciej Paweł (2016a): Do we need equivalence-based e-tools? In CORPAS PASTOR, Gloria, ed. *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. Geneva: Tradulex.
- JASKOT, Maciej Paweł (2016b): Equivalent Culture-Anchored Units Translation? The Phraseological Units Issue. *Cognitive Studies*. 16: 57-64.
- KADE, Otto (1968) *Zufall und Gesetzmässigkeit in der Übersetzung*. Leipzig: Verlag Enzyklopädie.
- KARPINSKA, Laura (2019): Specialized bilingual dictionaries as a translation tool: a case study of Latvian-English-Latvian dictionaries of legal terms. In OSUCHOWSKA, Dorota and HARMON, Lucyna, eds. *Contents, Use, Usability: Dictionaries from the Perspective of a Translator and a Language Teacher*. Berlin: Peter Lang.
- KNOWLES, Murray and MOON, Rosamund (2006): *Introducing Metaphor*. London: Routledge.
- KOLLER, Werner (1989): Equivalence in translation theory. In CHESTERMAN, Andrew, ed. *Reading in Translation Theory*. Helsinki: Oy Finn Lectura Ab.
- KRISTIANSEN, Marita (2014): Concept change, term dynamics and culture-boundness in economic-administrative domains. In TEMMERMAN, Rita and VAN CAMPENHOUDT, Marc, eds., *Dynamics and Terminology: an interdisciplinary perspective on monolingual and multilingual culture-bound communication*. Amsterdam: John Benjamins Publishing.
- KÜBLER, Natalie and PECMAN, Mojca (2012): The ARTES bilingual LSP dictionary: From collocation to higher order phraseology. In Granger, Sylviane and Paquot, Magali, eds. *Electronic Lexicography*. Oxford: Oxford University Press.
- L'HOMME, Marie-Claude (2005): Sur la notion de « terme ». *Meta : Journal des traducteurs*. 50: 1112–32.
- L'HOMME, Marie-Claude (2020): *Lexical Semantics for Terminology: An Introduction*. Amsterdam: John Benjamins Publishing Company.
- LEFEVERE, André and BASSNETT, Susan (1995): Introduction : Proust's grandmother and The thousand and one nights : the "cultural turn" in translation studies. In BASSNETT, Susan and LEFEVERE, André, eds. *Translation, History, and Culture*. New York: Cassell.
- LEONARDI, Vanessa (2000): Equivalence in translation: Between myth and reality. *Translation Journal*. 4.
- MARTÍNEZ DE SOUSA, José (1995): *Diccionario de lexicografía práctica*. Barcelona: Bibliograf.
- MEL'ČUK, Igor (1998): Collocations and Lexical Functions. In COWIE, Anthony Paul, ed. *Phraseology: Theory, Analysis and Applications*. Oxford: Oxford University Press.
- MEL'ČUK, Igor (2012): Phraseology in the language, in the dictionary, and in the computer. In KUIPER, Koenraad, ed. *Yearbook of Phraseology*. New York: De Gruyter Mouton.
- MEL'ČUK, Igor (2013) Tout ce que nous voulions savoir sur les phrasèmes, mais... *Cahiers de lexicologie*. 102: 129-49.
- MELLADO BLANCO, Carmen (2015): Parámetros específicos de equivalencia en las unidades fraseológicas (con ejemplos del español y el alemán). *Revista de Filología*, 33: 153-74.
- MOLINA, Lucía and HURTADO ALBIR, Amparo (2002). Translation Techniques Revisited: A Dynamic and Functionalist Approach. *Meta : Journal des traducteurs*. 47: 498-512.
- MORVAY, Károly (1996): Harri batez bi kolpe. Cuestiones de fraseología comparada. *Euskera: Euskaltzaindiaren lan eta agiriak*, 41: 719-67.
- MUNDAY, Jeremy (2008): *Introducing Translation Studies: Theories and Applications*. New York: Routledge.
- NIDA, Eugene Albert and TABER, Charles (1982): *The Theory and Practice of Translation*. Leiden: E. J. Brill.
- PANOU, Despoina. 2013. Equivalence in Translation Theories: A Critical Evaluation. *Theory and Practice in Language Studies*. 3: 1-6.
- PYM, Anthony (2014): *Exploring Translation Theories*. New York: Routledge.
- ROJAS DÍAZ, José Luis (forthcoming): 'Arm's length' phraseology? Building bridges from general language to specialized language phraseology – a study based on a specialized dictionary of International Commerce and Economics in Spanish and English. Manuscript submitted for publication.
- SAGER, Juan Carlos (1994): Reflections on Terminological Translation Equivalents. *Lebende Sprachen*. 39: 55-57.
- SINCLAIR, John and MOON, Rosamund, eds. (1997): *Collins COBUILD Dictionary of Idioms*. Glasgow: HarperCollins Publishers.
- SNELL-HORNBY, Mary (1995): *Translation Studies: An integrated approach*. Amsterdam: John Benjamins Publishing Company.
- TEMMERMAN, Rita (1997): Questioning the univocity ideal: the difference between sociocognitive terminology and traditional terminology. *Hermes*. 18: 51-91.
- TEMMERMAN, Rita (2000): *Towards New Ways of Terminology Description: the sociocognitive approach* (John Benjamins Publishing Company: Amsterdam).

- TOURY, Gideon (1981): *Translated Literature: System, Norm, Performance: Toward a TT-Oriented Approach to Literary Translation. Poetics Today. 2: 9-27.*
- TOURY, Gideon (1985): A Rationale for Descriptive Translation Studies. In HERMANS, Theo, ed. *The Manipulation of Literature: Studies in Literary Translation.* London: Croom Helm.
- TRAP-JENSEN, Lars (2018) Lexicography between NLP and Linguistics: Aspects of Theory and Practice. In ČIBEJ, Jaka, GORJANC, Vojko, KOSEM, Iztok and KREK, Simon, eds. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts.* Ljubljana: Ljubljana University Press.
- TYTLER, Alexander Fraser (1907): *Essay on the principle of translation.* London: J. M. Dent & Sons Ltd.
- VEISBERGS, Andrejs (1994): Borrowing of English idioms in Latvian. *Journal of Baltic Studies. 25: 43-52.*
- VERMEER, Hans Josef (2012): Skopos and comission in translational action. In VENUTI, Lawrence, ed. *The Translation Studies Reader.* New York: Routledge.
- VINAY, Jean-Paul and DARBELNET, Jean (1995): *Comparative Stylistics of French and English: A methodology for translation.* Amsterdam: John Benjamins Publishing Company.
- WILSS, Wolfram (1982): *The science of translation: problems and methods.* Tübingen: G. Narr.
- WÜSTER, Eugen (1998): *Introducción a la Teoría General de la Terminología y a la Lexicografía Terminográfica* (Translated from German by María Teresa, CABRÉ) Barcelona: Institut Universitari de Lingüística Aplicada.
- ZGUSTA, Ladislav (1971): *Manual of lexicography.* Prague: Academia.
- ZULUAGA, Alberto (1999): Traductología y Fraseología. *Paremia. 8: 537-49.*
- ZULUAGA, Alberto (2019) Translatology and Phraseology. *Paremia, 29: 61-76.*

## Appendices

### Appendix A

*Table 1A. Definitions and examples of PUs (Mel'čuk, 1998, pp. 6-8), and classification of idioms and collocations (Mel'čuk, 2012, pp. 37-40)*

<b>Definitions and examples of PUs, and classification of idioms and collocations</b>		
PU type (for this study)	Definition (Mel'čuk, 1998, 2012)	Example and PU type (Mel'čuk, 1998, 2012)
Idiom	An idiom <b>AB</b> is a full idiom iff its meaning ('AB') does not include the meaning of any of its lexical ('A' and 'B') components	kick <sub>A</sub> the bucket <sub>B</sub> (Full idiom)
<b>'AB' ⊄ 'A' and 'AB' ⊄ 'B'</b>		
Collocation	<p>Type 1 A collocation <b>AB</b> is a semantic phraseme such that its signified 'AB' is constructed out of the signified of one of its two constituent lexemes — say, of <b>A</b> — and a signified 'C' ['AB' = 'A' + 'C'] such that the lexeme <b>B</b> expresses 'C' contingent on <b>A</b>. The formulation "<b>B</b> expresses 'C' contingent on <b>A</b>" covers four major cases, which correspond to the following four major types of collocations:</p> <p><b>Either</b> 'C' ⊄ B', i.e., <b>B</b> does not have (in the dictionary) the corresponding signified;  <b>And</b> 'C' is empty, that is, the lexeme <b>B</b> is, so to speak, a semi-auxiliary selected by <b>A</b> to support it in a particular syntactic configuration</p> <p>AB' ⊃ 'A', and 'AB' ⊄ 'B', and 'AB' ⊃ 'C'            'AB' = 'A' + 'C'            'C' = ∅ (empty)</p>	take <sub>B</sub> a shower <sub>A</sub> <i>'take' is a support verb</i> (semi-idiom)
	<p>Type 2 <b>Either</b> 'C' ⊄ B', i.e., <b>B</b> does not have (in the dictionary) the corresponding signified;  <b>And</b> 'C' is not empty but the lexeme <b>B</b> expresses 'C' only in combination with <b>A</b> (or with a few other similar lexemes).</p> <p>'AB' ⊃ 'A', and 'AB' ⊄ 'B', and 'AB' ⊃ 'C'            'AB' = 'A' + 'C'            'B' ⊄ 'C'</p>	private <sub>A</sub> eye <sub>B</sub> <i>private<sub>A</sub> detective<sub>B</sub></i> (semi-idiom)
	<p>Type 3 <b>Either</b> 'C' ⊃ 'B', i.e., <b>B</b> has (in the dictionary) the corresponding signified;  <b>And</b> 'B' cannot be expressed with <b>A</b> by any otherwise possible synonym of <b>B</b>;</p> <p>'AB' ⊃ 'A', and 'AB' ⊃ 'B'            AB ⇔ A+B            AB ≠ A+ ≈B (No synonym of B is possible)</p>	strong <sub>B</sub> coffee <sub>A</sub> different from, powerful- <sub>B</sub> * coffee <sub>A</sub> (standard collocation)
	<p>Type 4 <b>Either</b> 'C' ⊃ 'B', i.e., <b>B</b> has (in the dictionary) the corresponding signified;  <b>And</b> 'B' includes (an important part of) the signified 'A', that is, it is utterly</p>	horse <sub>A</sub> neigh <sub>B</sub> (non-standard collocation)

		specific, and thus <b>B</b> is ‘bound’ by <b>A</b>	
		‘AB’ ⊃ ‘A’, <b>and</b> ‘AB’ ⊃ ‘B’ B ∈ A	
	Type 5	<b>AB</b> includes the signified of both constituent lexemes but contains an unpredictable addition ‘C’ as well.	start <sub>A</sub> a family <sub>B</sub> ‘conceive first child’ <sub>C</sub> (quasi-idiom)
		‘AB’ ⊃ ‘A’, <b>and</b> ‘AB’ ⊃ ‘B’, <b>and</b> ‘AB’ ⊃ ‘C’ ‘AB’ = ‘A’ + ‘B’ + ‘C’ ‘C’ ≠ ‘A’ <b>and</b> ‘C’ ≠ ‘B’	
Pragmateme	Type 1	<b>AB</b> includes the signified of both constituent lexemes, but it is <b>pragmatically</b> constraint i.e., the use of the pragmateme requires a specific situation of use.	return <sub>A</sub> to sender <sub>B</sub> <i>on a postal sending</i> (pragmateme)]
		‘AB’ ⊃ ‘A’, <b>and</b> ‘AB’ ⊃ ‘B’ ‘AB’ = ‘A’ + ‘B’	
	Type 2	<b>AB</b> is a compositional complex proper name. “The linguistic meaning of the complex proper name is literal, but idiosyncratically corresponds to one particular referent”.	City <sub>A</sub> of Lights <sub>B</sub> <i>nickname of Paris</i> (complex proper name)
		‘AB’ ⊃ ‘A’, <b>and</b> ‘AB’ ⊃ ‘B’ ‘AB’ = ‘A’ + ‘B’	

Table 2A. Denominations and definitions of LSP phraseology gathered in the work by (Rojas Díaz, forthcoming-a)

Denominations and definitions of SPU		
Author	Denomination	Definition
Picht (1987)	LSP phrase	“A phrase whose nucleus is a <b>term</b> with which linguistic elements are connected which, by entering a semantic relationship with the <b>term</b> , undergo a modification of meaning which renders them capable of collocation with the <b>term</b> and of forming together with the <b>term</b> a lexical unit with an <b>LSP meaning</b> ”. (p. 151 [emphasis added]).
Gouadec (1993)	Phraseological unit	“Every entity worthy of interest and bigger than the standard <b>terminological unit</b> is called phraseological unit” <sup>20</sup> (p. 550 [emphasis added, my translation]).
Blais (1993)	Phraseologism (Phraséologisme)	Combination of linguistic elements distinctive of a specific <b>field of specialty</b> , in which one of the elements, the head, is a <b>term</b> , and those elements are related syntactically and semantically, and there is a paradigmatic constraint for them” <sup>21</sup> (p. 52 [my translation]).
Pavel (1993)	LSP collocation	“..the inter-phrasal combinations of <b>terms</b> and words in actual <b>LSP discourse</b> ” (p. 29 [emphasis added]).
L’Homme (1998)	Specialized Lexical Combinations - SLC ( <i>Combinaison lexicale spécialisée - CLS</i> )	<b>Specialized lexical combinations</b> , like general language collocations, have a conventional character. They are subject of a consensus, but this time the consensus is established within a group of specialists. A non-specialist must learn to mobilize these uses to insert <b>terminological units</b> in idiomatic environments [...] specialized lexical combinations (CLS [in French]) are also made up of two lexemes. One of these lexemes is a <b>terminological unit</b> ; the other constitutes the co-occurring lexeme” <sup>22</sup> (p. 515 [emphasis added, my translation]).
Lorente Casafont (2002)	Specialized Phraseological Unit - SPU	“...units of <b>specialized knowledge</b> , which correspond to non-lexicalized phrasal or sentence structures, that present a certain

<sup>20</sup> Original in French: “toute entité digne d’intérêt et plus grande que l’unité terminologique standard est dite unité phraséologique”

<sup>21</sup> Original in French: “Combinaison d’éléments linguistiques propre à un domaine de spécialité, dont l’un est un terme noyau, qui sont liés sémantiquement et syntaxiquement et pour lesquels il existe une contrainte paradigmatique”

<sup>22</sup> Original in French: “Les CLS, tout comme les collocations, ont un caractère conventionnel. Elles font l’objet d’un consensus, mais cette fois-ci, le consensus s’établit au sein d’un groupe de spécialistes. Un non-spécialiste doit apprendre à mobiliser ces usages pour insérer les unités terminologiques dans des environnements idiomatiques [...] les combinaisons lexicales spécialisées (CLS) sont également composées de deux lexèmes. L’un de ces lexèmes est une unité terminologique; l’autre constitue le cooccurrent”

	<i>(Unidad fraseológica especializada - UFE)</i>	tendency to stereotyping to a certain level of fixation, and that contain, at least, a <b>term</b> <sup>23</sup> (2002 [emphasis added, my translation]).
Bevilacqua (2004)	Specialized Phraseological Unit - SPU <i>(Unidad fraseológica especializada - UFE)</i>	“Syntagmatic units of <b>specialized meaning</b> composed by a <b>terminological head</b> (simple or syntagmatic <b>terminological unit</b> ) and by an eventive head (verb, de-verbal noun, or participle derived from a verb) that represent specific processes in a certain field of knowledge” <sup>24</sup> (p. 28 [emphasis added, my translation]).

<sup>23</sup> Original in Spanish: “...unidades de conocimiento especializado, que corresponden con estructuras sintagmáticas u oracionales, no lexicalizadas, pero que presentan una cierta tendencia al estereotipo o un cierto grado de fijación, y que contienen como mínimo un término”.

<sup>24</sup> Original in Spanish: “son unidades de significación especializada sintagmáticas, que están formadas por un NT (UT simple o sintagmática) y por un NE (verbo, nombre deverbal o participio derivado del verbo), que representan las actividades y procesos específicos de un ámbito. Son, pues, dependientes de un área temática, poseen un determinado grado de fijación interna y tienen una frecuencia relevante en los textos de un ámbito especializado”.



*Table 3A. Some authors and definitions of equivalence in translation studies addressed by Halverson (1997), Leonardi (2000), and Panou (2013)*

<b>Equivalence in Translation Studies</b>	
Author (year)	Definition
Vinay & Dalbernet (1995)	A translation procedure, the result of which replicates the same situation as in the original, whilst using completely different wording (p. 342).
Jakobson (1959)	[...] on the level of interlingual translation, there is ordinarily no full equivalence between code-units [...] Thus translation involves two equivalent messages in two different codes (p. 233).
Nida & Taber (1982)	<i>dynamic equivalence</i> : quality of a translation in which the message of the original text has been so transported into the receptor language that the <i>response</i> of the <i>receptor</i> is essentially like that of the original receptors.  <i>formal correspondence</i> : quality of a translation in which the features of the form of the source text have been mechanically reproduced in the receptor language. Typically, formal correspondence distorts the grammatical and stylistic patterns of the receptor language, and hence distorts the message, so as to cause the receptor to misunderstand or to labor unduly hard; opposed to <i>dynamic equivalence</i> (pp. 200-201).
Catford (1965)	A <i>textual equivalent</i> is any TL text or portion of text which is observed on a particular occasion [...] to be the equivalent of a given SL text or portion of text.  A <i>formal correspondent</i> , on the other hand, is any TL category (unit, class, structure, element of structure, etc.) which can be said to occupy, as nearly as possible, the 'same' place in the 'economy' of the TL as the given SL category occupies in the SL (p. 27).
Kade (1968)	[...] equivalence at the level of the word or phrase comes in four modes: <i>one-to one</i> , as in the case of stable technical terms.  <i>one-to-several</i> , when translators have to choose between alternatives.  <i>one-to-part</i> , when the available equivalents are only partial matches.  <i>one-to-none</i> , when translators have to create a new solution (coining neologisms or perhaps borrowing the foreign term...) (pp. 72-90 [as cited in Pym, 2014, pp. 28-29])
House (1977)	[...] an adequate translation text (TT) is a semantically and pragmatically equivalent one. As a first requirement for this equivalence it is posited that a translation text have a function equivalent to that of its source text (ST) (p. 103).
Koller (1989)	The concept of equivalence postulates a relation between SL text (or text element) and TL text (or text element). five factors can be argued to play a relevant role in the specification of equivalence types:  1. The <i>extralinguistic content</i> transmitted by a text [...] <i>denotative equivalence</i> [...]  2. The <i>connotations</i> transmitted by means of the word choice [...], with respect to the level of style (register), the social and geographical dimension, frequency, [...] this is <i>connotative equivalence</i> [...]  3. The <i>text and language norms</i> (usage norms) for given text types: this kind of equivalence, [...] I call <i>text-normative equivalence</i> [...]

---

Toury (1981)	4. The <i>receiver</i> (reader) to whom the translation is directed [...] and to whom the translation is tuned in order e.g. to achieve a given effect; this is <i>pragmatic equivalence</i> [...]
	5. Certain formal-aesthetic features of the SL text, including word play, metalinguistic aspect, individual stylistic features; the kind of equivalence that relates to these textual characteristics I call <i>formal equivalence</i> [...] (pp. 100-101).
	It should be noted that this description which contrasted the picture as seen from ST's and from TT's point of view made use of the term "equivalence" in two different senses: (a) as a <i>theoretical</i> term, denoting an abstract, ideal relationship, or category of relationships between TTs and STs, translations and their sources; (b) as a <i>descriptive</i> term, denoting concrete objects - actual relationships between actual utterances in two different languages (and literatures), recognized as TTs and STs - which are subject to direct observation (p. 13).
Newmark (1981)	Communicative translation attempts to produce on its readers an effect as close as possible to that obtained on the readers of the original. Semantic translation attempts to render, as closely as the semantic and syntactic structures of the second language allow, the exact contextual meaning of the original (p. 39).
Wilss (1982)	An empirical phenomenon which carries with it problems which presently can be solved, if at all, only for each individual translation text (p. 145).
Baker (1992)	<i>Grammatical equivalence</i> , when referring to the diversity of grammatical categories across languages. Textual equivalence, when referring to the equivalence between a SL text and a TL text in terms of information and cohesion. <i>Pragmatic equivalence</i> , when referring to implicatures and strategies of avoidance during the translation process (pp. 11-12, 82-83, 217-218 [as cited in Leornardi, 2000, emphasis added]).
Pym (2014)	<i>Equivalence</i> is a relation of "equal value" between a source-text segment and a target-text segment [...] on any linguistic level, from form to function."
	<i>Natural equivalence</i> is presumed to exist between languages or cultures prior to the act of translating."
	<i>Directional equivalence</i> is an asymmetric relation where the creation of an equivalent by translating one way does not imply that the same equivalence will also be created when translating the other way. [...] Theories of <i>directional equivalence</i> allow that the translator has a choice between several translation strategies, and that those strategies are not dictated by the source text (pp. 6-7, 24-25 [emphasis added])

---

Table 4A. Some authors and definitions of equivalence in Lexicography

<b>Equivalence in Lexicography</b>	
Author (year)	Definition
Zgusta (1971)	We call equivalent such a lexical unit of the target language which has the same lexical meaning as the respective lexical unit of the source language (p. 312)
Kromann (1991)	[...] equivalence is a relation between the individual meanings of the lemmatized word and the equivalents. We are thus disregarding syntactic equivalence in this context (p. 2117).
Martínez de Sousa (1995)	word form that corresponds exactly or approximately with another form in a different language <sup>25</sup> (p. 193 [my translation]).
Hartmann & James (1998)	The relationship between words or phrases, from two or more languages, which share the same meaning. Because of the problem of anisomorphism, equivalence is 'partial' or 'relative' rather than 'full' or 'exact' for most contexts. Compilers of bilingual dictionaries often struggle to find and codify such translation equivalents, taking into account the directionality of the operation. In bilingual or multilingual terminological dictionaries, equivalence implies interlingual correspondence of designations for identical concepts (p. 51).
Atkins and Rundell (2008)	The equivalence relationship between a pair of words, SL and TL, varies from exact to very approximate, from perfect to just-adequate, and the skill of the dictionary writer lies in selecting the best TL match available, and second in making sure that the SL-speaking, encoding users are aware of the pitfalls that lie in wait for them (p. 468).
Adamska-Sałaciak (2010)	Adamska-Sałaciak [...] proposes a tentative classification of types of equivalents used in bilingual lexicography [...]. This classification comprises four equivalence categories:  <i>semantic or cognitive equivalents</i> that are established lexical items in the TL;  <i>explanatory or descriptive equivalents</i> provide explanations of a source language (SL) item into the TL when an established equivalent is not available;  <i>translation(al) or insertable equivalents</i> are related to contextual usage of a SL item, they present a translation of this item that can be inserted in the TL text;  <i>functional or situational equivalents</i> are also related to contextual usage of a SL item, but the grammatical category of the equivalent can differ from the SL item or it can be an idiom with different wording (pp. 397-399 [as cited in Karpinska 2019, pp. 38-39, emphasis added]).

<sup>25</sup> Original in Spanish: “*forma que corresponde exacta o aproximadamente con otra de un idioma diferente.*”

*Table 5A. Some authors and definitions of equivalence in Terminology*

Author (year)	Definition
Picht and Draskau (1985)	‘Synonymy’ and ‘equivalence’ denote very similar phenomena. The only difference lies in the fact that synonymy refers to identity of concept designated by different terms in the same language, while equivalence refers to the same phenomenon expressed in two or more languages (p. 103).
ISO 1087 (2019)	Relation between designations (3.4.1) in different natural languages (3.1.7) representing the same concept (3.2.7).
L’Homme (2020)	<i>Conceptual equivalence</i> : knowledge-driven approaches to terminology consider that terms are equivalent if they belong to different languages and denote the same concept within the same domain. <i>Terminological equivalence</i> : equivalence can also be established on the basis of the meaning they convey rather than on the basis of their potential to label a predefined concept (pp. 230-234).

Table 6A. Some authors and definitions of phraseological equivalence

Phraseological equivalence	
Author (year)	Definition
Corpas Pastor (2003)	<p><i>Total equivalence</i>: it is produced when a PU of the SL corresponds to another PU in the TL. This PU presents the same denotative and connotative meaning, the same metaphorical base, the same distribution and frequency of use, the same conventional implicatures. The same pragmatic charge, and similar connotations<sup>26</sup> (pp. 281-282 [my translation]).</p> <p><i>Partial equivalence</i>: it is produced generally due to the divergencies and overlapping of the semantic content or the diasystematic restriction of the involved PUs<sup>27</sup> (p. 282 [my translation]).</p> <p><i>False equivalence (Feigned equivalence)</i>: this happens when the constitutive elements of the units of the phraseologism in the source language present the same formal sameness than in the target language, but a semantic difference<sup>28</sup> (pp. 208-209 [my translation]).</p> <p><i>Non-equivalence</i>: All those idioms of the SL that do not present a translation equivalent in the TL. Those are cases of linguistic realities in the SL that have not been lexicalized in the TL (...) the translator must opt for a paraphrasis or any other transfer technique to express the meaning of the unit (PU)<sup>29</sup> (p. 208 [my translation]).</p>
Dobrovol'skij and Piirainen (2005)	<p><i>Functional (Cross-linguistic) equivalence</i>: there are several factors responsible for semantic, syntactic, or pragmatic differences between similar idioms in different languages. These factors can arise from the internal structure of a given language (cf., above all, the syntactic behavior of idioms), or they can be motivated cognitively and/or culturally (p. 78).</p>
Mellado Blanco (2015)	<p><i>Systemic equivalence</i>: it is independent from the deictic coordinates of time and space (...) neither cotext nor context exert any influence in the determination of a systemic equivalence (...) it is a generic equivalence that answers the question 'how do you say x of L1 in L2?'<sup>30</sup> (p. 154 [my translation]).</p>

<sup>26</sup> Original in Spanish: "Equivalencia plena: se produce cuando a una UF de la LO le corresponde otra UF de la LM, la cual presenta el mismo significado denotativo y connotativo, una misma base metafórica, una misma distribución y frecuencia de uso, las mismas implicaturas convencionales, la misma carga pragmática y similares connotaciones."

<sup>27</sup> Original in Spanish: "La equivalencia parcial se produce generalmente por las divergencias y solapamientos en cuanto al contenido semántico o las estrictiones diasistemáticas de las UFS implicadas."

<sup>28</sup> Original in Spanish: "equivalencias que se dan cuando los elementos constitutivos de las unidades del fraseologismo en la lengua origen y lengua meta presentan similitud formal, pero diferencia semántica."

<sup>29</sup> Original in Spanish: "todas aquellas locuciones de la LO que no presentan un equivalente de traducción en la LM. Se trata de casos de realidades lingüísticas en la LO que no se lexicalizan en la LM (...) el traductor debe optar por la paráfrasis o por alguna otra técnica de transferencia para expresar el significado de la unidad."

<sup>30</sup> Original in Spanish: "la equivalencia sistémica es independiente de las coordenadas deícticas de tiempo y espacio (...) tampoco el cotexto y el contexto ejercen influencia en la determinación de una equivalencia sistémica. (...) de una equivalencia genérica que podría responder a la pregunta "¿Cómo se dice x de la L1 en la L2?"

*Functional equivalence*: in the phraseological sphere, as we understand it, differentiates from the systemic equivalence that it is not limited to the description of decontextualized phraseological pairs, but it takes into consideration the context of use as well as the possibility of equivalence through a single lexeme.<sup>31</sup> (p. 155 my translation).

*Lexicographic equivalence*: on the one hand, it should be obtained through the analysis of the *prototypic* behavior of the phraseological units at a discourse level. It is necessary then, to know how to discriminate the typical from the peripheral or less representative uses (of the PUs). On the other hand, the lexicographic equivalence does not intend to record all the possible cases of translation of a phraseologism from L1 to L2, as the textual equivalence does<sup>32</sup> (p. 155 [my translation]).

---

<sup>31</sup> Original in Spanish: “*en el plano fraseológico, la equivalencia funcional según nosotros la entendemos se diferenciaría de la equivalencia sistémica en que no se limita a describir pares de unidades fraseológicas descontextualizadas, sino que considera las situaciones de uso y contempla además la posibilidad de equivalencia mediante un lexema simple.*”

<sup>32</sup> Original in Spanish: “*debe obtenerse mediante el análisis del comportamiento prototípico de las unidades fraseológicas en el nivel del discurso, para lo cual es esencial saber discriminar los usos típicos de los periféricos o poco representativos. Por otra parte, la equivalencia lexicográfica se diferencia de la equivalencia textual en que no pretende hacerse eco de toda la casuística posible de traducciones de un fraseologismo de la L1 en la L2.*”