

MON

Monografías de Traducción e Interpretación
Monografies de Traducció i d'Interpretació
Monographs in Translation and Interpreting
Monographies de Traduction et d'Interprétation
Monographien zur Translation

TI

2020

Special Issue 6
Número especial 6

Multidisciplinary
analysis
of the phenomenon
of phraseological
variation in translation
and interpreting

Análisis multidisciplinar
del fenómeno de la variación
en traducción e interpretación

Pedro Mogorrón Huerta (Ed.)

ISSN-e 1989-9335
ISSN 1889-4178


Universitat d'Alacant
Universidad de Alicante


UNIVERSITAT
JAUME I


UNIVERSITAT
ID VALÈNCIA

Universitat d'Alacant - Universitat Jaume I - Universitat de València

General Editor / Director

Albaladejo-Martínez, Juan Antonio (Universitat d'Alacant)

Managing Editor / Secretario

Agost Canós, Rosa (Universitat Jaume I)

Editorial Board / Comité de Redacción

Agost Canós, Rosa (Universitat Jaume I - Secretaria); Alarcón Navío, Esperanza (Universidad de Granada); Albaladejo-Martínez, Juan Antonio (Universitat d'Alacant - Director); Corpas Pastor, Gloria (AIETI - Universidad de Málaga); Farrés Puntí, Ramon (Universitat Autònoma de Barcelona); Hernández Sacristán, Carlos (Universitat de València); Iliescu Gheorghiu, Catalina (Universitat d'Alacant); Marco Borillo, Josep (Universitat Jaume I); Martínez-Gómez Gómez, Aida (John Jay College - City University of New York); Munday, Jeremy (University of Leeds); Pinilla Martínez, Julia (Subdirectora & Coordinadora de publicaciones) (Universitat de València).

Board of Advisors / Comité Científico

Baker, Mona (U. of Manchester); Chesterman, Andrew (U. of Helsinki); Delisle, Jean (U. d'Ottawa); Gambier, Yves (U. of Turku); Gile, Daniel (ESIT, Université Paris 3); Hatim, Basil (American U. of Sharjah); Ladmiral, Jean-René (U. Paris X - Nanterre); Pöckl, Wolfgang (Universität Innsbruck); Venuti, Lawrence (Temple U.); Wotjak, Gerd (U. Leipzig).

Board of Referees for this issue / Comité Evaluador para este número

Arrizabalaga, Carlos (U. de Plura); Candel-Mora, Miguel Ángel (U. Politécnica de Valencia); Castañeda Naranjo, Luz Stella (U. de Antioquia); Cortés Rodríguez, Francisco José (U. de La Laguna); del Río Zamudio, María Sagrario (U. degli Studi di Udine); Djabri, Souhila (U. de Valladolid); Echenique Elizondo, M^a Teresa (U. de València); Fraga-pane, Federica (U. di Catania); González Rey, María Isabel (U. de Santiago de Compostela); González Royo, Carmen (U. d'Alacant); Holzer, Peter (U. Innsbruck); Losey León, María Araceli (U. de Cádiz); Marín Hernández, David (U. de Málaga); Martí Sánchez, Manuel (U. de Alcalá); Mellado Blanco, Carmen (U. de Santiago de Compostela); Oltra Ripoll, María Dolores (U. Jaume I); Olza Moreno, Inés (U. de Navarra); Penadés Martínez, Inmaculada (U. de Alcalá); Pérez Durán, Marco Antonio (U. Autónoma de San Luis Potosí); Pulido Correa, Martha (U. de Antioquia); Rica Peromingo, Juan Pedro (U. Complutense de Madrid); Rodríguez Espinosa, Marcos (U. de Málaga); Tabares Plasencia, Encarnación (U. Leipzig); Timofeeva Timofeev, Larissa (U. d'Alacant); Vargas Sierra, Chelo (U. d'Alacant); Vega Cernuda, Miguel Ángel (U. d'Alacant).

Número de artículos propuestos para *MonTI* Especial 6 (2020) / Number of contributions submitted to *MonTI* Special Issue 6 (2020): Catorce / Fourteen.

Número de artículos aceptados en *MonTI* Especial 6 (2020) / Number of articles accepted in *MonTI* Special Issue 6 (2020): Once / Eleven (78,6%)

La revista *MonTI* está indexada en / *MonTI* is indexed in:

BITRA, Carhus Plus+, Dialnet, ESCI (Web of Science); DICE, ERIH-PLUS, FECYT, ISOC, Latindex, Redalyc, Scopus & TSB.

Website: <https://web.ua.es/es/monti>

MonTI ha recibido ayudas económicas de la Facultad de Filosofía y Letras (Universitat d'Alacant), del Vicerectorat d'Investigació i Doctorat (Universitat Jaume I) y de la Universitat de València (Departament de Filologia Anglesa i Alemanya, Departament de Filologia Francesa i Italiana, Departament de Teoria dels Llenguatges i Ciències de la Comunicació).

Publicacions de la Universitat d'Alacant
03690 Sant Vicent del Raspeig
publicaciones@ua.es
<https://publicaciones.ua.es>
Teléfono: 965 903 480

© d'aquesta edició: Universitat d'Alacant
Universitat Jaume I
Universitat de València

ISSN-e 1989-9335
ISSN 1889-4178
Dipòsit legal: A-257-2009

Composició:
Marten Kwinkelenberg

MonTI está editada por las universidades de Alicante (Departamento de Traducción e Interpretación), Jaume I (Departament de Traducció i Comunicació) y València (Departaments de Filologia Anglesa i Alemanya, de Filologia Francesa i Italiana i de Teoria dels llenguatges i Ciències de la Comunicació).

Reservados todos los derechos. No se permite reproducir, almacenar en sistemas de recuperación de la información, ni transmitir alguna parte de esta publicación, cualquiera que sea el medio empleado –electrónico, mecánico, fotocopia, grabación, etcétera–, sin el permiso previo de los titulares de la propiedad intelectual.

PEDRO MOGORRÓN HUERTA (ED.)

MONTI
Special Issue 6 (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

UNIVERSITAT D'ALACANT
UNIVERSITAT JAUME I
UNIVERSITAT DE VALÈNCIA

ÍNDICE

Pedro Mogorrón Huerta

Multidisciplinary analysis of the phenomenon of phraseological variation in translation and interpreting..... 7

Pedro Mogorrón Huerta

Análisis multidisciplinar del fenómeno de la variación fraseológica en traducción e interpretación 36

Silvia Cataldo

Approccio cognitivo alla variazione fraseologica: alcune concettualizzazioni della pazzia in italiano e implicazioni per la loro traduzione in spagnolo e tedesco 65

Walter Koza & Natalia Rivas Folch

Modelización computacional de las locuciones verbales del español chileno a partir de la propuesta de la léxico-gramática..... 94

Daniele Zuccalà & María Bernarda Espejo Olaya

Estudio de caso de la fraseología empleada en Colombia: análisis crítico de un corpus *ad hoc* de locuciones..... 121

Carlos Manuel Hidalgo-Ternero

Google Translate vs. DeepL: analysing neural machine translation performance under the challenge of phraseological variation 154

<i>M.ª Victoria Valencia Giraldo</i>	
Standardisation in translated language: A study of collocations in Peninsular and Colombian Spanish translations of <i>The Picture of Dorian Gray</i>	178
<i>Pilar León Araúz & Melania Cabezas García</i>	
Term and translation variation of multi-word terms.....	210
<i>María del Carmen López Ruiz</i>	
La recepción y traducción de unidades fraseológicas en el discurso político. Análisis de un corpus <i>ad hoc</i> (EN, FR > ES).....	248
<i>José Luis Rojas Díaz</i>	
From head to toe: A lexical, semantic, and morphosyntactic study of idioms in phraseological dictionaries in English and Spanish.....	287
<i>Lucía Navarro-Brotons</i>	
Traducción de unidades fraseológicas gastronómicas francés-español peninsular/español de México.....	327
<i>Analia Cuadrado Rey</i>	
Las lenguas originarias en la fraseología del español hablado en Argentina: un estudio basado en corpus	355
<i>M.ª Ángeles Recio Ariza</i>	
Equivalences of phraseological units in corpora.....	386
Aims / Objetivos / Objectius.....	416

Recibido / Received: 06/11/2020

Para enlazar con este artículo / To link to this article:

<http://dx.doi.org/10.6035/MonTI.2020.ne6.1>

Para citar este artículo / To cite this article:

Mogorrón Huerta, Pedro. (2020) "Multidisciplinary analysis of the phenomenon of phraseological variation in translation and interpreting." 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. 7-35.

MULTIDISCIPLINARY ANALYSIS OF THE PHENOMENON OF PHRASEOLOGICAL VARIATION IN TRANSLATION AND INTERPRETING

PEDRO MOGORRÓN HUERTA

pedro.mogorrón@ua.es
Universidad de Alicante

Abstract

Focusing on the Spanish language, this paper aims to introduce the phenomenon of phraseological variation, presenting, on the basis of the studies carried out in the context of Hispanic linguistics, the theoretical aspects, the quantitative frequency of the variations and the phenomenon of geolinguistic or diatopic variation. The difficulty that phraseological translation represents is pointed out emphasizing the fact that this type of translation is not impossible and that the user should be able to count on phraseological tools and databases that will allow to find phraseological equivalents in other languages before thinking about using recurrent translation techniques.

Keywords: Phraseology, fixing, phraseological variation, diatopy, translation.

Resumen

El presente trabajo, centrándose en la lengua española, pretende introducir el fenómeno de la variación fraseológica, presentando, a partir de las investigaciones llevadas a cabo en el marco de la lingüística hispánica, los aspectos teóricos, la frecuencia cuantitativa de las variaciones y el fenómeno de la variación geolingüística o diatópica. Se señala la dificultad de la traducción fraseológica incidiendo en que esta traducción no es imposible y que el usuario debería poder contar con herramientas y bases de datos fraseológicas que le permitan encontrar las equivalencias fraseológicas en otros idiomas antes de pensar en usar las recurrentes técnicas de traducción.

Palabras clave: Fraseología, fijación, variación fraseológica, diatopía, traducción.



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

1. Introduction

Given that this subject has little relevance, phraseology has become a recurrent research topic in a large number of languages and even in translation, because of the many issues regarding understanding and equivalents that this topic presents. *Prinzipien der Sprachgeschichte* (Paul 1880), *Essai de sémantique* (Breal 1897), *Cours de Linguistique Générale* (Saussure 1916) and *Précis de Stylistique* (Bally 1905) are considered the precursors of phraseology. However, in the end, the Russian structuralists were the ones who laid the foundations of modern phraseology.

Nowadays, we see that there is constant scientific activity in the area of phraseology around the world. In Spanish phraseology, on which we will focus, after the pioneering contributions of Casares (1950) and Zuluaga (1980), the advancement of phraseology undergoes a decisive expansion with the publication of Corpas (1996), a reference paper that generated an exponential growth of phraseological studies. This advancement was consolidated by García-Page (2008), giving Spanish phraseology a high level of maturity.

In the last forty years, the large number of studies in the area of phraseology has shown that this is not about linguistic anomalies, analyzing elements as diverse and, at the same time, as different as:

- (1) Theoretical classifications: Casares (1959), Zuluaga (1980), Corpas (1996), Ruiz Gurillo (1997), García-Page (2008), etc.
- (2) The properties of phraseological units (PU): idiomaticity, compositionality, opacity, iconicity, language registers, motivation, metaphorization, deautomatization, etc.
- (3) The phraseodidactic: it is born within applied linguistics and it is focused on the teaching/learning of the PU of both mother and foreign languages (González Rey 2012).

2. Phraseological variation

Interest in linguistic variation arises in the middle of the 20th century with the rise of sociolinguistics¹. Originally, in the phraseological field, most of

1. This analyzes the possible modification that can be made in the daily use of the language of a community of speakers, conditioned by elements of a temporal, sociocultural or

the studies indicated that the different types of phraseological units (PU) differed from free linguistic combinations, mainly because of their fixation and idiomaticity and because their components and their structure(s) did not allow changes or modifications². Researchers, in order to demonstrate such a possibility, used PU with a high degree of fixation in which no element could be replaced by another one (Montoro del Arco 2004). For example:

- (1) *salir del armario* (DUE³), [^{*}emerger, ^{*}aparecer] del [^{*}mueble, ^{*}ropero]; *estirar la pata* (RAE/DUE), [^{*}extender, ^{*}alargar, ^{*}desplegar] alguien la pata, ^{*}estirar alguien la ^{*}pierna;
- (2) *ladrar a la luna* (DUE), ladrar [al ^{*}sol, a las ^{*}estrellas];
- (3) *apretarse el cinturón* (DUE), [*apretarse la ^{*}cintura, la ^{*}correa*]; etc.

Phraseological variation (PV) has been stirring interest, as evidenced by the many existing studies in the area of Hispanic phraseology (Carneado Moré 1985; Corpas 1996, 2008, 2018; Corpas & Mena Martínez 2003; Cuadrado-Rey 2016, 2018; García-Page 2001, 2008; Koike 2001; Mogorrón 2010, 2014, 2015, 2020a, 2020b; Navarro-Brotons 2011, 2018; Penadés 2014, 2020; Zuluaga 1980)⁵. The distinctions established and the different denominations formulated show some confusion (Montoro del Arco 2004, 2008; Alvarado Ortega 2008; Sinner & Tabares Plasencia 2016). In these studies, it has been considered, as a general rule, that PV is not a fundamental property of PU, contrary to idiomaticity and fixation. However, at the end of the 20th century,

geolinguistic nature. All languages, as Gadet (2003:13) points out, are subjected to many changes: “Il n’est pas de langue que ses locuteurs ne manient sous des formes diversifiées... les sociolinguistes, la saisissent en parlant de variétés pour désigner différentes façons de parler, de variation pour les phénomènes diversifiés en synchronie, et de changement pour la dynamique en diachronie; et ce, à la fois pour les productions d’un individu, d’un group ou d’une communauté”.

2. In other words, the components could not be replaced by other ones of the same syntactic class without changing the meaning of the PU. These PU do admit other possible modifications or transformations such as the variation of person, tense, inclusion of adverbs, pronominalization, etc.
3. All the acronyms relate to dictionaries that appear in the bibliography.
4. The * will indicate that a multi-verbal sequence is not admitted as PV, losing the original meaning of the PU.
5. The issues of the ELUA (2020) and L.I (2020) journals related to phraseological variation and diatopic variation are currently being solved.

some other studies began to highlight the possible importance of PV: “son muy numerosos los fraseologismos que admiten variaciones” (Carneado Moré 1985: 269); “La lingüística hispánica apenas ha prestado atención a la variación fraseológica, debido fundamentalmente a la falta de datos fraseológicos que cubran toda el área hispana y con los que se puedan hacer investigaciones sobre el tema. Los escasos trabajos sobre los fraseologismos [...] no han sido suficientes para obtener una visión global de la variación fraseológica del mundo hispánico” (Koike 2001: 77); “[...] no es menos cierto que la inmensa mayoría de ellas experimenta algunos cambios o puede tener formulaciones diversas al ser actualizadas en el discurso” (Montoro del Arco 2004: 593). Thus, in the case of the verbal phrases *estar en la luna* and *coger el toro por los cuernos* we find the following variants:

- (1) *estar en la luna*, *estar en* [*Babia* (DUE), *Belén* (DUE), *en el limbo* (DUE), *en la higuera* (DUE), *la parra* (CREA), *las Batuecas* (DUE), *las nubes* (DUE)].
- (2) [*agarrar*, *coger*] *el toro por los cuernos* (DUE), *coger al toro por* [*las astas*, *los cuernos*] (RAE).

The main problem is that variation appears in unpredictable processes regarding potential candidates for variant(s). Indeed, there is no rule that allows neither to determine nor to guess the reason why some words can be used as switching elements in some PU and other words cannot. This is the case of the possible variants of: *estar hasta las narices*, for which we have found: *estar hasta* [*el coco* (DUE), *el (mismísimo) coño* (DFDEA), *el copete* (LARBI), *el culo* (DUE), *el gollete* (RAE), *el gorro* (DUE), *el moño* (DUE), *el nabo* (oral), *el pirri* (VV⁶), *el pito* (VV), *el punto com* (VV), *la coronilla* (VV), *la polla* (DUE), *la punta de los pelos* (DUE), *las cejas* (DUE), *las narices* (DUE), *las pelotas* (DUE), *los cojones* (DUE), *los huevos* (DUE), *los ovarios* (VV), *los pelos* (DUE), etc.] de alguien / algo. Why can we say *estar hasta las narices*, *hasta el moño*, *hasta las cejas* and not **estar hasta las orejas* o **hasta la frente*?

6. VV means Viva Voz and its use is attested on the Internet browsers. *Estar hasta el punto com* presents 1450 occurrences on Google.

This process of variation and its unpredictability (Mellado Blanco 2004: 159; Mogorrón 2010: 96) can be found in most types of PU⁷ (namely: verbal, adjectival, nominal, adverbial phrases, collocations, sayings⁸, pragmatemes⁹, etc.). However, taking into account the material impossibility of presenting the changes in each of them, in the context of this paper, we will analyze the variation in the PU that we call fixed verbal constructions (FVC) in which we include: *poner una pica en Flandes* (DUE); 2) verbal collocations: *guiñar un ojo* (DUE); 3) support verbs: *dar un paseo* (DUE); *dar ayotes a alguien* (DUE); 4) comparative verbal constructions: *dormir como un lirón* (DUE), (see Mogorrón 2010).

The possible variations are unpredictable and, in addition to that, the number of words that can be accepted as variants can fluctuate and lead to low-productive or highly productive variants, numerically speaking. As part of the low-productive variants, we find, for example:

- (1) [*acabar, terminar*] *como el rosario de la aurora* (DUE);
- (2) [*arrojar, echar, tirar*] *la casa por la ventana* (DUE);
- (3) *echar(le) margaritas a los [cerdos, puercos]* (RAE);
- (4) [*despedirse, marcharse*] *a la francesa* (DUE); etc.

On the other hand, other PU allow a much larger number of words accepted in the process of variation and modification of the canonical form. Thus, for example:

- (1) *irse algo a [freír espárragos (DEA), freír monas (DEA), hacer puñetas (DEA), hacer gárgaras (DEA), pique (DUE), tomar por culo (DEA), tomar por el culo (DEA), tomar viento (DUE)], irse algo al [carajo (DUE), cuerno (DUE), diablo (DUE), demonio (DUE), garete (DFDEA), infierno (VV), traste (DUE)], irse algo a la [mierda (DUE), porra (DUE), etc.]*.

7. See Corpas (1996), García-Page (2008).

8. See Sevilla Muñoz (2020), Navarro-Brotons (2011, 2018, 2020). E.g.: (1) *A Dios rogando y con el mazo dando / a Dios rogando y al macho dando*. (2) *[A/ para cada cerdo / puerco] [le llega / hay / le viene] su San Martín*. (3) *[Quién / el que / al que] a buen árbol se arrima, buena sombra [le cobija / tiene]*.

9. See Alvarado Ortega (2008). E.g.: *¡me cachis en la mar!, ¡me cago en la mar!, ¡me cagüen la mar!, ¡me cago en la puta [la hostia, la leche]!*

- (2) [*descalzarse, descojonarse, descoyuntarse, despedazarse, despelotarse, desternillarse, desperezarse, mearse, mondarse, morirse, partirse, retorcerse, reventar, troncharse*] *de risa* (RAE).

Finally, cases of simultaneous variations regarding several components of a PU, such as verbs and nouns, verbs and prepositions, nouns and prepositions, etc. may occur. E.g.:

- (1) [*dejarse, quitarse*] *de [cuentos, historias]* (DUE)
 (2) [*descubrir, inventar*] [*América (VV), el agua caliente (VV), el mediterráneo (DFDEA), la pólvora (DFDEA)*];

At this point and in order to deal with the concept of phraseological variation, a broad conception of phraseology will be used here (Gläser 1986: 42; Corpas Pastor 2003) and we will talk about PV when there is, in a given PU, a substitution of one or more of its lexical or grammatical components with another one which has the same characteristics (synonymous or not), the resulting PU having the same meaning and an identical or a very similar syntactic structure.

2.1. Geolinguistic or diatopic variants

In the case of languages such as Spanish, French, English, Arabic, etc., used as official languages in many countries, there are also numerous diatopic or geolinguistic variants. In the case of Spanish, most of the cited studies - with the exception of Zuluaga (1980); Carneado Moré (1985); Square-King (2016, 2018, 2020); Koike (2001); Mogorrón (2014, 2015, 2020); Navarro-Brotons (2011, 2018, 2020); Pamies (2017)– show the typology of phraseological variations adhering to Peninsular and/or Common Spanish (EspPC), leaving aside numerous creations used in *los españoles de América* (LaS), as recognized by García-Page (2008: 253): “La lista de variantes de una lengua ha de tener en cuenta las diferencias institucionalizadas que se producen en cualquier ámbito de su empleo, en especial las de índole sociolectal o distráctica y geolialectal o diatópica. Por esa razón, resulta arriesgado afirmar que los inventarios de variantes constituyen conjuntos cerrados, sobre todo cuando se trata de lenguas como el español de un radio de acción tan amplio”.

Going back to some of the examples belonging to the EspPC in point 2., we see that Latin-American dictionaries show that there are recognized variants in Latin-American countries. E.g.:

- (1) with examples of apparently fixed PU in EspPC¹⁰:
 - a) *apretarse el cinturón* (DUE) / *ajustarse el cinturón, los cinturones* (Ddmexicanismos, Mx), *amarrarse el cinturón* (DdAm, Pe).
 - b) *salir del armario* (DUE) / *salir del closet* (DdAm, ES, Ni, PR, Co, Ve, E, Pe, Ch, Py).
- (2) with new variants in Latin-America of FVC that already present variants in EspPC. For example:
 - a) *irse algo a* [*freír espárragos* (DEA), *freír monas* (DEA), *hacer puñetas* (DEA), *hacer gárgaras* (DEA), **a justa** (DdAm, PR), **a juste** (DdAm, PR), **a la chucha** (DdAm, Ch), *pique* (DUE), *tomar por culo* (DEA), *tomar por el culo* (DEA), *tomar viento* (DUE)]; *irse algo al* [*carajo* (DUE), **cachimbo** ((DdAm, RD), *cuerno* (DUE), *diablo* (DUE), *demonio* (DUE), *garete* (DFDEA), **hoyo** (DdAm, Bo, Ch, Ar), *infierno* (VV), **tacho** (DdAm, Bo, Ch, Py, Ar, Uy), **sipote** (DdAm, RD), *traste* (DUE)], *irse algo a la* [*mierda* (DUE), *porra* (DUE)].
 - b) *coger al toro por los cuernos*, [*agarrar, coger*] *el toro por los cuernos* (DUE), *coger al toro por* [*las astas, los cuernos*] (RAE) / ***agarrar al toro por los cachos*** (DdAm, Py, RD, Ve, Pe), ***tomar el toro por las guampas*** (DdAm, Ur).
 - c) [*lanzarse, tirarse*] *alguien a la piscina* (RAE) / ***tirarse a la pileta*** (DdAm, Ar, Ur).

10. LaS variants are indicated by separating them with a slash /. Usually, Spanish dictionaries include a large majority of words and expressions from EspPC and Latin American dictionaries include content from LaS. As an example of this, the two most representative dictionaries of the EspPC, the DRAE and the DUE, have only about 770 PU of the LaS.

2.2. Numerical variations in PU

The first contributions in which the variations of the FVC were analyzed and quantified have been carried out by the University of Alicante FRASYTRAM¹¹ research group (see Mogorrón 2010, 2015), thanks to a Database (DB) of 41,000 FVC collected thanks to more than 30 Spanish and Latin-American dictionaries, 26,000 of which belong to EspPC and 15,000 to LaS.

In the case of the EspPC, the analysis of the 26,000 entries highlights that 15,001 FVCs, that is a 58%, present some type of variation (see figure 1).

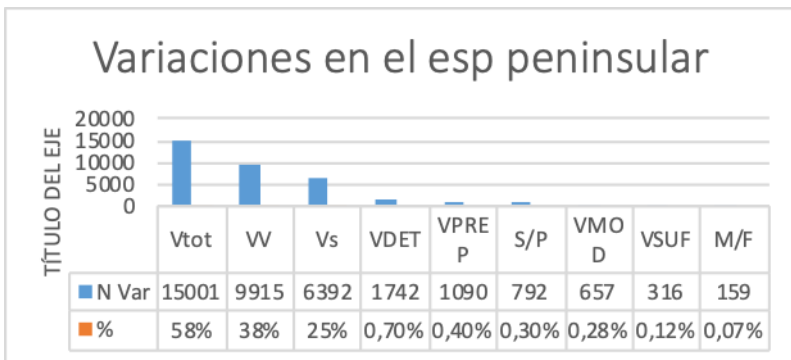


Figure 1. Variations in Peninsular Spanish

In the case of LaS, the analysis shows that 3359 CVF of LaS present some type of variation¹² (see figure 2).

11. Fraseología y Traducción multilingüe (Phraseology and Multilingual Translation)

12. This is an ongoing analysis.

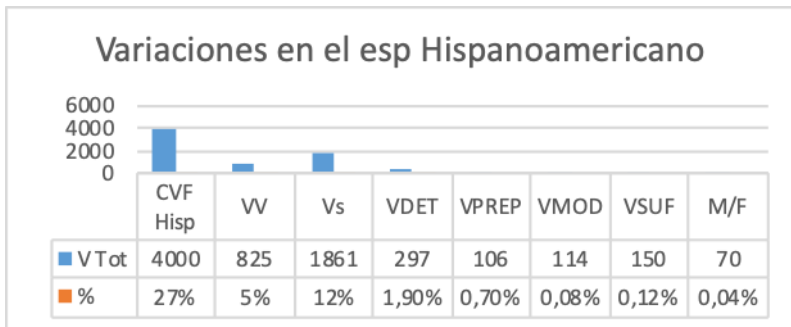


Figure 2. Variations in Latin-American Spanish

In both cases, the most frequent variations take place with lexical variants, either verbal or nominal. Now, while in the EspPC there are more verbal variations, in the LaS there are more nominal variations.

2.3. Actual use of phraseological variants

Finding the possible phraseological variants (PV) generally implies an extensive lexicographic compilation work, consulting different dictionaries, since the phraseological content of each one presents considerable differences (see table 1).

DUE	[agarrar / coger] el toro por los cuernos
RAE	coger al toro por [las astas / los cuernos]
Espasa Calpe	[agarrar / coger / tomar] el toro por los cuernos

Table 1: PV in dictionaries

However, is this collection of documentary sources complete? Currently, the possibility of searching for some lexical chains in textual corpora¹³, through

13. Since the 1960s, textual corpora have significantly changed linguistic studies, allowing researchers to have great volumes of data and to study the language through real examples using automatic processes (Corpas, 2008, 2013, 2018, etc.), which extract the requested results thanks to natural language processing.

automatic text processing, allows finding numerous VF that are not included in the analyzed peninsular dictionaries and that are still used by the speakers daily. In the case of the variants of *coger el toro por los cuernos*, according to the uses listed in the dictionaries, the forms *tomar al toro por los cuernos*, *agarrar al toro por las astas*, *tomar el toro por las astas* would not be acceptable despite being on Google.es on the Internet.

In order to find out the possible variants of *coger el toro por los cuernos* y de *echar margaritas a los cerdos* (see tables 3 and 4), a search was conducted in three textual corpora: CREA¹⁴, CORDE¹⁵ and Sketch-Engine (S-E)¹⁶, which bring together a large collection of representative texts. The search for the VF, alphabetically ordered starting with the verb, offers the following results:

		CREA	CORDE	S-E	EspPC	LaS
1	<i>agarrar al toro por los cachos</i>			4	3	1
2	<i>agarrar al toro por los cuernos</i>	2		194	153	41
3	<i>agarrar al toro por las astas</i>			14	10	4
4	<i>agarrar el toro por las astas</i>		2	35	14	21
5	<i>agarrar el toro por las guampas</i>			2		2
6	<i>agarrar el toro por los cachos</i>	1	1	8		8
7	<i>agarrar el toro por los cuernos</i>	1	1	184	82	102
8	<i>amarrar al toro por los cuernos</i>			1	1	
9	<i>asir al toro por las astas</i>			2		2
10	<i>asir al toro por los cuernos</i>			1		1

14. REAL ACADEMIA ESPAÑOLA: Database (CREA) [online]. Corpus de referencia del español actual. <http://www.rae.es>.

15. REAL ACADEMIA ESPAÑOLA: Database (CORDE) [online]. Corpus diacrónico del español. <http://www.rae.es>.

16. <https://www.sketchengine.eu/> el corpus esTenTen18 de S-E, consists of the *European Spanish Web* and *American Spanish Web* subcorpus and has 17.553.075.259 tokens, so it is a large corpus, with 49.32% of European Spanish domains and 46.46% of Latin-American Spanish.

11	<i>asumir al toro por los cuernos</i>			1		1
12	<i>atacar al toro por los cuernos</i>		1	1	1	
13	<i>atrapar al toro por las astas</i>			1		1
14	<i>atrapar al toro por los cuernos</i>			1	1	
15	<i>coger al toro por las astas</i>		1	6	3	3
16	<i>coger al toro por los cachos</i>			4		4
17	<i>coger al toro por los cuernos</i>	3	1	353	333	20
18	<i>coger el toro por los cachos</i>			2		2
19	<i>coger el toro por los cuernos</i>	39	1	2050	1980	70
20	<i>enfrentar al toro por los cuernos</i>			11	2	9
21	<i>entrar al toro por los cuernos</i>	2		59	38	21
22	<i>ir al toro por los cuernos</i>			4	1	3
23	<i>pillar al toro por los cuernos</i>			3	3	
24	<i>tener al toro por las astas</i>			4		4
25	<i>tener al toro por los cuernos</i>			11	6	5
26	<i>tomar al buey por los cuernos</i>			2	1	1
27	<i>tomar al toro por las astas</i>	2		191	31	160
28	<i>tomar al toro por los cachos</i>			8	3	5
29	<i>tomar al toro por los cuernos</i>	3	1	338	100	238
30	<i>tomar el toro por las astas</i>	8	2	793	20	773
31	<i>tomar el toro por las guampas</i>	1		16		16
32	<i>tomar el toro por los cachos</i>			57	10	47
33	<i>tomar el toro por los cuernos</i>	10		620	160	460

Table 2. VF *coger el toro por los cuernos*

		CREA	CORDE	S-E	EspPC	EspHisp
1	<i>arrojar margaritas a los cerdos</i>		1	8	4	4
2	<i>arrojar margaritas a los chanchos</i>			1		1
3	<i>arrojar margaritas a los puercos</i>	1	1	3	1	2
4	<i>dar de comer margaritas a los cerdos</i>			11	4	7
5	<i>dar margaritas a los cerdos</i>			33	29	4
6	<i>dar margaritas a los chanchos</i>			11		11
7	<i>dar margaritas a los cochinos</i>			1	1	
8	<i>dar margaritas a los gorrinos</i>			1	1	
9	<i>dar margaritas a los puercos</i>			4	1	3
10	<i> echar margaritas a los cerdos</i>	4	1	84	76	6
11	<i> echar margaritas a los puercos</i>		3	11	6	5
12	<i> lanzar margaritas a los cerdos</i>			4	2	2
13	<i> ofrecer margaritas a los cerdos</i>			3	3	
14	<i> tirar margaritas a los cerdos</i>			15	7	8
15	<i> tirar margaritas a los chanchos</i>			30		30

Table 3. VF *echar margaritas a los cerdos*

In both cases, we have highlighted in bold the VF collected in the dictionaries used for the elaboration of the DB. Although it is true that all the VF that are collected in the dictionaries appear in the textual corpora, it can be observed that they do not always include the most usual forms and that some added VF present very few occurrences in the three analyzed textual corpora. What is most outstanding is that they can be detected with contextualized examples:

- (1) uses of the same forms on the two sides of the Atlantic, corroborating the existence of a common Spanish language. The variants with the most occurrences in S-E, and surely the most common today, appear on both sides of the Atlantic. E.g.: *coger el toro por los cuernos* (DUE) *echar margaritas a los cerdos* (RAE); y otras muchas como *estar en la*

luna (RAE); *estirar la pata* (DUE); *comerse las uñas* (DUE); *hacer el agosto* (DFDEA); etc.

(2) new variations that did not appear in dictionaries.

However, the use of these magnificent tools does not imply that variants will be found in each and every PU, since many are still unalterable concerning its formulation. This would be the case of *hacer de tripas corazón* (RAE), *ladrar a la luna* (RAE), *ponerse el mundo por montera* (RAE), *estar entre Pinto y Valdemoro* (RAE); *dar el ala para comerse la pechuga* (DdAm, PR); that appear in many occurrences in EspPC and/or LaS in S-E.

In view of the data obtained, we think that it is convenient to reconsider not only the importance in the world of phraseology of variations, given that the two concepts of fixation and variation are surely the two opposite faces of the same phenomenon, but also its treatment and consideration as an essential property, and most likely with the same importance as fixation and idiomaticity. As anticipated by Corpas & Mena Rodríguez (2003: 183):

hay que admitir que el concepto de fijación tal y como fue planteado en las primeras investigaciones queda en entredicho. Si en un principio las UFs se definieron como unidades estables y fijas que mostraban rechazo a cualquier alteración léxica, semántica y morfosintáctica, en vista de la existencia de cambios reales y potenciales, no cabe hablar de la fijación como una propiedad absoluta (Burger 1998) sino como una cualidad relativa (Fleischer 1982; Gläser 1986; Wotjak 1992; Corpas 1996; Burger 1998) y variable.

2.4. Variation of meaning

Another type of variation that should be highlighted, in relation to the understanding and translation process of the PU, which has not been analyzed either, is related to their meaning(s) and geolinguistic polysemy. However, it is also a problem of great magnitude. The exhaustive analysis of the 41,000 FVC of the BD has made it possible to collect 14,238 cases of polysemic FVC so far and to identify polysemic FVC that present situations ranging from two meanings to ten (Mogorrón 2020) with the following results (see table 4).

polysemic FVC	
1 meaning	26652
2 meanings	4335
3 meanings	1006
4 meanings	328
5 meanings	120
6 meanings	31
7 meanings	19
8 meanings	6
9 meanings	1
10 meanings	4

Table 4. FVC, cases of polysemy

Furthermore, the analysis of the 14,238 cases of polysemy to determine the area of use, offers the following results:

- (1) 8.423 FVC belong to EspPC only, being the most frequent situation. E.g.:
 - a) *pegar un metido* (1. Regañar, reñir con rigor DUE; 2. Sonsacar dinero a alguien, DUE; 3. Pegar, dar una paliza a alguien, DUE);
 - b) *pedir la luna* (1. Pedir, pretender algo muy difícil o imposible de obtener, RAE; 2. Pedir un precio o sueldo exagerado, DFDEA), etc.
- (2) 1.358 FVC have meanings that belong to EspPC or LaS. E.g.:
 - a) *bañarse en agua de rosas* (1. Alegrarse de algún mal ajeno, DUE; 2. Sentirse alguien muy orgulloso por el éxito de un ser querido, DdAm, Ve, Ec; 3. Disfrutar de bonanza económica, DdAm, Gu);
 - b) *no tener abuela* (1. Se dice irónicamente de la persona que se alaba a sí misma, DUE; 2. Ser malo, DdAm; 3. Mantener una conducta incorregible, DdAm, Mx, ES, Ni, Ar); etc.
- (3) 4.454 FVC have meanings that belong only to LaS. E.g.:
 - a) *enredar la pita* (1. Complicar las cosas, DdAm, Ho, Ni, Co, RD, PR; 2. Hacer difícil de comprender algo, DdAm, Cu, RD; 3. Confundirse, DdAm, RD; 4. Irse por las ramas, DdAm, Ni);

- b) *acabarse el veinte* (1. Agotarse el tiempo, DDMex, Mx; 2. Agotarse la oportunidad de hacer algo, DDMex, Mx; 3. Acabarse la suerte, DdAm, Mx); etc.

2.5. Phraseological parasynonymy

The last type of variation that we would like to talk about in this paper, also due to its great impact on translation, is the one about phraseological parasynonymy. In addition to the idiosyncratic culturemes associated with a well-determined culture (Pamies 2008), there are many concepts in many languages that we could qualify as polycultural, or even as phraseological universals (Dobrovol'skij 1988; Corpas 1995, 1998 ; Luque & Majón 2002; Mogorron 2002), which are so usual that speakers of different languages have created many expressions with the same meaning to refer to these concepts. In this case we can find parasynonymic groups ranging from a couple of expressions to groups formed by several dozens of PU. These usual polycultural concepts can refer to any activity of daily life, for example: sex, food, drink, death, work, pregnancy, skipping class, being fat, skinny, being handsome, ugly, short, small, brave, fearful, eating a lot, homosexuality, wealth, poverty, distraction, etc.

The 41,000 entries in the FRASYTRAM database have been semantically classified according to their definition(s). As an example, the following parasynonymic FVC for *being distracted* have been collected on the basis of the UF *estar en la luna*:

- (1) **in Spain:** *andar en las nubes* (RAE); *andar por las nubes* (DUE); *cazar tilingos* (RAE); [*contar, estar contando, ponerse a contar*] *las vigas* (RAE); [*estar, quedar*] *a uvas* (EPM); *estar en Babia* (DUE); [*estar, estar bailando*] *en Belén* (RAE); *estar con la torta* (DTDFH); *estar con los angelitos* (RAE); *estar en el limbo* (RAE); *estar en la higuera* (RAE); *estar en la luna* (DUE); *estar en la parra*; *estar en las Batuecas* (DUE); *estar en las nubes* (DUE); *estar pensando en la mona de Pascua* (DTDFH); *ir a por uvas* (Akal); [*mirar a, pensar en*] *las musarañas* (RAE); *mirar las telarañas* (LARBI); *quedarse encantado* (DUE); *tener la cabeza a las once* (RAE); *tener la cabeza a pájaros* (RAE); *tener la cabeza en el aire* (VV) *tener la cabeza en la luna* (VV); *tener la cabeza*

- en las nubes* (RAE); *tocar el violón* (RAE); *tocar el violón a dos manos* (DTDFH); *vivir en las nubes* (DUE); *etc.*
- (2) **in Latin-America:** [*andar, vivir, pasársela*] *en la luna* (DEUEM, Mx, Ch); [*andar, estar*] *fuera de onda* (DTDFH, Mx); *cortar varas* (DdAm, Gu, Ho, Nicaragua); *estar* (alguien) *en el aire* (RAE; Cu); *estar en la estratosfera* (DFHA, Ar); *estar en otra* (DFHA, Ar); *estar papando moscas* (GDLA, Ar); *estar en la luna de Paíta* (Bo, Ec, Pe); *estar en la luna de Paíta y el sol de Colán* (DdAm, Pe); *estar en las nebulosas* (RAE, Ve); *estar fuera de onda* (DTDFH, Mx); *estar pensando en los pajaritos de colores* (DdAm, Ar); *pensar en la inmortalidad del cangrejo* (DdAm, Mx, Gu, Ho, ES, Ni, Py, Cu, RD, Ve, Ec, Bo, Ur); *pensar en la inmortalidad del mosquito* (DdAm, Pe); *pensar en las muelas del gallo* (DdAm, Gu); *pensar en los anteojos del gallo* (DdAm, Gu); *pensar en los huevos del gallo* (DdAm, CR); *quedarse en China* (DdAm, Cu); [*ser, tener*] *cabeza de novia* (GDLA, Ar); *tener la cabeza en los pies* (DEUEM, Mx); *vivir en el limbo* (GDLA, Ar).

3. Translation of the PU

3.1. Translation and phraseology

Translation is a complicated multifunctional process which is much more complex than the simple fact of searching for equivalences between words or sentences. In this process, the purpose and the context that will give rise to different translation methods (Hurtado Albir 2001: 252) should not be ignored. However, PU represent the desire of the speaker to use in the act of communication, written or spoken, a specific wording endorsed by a large part of society, adding a cultural, socializing, generational nuance and performing stylistic, discursive, argumentative and even pragmatic functions instead of using a neutral word or a verbal paraphrase. Furthermore, its use presupposes that the interlocutors have a common cultural and linguistic background, otherwise they will find groups of words that they will not understand and will have to rely on other users' help or dictionaries to try to decode this possible PU.

What would the spoken language be without PU or even the literary works of some authors as expressive in their discourse as:

- (1) Pérez Reverte (*El capitán Alatriste*, 1996: 123 y 133): “pueden más dos tetas que dos carretas”, “dormir a pierna suelta”, “puso pies en polvorosa”, “dar tres cuartos al pregonero”, “la pasó en vela”, “estar al tanto”);
- (2) Torres (¡Oh es Él!, 1998: 33 y 44): “echar las campanas al vuelo”, “caer en manos de”, “estás como un cencerro”; “tenía el corazón en un puño”, “estaba hecha un manojo de nervios”;
- (3) Ruiz Zafón (*El juego del ángel*, 2008: 23-24): “ver ni en pintura”, “subsistir por los pelos”, “otro gallo nos cantaría”, “de medio pelo”, etc.);
- (4) Gala (*El dueño de la herida*, 2003: 24-25): “le echó en cara”, “la gata flora, que si se la meten chilla, y si se la sacan llora”, “quitarme de en medio”, “soy de ordeno y mando”).

What would their translation become if these forms that the author has consciously used on the basis of their inherent phraseological function were omitted or eliminated? It is true that the phraseological competence of every user, already limited in their mother tongue, decreases considerably in other languages (Corpas 2003, Tabares & Batista (2019)). It is also true that the phraseological content of dictionaries is far from being complete and it is also true that the translation of the PU represents a real challenge for translators, but no matter how difficult it is, it is an acceptable challenge (Corpas 2003; Richart Maset 2008) as indicated by Mona Baker (1992: 65): “the main problems that idiomatic and fixed expressions pose in translation relate to two main areas: the ability to recognize and interpret an idiom correctly; and the difficulties involved in rendering the various aspects of meaning that an idiom or a fixed expression conveys into the target language”, and Ballard (1998: 247): “le jugement de fidélité et de qualité que l’on porte sur la traduction ne peut s’exercer qu’au travers des formes, porteuses de sens et d’effets stylistiques”, having then to forge a work of translational goldsmithing.

3.2. Steps of the phraseological translation process

Several authors have analyzed phraseological translation, highlighting that the different steps of the process correspond, according to Corpas (2003), with the identification, interpretation and search for equivalence (at the

lexical and later textual level) (Sevilla & Sevilla 2005, 2009; Mogorrón 2008; Timofeeva 2008, etc.).

The identification and interpretation-understanding steps of the PU are closely related to idiomaticity/opacity, as well as to the phraseological competence of the translator, who will be facing several prototypical situations.

- (1) If the PU belongs to the translator's phraseological competence, he or she should try to find another equivalent of meaning in the target language.
- (2) If it is not within the translator's phraseological competence, he or she may identify it or not.
- (3) If he or she does not identify it as a PU, it will be detrimental to the translation, no matter how much the overall meaning of the text is preserved.
- (4) If he or she does identify it as a PU, again, he or she will have to search for it in monolingual or bilingual dictionaries to see if the dictionary hopefully includes that form, and then he or she will have to be able to identify, interpret and translate the PU.

It has become evident that the process of understanding/interpreting the PU is much more complex than it might seem, as evidenced by the examples shown below, due to the limited phraseological competence of each user, the deficient phraseological treatment of dictionaries¹⁷ (Corpas 2003: 155-169; Mogorrón 2011, 2015) and the numerous phraseological and paremiological publications in the publishing market. E.g.:

- (1) *cerrar los ojos* 1. Dormir, RAE; 2. Morir, RAE; 3. Sujetar el entendimiento al dictamen de otro, RAE; 4. Obedecer sin examen ni réplica, RAE; 5. Arrojar temerariamente a hacer algo sin reparar en inconvenientes, RAE; 6. No apartarse de un enfermo hasta que expire, RAE; 7. Permitir cierta cosa haciendo como que no se advierte, DUE; 8. No querer enterarse de ello para no disgustarse o por terquedad, DUE; 9. Hacerlo como gesto de miedo, DUE; 10. Alucinarle, para que no oiga, vea lo que le conviene, RAE;

17. The expression *salir del armario* does not appear in the DRAE in October 2020.

- (2) *meter el diente* 1. Morder o comer algo difícil de mascar, DFDEA; 2. Empezar con interés una cosa que implica dificultad, DUE; 3. Apropiarse de algo que pertenece a otra persona, RAE; 4. Criticar hablar mal de alguien, RAE; 5. Cobrar más de lo establecido, DdAm, ES, CR, Pa;

3.3. Existence of phraseological correspondences

After the identification and understanding/interpretation process, the translator has to search for correspondences. The semantic analysis of the 41,000 entries in the DB and the search for translational and phraseological equivalents has allowed us to see that there are many cases of parasynonymy and that in the translation process the translator is going to find three well-differentiated situations.

- (1) The one in which in the source and target language there is a single PU to express a situation (see table 5):

Spanish UF	French UF	English UF
[marcharse, despedirse] a la francesa (DUE)	filer à l'anglaise (GR)	To take french leave ((DCU) to make an Irish exit ¹⁸ .
cantar victoria (RAE)	chanter victoire (GR)	to claim/cry victory (DCU)
poner una pica en Flandes (RAE)	mettre 1 grain de sel sur la queue d'un oiseau (GR)	to bring off a real coup (DCU)

Table 5. Only PU in source and target language

- (2) The one in which for the same situation we find several PU in different languages (see table 6):

18. In the same way that we are collecting the Latin-American Spanish PU, it would be very important to be able to differentiate the British ones from those of the United States or from other English-speaking countries in order to be able to distinguish between translation and understanding.

Empezar la casa por el tejado (RAE) Poner el carro antes de los bueyes (VV)	mettre la charrue avant les boeufs (GR) mettre la farine avant le beurre (VV)	To put the cart before the horse (DCU)
estar en la luna (RAE) estar en las nubes (DUE) estar en el limbo (RAE) estar en las Batuecas (DUE) tener la cabeza a las once (RAE)	être dans la lune (GR) avoir la tête dans les nuages (GR) être dans les nuages (GR) Bayer aux corneilles (GR)	to have one's head in the clouds (CDU) to be lost in reverie (CDU)

Table 6. Several PU in many languages

- (3) The one in which there is a PU to express a situation in one language but not in (an)other(s) (see table 7).

dar gato por liebre (RAE)	rouler, tromper qq'un (LBI)	<i>[buy/accept] a pig in a poke to pull the wool over someone's eyes to take someone for a ride</i>
armarse la de San Quintín (DUE)	il va y avoir du grabuge	There will be an almighty row All hell will break loose
salir por la puerta grande (DFDEA)	sortir par la grande porte	to make a triumphant exit to do something in grand style

Table 7. The PU does not exist in all languages

3.4. Phraseological equivalence degrees

Although the concept of equivalence is recurrent in translation (Nida 1964, Newmark 1982, Rabadán 1991, Elena García 1994, Hatim & Mason 1995; Jakobson 1959¹⁹), phraseological equivalence has been fundamentally researched in contrastive linguistics, analyzing the lexical structures and components, looking for parameters that, based on formal and semantic

19. Jakobson (1959 [1975]: 70): "La equivalencia en la diferencia es el problema cardinal del lenguaje y la cuestión central de la Lingüística".

similarities and differences, would allow to set types of equivalences, such as isomorphisms²⁰ (syntactic, lexical-semantic, etc.). These studies did not take into account the level of the language, the frequency and the area of use, or the possible connotations; which is, in the opinion of many translators, essential information in order to obtain a good translation.

On the other hand, the study of phraseological correspondences establishes that there are three degrees of translation equivalence (Corpas 2003; Mogorrón 2008, 2014; J. Sevilla 1997; M. Sevilla 2015; Mellado Blanco 2015). That is:

- (1) total or full equivalence (both units have the same denotative and connotative meaning. E.g: *quemar las naves* (RAE) / *to burn one's boats* (CDU) / *brûler ses vaisseaux* (GR).
- (2) partial equivalence (when a locution of the source language is translated by a locution in the target language that presents differences of denotative or connotative meaning); or it is translated by a lexical unit without the expressive values of the locution in question: *quemar las naves* (RAE) = *couper les ponts*²¹ (GR); *dar gato por liebre* (RAE) = *to pull the wool over someone's eyes/rouler, tromper qq'un*;
- (3) null equivalence (these are cases of sociolinguistic realities in the source language that are not lexicalized in the target language. The translator has to choose some other transfer technique to express the meaning of the unit, often losing the corresponding connotations); *armarse la de San Quintín* (RAE) = *There will be an almighty row / il va y avoir du grabuge*; *salir por la puerta grande* = *to make a triumphant exit / sortir triomphalement*.

The most frequent type of equivalence between phraseological systems is generally the partial one. We agree with Corpas (2003: 254):

20. Giving way to a large number of false friends like *salir por la puerta grande / sortir par la grande porte* (reach a good situation, TLF1).

21. In French, *brûler ses vaisseaux* and *couper les ponts* could be used with the same meaning, but with different connotations. When searching for the form "cortar los puentes" in Spanish, we see that: (1) it does not appear in any of the major reference dictionaries (DRAE, DUE); (2) one occurrence appears in CREA, two in CORDE, more than 500 in Sketch-Engine and more than 2,000 in the Google search engine.

Las correspondencias fraseológicas que se establecen entre dos lenguas distintas suelen ser relaciones de equivalencia parcial por cuanto las UF implicadas presentan ciertas diferencias y solapamientos con respecto a uno o varios de los siguientes aspectos: contenido semántico, base figurativa, características morfosintácticas, restricciones diasistemáticas y rasgos pragmáticos.

Searching for phraseological equivalences in a thematic database of 2,482 Spanish PU and 2,031 French PU classified semantically (Mogorrón 2002) makes it possible to demonstrate this statement and proves that there are very few cases of null phraseological equivalence, at least between phraseological systems of nearby languages and cultures, with less than 5% of the cases²².

We will not deal here with the techniques and strategies of translation, nor with the difficulties of phraseological translation that may arise at the lexical or textual level, which were already discussed in many studies and that should be used when an suitable phraseological equivalent is not found (Capra 2012; Corpas 2000, 2003; J. Sevilla 1997; Sevilla & Sevilla 2000; M. Sevilla 2009, 2015).

The translator should know or be able to quickly know the different possible meanings of the PU (see 2.4). PU like *pedir la luna*, *pegar un metido* have several meanings that should be distinguished in each case. PU such as *no tener abuela*, *bañarse en agua de rosas*, *enredar la pita*, can have many different meanings depending on the geographical origin of the text in which we find them. We are in an endless circle where we go back to the starting point again and again, because if the translator detects the PU, but does not have all the necessary tools to interpret them correctly, he or she will be powerless and will have to interpret them blindly.

It is extremely important that these tools also include all possible variants in order to facilitate their location and correct interpretation. Thus, in the case of the PU *no pegar ojo* we find in the DRAE *no pegar [ojo, el ojo, los ojos]* with the meaning of not being able to sleep and in the DFDEM: *no pegar [ojo, los ojos]*. If with these data we see the subtitled version of the French film *Les femmes du sixième étage*²³ in which we see on screen: *no pude pegar*

22. Obviously, in the future it would be useful to find bilingual and bicultural experts who would carry out this type of analysis among PU belonging, for example, to Chinese, Arabic or Japanese, to mention linguistically and culturally distant languages.

23. By Philippe le Guay (2010). Spanish title: *Las chicas de la sexta planta*.

un ojo, we can then think that it is an error or a wrong literal translation (see screenshot 1).


Subt: No pude pegar un ojo
D: No he podido pegar ojo
O: J'ai pas pu fermer l'oeil 00:03:56 – 00:04:05

Screenshot 1

However, the DUE says: no pegar [ojo, el ojo, los ojos, *un ojo*], no cerrar los ojos and the DFDEA: [cerrar, pegar] [ojo, el ojo, los ojos, *un ojo*], showing that it is a correct variant accepted by the Hispanic lexicography. The translation of PU is therefore not impossible, but very complex, but it is due to: the large number of PU in each language; fixation and, on its opposite side, variation; the limited phraseological competence of any user and any translator; the limited phraseological content of reference or specialized dictionaries; the phraseological universals (2.5); the several cases of parasynonymous PU.

Though they themselves could be the subject of all kinds of very interesting research papers, we will not address the topic of translation techniques. Indeed, before using translation techniques to translate a PU, the first step should be a DB that helps to confirm the existence of a PU, its variant(s) and possible meaning(s), as well as parasynonymous expressions and their phraseological equivalents. This DB would allow us to visualize the meanings of numerous PU which are not part of the phraseological competence of the

It allows us to know the phraseological correspondences in many languages²⁵, and also to select them thanks to filters according to diatopic values, language levels, frequency of use, and it quickly provides the PU with the required lexical and textual values.

With regard to translation, this type of application, which incorporates the possible variations, the PU and their possible phraseological equivalent(s) has multiple uses:

- (1) It will allow machine translation programs not to make verbatim versions of variants that have not been included in their terminology bases.
- (2) It will allow in some translation modalities to select the equivalents according to the textual parameters, for example, in literary translation.
- (3) It will allow to select the closest phraseological equivalent according to the source language depending on technical aspects that must be taken into account, such as the correct oral synchronization in audiovisual translation.

Conclusions

Phraseological variation is a complex phenomenon whose numerical importance shows that it must be given the relevance it deserves. The development of corpus linguistics will undoubtedly make it possible to further highlight its importance in terms of users and geolinguistics. Regarding phraseological translation, it is convenient to repeat what has already been indicated: it is not impossible, since (very) often the same concepts are used as PU in different languages. It is complex, but it is due to: the large number of PU in each language; fixation and variation; the limited phraseological competence of any user; the limited phraseological content of reference or specialized dictionaries; the phraseological universals (2.5); the several cases of parasyonymous PU, etc.

Before using recurrent translation techniques, it is essential for human and machine translation to develop and have huge databases that will allow

25. <https://dti.ua.es/es/frasytram/grupo-de-investigacion-frasytram.html>

to decide whether to use an available phraseological equivalent or to opt for another solution.

Although we should always bear in mind that translation involves a more complex process than simply searching for correspondences, it should not be forgotten that the step of going from a lexical level to a contextualized discourse leads to consider the possible associated pragmatic values.

References

- BAKER, Mona. (1992) *In Other Words. A Coursebook on Translation*. London & New York: Routledge.
- BALLARD, Michel. (1998) "Comparatisme et didactique de la traduction." In: García Izquierdo, Isabel & Joan M. Verdegel (eds.) *Los estudios de traducción: un reto didáctico*. Castellón: Universidad Jaume I, pp. 55-68.
- CARNEADO MORÉ, Zoila. (1985) "Notas sobre las variantes fraseológicas." *Anuario L/L* 16, pp. 269-277.
- CORPAS, Gloria. (1996) *Manual de fraseología española*. Madrid: Gredos.
- CORPAS, Gloria. (2003) *Diez años de investigación en fraseología: Análisis sintáctico-semántico, contrastivos y traductológicos*. Madrid: Iberoamericana & Vervuert.
- CORPAS, Gloria. (2008) *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Berlin & Frankfurt am Main: Peter Lang
- CORPAS, Gloria. (2018) "Laughing one's head off in Spanish subtitles: a corpus-based study on diatopic variation and its consequences for translation." In: Mogorrón, Pedro & Juan Antonio Albaladejo-Martínez (eds.) 2018. *Fraseología, Diatopía y Traducción*. Amsterdam & Philadelphia: John Benjamins, pp. 31-70.
- CORPAS, Gloria & Florentina Mena Martínez. (2003) "Aproximación a la variabilidad fraseológica de las lenguas alemana, inglesa y española." *ELUA* 17, pp.181-201.
- CUADRADO-REY, Analía. (2018) "La variación fraseológica en el español hablado en Argentina: propuesta de clasificación." In: Mogorrón, Pedro & Juan Antonio Albaladejo-Martínez (eds.) 2018. *Fraseología, Diatopía y Traducción*. Amsterdam & Philadelphia: John Benjamins, pp. 334-352.
- GARCÍA-PAGE, Mario. (2008) *Introducción a la fraseología española: Estudio de locuciones*. Barcelona: Anthropos.

- GADGET, Françoise. (2003) *La variation sociale en français*. Paris & Gap: Ophrys.
- KOIKE, Kazumi. (2001) "Variación fraseológica del español." *Varilex* 9, pp. 77-92.
- MOGORRÓN, Pedro. (2002) *La expresividad en las locuciones verbales en francés y en español*. Alicante: Publicaciones Universidad de Alicante.
- MOGORRÓN, Pedro. (2010) "Analyse du figement et de ses possibles variations dans les constructions verbales espagnoles." *Linguisticae Investigationes* 33:1, pp. 86-152.
- MOGORRÓN, Pedro. (2015) "La variation dans les Constructions Verbales Figées de l'espagnol d'Amérique." *Linguisticae Investigationes* 3:2, pp. 276-301
- MOGORRÓN, Pedro. (2019) "Les phrasèmes et les outils lexicographiques des traducteurs." *Phrasis* 3, pp. 12-37.
- MOGORRÓN, Pedro. (2020) "Création, variabilité, variantes phraséologiques et diatopiques." *Cahiers de Lexicologie* 116:1, pp. 173-196.
- MONTORO DEL ARCO, Esteban. (2004) "La variación fraseológica y el diccionario." In: Battaner, Paz & Janet DeCesaris (eds.) 2004. *De Lexicografía*. Barcelona: Institut Universitari de Lingüística Aplicada, pp. 591-604.
- NAVARRO-BROTONS, Lucía. (2018) "Refranes españoles y sus equivalentes mexicanos." In: Mogorrón, Pedro & Juan Antonio Albaladejo-Martínez (eds.) 2018. *Fraseología, Diatopía y Traducción*. Amsterdam & Philadelphia: John Benjamins, pp. 263-177.
- SEVILLA, Julia & Manuel Sevilla. (2000) "Técnicas de la traducción paremiológica (francés-español)." *Proverbium* 17, pp. 369-386.
- ZULUAGA OSPINA, Alberto. (1980) *Introducción al estudio de las expresiones fijas*. Frankfurt am Main: Peter Lang.

DICTIONARIES

- ACADEMIA MEXICANA DE LA LENGUA. (2010) *Diccionario de Mexicanismos*. México: Siglo XXI. [DDMexicanismos]
- ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA. (2010) *Diccionario de Americanismos*. Madrid: Santillana. [DDAM]
- GARCÍA-PELAYO, Ramon & Jean Testas. (1993 [1967]) *Larousse moderno français-espagnol español-francés*. Paris: Larousse. [LBI]
- LARA, Luis Fernando (dir.) (1996) *Diccionario del español usual en México*. (1996). México: El Colegio de México. [DEUEM]
- MOLINER, María. (1999 [1977]) *Diccionario de uso del español*. Madrid: Gredos. [DUE]

- RAE. (1992) *Diccionario de la Real Academia Española* (21^a ed.). Madrid: Espasa-Calpe. [DRAE]
- REY, Alain. (2001) *Le Grand Robert: Le dictionnaire le plus complet de la langue française*. Paris: Société du nouveau Littre. [GR]
- SECO, Manuel; Olimpia de Andrés & Gabino Ramos. (2004) *Diccionario fraseológico documentado del español actual*. Madrid: Aguilar. [DFDEA]
- VARELA, Fernando. (1996) *Diccionario fraseológico del español moderno*. Madrid: Gredos. [DFDEM]
- VV. AA. (2004) *Diccionario Collins Español-Inglés / English-Spanish*. Barcelona: Grijalbo. [DCU]

BIONOTE / BIONOTA

Professor at the University of Alicante (UA). He has been director of the Department of Translation and Interpreting (2009-2017), coordinator of the PhD Programme in Translation distinguished with a quality mention (2006-2009) and is currently coordinator of the inter-university PhD Programme in Translation Theory, Professional and Audiovisual Translation. He was the principal researcher on many national, regional and local research projects and directed numerous Final Degree Projects, Master's final projects and supervised 16 doctoral theses on terminology, audiovisual translation, translation of cultural elements, phraseology, didactic phraseology and diatopic variations. Since 2006 he is director of the FRASYTRAM research group (Phraseology and Multilingual Translation) and has published numerous papers in journals and books by publishers listed in the Scholar Publishing Index (SPI). He has been a visiting professor at the Universities of Paris 13, Paris Marne-la-Vallée, Paris-Ivry, Paris-Diderot, U. Pedagogie de Cracow, U. de Antananarivo, U. R. Palma, etc. His research and study lines are: terminology and specialized translation, audiovisual translation and humanistic-literary translation (literary, cultural, journalistic field).

Catedrático de Universidad en la U de Alicante (UA). Ha sido Director del Dto. de Traducción e Interpretación (2009-2017), coordinador del Doctorado en Traducción con mención de calidad (2006-2009) de la UA y es coordinador del Doctorado interuniversitario en Traductología, Traducción Profesional y

Audiovisual. IP de numerosos proyectos de investigación a nivel nacional, autonómico y local. Ha dirigido numerosos TFGs, TFMs y 16 tesis doctorales sobre terminología, traducción audiovisual, traducción de elementos culturales, fraseología, fraseodidáctica y variaciones diatópicas. Director del grupo de investigación FRASYTRAM desde 2006 (Fraseología y Traducción Multilingüe). Cuenta con numerosas publicaciones en revistas y editoriales reconocidas en el Scholar Publishing Index (SPI). Ha sido profesor invitado en las Universidades de Paris 13, Paris Marne-la-Vallée, Paris-Ivry, Paris-Diderot, U. Pedagógica de Cracovia, U. de Antananarivo, U. R. Palma, etc. Líneas de trabajo e investigación: terminología y traducción especializada, traducción audiovisual, traducción humanístico-literaria (ámbito literario, cultural, periodístico).

Recibido / Received: 12/10/2020

Para enlazar con este artículo / To link to this article:

<http://dx.doi.org/10.6035/MonTI.2020.ne6.1>

Para citar este artículo / To cite this article:

Apellido(s), Nombre & Nombre Apellido(s). (2020) “Análisis multidisciplinar del fenómeno de la variación fraseológica en traducción e interpretación.” En: 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. 36-64.

ANÁLISIS MULTIDISCIPLINAR DEL FENÓMENO DE LA VARIACIÓN FRASEOLÓGICA EN TRADUCCIÓN E INTERPRETACIÓN

MULTIDISCIPLINARY ANALYSIS OF THE PHENOMENON OF PHRASEOLOGICAL VARIATION IN TRANSLATION AND INTERPRETING

PEDRO MOGORRÓN HUERTA

pedro.mogorron@ua.es
Universidad de Alicante

Resumen

El presente trabajo, centrándose en la lengua española, pretende introducir el fenómeno de la variación fraseológica, presentando, a partir de las investigaciones llevadas a cabo en el marco de la lingüística hispánica, los aspectos teóricos, la frecuencia cuantitativa de las variaciones y el fenómeno de la variación geolingüística o diatópica. Se señala la dificultad de la traducción fraseológica incidiendo en que esta traducción no es imposible y que el usuario debería poder contar con herramientas y bases de datos fraseológicas que le permitan encontrar las equivalencias fraseológicas en otros idiomas antes de pensar en usar las recurrentes técnicas de traducción.

Palabras clave: Fraseología, fijación, variación fraseológica, diatópía, traducción.

Abstract

Focusing on the Spanish language, this paper aims to introduce the phenomenon of phraseological variation, presenting, on the basis of the studies carried out in the



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

context of Hispanic linguistics, the theoretical aspects, the quantitative frequency of the variations and the phenomenon of geolinguistic or diatopic variation. The difficulty that phraseological translation represents is pointed out emphasizing the fact that this type of translation is not impossible and that the user should be able to count on phraseological tools and databases that will allow to find phraseological equivalents in other languages before thinking about using recurrent translation techniques.

Keywords: Phraseology, fixing, phraseological variation, diatopy, translation.

1. Introducción

De ser un tema poco relevante, la fraseología ha pasado a ser un tema de investigación recurrente en un gran número de lenguas e incluso en traducción debido a los numerosos problemas de comprensión y de equivalencia que plantea. *Prinzipien der Sprachgeschichte* (Paul 1880); *Essai de sémantique* (Breal 1897), *Cours de Linguistique Générale* (Saussure 1916) y *Précis de Stylistique* (Bally 1905) son considerados los precursores de la fraseología, sin embargo, a la postre, fueron los estructuralistas rusos los que establecieron las bases de la fraseología moderna.

Hoy en día, se puede observar en todo el mundo una incesante actividad científica en el campo de la fraseología. En la fraseología española, en la que nos centraremos, tras las aportaciones pioneras de Casares (1950) y Zuluaga (1980), el avance de la fraseología experimenta una expansión decisiva con la publicación de Corpas (1996), obra de referencia que generó un crecimiento exponencial de las investigaciones fraseológicas. Avance que consolidó García-Page (2008), otorgándole a la fraseología española sus letras de madurez.

En los últimos cuarenta años los numerosos estudios en el campo de la Fraseología han mostrado que no se trata de anomalías lingüísticas, analizando elementos tan diversos y a la vez tan dispares como:

- (1) Las clasificaciones teóricas: Casares (1959), Zuluaga (1980), Corpas (1996), Ruiz Gurillo (1997), García-Page (2008), etc.
- (2) Las propiedades de las UF: idiomatidad, composicionalidad, opacidad, iconicidad, registros de lengua, motivación, metaforización, desautomatización, etc.

- (3) La fraseodidáctica: nace en el seno de la lingüística aplicada y está enfocada a la enseñanza / aprendizaje de las UF tanto de lenguas maternas como extranjeras (González Rey 2012).

2. Variación fraseológica

El auge del interés por la variación lingüística surge a mediados del s. XX con la aparición de la sociolingüística¹. En el ámbito fraseológico, inicialmente, la mayoría de los estudios señalaba que los diferentes tipos de unidades fraseológicas (UF) se diferenciaban de las combinaciones lingüísticas libres, principalmente por su fijación e idiomatidad y porque sus componentes y su(s) estructura(s) no permitían cambios ni modificaciones². Los investigadores, para evidenciar tal posibilidad, utilizaban UF con un alto grado de fijación en las que ningún elemento podía ser reemplazado por otro (Montoro del Arco 2004). Por ejemplo:

- (1) *salir del armario* (DUE³), [^{*}emerger, ^{*}aparecer] del [^{*}mueble, ^{*}ropero]; *estirar la pata* (RAE/DUE), [^{*}extender, ^{*}alargar, ^{*}desplegar] alguien la pata, ^{*}estirar alguien la ^{*}pierna;
- (2) *ladrar a la luna* (DUE), ladrar [al ^{*}sol, a las ^{*}estrellas];
- (3) *apretarse el cinturón* (DUE), [*apretarse la ^{*}cintura, la ^{*}correa*]; etc.

La variación fraseológica (VF) viene suscitando interés, como prueban los numerosos trabajos existentes en el ámbito de la fraseología hispánica

-
1. Esta analiza la posible modificación, que se puede producir en el uso diario de la lengua de una comunidad de hablantes, condicionada por elementos de índole temporal, sociocultural o geolingüísticos. Todas las lenguas, como señala Gadet (2003: 13) se ven sometidas a numerosos cambios: “Il n’est pas de langue que ses locuteurs ne maintient sous des formes diversifiées... les sociolinguistes, la saisissent en parlant de variétés pour désigner différentes façons de parler, de variation pour les phénomènes diversifiés en synchronie, et de changement pour la dynamique en diachronie; et ce, à la fois pour les productions d’un individu, d’un groupe ou d’une communauté”.
 2. Es decir que no podían ser sustituidos por otros componentes de la misma clase sintáctica sin cambiar el significado de la UF. Estas UF sí admiten otras posibles modificaciones o transformaciones como la variación de persona, de tiempo verbal, inclusión de adverbios, pronominalización, etc.
 3. Todas las siglas corresponden a diccionarios que aparecen en la bibliografía.
 4. El * indicará que una secuencia pluriverbal no es admitida como VF, al perder el significado original de la UF.

(Carneado Moré 1985; Corpas 1996, 2008, 2018; Corpas & Mena Martínez 2003; Cuadrado-Rey 2016, 2018; García-Page 2001, 2008; Koike 2001; Mogorrón 2010, 2014, 2015, 2020a, 2020b; Navarro-Brotons 2011, 2018; Penadés 2014, 2020; Zuluaga 1980)⁵. Las distinciones establecidas y las diferentes denominaciones formuladas evidencian cierta confusión (Montoro del Arco 2004, 2008; Alvarado Ortega 2008; Sinner & Tabares Plasencia 2016). En estos estudios, se ha considerado, por regla general, que la VF no es una propiedad fundamental de las UF; al contrario de la idiomatidad y la fijación. Sin embargo, a finales del s. XX algunas investigaciones empezaron a destacar la posible importancia de la VF: “son muy numerosos los fraseologismos que admiten variaciones” (Carneado Moré 1985: 269);

“La lingüística hispánica apenas ha prestado atención a la variación fraseológica, debido fundamentalmente a la falta de datos fraseológicos que cubran toda el área hispana y con los que se puedan hacer investigaciones sobre el tema. Los escasos trabajos sobre los fraseologismos [...] no han sido suficientes para obtener una visión global de la variación fraseológica del mundo hispánico” (Koike 2001: 77).

“[...] no es menos cierto que la inmensa mayoría de ellas experimenta algunos cambios o puede tener formulaciones diversas al ser actualizadas en el discurso” (Montoro del Arco 2004: 593). Así, en el caso de las locuciones verbales *estar en la luna* y *coger el toro por los cuernos* encontramos las siguientes variantes:

- (1) *estar en la luna*, *estar en* [Babia (DUE), Belén (DUE), en el limbo (DUE), en la higuera (DUE), la parra (CREA), las Batuecas (DUE), las nubes (DUE)].
- (2) [agarrar, coger] *el toro por los cuernos* (DUE), *coger al toro por* [las astas, los cuernos] (RAE).

El principal problema estriba en que la variación aparece en procesos imprevisibles en cuanto a los posibles candidatos a variante(s). En efecto, no hay ninguna regla que permita ni determinar ni intuir el motivo por el cual algunos términos pueden ser utilizados como elementos de conmutación en algunas

5. Actualmente se están terminando los números de las revistas ELUA (2020) y L.I (2020) relacionados con la variación fraseológica y la diatopía.

UF y otros no. Es el caso de las posibles variantes de: *estar hasta las narices*, para la que hemos encontrado: *estar hasta [el coco (DUE), el (mismísimo) coño (DFDEA), el copete (LARBI), el culo (DUE), el gollete (RAE), el gorro (DUE), el moño (DUE), el nabo (oral), el pirri (VV⁶), el pito (VV), el punto com (VV), la coronilla (VV), la polla (DUE), la punta de los pelos (DUE), las cejas (DUE), las narices (DUE), las pelotas (DUE), los cojones (DUE), los huevos (DUE), los ovarios (VV), los pelos (DUE), etc.] de alguien / algo. ¿Por qué se puede decir *estar hasta las narices*, *hasta el moño*, *hasta las cejas* y no **estar hasta las orejas* o **hasta la frente*?*

Este proceso de variación y su imprevisibilidad (Mellado Blanco 2004: 159; Mogorrón 2010: 96) está presente en la mayoría de las tipologías de UF⁷ (a saber: locuciones verbales, adjetivales, nominales, adverbiales, colocaciones, refranes⁸, pragmatemas⁹, etc.). Sin embargo, teniendo en cuenta la imposibilidad material de presentar los cambios en cada una de ellas, en el marco de este trabajo, analizaremos la variación en las UF que denominamos construcciones verbales fijas (CVF) en las que incluimos: 1) locuciones verbales: *poner una pica en Flandes* (DUE); 2) colocaciones verbales: *guiñar un ojo* (DUE); 3) verbos soportes: *dar un paseo* (DUE); *dar ayotes a alguien* (DUE); 4) construcciones verbales comparativas: *dormir como un lirón* (DUE), (véase Mogorrón 2010).

Además de ser imprevisibles las posibles variaciones, el número de los términos que pueden ser aceptados como variantes, puede oscilar y dar lugar a variantes poco o muy productivas numéricamente. Dentro de las variantes poco productivas se pueden citar p. ej.:

- (1) [*acabar, terminar*] *como el rosario de la aurora* (DUE);
- (2) [*arrojar, echar, tirar*] *la casa por la ventana* (DUE);

6. VV significa Viva Voz y su uso está atestiguado en buscadores de internet. *Estar hasta el punto com* presenta 1450 ocurrencias en Google.

7. Véase Corpas (1996), García-Page (2008).

8. Véase Sevilla Muñoz (2020), Navarro-Brotons (2011, 2018, 2020). P. ej.: (1) *A Dios rogando y con el mazo dando / a Dios rogando y al macho dando*. (2) [A/ para cada cerdo / puerco] [le llega / hay / le viene] su San Martín. (3) [Quien / el que / al que] a buen árbol se arrima, buena sombra [le cobija / tiene].

9. Véase Alvarado Ortega (2008): P. ej.: ¡me cachis en la mar!, ¡me cago en la mar!, ¡me cagüen la mar!, ¡me cago en la puta [la hostia, la leche]!

- (3) *echar(le) margaritas a los [cerdos, puercos]* (RAE);
- (4) [*despedirse, marcharse*] *a la francesa* (DUE); etc.

Otras UF permiten en cambio un número mucho mayor de términos aceptados en el proceso de variación y modificación de la forma canónica. Así, p. ej.:

- (1) *irse algo a [freír espárragos (DEA), freír monas (DEA), hacer puñetas (DEA), hacer gárgaras (DEA), pique (DUE), tomar por culo (DEA), tomar por el culo (DEA), tomar viento (DUE)], irse algo al [carajo (DUE), cuerno (DUE), diablo (DUE), demonio (DUE), garete (DFDEA), infierno (VV), traste (DUE), irse algo a la [mierda (DUE), porra (DUE), etc.]*.
- (2) [*descalzarse, descojonarse, descoyuntarse, despedazarse, despelotarse, desternillarse, desperezarse, mearse, mondarse, morirse, partirse, retorcerse, reventar, troncharse*] *de risa* (RAE).

Finalmente, también se pueden producir casos de variaciones simultáneas de varios componentes de la UF como verbos y sustantivos, verbos y preposiciones, de sustantivos y preposiciones, etc. P. ej.:

- (1) [*dejarse, quitarse*] *de [cuentos, historias]* (DUE)
- (2) [*descubrir, inventar*] [*América (VV), el agua caliente (VV), el mediterráneo (DFDEA), la pólvora (DFDEA)*];

Llegados a este punto y con el fin de tratar el concepto de variación fraseológica, se empleará aquí una concepción ancha o amplia de la fraseología (Gläser 1986: 42; Corpas Pastor 2003) y hablaremos de VF cuando en una determinada UF se produzca la sustitución de uno o varios de sus componentes léxicos o gramaticales por otro de iguales características (sinónimo o no) conservando la UF resultante el mismo significado y una estructura sintáctica idéntica, o muy similar.

2.1. Variantes geolingüísticas o diatópicas

En el caso de lenguas como el español, el francés, el inglés, el árabe, etc. utilizadas como lenguas oficiales en numerosos países, también existen numerosas variantes diatópicas o geolingüísticas. En el caso del español, la mayoría de las investigaciones citadas –con la excepción de Zuluaga (1980); Carneado Moré

(1985); Cuadrado-Rey (2016, 2018, 2020); Koike (2001); Mogorrón (2014, 2015, 2020); Navarro-Brotóns (2011, 2018, 2020); Pamies (2017)– muestran la tipología de variaciones fraseológicas ciñéndose al español peninsular y/o común (EspPC), dejando de lado numerosas creaciones usadas en *los españoles de América* (EspHis), como así lo reconoce García-Page (2008: 253):

La lista de variantes de una lengua ha de tener en cuenta las diferencias institucionalizadas que se producen en cualquier ámbito de su empleo, en especial las de índole sociolectal o distráctica y goepectal o diatópica. Por esa razón, resulta arriesgado afirmar que los inventarios de variantes constituyen conjuntos cerrados, sobre todo cuando se trata de lenguas como el español de un radio de acción tan amplio.

Si retomamos, ahora algunos de los ejemplos pertenecientes al EspPC en el punto 2., podemos observar que los diccionarios hispanoamericanos señalan que existen variantes reconocidas en países hispanoamericanos. P. ej.:

- (1) con ejemplos de UF supuestamente fijas en EspPC¹⁰:
 - a) *apretarse el cinturón* (DUE) / *ajustarse el cinturón, los cinturones* (Ddmexicanismos, Mx), *amarrarse el cinturón* (DdAm, Pe).
 - b) *salir del armario* (DUE) / *salir del closet* (DdAm, ES, Ni, PR, Co, Ve, E, Pe, Ch, Py).
- (2) con nuevas variantes en Hispanoamérica de CVF que ya presentan variantes en EspPC. P. ej.:
 - a) *irse algo a* [*freír espárragos* (DEA), *freír monas* (DEA), *hacer puñetas* (DEA), *hacer gárgaras* (DEA), **a justa** (DdAm, PR), **a juste** (DdAm, PR), **a la chucha** (DdAm, Ch), *pique* (DUE), *tomar por culo* (DEA), *tomar por el culo* (DEA), *tomar viento* (DUE)]; *irse algo al* [*carajo* (DUE), **cachimbo** ((DdAm, RD), *cuerno* (DUE), *diablo* (DUE), *demonio* (DUE), *garete* (DFDEA), **hoyo** (DdAm, Bo, Ch, Ar), *infierno* (VV), **tacho** (DdAm, Bo, Ch, Py, Ar, Uy),

10. Se indican las variantes del EspHis separándolas con una barra oblicua /. Por lo general los diccionarios españoles incluyen una enorme mayoría de términos y de expresiones del EspPC y los diccionarios Hispanoamericanos incluyen un contenido del EspHis. Como botón de muestra, los dos diccionarios más referenciales del EspPC, el DRAE y el DUE solamente incluyen unas 770 UF del EspHis.

- sipote** (DdAm, RD), **traste** (DUE), **irse algo a la [mierda (DUE), porra (DUE)]**.
- b) **coger al toro por los cuernos**, [**agarrar, coger**] **el toro por los cuernos** (DUE), **coger al toro por [las astas, los cuernos]** (RAE) / **agarrar al toro por los cachos** (DdAm, Py, RD, Ve, Pe), **tomar el toro por las guampas** (DdAm, Ur).
- c) [**lanzarse, tirarse**] **alguien a la piscina** (RAE) / **tirarse a la pileta** (DdAm, Ar, Ur).

2.2. Variaciones numéricas en las UF

Las primeras investigaciones que han analizado y cuantificado las variaciones de las CVF en la lengua española han sido realizadas por el grupo de investigación FRASYTRAM¹¹ de la Universidad de Alicante, (véase Mogorrón 2010, 2015) a partir de una Base de Datos (BD) de 41.000 CVF recopiladas en más de 30 diccionarios españoles e hispanoamericanos de las que 26.000 pertenecen al EspPC y 15.000 al EspHisp.

En el caso del EspPC, el análisis de las 26.000 entradas destaca que 15.001 CVF, es decir un 58%, presentan algún tipo de variación (ver gráfico 1).

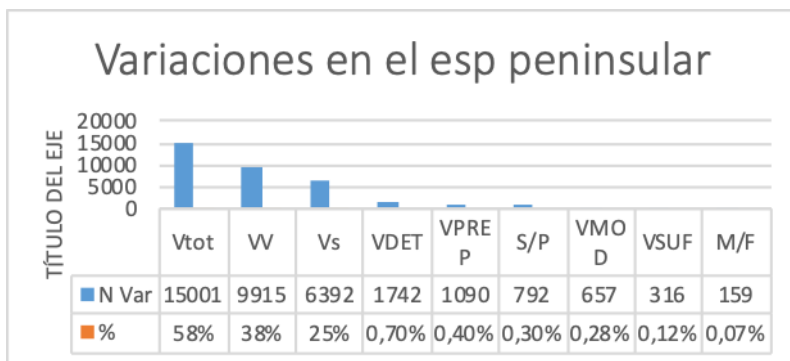


Gráfico nº 1. Variaciones en Esp. peninsular

En el caso del EspHisp, el análisis, nos indica que 3359 CVF del EspHisp presentan algún tipo de variación¹² (ver gráfico 2).

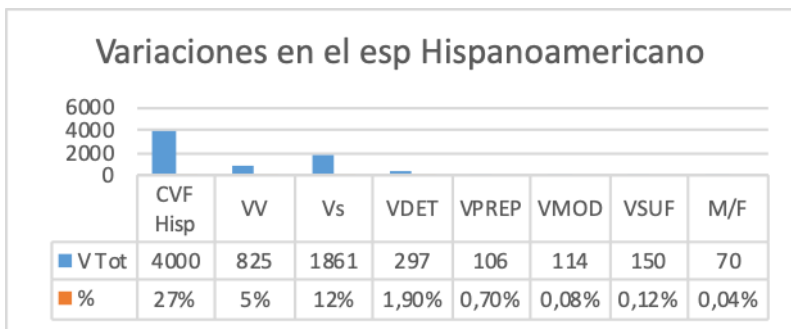


Gráfico nº 2. Variaciones en el EspHisp

En los dos casos, las variaciones más frecuentes se producen con variantes léxicas, bien verbales, bien nominales. Ahora bien, mientras que en el EspPC se producen más variaciones verbales, en el EspHisp se producen más variaciones nominales.

2.3. *Uso real de las variantes fraseológicas*

Encontrar las posibles variantes fraseológicas (VF) implica, por lo general, un largo trabajo de recopilación lexicográfica, de consulta de diferentes diccionarios, pues el contenido fraseológico de cada uno presenta diferencias considerables (ver tabla nº 1).

DUE	[agarrar / coger] el toro por los cuernos
RAE	coger al toro por [las astas / los cuernos]
Espasa Calpe	[agarrar / coger / tomar] el toro por los cuernos

Tabla nº 1: VF en diccionarios

Pero, ¿esta recopilación de fuentes documentales es completa? Actualmente, la posibilidad de realizar labores de búsqueda de determinadas cadenas léxicas

12. Este análisis sigue en curso.

en corpus textuales¹³, a través del tratamiento automático de textos, permite encontrar numerosas VF que no vienen recogidas en los diccionarios peninsulares analizados y que, sin embargo, son usadas a diario por los hablantes. En el caso de las variantes de *coger el toro* por los cuernos, según los usos repertoriados en los diccionarios, las formas *tomar al toro por los cuernos*, *agarrar al toro por las astas*, *tomar el toro por las astas* no serían aceptables a pesar de localizarse en internet en el buscador de Google.es.

Para averiguar las posibles variantes de *coger el toro por los cuernos* y de *echar margaritas a los cerdos* (ver tabla 3 y 4) se ha realizado una búsqueda en tres corpus textuales: CREA¹⁴, CORDE¹⁵ y Sketch-Engine (S-E)¹⁶ que reúnen una gran colección de textos representativos. La búsqueda de las VF, ordenadas alfabéticamente a partir del verbo, ofrece los resultados siguientes:

		CREA	CORDE	S-E	EspPC	EspHisp
1	<i>agarrar al toro por los cachos</i>			4	3	1
2	<i>agarrar al toro por los cuernos</i>	2		194	153	41
3	<i>agarrar al toro por las astas</i>			14	10	4
4	<i>agarrar el toro por las astas</i>		2	35	14	21
5	<i>agarrar el toro por las guampas</i>			2		2
6	<i>agarrar el toro por los cachos</i>	1	1	8		8

13. Desde los años 1960, los corpus textuales han cambiado decisivamente las investiga-

ciones lingüísticas, permitiendo a los investigadores disponer de grandes volúmenes de datos y estudiar la lengua a través de ejemplos reales mediante procesos automáticos (Corpas, 2008, 2013, 2018, etc.), que extraen los resultados solicitados gracias al procesamiento del lenguaje natural.

14. REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. <http://www.rae.es>.
15. REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) [en línea]. *Corpus diacrónico del español*. <http://www.rae.es>.
16. <https://www.sketchengine.eu/> el corpus esTenTen18 de S-E, está compuesto por los subcorpus *European Spanish Web* y *American Spanish Web* y contiene 17.553.075.259 tokens, por lo que se trata de un corpus de gran tamaño, con un 49,32% de dominios de español europeo y un 46,46% del español americano.

7	<i>agarrar el toro por los cuernos</i>	1	1	184	82	102
8	<i>amarrar al toro por los cuernos</i>			1	1	
9	<i>asir al toro por las astas</i>			2		2
10	<i>asir al toro por los cuernos</i>			1		1
11	<i>asumir al toro por los cuernos</i>			1		1
12	<i>atacar al toro por los cuernos</i>		1	1	1	
13	<i>atrapar al toro por las astas</i>			1		1
14	<i>atrapar al toro por los cuernos</i>			1	1	
15	<i>coger al toro por las astas</i>		1	6	3	3
16	<i>coger al toro por los cachos</i>			4		4
17	<i>coger al toro por los cuernos</i>	3	1	353	333	20
18	<i>coger el toro por los cachos</i>			2		2
19	<i>coger el toro por los cuernos</i>	39	1	2050	1980	70
20	<i>enfrentar al toro por los cuernos</i>			11	2	9
21	<i>entrar al toro por los cuernos</i>	2		59	38	21
22	<i>ir al toro por los cuernos</i>			4	1	3
23	<i>pillar al toro por los cuernos</i>			3	3	
24	<i>tener al toro por las astas</i>			4		4
25	<i>tener al toro por los cuernos</i>			11	6	5
26	<i>tomar al buey por los cuernos</i>			2	1	1
27	<i>tomar al toro por las astas</i>	2		191	31	160
28	<i>tomar al toro por los cachos</i>			8	3	5
29	<i>tomar al toro por los cuernos</i>	3	1	338	100	238
30	<i>tomar el toro por las astas</i>	8	2	793	20	773
31	<i>tomar el toro por las guampas</i>	1		16		16
32	<i>tomar el toro por los cachos</i>			57	10	47
33	<i>tomar el toro por los cuernos</i>	10		620	160	460

Tabla n° 2. VF coger el toro por los cuernos

		CREA	CORDE	S-E	EspPC	EspHisp
1	arrojar margaritas a los cerdos		1	8	4	4
2	arrojar margaritas a los chanchos			1		1
3	arrojar margaritas a los puercos	1	1	3	1	2
4	dar de comer margaritas a los cerdos			11	4	7
5	dar margaritas a los cerdos			33	29	4
6	dar margaritas a los chanchos			11		11
7	dar margaritas a los cochinos			1	1	
8	dar margaritas a los gorrinos			1	1	
9	dar margaritas a los puercos			4	1	3
10	echar margaritas a los cerdos	4	1	84	76	6
11	echar margaritas a los puercos		3	11	6	5
12	lanzar margaritas a los cerdos			4	2	2
13	ofrecer margaritas a los cerdos			3	3	
14	tirar margaritas a los cerdos			15	7	8
15	tirar margaritas a los chanchos			30		30

Tabla nº 3. VF *echar margaritas a los cerdos*

En ambos casos, hemos destacado en negrita las VF recopiladas en los diccionarios utilizados para la elaboración de la BD. Si bien es verdad que todas las VF que vienen recogidas en los diccionarios aparecen en los corpus textuales, se puede observar que no siempre incluyen las formas más usuales y que algunas VF incorporadas presentan muy pocas ocurrencias en los tres corpus textuales analizados. Lo más destacable es que se puedan detectar con ejemplos contextualizados:

- (1) usos de las mismas formas en las dos orillas atlánticas, corroborando la existencia de un español común. Las variantes con más ocurrencias en S-E, y seguramente las más usuales en la actualidad, aparecen en las dos orillas del Atlántico. Ej: *coger el toro por los cuernos* (DUE) *echar margaritas a los cerdos* (RAE); y otras muchas como *estar en la*

luna (RAE); *estirar la pata* (DUE); *comerse las uñas* (DUE); *hacer el agosto* (DFDEA); etc.

(2) nuevas variaciones que no aparecían en los diccionarios.

Ahora bien, el uso de estas magníficas herramientas no implica que se vayan a encontrar variantes en todas y cada una de las UF, pues muchas siguen conservando la inalterabilidad de su formulación. Sería el caso de *hacer de tripas corazón* (RAE), *ladrar a la luna* (RAE), *ponerse el mundo por montera* (RAE), *estar entre Pinto y Valdemoro* (RAE); *dar el ala para comerse la pechuga* (DdAm, PR); que aparecen en numerosas ocurrencias en EspPC y/o en EspHispc en S-E.

A la luz de los datos obtenidos, pensamos que conviene replantearse no solamente la importancia en el mundo de la fraseología de las variaciones, ya que a buen seguro los dos conceptos fijación y variación son las dos caras opuestas de un mismo fenómeno, sino su tratamiento y su consideración como propiedad esencial posiblemente con la misma importancia que la fijación y la idiomática. Tal y como adelantan (Corpas & Mena Rodríguez (2003: 183):

hay que admitir que el concepto de fijación tal y como fue planteado en las primeras investigaciones queda en entredicho. Si en un principio las UFs se definieron como unidades estables y fijas que mostraban rechazo a cualquier alteración léxica, semántica y morfosintáctica, en vista de la existencia de cambios reales y potenciales, no cabe hablar de la fijación como una propiedad absoluta (Burger 1998) sino como una cualidad relativa (Fleischer 1982; Gläser 1986; Wotjak 1992; Corpas 1996; Burger 1998) y variable.

2.4. Variación de significado

Otro tipo de variación que conviene destacar, en relación con el proceso de comprensión y de traducción de las UF y que tampoco ha sido muy analizado, está relacionado con su(s) significado(s) y la polisemia geolingüística. Sin embargo, se trata también de un problema de gran magnitud. El análisis exhaustivo de las 41.000 CVF de la BD ha permitido recopilar, hasta la fecha, 14.238 casos de CVF polisémicas y detectar CVF polisémicas que presentan situaciones que van desde los dos significados hasta los diez (Mogorrón 2020) con los siguientes resultados (ver tabla 4).

CVF polisémicas	
1 Sdo	26652
2 Sdos	4335
3 Sdos	1006
4 Sdos	328
5 Sdos	120
6 Sdos	31
7 Sdos	19
8 Sdos	6
9 Sdos	1
10 Sdos	4

Tabla 4. CVF, casos de polisemia

Además, el análisis de los 14.238 casos de polisemia para determinar la zona de uso, ofrece los siguientes resultados:

- (1) 8.423 CVF pertenecen únicamente al EspPC, siendo la situación más frecuente. P. ej.:
 - a) *pegar un metido* (1. Regañar, reñir con rigor DUE; 2. Sonsacar dinero a alguien, DUE; 3. Pegar, dar una paliza a alguien, DUE);
 - b) *pedir la luna* (1. Pedir, pretender algo muy difícil o imposible de obtener, RAE; 2. Pedir un precio o sueldo exagerado, DFDEA), etc.
- (2) 1.358 CVF presentan significados que pertenecen al EspPC o al EspHisp. P. ej.:
 - a) *bañarse en agua de rosas* (1. Alegrarse de algún mal ajeno, DUE; 2. Sentirse alguien muy orgulloso por el éxito de un ser querido, DdAm, Ve, Ec; 3. Disfrutar de bonanza económica, DdAm, Gu);
 - b) *no tener abuela* (1. Se dice irónicamente de la persona que se alaba a sí misma, DUE; 2. Ser malo, DdAm; 3. Mantener una conducta incorregible, DdAm, Mx, ES, Ni, Ar); etc.
- (3) 4.454 CVF tienen significados que pertenecen solamente al EspHisp. P. ej.:

- a) *enredar la pita* (1. Complicar las cosas, DdAm, Ho, Ni, Co, RD, PR; 2. Hacer difícil de comprender algo, DdAm, Cu, RD; 3. Confundirse, DdAm, RD; 4. Irse por las ramas, DdAm, Ni);
- b) *acabarse el veinte* (1. Agotarse el tiempo, DDMex, Mx; 2. Agotarse la oportunidad de hacer algo, DDMex, Mx; 3. Acabarse la suerte, DdAm, Mx); etc.

2.5. *Parasinonimia fraseológica*

El último tipo de variación al que deseamos hacer referencia en este estudio, también por su gran repercusión en la traducción, es el de la parasinonimia fraseológica. Además de los culturemas idiosincrásicos asociados a una cultura bien determinada (Pamies 2008), existen en muchas lenguas numerosos conceptos, que podríamos calificar de policulturales, o incluso de universales fraseológicos (Dobrovolskij 1988; Corpas 1995, 1998; Luque & Majón 2002; Mogorrón 2002), tan usuales que los hablantes de lenguas diferentes han creado numerosas expresiones, con el mismo significado, para referirse a esos conceptos. Podemos encontrar en ese caso grupos parasinónimos que van de un par de expresiones hasta grupos formados por varias decenas de UF. Esos conceptos policulturales usuales pueden referirse a cualquier actividad de la vida cotidiana, p. ej.: sexo, comida, bebida, muerte, trabajo, embarazo, faltar a clase, estar gordo, flaco, ser guapo, feo, bajo, pequeño, valiente, miedoso, comer mucho, homosexualidad, riqueza, pobreza, distracción, etc.

Las 41.000 entradas de la BD de FRASYTRAM han sido catalogadas semánticamente a partir de su(s) definición(es). A modo de ejemplo, se han recopilado las siguientes CVF parasinónimas de estar distraído partiendo de la UF *estar en la luna*:

- (1) **en España:** *andar en las nubes* (RAE); *andar por las nubes* (DUE); *cazar tilingos* (RAE); [*contar, estar contando, ponerse a contar*] *las vigas* (RAE); [*estar, quedar*] *a uvas* (EPM); *estar en Babia* (DUE); [*estar bailando*] *en Belén* (RAE); *estar con la torta* (DTDFH); *estar con los angelitos* (RAE); *estar en el limbo* (RAE); *estar en la higuera* (RAE); *estar en la luna* (DUE); *estar en la parra*; *estar en las Batuecas* (DUE); *estar en las nubes* (DUE); *estar pensando en la mona de Pascua* (DTDFH); *ir a por uvas* (Akal); [*mirar a, pensar en*] *las musarañas*

(RAE); *mirar las telarañas* (LARBI); *quedarse encantado* (DUE); *tener la cabeza a las once* (RAE); *tener la cabeza a pájaros* (RAE); *tener la cabeza en el aire* (VV) *tener la cabeza en la luna* (VV); *tener la cabeza en las nubes* (RAE); *tocar el violón* (RAE); *tocar el violón a dos manos* (DTDFH); *vivir en las nubes* (DUE); etc.

- (2) **en Hispanoamérica:** [*andar, vivir, pasársela*] *en la luna* (DEUEM, Mx, Ch); [*andar, estar*] *fuera de onda* (DTDFH, Mx); *cortar varas* (DdAm, Gu, Ho, Nicaragua); *estar* (alguien) *en el aire* (RAE; Cu); *estar en la estratosfera* (DFHA, Ar); *estar en otra* (DFHA, Ar); *estar papando moscas* (GDLA, Ar); *estar en la luna de Paita* (Bo, Ec, Pe); *estar en la luna de Paita y el sol de Colán* (, DdAm, Pe); *estar en las nebulosas* (RAE, Ve); *estar fuera de onda* (DTDFH, Mx); *estar pensando en los pajaritos de colores* (DdAm, Ar); *pensar en la inmortalidad del cangrejo* (DdAm, Mx, Gu, Ho, ES, Ni, Py, Cu, RD, Ve, Ec, Bo, Ur); *pensar en la inmortalidad del mosquito* (DdAm, Pe); *pensar en las muelas del gallo* (DdAm, Gu); *pensar en los anteojos del gallo* (DdAm, Gu); *pensar en los huevos del gallo* (DdAm, CR); *quedarse en China* (DdAm, Cu); [*ser, tener*] *cabeza de novia* (GDLA, Ar); *tener la cabeza en los pies* (DEUEM, Mx); *vivir en el limbo* (GDLA, Ar).

3. Traducción de las UF

3.1. Traducción y fraseología

La traducción es un complicado proceso multifuncional mucho más complejo que la simple búsqueda de correspondencias entre palabras u oraciones en el que no conviene obviar ni la finalidad ni el contexto que darán pie a diferentes métodos traductores (Hurtado Albir 2001: 252). Ahora bien, las UF representan el deseo del locutor de utilizar en el acto de comunicación, escrito u oral, una determinada fórmula refrendada por gran parte de la sociedad, sumándole un matiz cultural, socializador, generacional y desempeñando funciones, estilísticas, discursivas, argumentativas e incluso pragmáticas en vez de utilizar un término neutro o una perífrasis verbal. Además, su uso presupone que los interlocutores poseen un fondo cultural y lingüístico común, pues en caso contrario, se encontrarán con grupos de palabras que no comprenderán

y deberán recurrir a la ayuda de los demás usuarios o de diccionarios para intentar descifrar esa posible UF.

Qué sería de la lengua oral sin UF o incluso de las obras literarias de algunos autores tan expresivos en su discurso como:

- (1) Pérez Reverte (*El capitán Alatriste*, 1996: 123 y 133): “pueden más dos tetas que dos carretas”, “dormir a pierna suelta”; “puso pies en polvorosa”, “dar tres cuartos al pregonero”, “la pasó en vela”, “estar al tanto”);
- (2) Torres (*¡Oh es Él!*, 1998: 33 y 44): “echar las campanas al vuelo”, “caer en manos de”, “estás como un cencerro”; “tenía el corazón en un puño”, “estaba hecha un manojo de nervios”;
- (3) Ruiz Zafón (*El juego del ángel*, 2008: 23-24): “*ver ni en pintura*”, “*subsistir por los pelos*”, “*otro gallo nos cantaría*”, “*de medio pelo*”, etc.);
- (4) Gala (*El dueño de la herida*, 2003: 24-25): “le echó en cara”, “la gata flora, que si se la meten chillar, y si se la sacan llorar”, “quitarme de en medio”, “soy de ordeno y mando”).

¿Qué sería de la traducción si se omitiesen o eliminasen estas formas que el autor ha usado conscientemente en base a su función fraseológica inherente? Es verdad que la competencia fraseológica de todo usuario, ya de por sí limitada en su lengua materna, disminuye considerablemente en otros idiomas (Corpas 2003, Tabares & Batista (2019). También es verdad que el contenido fraseológico de los diccionarios dista mucho de ser completo y no menos cierto es que la traducción de las UF representa un verdadero reto para los traductores, pero por muy difícil que sea es un reto asumible (Corpas 2003; Richart Maset 2008) como indican Mona Baker (1992: 65):

the main problems that idiomatic and fixed expressions pose in translation relate to two main areas: the ability to recognize and interpret an idiom correctly; and the difficulties involved in rendering the various aspects of meaning that an idiom or a fixed expression conveys into the target language,

y Ballard (1998: 247): “le jugement de fidélité et de qualité que l’on porte sur la traduction ne peut s’exercer qu’au travers des formes, porteuses de sens et d’effets stylistiques“, teniendo entonces que forjar un ejercicio de orfebrería traductológica.

3.2. Fases del proceso de traducción fraseológica

Varios autores han analizado la traducción fraseológica, destacando que las diferentes fases del proceso corresponden, siguiendo a Corpas (2003), a la identificación, interpretación y búsqueda de equivalencia (a nivel léxico y posteriormente textual) (Sevilla & Sevilla 2005, 2009; Mogorrón 2008; Timofeeva 2008, etc.).

Las fases de identificación e interpretación-comprensión de la UF están íntimamente relacionadas con la idiomaticidad / opacidad, así como con la competencia fraseológica del traductor que se encontrará frente a varias situaciones prototípicas.

- (1) Si la UF pertenece a la competencia fraseológica del traductor, este deberá intentar encontrar otra equivalente en significado en la lengua meta.
- (2) Si no es de la competencia fraseológica del traductor, podrá identificarla o no.
- (3) Si no la identifica como UF será perjudicial para la traducción, por mucho que se conserve el significado global del texto.
- (4) Si la identifica como UF, otra vez, deberá buscarla irremediablemente en diccionarios monolingües o bilingües para ver si con suerte el diccionario incluye esa forma, y entonces poder identificarla, interpretarla y traducirla.

Ha quedado patente que el proceso de comprensión / interpretación de las UF es mucho más complejo de lo que podría parecer, como evidencian los ejemplos que se muestran a continuación, debido a la competencia fraseológica limitada de cada usuario, al tratamiento fraseológico deficitario de los diccionarios¹⁷ (Corpas 2003: 155-169; Mogorrón 2011, 2015) y a las numerosas obras fraseológicas y paremiológicas presentes en el mercado editorial. P. ej.:

- (1) *cerrar los ojos* 1. Dormir, RAE; 2. Morir, RAE; 3. Sujetar el entendimiento al dictamen de otro, RAE; 4. Obedecer sin examen ni réplica, RAE; 5. Arrojar temerariamente a hacer algo sin reparar en inconvenientes, RAE; 6. No apartarse de un enfermo hasta que expire, RAE;

17. La expresión *salir del armario* no aparece en octubre 2020 en el DRAE.

7. Permitir cierta cosa haciendo como que no se advierte, DUE; 8. No querer enterarse de ello para no disgustarse o por terquedad, DUE; 9. Hacerlo como gesto de miedo, DUE; 10. Alucinarle, para que no oiga, vea lo que le conviene, RAE;
- (2) *meter el diente* 1. Morder o comer algo difícil de mascar, DFDEA; 2. Empezar con interés una cosa que implica dificultad, DUE; 3. Apropiarse de algo que pertenece a otra persona, RAE; 4. Criticar hablar mal de alguien, RAE; 5. Cobrar más de lo establecido, DdAm, ES, CR, Pa;

3.3. Existencia de correspondencias fraseológicas

Tras el proceso de identificación y comprensión / interpretación, el traductor debe proceder a la búsqueda de correspondencias. El análisis semántico de las 41.000 entradas de la BD y la búsqueda de equivalentes traductológicos y fraseológicos nos ha permitido observar que existen numerosos casos de parasonimia y que en el proceso de traducción el traductor se va a encontrar con tres situaciones bien diferenciadas.

- (1) Aquella en la que en la lengua de Origen y Meta existe una única UF para expresar una situación, (ver tabla nº 5):

UF español	UF francés	UF inglés
[marcharse, despedirse] a la francesa (DUE)	filer à l'anglaise (GR)	To take french leave ((DCU) to make an Irish exit ¹⁸ .
cantar victoria (RAE)	chanter victoire (GR)	to claim/cry victory (DCU)
poner una pica en Flandes (RAE)	mettre 1 grain de sel sur la queue d'un oiseau (GR)	to bring off a real coup (DCU)

Tabla nº 5 única UF en lengua Origen y Meta

- (2) Aquella en la que para una misma situación encontramos varias UF en diferentes lenguas, (ver tabla nº 6).

18. Al igual que estamos recopilando las UF hispanoamericanas, sería muy importante poder diferenciar las británicas de las de EE. UU. o de otros países anglófonos para poder discriminar la traducción y la comprensión.

Empezar la casa por el tejado (RAE) Poner el carro antes de los bueyes (VV)	mettre la charrue avant les boeufs (GR) mettre la farine avant le beurre (VV)	To put the cart before the horse (DCU)
estar en la luna (RAE) estar en las nubes (DUE) estar en el limbo (RAE) estar en las Batuecas (DUE) tener la cabeza a las once (RAE)	être dans la lune (GR) avoir la tête dans les nuages (GR) être dans les nuages (GR) Bayer aux corneilles (GR)	to have one's head in the clouds (CDU) to be lost in reverie (CDU)

Tabla nº 6. Varias UF en varios idiomas

- (3) Aquella en la que en una lengua existe una UF para expresar una situación y en otra(s) no, (ver tabla nº 7).

dar gato por liebre (RAE)	rouler, tromper qq'un (LBI)	[buy/accept] a pig in a poke to pull the wool over someone's eyes to take someone for a ride
armarse la de San Quintín (DUE)	il va y avoir du grabuge	There will be an almighty row All hell will break loose
salir por la puerta grande (DFDEA)	sortir par la grande porte	to make a triumphant exit to do something in grand style

Tabla nº 7. La UF no existe en todas las lenguas

3.4. Grados de equivalencia fraseológicos

Si bien el concepto de equivalencia es recurrente en traducción (Nida 1964, Newmark 1982, Rabadán 1991, Elena García 1994, Hatim & Mason 1995; Jakobson 1959¹⁹), la equivalencia fraseológica se ha investigado fundamentalmente en lingüística contrastiva, analizando las estructuras y los componentes léxicos, buscando parámetros que a partir de similitudes y diferencias formales y semánticas permitieran fijar tipos de equivalencias, como los

19. Jakobson (1959 [1975]: 70): "La equivalencia en la diferencia es el problema cardinal del lenguaje y la cuestión central de la Lingüística".

isomorfismos²⁰ (sintácticos, léxico-semánticos, etc.). Esos estudios, no tenían en cuenta ni el nivel de lengua, ni la frecuencia y la zona de uso, ni tampoco posibles connotaciones, datos a juicio de muchos traductólogos imprescindibles para obtener una traducción propicia.

Por otro lado, la investigación de correspondencias fraseológicas, establece que existen tres grados de equivalencia traductológica (Corpas 2003; Mogorrón 2008, 2014; J. Sevilla 1997; M. Sevilla 2015; Mellado Blanco 2015). Se trata de:

- (1) equivalencia total o plena (ambas unidades presentan el mismo significado denotativo y connotativo. P. ej.: *quemar las naves* (RAE) / *to burn one's boats* (CDU) / *brûler ses vaisseaux* (GR).
- (2) equivalencia parcial (cuando una locución de la LO se traduce por una locución en la LM que presenta diferencias de significado denotativo o connotativo; o es traducida por una unidad léxica carente de los valores expresivos de la locución en cuestión: *quemar las naves* (RAE) = *couper les ponts*²¹ (GR); *dar gato por liebre* (RAE) = *to pull the wool over someone's eyes/rouler, tromper qq'un*;
- (3) equivalencia nula (se trata de casos de realidades sociolingüísticas en la LO que no se lexicalizan en la LM. El traductor debe optar por alguna otra técnica de transferencia para expresar el significado de la unidad, perdiéndose, generalmente, las connotaciones correspondientes); *armarse la de San Quintín* (RAE) = *There will be an almighty row / il va y avoir du grabuge*; *salir por la puerta grande* = *to make a triumphant exit / sortir triomphalement*.

El tipo de equivalencia más frecuente entre sistemas fraseológicos es por regla general el parcial. Coincidimos con Corpas (2003: 254):

20. Dando lugar a numerosos falsos amigos como *salir por la puerta grande / sortir par la grande porte* (acceder a una buena situación, TLFi).

21. En francés se podrían usar para *quemar las naves*, *brûler ses vaisseaux* y *couper les ponts* con el mismo significado, pero connotaciones diferentes. Al buscar la forma "cortar los puentes" en español, observamos que: (1) no aparece en ninguno de los grandes diccionarios referenciales (DRAE, DUE); (2) aparece una ocurrencia en el CREA, dos en el CORDE, más de 500 en Sketch-Engine y más de 2.000 en el buscador de Google.

Las correspondencias fraseológicas que se establecen entre dos lenguas distintas suelen ser relaciones de equivalencia parcial por cuanto las UF implicadas presentan ciertas diferencias y solapamientos con respecto a uno o varios de los siguientes aspectos: contenido semántico, base figurativa, características morfosintácticas, restricciones diasistemáticas y rasgos pragmáticos.

La búsqueda de equivalencias fraseológicas en un BD temática de 2482 UF españolas y de 2031 francesas clasificadas semánticamente (Mogorrón 2002), permite evidenciar esa afirmación así como que los casos de equivalencia fraseológica nula son muy escasos, al menos entre sistemas fraseológicos de lenguas y culturas próximas, con menos del 5% de los casos²².

No abordaremos aquí, las técnicas y estrategias de traducción, ni tampoco las dificultades de traducción fraseológica que pueden surgir en el plano léxico o textual, ya tratadas en numerosos trabajos y que deberían emplearse cuando no se encuentre un equivalente fraseológico adecuado (Capra 2012; Corpas 2000, 2003; J. Sevilla 1997; Sevilla & Sevilla 2000; M. Sevilla 2009, 2015).

El traductor debería conocer o poder conocer rápidamente los diferentes significados posibles de las UF (ver 2.4). UF como *pedir la luna*, *pegar un metido* tienen varios significados que convendría discriminar en cada caso. Las UF como *no tener abuela*, *bañarse en agua de rosas*, *enredar la pita*, pueden tener varios significados diferentes en función del origen geográfico del texto en el que figuren. Estamos en un círculo sin fin en el que se vuelve al punto de partida una y otra vez, pues si el traductor detecta la UF, pero no dispone de todas las herramientas necesarias para interpretarlas correctamente se ve impotente y obligado a interpretar a ciegas.

Es sumamente importante que estas herramientas incluyan también todas las posibles variantes de cara a facilitar la localización y la correcta interpretación de las mismas. Así, en el caso de la UF *no pegar ojo* encontramos en el DRAE *no pegar [ojo, el ojo, los ojos]* con el significado de no poder dormir y en el DFDem: *no pegar [ojo, los ojos]*. Si con esos datos vemos la versión subtitulada de la película francesa *Les femmes du sixième étage*²³ en la que

22. Evidentemente, en un futuro convendría encontrar especialistas bilingües y biculturales que realizasen este tipo de análisis entre UF pertenecientes por ejemplo al chino, árabe o japonés, por citar lenguas alejadas lingüística y culturalmente.

23. De Philippe le Guay (2010). Título español: *Las chicas de la sexta planta*.

aparece en pantalla: *no pude pegar un ojo*, podemos pensar que se trata de un error o de una mala traducción literal (ver captura de imagen 1).



Captura de imagen nº 1

Sin embargo, en el DUE figura: no pegar [ojo, el ojo, los ojos, *un ojo*], no cerrar los ojos y en el DFDEA: [*cerrar, pegar*] [*ojo, el ojo, los ojos, un ojo*], mostrando que se trata de una variante correcta aceptada por la lexicografía hispánica. La traducción de las UF no es, pues, imposible, es compleja, pero lo es debido a: la enorme cantidad de UF existentes en cada lengua; la fijación y a su cara opuesta, la variación; la competencia fraseológica limitada de todo usuario y traductor; al contenido fraseológico limitado de los diccionarios referenciales o especializados; los universales fraseológicos (2.5); los numerosos casos de UF parasinónimas.

No abordaremos las técnicas de traducción, las cuales por sí solas podrían ser objeto de todo tipo de artículos especialmente interesantes. En efecto, pensamos que antes de llegar a usar las técnicas de traducción para traducir una UF, el primer paso pasaría por disponer de una BD que permita conocer la existencia de una UF, de su(s) variante(s) y de su(s) posible(s) significado(s), así como de las expresiones parasinónimas y de sus equivalentes

fraseológicos. Esa BD permitiría visualizar el significado de numerosas UF que no pertenecen a la competencia fraseológica del usuario / traductor, o que no vienen incluidas en los fondos de los diccionarios monolingües o bilingües. Así, cuando existen equivalentes fraseológicos claros, se evitaría tener que usar técnicas o estrategias de traducción.

Tal herramienta, sea BD fraseológica o no, que se precie ser exhaustiva y útil para la traducción, debería incluir una clasificación semántica que permita conocer las UF con determinado significado, sus posibles variantes, así como indicaciones acerca de su frecuencia de uso, el nivel de lengua, el registro al que pertenecen, datos a nuestro entender indispensables para conseguir realizar una buena traducción y que se echaban de menos en los trabajos de lingüística contrastiva.

Si retomamos el caso de las CVF parasinónimas de *estar en la luna*, (2.5), el uso de todas las informaciones que vienen exponiéndose pueden aplicarse actualmente a aplicaciones informáticas. Un ejemplo podría ser la siguiente aplicación fraseológica (captura de pantalla 2²⁴):

JAMPOS SEMÁNTICOS: Carácter-forma de ser/Comportamiento-actitud-conducta >> distracción-descuido (antes atención-vigilancia-distracción), cambiado. >> estar alguien distraído

Niveles de uso: Marcas dialectales: Frecuencia de uso: Filtros:

111 Català Deutsch English Español Français Italiano Polski Português العربية Otros Enviar

Español	Français	English
<input type="checkbox"/> andar avanzado <input checked="" type="checkbox"/> CA V <input type="checkbox"/> andar en las nubes <input type="checkbox"/> andar fuera de onda <input type="checkbox"/> andar por las nubes <input type="checkbox"/> andar por las nubes <input checked="" type="checkbox"/> ES ES CA CA CEI, CEY CEV V <input type="checkbox"/> andar ríngos <input type="checkbox"/> andar las vigas <input type="checkbox"/> andar varas <input type="checkbox"/> andar varas <input type="checkbox"/> andar globos <input type="checkbox"/> andar a loaves <input type="checkbox"/> andar balanceado en Balén <input type="checkbox"/> andar con la beaza <input type="checkbox"/> andar con la tortá <input type="checkbox"/> andar con los angelitos <input type="checkbox"/> andar contando las vigas <input type="checkbox"/> andar desplazado <input type="checkbox"/> andar dentro del palo <input type="checkbox"/> andar en Babia <input type="checkbox"/> andar en Balén <input type="checkbox"/> andar en China <input type="checkbox"/> andar en el aire <input type="checkbox"/> andar en el camino de Emaús	<input type="checkbox"/> avoir la tête ailleurs <input checked="" type="checkbox"/> ES ES CA CA CEI, CEY CEV V <input type="checkbox"/> avoir la tête dans les nuages <input type="checkbox"/> avoir la tête en l'air <input type="checkbox"/> FR ES CA V <input type="checkbox"/> avoir une tête de linotte <input type="checkbox"/> FR ES CA V <input type="checkbox"/> avoir sa cornette <input type="checkbox"/> être acblimé dans ses pensées <input type="checkbox"/> être acblimé <input type="checkbox"/> être absorbé dans ses pensées <input type="checkbox"/> être dans la lune <input type="checkbox"/> être dans la lune <input type="checkbox"/> être dans les limbes <input checked="" type="checkbox"/> ES ES CA CA CEI, CEY CEV V <input type="checkbox"/> être dans les nuages <input type="checkbox"/> ES CA CEI V <input type="checkbox"/> être perdu dans ses pensées <input type="checkbox"/> être perché dans ses pensées <input type="checkbox"/> ES CEI V <input type="checkbox"/> être sur sa planète <input type="checkbox"/> ES CEI V <input type="checkbox"/> être tout à ses pensées <input type="checkbox"/> ES ES CA V <input type="checkbox"/> être tout à ses pensées <input checked="" type="checkbox"/> ES ES CA CA CEI, CEY CEV V <input type="checkbox"/> regarder les hochets <input checked="" type="checkbox"/> ES ES CA CA CEI, CEY CEV V <input type="checkbox"/> être égaré dans ses pensées <input type="checkbox"/> ES V <input type="checkbox"/> être à ses pensées <input type="checkbox"/> ES CA V <input type="checkbox"/> être dans les nuages <input type="checkbox"/> ES CEI V	<input type="checkbox"/> to be absent-minded <input type="checkbox"/> FR ES V <input type="checkbox"/> to be distracted <input type="checkbox"/> FR ES V <input type="checkbox"/> to be in a brown study <input type="checkbox"/> FR ES V <input type="checkbox"/> to be inattentive <input type="checkbox"/> FR ES V <input type="checkbox"/> to be lost in thought <input type="checkbox"/> FR ES V <input type="checkbox"/> to be out of step <input type="checkbox"/> FR ES V <input type="checkbox"/> to be out of vogue <input type="checkbox"/> FR ES V <input type="checkbox"/> to be scattered <input type="checkbox"/> FR ES V <input type="checkbox"/> to build castles in air <input type="checkbox"/> FR ES V <input type="checkbox"/> to be lost <input type="checkbox"/> FR ES CEI V <input type="checkbox"/> to have head in the clouds <input type="checkbox"/> FR ES V <input type="checkbox"/> to have high-flying ideas <input type="checkbox"/> FR ES V <input type="checkbox"/> to live in cloud cuckoo-land <input type="checkbox"/> FR ES V

Captura de pantalla n° 2

Permite conocer las correspondencias fraseológicas en varios idiomas²⁵, así como seleccionarlas a partir de filtros en función de valores diatópicos, de niveles de lengua, de frecuencia de uso y proporciona rápidamente las UF con los valores léxicos y textuales requeridos.

De cara a la traducción, este tipo de aplicación, que incorpora las posibles variaciones, las UF y su(s) posible(s) equivalente(s) fraseológico(s) tiene a su vez múltiples utilidades:

- (1) permitiría a los programas de traducción automática no realizar versiones literales de las variantes que no han sido incluidas en sus bases terminológicas.
- (2) permitiría en algunas modalidades de traducción seleccionar los equivalentes en función de los parámetros textuales, por ejemplo, en traducción literaria.
- (3) permitiría seleccionar el equivalente fraseológico más próximo al de la lengua origen en función de aspectos técnicos que deben tenerse en cuenta, como la correcta sincronización bucal en traducción audiovisual.

Conclusiones

La variación fraseológica es un fenómeno complejo cuya importancia numérica muestra que se le debe dar la relevancia que se merece. El desarrollo de la lingüística de corpus permitirá sin duda alguna destacar más si cabe su importancia a nivel de usuarios y de geolingüística. En cuanto a la traducción fraseológica, conviene repetir lo ya señalado: no es imposible, pues (muy) a menudo los mismos conceptos se usan como UF en diferentes lenguas. Es compleja, pero lo es debido a: la enorme cantidad de UF existentes en cada lengua; la fijación y a la variación; la competencia fraseológica limitada de todo usuario; al escaso contenido de los diccionarios referenciales o especializados; a los universales fraseológicos (2.5); a los numerosos casos de UF parasinónimas, etc.

Antes de utilizar las técnicas de traducción recurrentes, es indispensable para la traducción humana y automática desarrollar y disponer de enormes

25. <https://dti.ua.es/es/frasytram/grupo-de-investigacion-frasytram.html>

bases de datos que permitirán poder decidir si se usa un equivalente fraseológico disponible o se opta por otra solución.

Si bien convendrá tener siempre en cuenta que la traducción es un proceso mucho más complejo que una simple búsqueda de correspondencias, no conviene olvidar que el paso del plano léxico al discurso contextualizado lleva a considerar entonces los posibles valores pragmáticos asociados.

Referencias bibliográficas

- BAKER, Mona. (1992) *In Other Words. A Coursebook on Translation*. Londres & Nueva York: Routledge.
- BALLARD, Michel. (1998) "Comparatisme et didactique de la traduction." En: García Izquierdo, Isabel & Joan M. Verdegal (eds.) *Los estudios de traducción: un reto didáctico*. Castellón: Universidad Jaume I, pp. 55-68.
- CARNEADO MORÉ, Zoila. (1985) "Notas sobre las variantes fraseológicas." *Anuario L/L* 16, pp. 269-277.
- CORPAS, Gloria. (1996) *Manual de fraseología española*. Madrid: Gredos.
- CORPAS, Gloria. (2003) *Diez años de investigación en fraseología: Análisis sintáctico-semántico, contrastivos y traductológicos*. Madrid: Iberoamericana & Vervuert.
- CORPAS, Gloria. (2008) *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Berlín & Fráncfort del Meno: Peter Lang.
- CORPAS, Gloria. (2018) "Laughing one's head off in Spanish subtitles: a corpus-based study on diatopic variation and its consequences for translation." En: Mogorrón, Pedro & Juan Antonio Albaladejo-Martínez (eds.) 2018. *Fraseología, Diatopía y Traducción*. Amsterdam & Filadelfia: John Benjamins, pp. 31-70.
- CORPAS, Gloria & Florentina Mena Martínez. (2003) "Aproximación a la variabilidad fraseológica de las lenguas alemana, inglesa y española." *ELUA* 17, pp.181-201.
- CUADRADO-REY, Analía. (2018) "La variación fraseológica en el español hablado en Argentina: propuesta de clasificación." En: Mogorrón, Pedro y Juan Antonio Albaladejo-Martínez (eds.) 2018. *Fraseología, Diatopía y Traducción*. Amsterdam & Filadelfia: John Benjamins, pp. 334-352.
- GARCÍA-PAGE, Mario. (2008) *Introducción a la fraseología española: Estudio de locuciones*. Barcelona: Anthropos.

- GADGET, Françoise. (2003) *La variation sociale en français*. Paris & Gap: Ophrys.
- KOIKE, Kazumi. (2001) “Variación fraseológica del español.” *Varilex* 9, pp. 77-92.
- MOGORRÓN, Pedro. (2002) *La expresividad en las locuciones verbales en francés y en español*. Alicante: Publicaciones Universidad de Alicante.
- MOGORRÓN, Pedro. (2010) “Analyse du figement et de ses possibles variations dans les constructions verbales espagnoles.” *Linguisticae Investigationes* 33:1, pp. 86-152.
- MOGORRÓN, Pedro. (2015) “La variation dans les Constructions Verbales Figées de l’espagnol d’Amérique.” *Linguisticae Investigationes* 3:2, pp. 276-301
- MOGORRÓN, Pedro. (2019) “Les phrasèmes et les outils lexicographiques des traducteurs.” *Phrasis* 3, pp. 12-37.
- MOGORRÓN, Pedro. (2020) “Création, variabilité, variantes phraséologiques et diatopiques.” *Cahiers de Lexicologie* 116:1, pp. 173-196.
- MONTORO DEL ARCO, Esteban. (2004) “La variación fraseológica y el diccionario.” En: Battaner, Paz & Janet DeCesaris (eds.) 2004. *De Lexicografía*. Barcelona: Institut Universitari de Lingüística Aplicada, pp. 591-604.
- NAVARRO-BROTONS, Lucía. (2018) “Refranes españoles y sus equivalentes mexicanos.” En: Mogorrón, Pedro & Juan Antonio Albaladejo-Martínez (eds.) 2018. *Fraseología, Diatopía y Traducción*. Amsterdam & Filadelfia: John Benjamins, pp. 263-177.
- SEVILLA, Julia & Manuel Sevilla. (2000) “Técnicas de la traducción paremiológica (francés-español).” *Proverbium* 17, pp. 369-386.
- ZULUAGA OSPINA, Alberto. (1980) *Introducción al estudio de las expresiones fijas*. Fráncfort del Meno: Peter Lang.

DICCIONARIOS

- ACADEMIA MEXICANA DE LA LENGUA. (2010) *Diccionario de Mexicanismos*. México: Siglo XXI. [DDMexicanismos]
- ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA. (2010) *Diccionario de Americanismos*. Madrid: Santillana. [DDAM]
- GARCÍA-PELAYO, Ramon & Jean Testas. (1993 [1967]) *Larousse moderno français-espagnol español-francés*. París: Larousse. [LBI]
- LARA, Luis Fernando (dir.) (1996) *Diccionario del español usual en México*. (1996). México: El Colegio de México. [DEUEM]
- MOLINER, María. (1999 [1977]) *Diccionario de uso del español*. Madrid: Gredos. [DUE]

- RAE. (1992) *Diccionario de la Real Academia Española* (21ª ed.). Madrid: Espasa-Calpe. [DRAE]
- REY, Alain. (2001) *Le Grand Robert: Le dictionnaire le plus complet de la langue française*. Paris: Société du nouveau Littre. [GR]
- SECO, Manuel; Olimpia de Andrés & Gabino Ramos. (2004) *Diccionario fraseológico documentado del español actual*. Madrid: Aguilar. [DFDEA]
- VARELA, Fernando. (1996) *Diccionario fraseológico del español moderno*. Madrid: Gredos. [DFDEM]
- VV. AA. (2004) *Diccionario Collins Español-Inglés / English-Spanish*. Barcelona: Grijalbo. [DCU]

NOTA BIOGRÁFICA / BIONOTE

Catedrático de Universidad en la Universidad de Alicante (UA). Ha sido Director del Dto. de Traducción e Interpretación (2009-2017), coordinador del Doctorado en Traducción con mención de calidad (2006-2009) de la UA y es coordinador del Doctorado interuniversitario en Traductología, Traducción Profesional y Audiovisual. IP de numerosos proyectos de investigación a nivel nacional, autonómico y local. Ha dirigido numerosos TFGs, TFM y 16 tesis doctorales sobre terminología, traducción audiovisual, traducción de elementos culturales, fraseología, fraseodidáctica y variaciones diatópicas. Director del grupo de investigación FRASYTRAM desde 2006 (Fraseología y Traducción Multilingüe). Cuenta con numerosas publicaciones en revistas y editoriales reconocidas en el Scholar Publishing Index (SPI). Ha sido profesor invitado en las Universidades de Paris 13, Paris Marne-la-Vallée, Paris-Ivry, Paris-Diderot, U. Pedagógica de Cracovia, U. de Antananarivo, U. R. Palma, etc. Líneas de trabajo e investigación: terminología y traducción especializada, traducción audiovisual, traducción humanístico-literaria (ámbito literario, cultural, periodístico).

Professor at the University of Alicante (UA). He has been director of the Department of Translation and Interpreting (2009-2017), coordinator of the PhD Programme in Translation distinguished with a quality mention (2006-2009) and is currently coordinator of the inter-university PhD Programme in Translation Theory, Professional and Audiovisual Translation. He was the

principal researcher on many national, regional and local research projects and directed numerous Final Degree Projects, Master's final projects and supervised 16 doctoral theses on terminology, audiovisual translation, translation of cultural elements, phraseology, didactic phraseology and diatopic variations. Since 2006 he is director of the FRASYTRAM research group (Phraseology and Multilingual Translation) and has published numerous papers in journals and books by publishers listed in the Scholar Publishing Index (SPI). He has been a visiting professor at the Universities of Paris 13, Paris Marne-la-Vallée, Paris-Ivry, Paris-Diderot, U. Pedagogie de Cracow, U. de Antananarivo, U. R. Palma, etc. His research and study lines are: terminology and specialized translation, audiovisual translation and humanistic-literary translation (literary, cultural, journalistic field).

Recibido / Received: 24/08/2019
Aceptado / Accepted: 25/11/2019

Para enlazar con este artículo / To link to this article:
<http://dx.doi.org/10.6035/MonTI.2020.ne6.2>

Para citar este artículo / To cite this article:

Cataldo, Silvia. (2020) "Approccio cognitivo alla variazione fraseologica: alcune concettualizzazioni della pazzia in italiano e implicazioni per la loro traduzione in spagnolo e tedesco." 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. 65-93.

APPROCCIO COGNITIVO ALLA VARIAZIONE FRASEOLOGICA: ALCUNE CONCETTUALIZZAZIONI DELLA PAZZIA IN ITALIANO E IMPLICAZIONI PER LA LORO TRADUZIONE IN SPAGNOLO E TEDESCO

COGNITIVE APPROACH TO THE PHRASEOLOGICAL VARIATION:
SOME CONCEPTUALIZATIONS OF MENTAL ILLNESS IN ITALIAN
AND THE IMPLICATIONS FOR THEIR TRANSLATION INTO
SPANISH AND GERMAN

SILVIA CATALDO
silvia.cataldo90@gmail.com
Universidad de Alicante

Riassunto

La linguistica cognitiva ha messo in luce la tendenza degli esseri umani a comprendere la realtà mediante trasferimenti concettuali tra domini diversi, che possono essere convenzionali nella mente e dar vita a espressioni lessicalizzate con valore figurato in ogni lingua. Nel presente lavoro si propone un approccio cognitivo allo studio della variazione fraseologica, considerando, nello specifico, alcune delle unità fraseologiche italiane usate per riferirsi a persone che si trovano in uno stato di alterazione delle proprie facoltà mentali, al fine di determinarne i reciproci rapporti concettuali. In una fase successiva, poi, si sviluppa una riflessione sulle modalità in cui l'esistenza delle varianti individuate può influire sull'intero processo traduttivo, dal loro riconoscimento e dalla loro interpretazione, alla loro vera e propria traduzione, nonché sullo sforzo cognitivo richiesto al traduttore, prendendo in esame lo spagnolo e il tedesco come ipotetiche lingue meta.



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

Parole chiave: Variazione fraseologica; Processo traduttivo; Equivalenza cognitiva; Equivalenza concettuale; Metafora.

Abstract

Cognitive linguistics has highlighted the tendency of human beings to understand reality through conceptual transfers between different domains, that can be conventional in the mind and create lexicalized expressions with a figurative value in every language. A cognitive approach to the study of phraseology variation will be proposed by taking into consideration a number of Italian phraseological units that refer to people whose mental faculties are in an altered condition, in order to determine their mutual conceptual correlation. Subsequently, the way in which the existence of the identified variants can influence the whole translation process will be studied, from their recognition and interpretation to their actual translation, as well as the translator's cognitive effort. In this phase, Spanish and German will be considered as hypothetical target languages.

Keywords: Phraseology variation; Translation process; Cognitive equivalence; Conceptual equivalence; Metaphor.

1. Introduzione

Nel presente contributo si delinea un approccio cognitivo alla variazione fraseologica e al processo traduttivo di varianti fraseologiche. Lo stato dell'arte illustrato nel seguente paragrafo abbraccia diverse prospettive di analisi del fenomeno, mediante la breve descrizione di: (i) alcuni studi che lo osservano e classificano da un punto di vista principalmente linguistico, (ii) altri che indagano il comportamento delle unità fraseologiche (UF) e la loro eventuale metaforicità, (iii) ricerche nel campo della linguistica cognitiva, tese a chiarire i trasferimenti concettuali alla base di usi figurati nella lingua attivati nella mente umana per comprendere determinate realtà, con esempi di trattamenti della variazione fraseologica in termini cognitivi, e (iv) nell'ambito dei *cognitive translation studies*, mirate a quantificare lo sforzo cognitivo messo in campo nel processo traduttivo.

Nel terzo paragrafo si seleziona una rosa di espressioni italiane figurate lessicalizzate designanti la pazzia e si suggeriscono le possibili relazioni concettuali che intercorrono tra le une e le altre, proponendo l'idea di una

variazione cognitiva consistente nell'esistenza di UF sinonimiche che rispondono a trasferimenti mentali distinti e/o che riflettono trasferimenti concettuali più specifici riconducibili a una metafora comune, pur senza negare il peso di eventuali aspetti pragmatici caratterizzanti alcune varianti esaminate.

Una volta chiariti i reciproci rapporti concettuali tra le espressioni italiane, nella sezione successiva si ipotizzano gli effetti della variazione cognitiva commentata sul processo di traduzione, prendendo in esame il tedesco e lo spagnolo come ipotetiche lingue meta: considerando alcune espressioni disponibili in queste ultime contenenti trasferimenti concettuali in qualche modo relazionati con quelli riscontrati in italiano, si osservano possibili analogie interlinguistiche, dirette e indirette, e si presentano delle riflessioni su come l'esistenza nelle lingue fonte e meta di varianti cognitive e le loro specifiche concettualizzazioni possano guidare verso l'individuazione e la corretta interpretazione di ciascuna variante fraseologica italiana e sulle valutazioni che dovrebbero o potrebbero precedere la ricerca e la proposta di corrispondenze.

2. Stato dell'arte

In questa prima parte del presente lavoro si illustrano, raggruppati per area di ricerca, alcuni degli studi sviluppati attorno al concetto di variazione fraseologica, allo scopo di definire meglio quello che sarà l'oggetto di osservazione e analisi e in cosa l'approccio delineato si discosta dalle proposte della linguistica e si conforma ad alcuni orientamenti cognitivisti.

2.1. *Variazione fraseologica: alcuni modelli*

La variazione rientra tra le possibili caratteristiche riscontrate nelle UF. Oltre ad essa, Corpas (1996: 19-32) ne elenca altre cinque: la frequenza d'uso e di co-apparizione dei componenti costitutivi; la convenzionalizzazione o istituzionalizzazione, concernente l'abitudine dei parlanti di utilizzare combinazioni già disponibili nella lingua senza doverne creare di originali; la conseguente stabilità, data dalla fissazione interna, che riduce la possibilità di manipolazione dei componenti dell'espressione, ed esterna, relativa al suo uso in contesti ben determinati, e dalla specializzazione semantica o lessicizzazione, ossia dall'acquisizione di un significato aggiuntivo o sostitutivo;

l'idiomaticità, cioè il valore metaforico dell'UF, che non va quindi interpretata letteralmente in base ai significati dei singoli elementi; la gradazione, vale a dire la misura in cui tutte le proprietà menzionate si manifestano. Il fatto che insieme a queste caratteristiche sia ammessa una certa variabilità parrebbe confermare e non contraddire il principio della fissazione nel sistema linguistico, dal momento che in assenza di una stabilità dell'espressione non avrebbe senso parlare di variazione e quindi di formulazioni alternative (cf. Corpas Pastor e Mena Martínez 2003: 183) e visto e considerato che, nel caso di modificazioni creative, si approfitta proprio della fissazione per raggiungere determinati scopi comunicativi (Zuluaga 1999: 542).

Riguardo al concetto di variazione, in Corpas (1996: 27-30, 235-250) e in Corpas e Mena (2003), si distingue tra varianti e modificazioni: le prime sono UF connesse a un'altra espressione fissata nella lingua in quanto frutto di cambiamenti lessicali o morfosintattici attuati su di essa, e non vanno confuse con le variazioni per derivazione o trasformazione, con le variazioni diatopiche e diafasiche, con le serie fraseologiche, che presentano significati opposti, e con espressioni aventi caselle vuote, cioè spazi da riempire con elementi non fissi; le seconde consistono in manipolazioni che il parlante mette in atto di proposito per perseguire specifiche finalità comunicative, riproducendo e producendo al contempo, poiché resta nei limiti di un linguaggio prestabilito pur personalizzandolo. Queste ultime, inoltre, possono essere interne, se interessano la struttura formale agendo sui componenti dell'espressione mediante sostituzioni, aggiunte, riduzioni, alterazioni dell'ordine degli elementi o combinazioni di più procedimenti, come avviene nella fusione di UF, oppure esterne, se riguardano solo il piano del significato per effetto di fattori semantico-contestuali, legati allo sfruttamento di polisemie, e pragmatici.

Quanto riportato da Corpas richiama la visione di Zuluaga (1980), il quale distingue tra varianti in senso stretto da una parte, intese come formulazioni alternative di una stessa UF che hanno significato analogo e parziale identità di struttura e componenti, non differiscono nei contesti d'uso, appartengono alla stessa lingua funzionale e a un gruppo limitato e stabile di realizzazioni linguistiche alternative, e varianti in senso ampio dall'altra, comprendenti trasformazioni, fraseologismi diversi nei loro componenti da quello di riferimento ma analoghi ad esso nel significato, serie fraseologiche, varianti diatopiche, diastratiche e diafasiche e UF con caselle vuote. Lo studioso, infine,

osserva che in una prospettiva diacronica determinate alterazioni innovative possono finire per stabilirsi nel sistema linguistico come varianti lessicalizzate (Zuluaga 1999: 542).

Montoro Del Arco (2004: 595-598), invece, suggerisce una distinzione tra varianti e variazioni, indicando come condizioni comuni un significato analogo e un'identità parziale di struttura e componenti rispetto all'UF di riferimento, l'indipendenza dal contesto e la fissità della sostituzione. Le prime possono essere interne, se riguardano alterazioni lessicali del nucleo verbale, del nucleo nominale o di entrambi, e morfo-lessicali, consistenti in un cambiamento morfologico di un componente lessicale dell'espressione, oppure esterne, se sono di carattere diatopico, diastratico, diafasico o diacronico. Le variazioni, invece, interessano l'aspetto strutturale e includono alterazioni dovute a cambiamenti lessicali o di categoria morfologica che ampliano o riducono l'espressione senza incidere sul significato, manipolazioni morfologiche di genere o di numero o sintattiche (relative all'ordine degli elementi), nominalizzazioni (se, ad esempio, una locuzione verbale diviene locuzione nominale), variazioni grafiche (quando alcuni componenti di un'UF possono essere scritti in un'unica parola o separatamente), lievi variazioni grammaticali (in caso di preposizioni intercambiabili) e UF con caselle vuote.

A proposito di espressioni con caselle vuote, che consentono di rimpiazzare più elementi, García Page (2016) individua e descrive quelle da lui denominate *variantes léxicas correlativas*, in cui la sostituzione è multipla e simultanea e avviene secondo una restrizione semantica prestabilita, mantenendo lo stesso significato metaforico e anche una certa coerenza semantica tra i componenti lessicali di ciascuna sequenza, come si osserva in *ese bollo no se ha cocido en su horno* ed *ese garbanzo no se ha cocido en su olla*, o *ser el garbanzo negro del cocido* e *ser la oveja negra del rebaño*.

Concentrandosi sulle UF con valore metaforico (*idiom*), e riferendosi al grado di fissazione delle loro possibili formulazioni alternative, Langlotz (2006: 224) riconosce, infine, più livelli di variazione: parte dalla massima cristallizzazione, procede attraverso varianti usuali, piuttosto ricorrenti e riportate nei dizionari o date da cambiamenti morfosintattici, varianti occasionali, motivate da specifiche finalità discorsive, talvolta basate su giochi di parole, e in alcuni casi erronee e quindi non intenzionali, per finire con le pseudo varianti, da interpretarsi letteralmente in quanto non richiedono l'attivazione

del significato figurato nel contesto in cui sono usate. Descrivendo le modalità in cui da una data espressione lessicalizzata possono nascere delle formulazioni alternative, lessicalizzate o creative, questo stesso autore (ibid.: 179-182) segnala eventuali trasformazioni formali di carattere morfosintattico, sintattico e lessicale, o modifiche di tipo semantico, relative al valore polisemico di un'espressione, ad adattamenti mediante l'uso di intensificatori, specificazioni, antonimi o cambi di prospettiva, o all'inserimento dell'UF in un nuovo contesto.

2.2. *UF metaforiche: gli idiom*

Dal momento che, se si escludono le modificazioni creative, le variazioni o varianti fraseologiche finora descritte (da adesso in avanti si useranno indistintamente questi termini per fare riferimento allo stesso fenomeno in senso generale) costituiscono delle UF a tutti gli effetti, in quanto espressioni alternative preconcepite nella lingua di appartenenza, è possibile attribuire loro le tre fondamentali caratteristiche che, nel tempo, sono state assegnate alle espressioni istituzionalizzate studiate nell'ambito della fraseologia, le quali possono anche non cooccorrere o non presentarsi con evidenza in ogni occasione: quella sintattica della stabilità o fissazione, quella semantica dell'idiomaticità o della non composizionalità di significati dei singoli componenti, e quella della motivazione, corrispondente alla possibilità di identificare l'immagine che ha dato origine all'espressione, piuttosto alta qualora esistesse o fosse ancora evidente il suo omofono letterale (cf. Arribas 2006: 166-168). Di conseguenza, alcune UF, con relative varianti lessicalizzate, possono avere una base metaforica e perché ciò avvenga è sufficiente che anche solo uno degli elementi che le compongono sia idiomatico (Castillo Carballo 1997-1998: 72). L'idiomaticità, inoltre, può manifestarsi in due modi: mediante combinazioni fisse derivate da combinazioni libere, in caso di assenza di incompatibilità semantica tra le parole di cui sono costituite, come si osserva in espressioni arcaiche che hanno avuto in passato un significato letterale poi andato perduto o in espressioni omonime di formulazioni libere, o attraverso combinazioni fisse non derivate da combinazioni libere, in presenza di incompatibilità semantica tra i componenti e quindi di impossibilità di interpretazione letterale (cf. González Rey 1998: 61-62). Alle due sottocategorie del

primo gruppo appartenerebbero rispettivamente UF italiane come *mettere alla berlina* e *tagliare la corda*, mentre un esempio per il secondo potrebbe essere *avere la testa fra le nuvole*. In linea con tale osservazione, Vega Moreno (2005: 398-399) ritiene che la metaforicità degli *idiom* possa essere sorta insieme all'espressione stessa, oppure aver avuto origine in un momento successivo. In un caso, si tratta di espressioni nate direttamente come usi figurati privi di interpretazione letterale che in seguito si sono standardizzati, mentre nell'altro si fa riferimento a espressioni letterali che hanno poi assunto un significato figurato, come si osserva negli *idiom* che rappresentano estensioni semantiche nuove di frasi familiari inizialmente letterali andate via via fissandosi nel sistema.

2.3. *Linguistica cognitiva e fraseologia*

Sono molti gli studiosi che si sono interessati alla dimensione semantico-cognitiva delle UF figurate al fine di determinare le rappresentazioni mentali dalle quali sono sorte: quest'approccio ha dimostrato un'interconnessione tra le dimensioni letterale e figurata all'interno del lessico mentale dell'essere umano, la quale fa sì che il significato di un'intera espressione sia fortemente vincolato alla sua motivazione metaforica (Corpas Pastor 2001: 32-33). Richards (1936) e Black (1962) sono tra i primi a credere nella natura concettuale della metafora, che da fenomeno linguistico con funzione puramente estetica inizia ad essere intesa come interazione tra il dominio metaforizzato e il dominio che permette di concettualizzarlo proiettando su di esso alcune delle sue caratteristiche. A partire da questa proposta, Lakoff e Johnson (1980) elaborano la *Conceptual Metaphor Theory*, secondo la quale nella metafora si manifestano dei processi cognitivi che associano realtà distinte mediante trasferimenti strutturali, ontologici o direzionali e la selezione di solo alcune delle loro proprietà.

Legato alla dimensione cognitiva della metafora è il grado di convenzionalità delle concettualizzazioni, relativo a quanto esse sono comuni o originali nel pensiero all'interno del contesto in cui si manifestano. Kövecses (2005; 2010a; 2010b; 2010c) ha proposto una distinzione tra metafore concettuali universali o quasi universali, diffuse oltre i limiti di specifiche culture e comunità linguistiche perché basate generalmente sull'esperienza fisica dell'essere

umano, e metafore vincolate (i) al contesto globale, relazionato con il territorio e quindi con la sua geografia, i suoi paesaggi, la sua flora e la sua fauna, con i gruppi sociali che le utilizzano e con la storia della comunità linguistica di appartenenza, e (ii) al contesto locale, relativo a condizioni fisiche come, ad esempio, la stagione dell'anno in cui ci si trova, alla conoscenza, da parte di chi parla, delle principali entità coinvolte nel discorso (sé stesso, chi ascolta e i concetti associati nel trasferimento metaforico), il contesto culturale immediato, legato alla rilevanza di un evento per la comunità nel momento dell'enunciazione, le relazioni sociali tra gli interlocutori e il ruolo sociale attribuito loro, la situazione enunciativa e il contesto linguistico nel caso di metafore nuove che ne richiamano altre già espresse all'interno dello stesso discorso.

Nell'ultimo ventennio si è assistito a un'interazione sempre maggiore tra studi di carattere linguistico-cognitivo e fraseologia: da una parte, ci si è serviti delle UF metaforiche per indagare i meccanismi di concettualizzazione della realtà negli esseri umani e, dall'altra, la linguistica cognitiva ha supportato spesso l'osservazione delle UF idiomatiche cercando di spiegarne l'origine. Tuttavia, nell'ambito della variazione fraseologica, si osserva un'abbondanza di ricerche mirate a presentare delle definizioni e delle classificazioni sulla base di aspetti linguistici e pragmatico-contestuali (quelle riportate nel paragrafo 2.1. ne sono un esempio) a fronte di una probabilmente più ridotta quantità di proposte di approcci cognitivi, benché la natura figurata di molte espressioni istituzionalizzate, nonché delle loro formulazioni alternative, motiverebbe a considerare la dimensione concettuale di queste ultime. L'importanza di trattare la variazione fraseologica anche da un punto di vista cognitivo parrebbe confermata da Kövecses (1986: 14), il quale riconosce una reciproca influenza tra dimensione concettuale e dimensione lessicale: una metafora nel pensiero, infatti, può essere espressa mediante molteplici parole o espressioni fisse, le quali, a loro volta, permettono di elaborare la metafora concettuale centrale. Tale considerazione pare suggerire che espressioni o scelte lessicali alternative per una stessa UF, oltre ad interessare la sfera linguistica, mettono in luce modi di comprendere una data realtà nel pensiero, evocando in chi parla, ascolta o legge determinate associazioni mentali.

Tra gli esempi di approcci cognitivi alla variazione fraseologica, si segnala uno studio di Dobrovolskij (2016) nel quale l'autore spiega la natura di certe

regolarità caratterizzanti questo fenomeno basandosi sull'analisi di alcune UF russe: nello specifico, si ipotizza che la possibilità di formare coppie causative per degli stessi *idiom*, costituite da un'espressione in cui l'attività è autoindotta e da un'altra in cui essa è provocata, come si osserva nella coppia *someone's hair stands on end* e *make someone's hair stand on end* ('rizzarsi i capelli a qualcuno' e 'far rizzare i capelli a qualcuno') per l'inglese, e in *ans Licht kommen* e *ans Licht bringen* ('venire alla luce' e 'portare alla luce') per il tedesco, dipenda dalle loro proprietà semantico-concettuali, e a dimostrazione di ciò lo studioso indica diverse aree semantiche che ammettono tale meccanismo, apportando per ognuna degli esempi tratti dalla lingua russa.

Langlotz (2006: 11), invece, all'interno di un lungo lavoro di ricerca sulla dimensione cognitiva della modificazione creativa degli *idiom*, propone l'idea che questa sia vincolata alle corrispondenze concettuali soggiacenti alla loro struttura semantica interna.

Lo stesso Kövecses (2010a: 666-667), infine, parla di variabilità in termini concettuali, riferendosi al modo in cui metafore generali possono essere elaborate per effetto del contesto per formare metafore complesse creative o convenzionali.

2.4. Processo traduttivo e sforzo cognitivo nella traduzione della metafora

La traduzione non consiste solo nel prodotto rappresentato dal testo meta, ma anche nell'intero processo che dalla lettura del testo fonte conduce alla sua ricostruzione in un'altra lingua. Tale concezione ha portato allo sviluppo di numerosi studi tesi a indagare i comportamenti cognitivi del traduttore nel trattamento di specifici fenomeni linguistico-cognitivi, basati su protocolli di *think-aloud* o sull'impiego di programmi di *eye-tracking* e *key-logging* capaci di misurare i movimenti oculari nell'osservazione dello schermo del computer in un caso e quelli manuali nell'uso della tastiera nell'altro. In termini generali, si riconoscono come fasi principali del processo traduttivo quella di comprensione del testo fonte e quella di produzione del testo meta, le quali sembrerebbero presentarsi non sempre in maniera sequenziale, ma anche parallela (cf. Sjørup 2013: 60-61). Nella sua tesi dottorale, Sjørup (2013) presenta diverse ricerche mirate a determinare lo sforzo cognitivo nella traduzione della metafora: di regola, l'entità di questo sforzo nella comprensione

è calcolata in base ai tempi in cui lo sguardo del traduttore si sofferma su una data parola, alla frequenza con cui ciò avviene e a regressioni su parti del testo già lette, mentre l'entità dello sforzo nella produzione è valutata in virtù delle pause realizzate nel corso della digitazione. I risultati degli studi esaminati rivelano, tra le altre cose, che il fatto stesso di leggere al fine di tradurre implicherebbe un maggior dispendio di energie rispetto a una normale lettura per la necessità di accedere contemporaneamente a due lessici mentali distinti e che il contesto aiuta ad accedere direttamente all'interpretazione figurata senza passare per quella letterale.

Particolarmente utile ai fini del presente lavoro è la ricerca condotta da Mandelblit (1995), che calcolando i tempi di traduzione tra francese e inglese di alcune espressioni metaforiche lessicalizzate, ha riscontrato che questi sono più limitati se i trasferimenti concettuali e le loro realizzazioni linguistiche disponibili nelle due lingue corrispondono, aumentano di poco nel caso di differenze unicamente lessicali, e in maniera considerevole se le differenze riguardano la dimensione concettuale.

3. Proposta di analisi cognitiva della variazione fraseologica: alcune concettualizzazioni lessicalizzate dello stato di pazzia in italiano

L'idea che si intende presentare come punto di partenza di questo lavoro è che, in termini cognitivi, la variazione fraseologica può essere intesa come l'esistenza di concettualizzazioni distinte e/o interconnesse per esprimere uno stesso significato. Per dimostrare tale teoria, inizialmente, si raccoglieranno alcune UF metaforiche sinonimiche usate in italiano per riferirsi a uno stato di alterazione delle facoltà mentali di un individuo, le quali saranno classificate secondo le loro reciproche relazioni concettuali. Nelle sezioni successive si svilupperanno poi delle riflessioni circa gli effetti che tali rapporti concettuali possono avere sul processo traduttivo, costituito da una fase di riconoscimento e interpretazione dell'UF, seguita dalla ricerca e dalla proposta di soluzioni traduttive.

Le espressioni italiane selezionate per l'analisi sono *essere fuori di testa*, *essere via di testa*, *essere uscito di testa*, *non starci con la testa*, *essere fuori di capoccia*, *essere fuori di melone*, *essere fuori di cocomero*, *essere fuori di zucca*, *essere fuori di cocuzza/cucuzza*, *essere fuori di coccia*, *essere fuori di brocca*,

essere fuori di mente, essere fuori di senno, essere fuori di cervello, essere via di cervello, essere uscito di cervello, non starci col cervello, essere fuori, essere fuori di sé, essere fuori come un balcone, essere più di là che di qua, avere il cervello in fondo ai piedi, avere il cervello nelle scarpe, avere il giudizio sotto (la suola del) le scarpe, ragionare coi/con i piedi, mancare qualche rotella (a qualcuno), avere qualche rotella fuori posto / non avere tutte le rotelle a posto, essersi bevuto il cervello, essersi fumato il cervello, mancare un venerdì (a qualcuno), aver (mandato) il cervello in vacanza, avere il cervello (andato) in pappa, essere (matto/pazzo) da legare, essere matto/pazzo da catena. Queste UF sono state ricondotte a cinque possibili macro-gruppi: ognuno di essi rappresenta una variante cognitiva di una tendenza generale alla metaforizzazione, mentre più espressioni inserite in uno stesso macro-gruppo costituiscono varianti cognitive di un unico trasferimento concettuale di livello superiore. Per alcune di esse, inoltre, sono riscontrabili ulteriori variazioni cognitive, nelle quali si mantiene sia l'immagine generale che quella più specifica aggiungendo o modificando un determinato trasferimento concettuale. Nel dettaglio:

A) metafora direzionale:

A.1) metafora direzionale dentro-fuori:

- a. *essere fuori di testa / essere uscito di testa / non starci con la testa / essere via di testa, essere fuori di capoccia* (variante dialettale romana): la testa è concepita e va intesa in chiave metonimica, come sede del cervello che regola e controlla le facoltà intellettive e la pazzia è vista e percepita come un trovarsi con il proprio cervello all'esterno.
 - i. testa concettualizzata in termini di frutto o ortaggio:
 - 1° *essere fuori di melone* (variante regionale settentrionale)
essere fuori di cocomero (variante regionale meridionale)
 - 2° *essere fuori di zucca*
essere fuori di cocuzza/cucuzza (variante regionale meridionale)
 - ii. testa concettualizzata in termini di contenitore:
 - 1st. *essere fuori di coccia* (variante regionale centro-meridionale);
 - 2nd. *essere fuori di brocca* (variante dialettale romana)
 - b. *essere fuori di mente*
 - c. *essere fuori di cervello / essere via di cervello / essere uscito di cervello / non starci col cervello*
 - d. *essere fuori di senno*
 - e. *essere fuori*
 - f. *essere fuori di sé*
 - g. *essere fuori come un balcone*: l'idea del trovarsi con il cervello al di fuori del luogo in cui si dovrebbe essere è mantenuta implicita e appare arricchita da una similitudine che associa la sporgenza del balcone al posizionamento esterno dell'individuo.
- A.2) la persona pazzo è distante da chi parla:
- a. *essere più di là che di qua*

B) il cervello è collocato nella parte inferiore del corpo
<ul style="list-style-type: none"> a. <i>avere il cervello in fondo ai piedi</i> b. <i>avere il cervello nelle scarpe</i> c. <i>avere il giudizio sotto (la suola del)le scarpe</i> d. <i>ragionare coi/con i piedi</i>
C) condizione in cui la persona considerata pazza difetta di o in qualcosa:
<ul style="list-style-type: none"> a. il cervello è un sistema malfunzionante di ingranaggi, con componenti mancanti o difettosi: <ul style="list-style-type: none"> i. <i>mancare qualche rotella (a qualcuno)</i> ii. <i>avere qualche rotella fuori posto / non avere tutte le rotelle a posto</i> b. il pazzo non è in possesso del proprio cervello e tale mancanza è associata a un vizio: <ul style="list-style-type: none"> i. <i>essersi bevuto il cervello</i>: da una parte è percepibile un'associazione tra la pazzia e lo stato di ebrezza e dall'altra il cervello è inteso in termini di bevanda, forse alcolica. ii. <i>essersi fumato il cervello</i>: da una parte è percepibile una possibile associazione tra la pazzia e uno stato di allucinazione causato probabilmente dal consumo di droga, e dall'altra il cervello è inteso in termini di sostanza che si può fumare, forse stupefacente. c. <i>aver (mandato) il cervello in vacanza</i> d. <i>mancare un venerdì (a qualcuno)</i>: la persona ritenuta pazza appare incompleta come lo sarebbe la settimana senza uno dei suoi giorni.
D) il cervello è in uno stato di decomposizione:
<ul style="list-style-type: none"> a. <i>avere il cervello (andato) in pappa</i>
E) la pazzia è descritta con riferimenti a trattamenti psichiatrici obsoleti:
<ul style="list-style-type: none"> a. <i>essere (matto/pazzo) da legare</i> b. <i>essere matto/pazzo da catena</i>

Come indicato, alcune delle varianti cognitive (e linguistiche) di *essere fuori di testa* sono legate a condizioni per lo meno diatopiche distinte, dal momento che *essere fuori di melone* parrebbe diffusa nelle regioni settentrionali dell'Italia, *essere fuori di cocomero* ed *essere fuori di cocuzza/cucuzza* sembrerebbe più usata nell'Italia meridionale, *essere fuori di cocchia* sarebbe propria delle regioni centro-meridionali e, infine, *essere fuori di brocca* ed *essere fuori di capoccia* apparterrebbero a un gergo romano, sebbene, come osserva Albaladejo (2016: 181) facendo riferimento alla fraseologia in lingua tedesca, risulti spesso impossibile stabilire a livello lessicografico dei limiti geografici netti nell'uso di determinate varianti regiolettali.

Molte delle varianti e sotto-varianti cognitive di *essere fuori di testa* proposte, inoltre, costituiscono anche delle varianti diafasiche, dal momento che potrebbero rispondere a esigenze di registro diverse rispetto all'UF di riferimento.

Si segnalano, infine, differenze nella frequenza d'uso, probabilmente proprio in virtù di alcune limitazioni geografiche, oltre che situazionali: *essere fuori di cocuzza/cucuzza*, *essere fuori di coccia*, *essere fuori di brocca* e soprattutto *essere fuori di cocomero* occorrono molto meno spesso in rete rispetto a *essere fuori di testa*, *essere fuori di zucca* ed *essere fuori di melone*. Tale fenomeno appare in linea con la visione di Kövecses illustrata nel paragrafo 2.3. riguardo all'influenza del contesto nell'attivazione di determinate concettualizzazioni: la selezione di una delle varianti diatopiche e diafasiche proposte, infatti, potrebbe dipendere dal contesto sia globale, riguardante la comunità territoriale di appartenenza, che locale, concernente la situazione enunciativa, le relazioni sociali tra i partecipanti alla comunicazione e il ruolo sociale che viene loro attribuito.

Le differenze pragmatiche osservate in queste espressioni riportanti concettualizzazioni specifiche, sono rilevabili anche nel confronto tra UF relative a trasferimenti mentali più generali: *essere fuori di testa*, ad esempio, non è sempre intercambiabile con *essersi bevuto il cervello*, così come quest'ultima potrebbe non funzionare in sostituzione a *mancare un venerdì (a qualcuno)* in determinate circostanze enunciative.

Le varianti cognitive *mancare qualche rotella (a qualcuno)* e *avere qualche rotella fuori posto / non avere tutte le rotelle a posto*, a differenza delle precedenti, non presentano differenze pragmatico-contestuali, ma solo concettuali: nella prima il cervello è inteso come congegno meccanico carente di alcuni elementi, mentre nelle due formulazioni della seconda gli elementi ci sono tutti, ma una parte di essi non è collocata nel posto giusto. Allo stesso modo, *essersi bevuto il cervello* ed *essersi fumato il cervello*, riflettono due trasferimenti metaforici specifici distinti riconducibili a un solo meccanismo di comprensione con tratti pragmatici analoghi.

In base a quanto esaminato, in termini generali e a seconda dei casi, sono individuabili più livelli di variazione cognitiva nel seguente ordine gerarchico: (i) concettualizzazioni completamente diverse che definiscono una stessa realtà (A, B, C, D, E); (ii) concettualizzazioni generali di una ancor più generale (A.1 e A.2); (iii) variazioni cognitive più specifiche di una concettualizzazione generale (ad esempio, a-g di A.1 o a-d di B); (iv) ulteriori varianti cognitive di una concettualizzazione più specifica (tutte le espressioni i e ii della tabella), e così via (si considerino 1° e 2°). In un'ottica cognitiva, quindi,

la variazione fraseologica qui illustrata appare come un sistema peculiare di matrioske, tutte con valore metaforico e tra loro sinonimiche, contenenti ognuna un numero variabile di pezzi, in cui le madri rappresentano le concettualizzazioni generiche e i semi quelle aventi il massimo grado di specificità; ogni bambolina, in più, ha la facoltà di contenere due o più pezzi della stessa grandezza nel caso di concettualizzazioni alternative parimenti specifiche. Alcune varianti cognitive, inoltre, possono disporre a loro volta di alternative puramente linguistiche (*essere fuori/via/uscito di cervello* o *avere qualche rotella fuori posto* e *non avere tutte le rotelle a posto*, tra le altre) o possedere valori pragmatico-contestuali differenti (*essere fuori di testa/melone/cocomero/zucca/cocuzza/coccia/brocca*). In quest'ultimo caso, si rilevano gradi distinti di convenzionalità nel pensiero e di lessicalizzazione nella lingua, vista la diversa diffusione delle variazioni cognitive in questione entro o oltre i confini di certe culture o comunità linguistiche e in o al di fuori di determinate circostanze comunicative.

4. Dimensione cognitiva della variazione fraseologica e implicazioni per la traduzione

Nella presente sezione si sviluppa una riflessione circa i possibili effetti della variazione cognitiva descritta sul processo traduttivo, ossia sul riconoscimento dell'UF, nonché della variante cognitiva in questione, sulla sua interpretazione, sulla ricerca e la proposta di una traduzione e, in generale, sullo sforzo mentale richiesto al traduttore. Tale riflessione si basa sull'esame delle espressioni italiane precedentemente condotto e sul confronto tra i sistemi concettuali italiano, spagnolo e tedesco relativi all'essere pazzo.

4.1. Riconoscimento e interpretazione delle varianti fraseologiche

Per quanto concerne il riconoscimento e l'interpretazione delle UF, Timofeeva (2012: 409-410), riformulando Corpas (2003) ne descrive i possibili rapporti cronologici, segnalando che l'interpretazione di un'espressione lessicalizzata può essere simultanea alla sua identificazione, nel caso in cui il traduttore la conoscesse e ne comprendesse il senso nel momento stesso in cui la individua, precederla, se attraverso l'interpretazione si riconosce l'UF come tale, oppure succedere ad essa, qualora il traduttore la individuasse ma non

ne conoscesse il significato e avesse bisogno di consultare un dizionario, e aggiunge la condizione in cui l'identificazione non avviene. Se parlando di identificazione ci si riferisce anche al riconoscimento del valore figurato di un *idiom*, inoltre, si può supporre che espressioni metaforiche altamente lessicalizzate possano essere interpretate in modo adeguato senza che vengano percepite come figurate.

Non potendo prevedere ordine e modalità in cui riconoscimento e interpretazione si verificano, sembra ragionevole, nell'ambito del presente lavoro, trattare tali processi generalmente insieme, considerando la traduzione come l'unione di una fase orientata alla comprensione del testo fonte e una alla produzione del testo meta, le quali, come constatato da vari autori, potrebbero talvolta anche svilupparsi in parallelo (cf. Sjørup 2013: 61, 96-97).

Goatly (1997), riferendosi a usi metaforici non per forza lessicalizzati, afferma che il loro riconoscimento potrebbe dipendere, oltre che da aspetti formali, dalla categoria grammaticale del dominio sfruttato per la concettualizzazione di una data realtà e da scelte linguistiche che esplicitano la presenza dell'uso figurato, quali il *come* nelle similitudini o le espressioni *in senso figurato* o *una specie di*, anche dalla distanza semantica tra i domini associati, nonché dalla contraddittorietà dell'immagine: ciò permette di ipotizzare che le varianti italiane proposte nel precedente paragrafo siano in buona parte identificabili come idiomatiche, in quanto, se intese in chiave letterale e fatte salve possibili occasioni in cui sia ammissibile anche un'interpretazione di questo tipo, descrivono scene inverosimili (*bersi/fumarsi il cervello*) e situazioni incoerenti con il contesto in cui compaiono (ad esempio, *essere fuori come un balcone*, *essere fuori di testa* e relative concettualizzazioni più specifiche).

Riguardo all'interpretazione delle UF con base metaforica, questa di norma non avviene sommando i significati dei suoi componenti, benché in certi contesti sia plausibile una lettura sia metaforica che letterale o sia possibile accedere al significato figurato relazionandolo con il senso letterale dell'espressione. A tal proposito, Vega Moreno (2005: 390, 394, 401) sostiene che non esiste una netta distinzione tra letteralità e metaforicità, bensì un continuum composto da gradi diversi di dipendenza tra queste due dimensioni, e reputa complicato stabilire il livello di opacità di un *idiom*, dal momento che la percezione di quanto il suo significato sia trasparente dipende, oltre che dai

valori semantici interni alla stringa, anche da fattori contestuali e dal sapere enciclopedico che ognuno possiede in merito alle parole che lo compongono. Tale percezione, inoltre, può essere diretta oppure retrospettiva, se deriva dalla previa conoscenza dell'accezione figurata (ibid.: 405; Keysar e Bly 1995, in Cacciari 2001: 315). Fermo restando che il riconoscimento delle condizioni di fissazione e idiomacità di un'espressione e la sua corretta interpretazione sono fortemente vincolati al contesto nel quale essa è inserita (vari autori, in Sjørup 2013: 31), alla competenza fraseologica del traduttore (Quiroga 2006: 144), alla composizione interna dell'espressione, nel caso di espressioni agrammaticali, e all'incoerenza tra l'eventuale significato letterale e il contesto in cui essa compare (Baker 1992: 65), si può pensare che anche la distanza tra i sistemi concettuali della lingua fonte e della lingua meta giochi un ruolo di primaria importanza, come confermato dallo studio di Mandelblit (1995) precedentemente illustrato. Da quest'ultima, di conseguenza, dipenderanno la traducibilità dell'UF e il suo grado di equivalenza con la soluzione traduttiva.

L'idea che si propone in questo frangente è che, a livello cognitivo, nell'individuare e nell'interpretare adeguatamente un'UF, il traduttore possa essere condizionato, oltre che da conoscenze personali, innanzitutto dall'esistenza, nella lingua meta, di sfruttamenti concettuali convenzionali per riferirsi a quel determinato dominio, in seconda istanza dalla tendenza di quest'ultima a sfruttare la stessa concettualizzazione generale e, infine, dal grado di specificità e dalla corrispondenza di immagine con la lingua fonte che essa è in grado di raggiungere.

Dall'analisi cognitiva delle variazioni fraseologiche condotta è stato osservato un comune ricorso dell'italiano a trasferimenti concettuali per riferirsi a uno stato di squilibrio mentale: se si considera la tendenza degli esseri umani ad associare concetti astratti a esperienze fisiche (cf. Lakoff e Johnson 1980), si può supporre che un parlante non nativo di italiano abbia buone possibilità di riconoscere e interpretare in maniera accettabile una delle UF esaminate anche senza averla mai sentita o letta in quanto sicuramente abituato a pensare alla pazzia in termini figurati; se tale ipotesi fosse valida, individui di madrelingua tedesca o spagnola che conoscono l'italiano potrebbero intuire il senso di espressioni come *essere fuori di testa*, *avere qualche rotella fuori posto* o *essersi bevuto il cervello* all'interno di un contesto appropriato senza esserne mai venuti a contatto prima, per il solo fatto che nelle loro lingue si tende ad

associare la pazzia a realtà più concrete, benché non sempre verosimili (l'atto del bersi il cervello è irrealizzabile, ma comunque appare più fisico rispetto a uno stato psicologico alterato). La probabilità che ciò avvenga cresce con l'aumentare della corrispondenza tra i trasferimenti concettuali prototipici disponibili nelle lingue in gioco. Kövecses (2010a: 666-667; 2010b: 199-204; 2010c: 735-736), a tal proposito, ipotizza l'esistenza di metafore universali derivanti da un'esperienza fisica condivisa dagli esseri umani usate per comprendere realtà intangibili, come ad esempio gli stati d'animo, le quali potrebbero manifestarsi attraverso espressioni linguistiche diverse da lingua a lingua, per cui queste condividerebbero la metafora primaria, più generale, per poi formare delle metafore più specifiche proprie.

Per fare qualche esempio, alcune UF tedesche rappresentanti modalità convenzionali di concettualizzare la pazzia riscontrate in varie fonti lessicografiche e relazionabili in qualche modo e più o meno direttamente con le concettualizzazioni italiane descritte sono: *nicht alle Daten im Speicher haben* ('non avere tutti i dati in memoria'), *nicht alle Tassen im Schrank haben* ('non avere tutte le tazze nella credenza'), *ein Rad abhaben* ('non avere su una ruota'), *eine Schraube locker haben* ('avere una vite traballante/allentata'), *bei jemandem ist ein Rad/Rädchen locker* ('qualcuno ha una ruota/rotella allentata'), *nicht alle Latten am Zaun haben* ('non avere tutte le assicelle al recinto'), *nicht mehr alle Nadeln an der Tanne haben* ('non avere più tutti gli aghi sull'abete'), *nicht mehr alle Kekse in der Dose haben* ('non avere più tutti i biscotti nella scatola'), *sie nicht (mehr) alle auf dem Christbaum haben* ('non averli/e più tutti/e sull'albero di Natale'), *sie nicht alle haben* ('non averli/e tutti/e'), *einen Dachschaden haben* ('avere un danno al tetto'), *nicht ganz dicht sein* ('non essere completamente impermeabile'), *nicht richtig ticken* ('non ticchettare correttamente'), *einen Sprung in der Schüssel/Tasse haben* ('avere una crepa nella ciotola/tazza'), *einen weichen Keks haben* ('avere un biscotto molle'), *einen Kopfschuss haben* ('avere un colpo di arma da fuoco in testa'). Per lo spagnolo, senza distinguere tra varianti peninsulari e americane, invece, si possono considerare: *estar mal de la cabeza* ('stare male di testa'), *estar mal del coco* ('stare male di cocco'), *estar mal de la olla* ('stare male di pentola'), *estar mal del tarro* ('stare male di barattolo'), *estar mal del quinto piso* ('stare male del quinto piano'), *estar mal de la azotea* ('stare male di terrazza'), *tener pájaros en la azotea* ('avere uccelli in terrazza'), *estar fuera de juicio* ('essere fuori di giudizio'), *estar alguien fuera*

de su acuerdo ('stare fuori dal proprio accordo'), *estar alguien fuera de sus cabales / no estar alguien en sus cabales* ('essere fuori dalla propria interezza / non essere nella propria interezza'), *estar más para allá que para acá* ('essere più di là che di qua'), *faltarle a alguien caramelos en el frasco* ('mancare caramelle dal vasetto a qualcuno'), *faltarle a alguien palos para el puente* ('mancare pali per il ponte a qualcuno'), *faltarle a alguien un tornillo a alguien* ('mancare una vite a qualcuno'), *aflojarsele a alguien los tornillos/un tornillo* ('allentarsi le viti/una vite a qualcuno'), *patinarle a alguien el embrague* ('slittare la frizione a qualcuno'), *tener seco el cráneo* ('avere il cranio secco'), *estar podrido de la cabeza* ('essere marcio di testa'), *estar loco de atar* ('essere matto da legare'), *estar alguien para encerrar / para que lo encierren* ('essere da rinchiudere').

In virtù di quanto sostenuto in precedenza, a prescindere dalla corrispondenza concettuale tra ognuna di queste espressioni istituzionalizzate e quelle italiane proposte, in nessun caso da intendersi come un elenco completo, si ipotizza che, essendo la mente dei madrelingua tedeschi e spagnoli già allenata a pensare metaforicamente alla pazzia, questi potrebbero essere facilitati nel cogliere la presenza e il senso delle UF italiane selezionate anche senza aver mai familiarizzato con esse, soprattutto se supportati dal contesto che le accoglie. Le possibilità di riconoscimento e corretta interpretazione aumentano probabilmente con l'avvicinarsi dei trasferimenti concettuali rappresentati dalla singola espressione italiana e quelle disponibili nelle lingue meta. Se un primo approccio alla comprensione e al trattamento traduttivo dell'UF dipende dalla presenza di usi metaforici per la pazzia in tedesco e in spagnolo, quindi, un secondo passo è dato dall'esistenza di concettualizzazioni generali comuni. Nello specifico, sono state individuate (i) corrispondenze tra italiano e spagnolo in quelle generiche A, D ed E, per cui si suppone che un madrelingua spagnolo abbia meno difficoltà nel riconoscere e interpretare espressioni come *essere fuori di testa*, *avere il cervello in pappa* ed *essere matto da legare* e simili, (ii) corrispondenze sia con lo spagnolo che con il tedesco nella concettualizzazione generica C, e (iii) nessuna corrispondenza per la concettualizzazione generica B.

Di seguito si ripropongono le variazioni cognitive generali e specifiche individuate e si presentano le relative corrispondenze concettuali interlinguistiche, indicando con (IT), (ES) e (DE) rispettivamente le varianti italiane, spagnole e tedesche:

<p>A) metafora direzionale: concettualizzazione riscontrata solo in spagnolo</p>
<p>A.1) metafora direzionale dentro-fuori:</p> <ul style="list-style-type: none"> a. (IT) <i>essere fuori di testa / essere uscito di testa / non starci con la testa / essere via di testa / essere fuori di capoccia</i> b. (IT) <i>essere fuori di mente</i> c. (IT) <i>essere fuori di cervello / essere via di cervello / essere uscito di cervello / non starci col cervello</i> d. (IT) <i>essere fuori di senno</i>: corrispondenza con (ES) <i>estar fuera de juicio</i> e. (IT) <i>essere fuori</i> f. (IT) <i>essere fuori di sé</i> g. (IT) <i>essere fuori come un balcone</i> h. (ES) <i>estar alguien fuera de su acuerdo</i> i. (ES) <i>estar alguien fuera de sus cabales / no estar alguien en sus cabales</i> <p>A.2) la persona pazzo è distante da chi parla:</p> <ul style="list-style-type: none"> a. (IT) <i>essere più di là che di qua</i>: corrispondenza con (ES) <i>estar más para allá que para acá</i>
<p>B) il cervello è collocato nella parte inferiore del corpo: nessuna corrispondenza riscontrata in tedesco e in spagnolo</p>
<p>C) condizione in cui la persona considerata pazzo difetta di o in qualcosa: corrispondenze in spagnolo e tedesco</p>
<ul style="list-style-type: none"> a. il cervello è un sistema malfunzionante di ingranaggi con componenti mancanti o difettosi: concettualizzazione riscontrata sia in tedesco che in spagnolo <ul style="list-style-type: none"> i. (IT) <i>mancare qualche rotella (a qualcuno)</i>: corrispondenze con (ES) <i>faltarle un tornillo a alguien</i> e (DE) <i>ein Rad abhaben</i> ii. (IT) <i>avere qualche rotella fuori posto / non avere tutte le rotelle a posto</i>: corrispondenze parziali con (ES) <i>aflojarse a alguien los tornillos/un tornillo</i>, (ES) <i>patinarle a alguien el embrague</i>, (DE) <i>eine Schraube locker haben</i>, (DE) <i>bei jemandem ist ein Rad/Rädchen locker</i> e (DE) <i>nicht richtig ticken</i>. b. il pazzo non è in possesso del proprio cervello e tale mancanza è associata ad un vizio: concettualizzazione non riscontrata in tedesco e in spagnolo c. (IT) <i>aver (mandato) il cervello in vacanza</i> d. (IT) <i>mancare un venerdì (a qualcuno)</i> e. (ES) <i>faltarle a alguien caramelos en el frasco</i> f. (ES) <i>faltarle a alguien palos para el puente</i> g. (ES) <i>tener seco el cráneo</i>: il cranio sembra essere asciutto perché manca il cervello al suo interno h. (DE) <i>nicht alle Daten im Speicher haben</i> i. (DE) <i>nicht alle Tassen im Schrank haben</i> j. (DE) <i>nicht alle Latten am Zaun haben</i> k. (DE) <i>nicht mehr alle Nadeln an der Tanne haben</i> l. (DE) <i>nicht mehr alle Kekse in der Dose haben</i> m. (DE) <i>sie nicht (mehr) alle auf dem Christbaum haben</i> n. (DE) <i>sie nicht alle haben</i> o. (DE) <i>einen Dachschaden haben</i> p. (DE) <i>nicht ganz dicht sein</i>: l'essere non del tutto impermeabile, presumibilmente riferito a un tetto, deriva da un difetto di quest'ultimo q. (DE) <i>einen Sprung in der Schüssel/Tasse haben</i> r. (DE) <i>einen weichen Keks haben</i> s. (DE) <i>einen Kopfschuss haben</i>: la testa non funziona in quanto ferita

D) il cervello è in uno stato di decomposizione: concettualizzazione riscontrata solo in spagnolo
a. (IT) <i>avere il cervello (andato) in pappa</i> : corrispondenza con (ES) <i>estar podrido de la cabeza</i>
E) la pazzia è descritta con riferimenti a trattamenti psichiatrici obsoleti: concettualizzazione riscontrata solo in spagnolo
a. (IT) <i>essere (matto/pazzo) da legare</i> : corrispondenza con (ES) <i>estar loco de atar</i>
b. (IT) <i>essere matto/pazzo da catena</i>
c. (ES) <i>estar alguien para encerrar / para que lo encierren</i>

Date le corrispondenze concettuali presentate, inoltre, se ne osservano anche altre incrociate, le quali, con certa probabilità, contribuiscono al riconoscimento e alla giusta interpretazione degli *idiom* italiani.

Per quanto concerne lo spagnolo, ad esempio, benché non siano stati rilevati meccanismi cognitivi analoghi a quelli caratterizzanti le variazioni fraseologiche dei punti i. e ii. relativi ad a. di A.1, la presenza di espressioni alternative appartenenti allo stesso trasferimento concettuale di *estar mal de la cabeza* come *estar mal del coco / del tarro / de la olla*, potrebbe facilitare un ispanofono nell'adeguata decodifica di UF quali *essere fuori di melone/cocomero/zucca/cocuzza*, per l'allusione al mondo vegetale, ed *essere fuori di coccia/brocca* per il rimando al contenitore, oltre che per l'associazione al trovarsi fuori trasmessa da *estar fuera de juicio* ed espressioni concettualmente affini. Si tenga inoltre conto del fatto che in spagnolo *calabaza* (zucca) può designare metaforicamente proprio la testa, il che rappresenta un'analogia concettuale con l'italiano che favorirebbe ulteriormente il processo interpretativo dell'essere fuori di zucca/cocuzza. In generale, la sola tendenza a intendere in termini figurati la testa osservata anche in *estar mal del quinto piso / de la azotea* e *tener pájaros en la azotea*, potrebbe contribuire a cogliere la metaforicità di alcune espressioni italiane.

Dicasi lo stesso per il tedesco, dove, nonostante non sia stata rilevata un'espressione istituzionalizzata contenente una concettualizzazione direzionale dentro-fuori, il fatto che il sostantivo *Kürbis* (zucca) possa indicare in senso figurato il capo, consente di non escludere la possibilità che un traduttore di madrelingua tedesca comprenda con successo l'UF *essere fuori di zucca/cocuzza* pur senza esserne mai entrato in contatto prima. Anche nello sfruttamento da parte degli italofoeni di contenitori di vario tipo per riferirsi alla testa, inoltre, si riscontra una certa vicinanza con la lingua tedesca, seppure

all'interno di trasferimenti generali distinti: *Schrank*, *Schüssel*, *Tasse* e *Speicher*, del resto, hanno proprio lo scopo di contenere cose più o meno tangibili. L'associazione della testa al biscotto (*einen weichen Keks haben*) è un'ulteriore prova dell'abitudine generalizzata alla sua metaforizzazione, mediante l'accostamento, ancora una volta, ad alimenti.

Non va tuttavia sottovalutato il rischio che talune corrispondenze incrociate possono comportare: da una parte, un'espressione come *essere fuori di zucca*, che combina concettualmente *estar fuera de juicio* e la comprensione metaforica della testa in termini di *calabaza*, potrebbe essere interpretata correttamente, ma riconosciuta come variazione fraseologica creativa, così come in tedesco potrebbe essere vista come una metafora originale contenente comuni trasferimenti di caratteristiche dal capo dell'essere umano all'ortaggio in questione; dall'altra, eventuali varianti linguisticamente libere e con tratti innovativi a un livello concettuale specifico, come le ipotetiche *avere una vite allentata* o *avere una vite traballante*, potrebbero essere identificate come lessicalizzate proprio per via di una corrispondenza concettuale completa con espressioni istituzionalizzate in spagnolo e in tedesco. In queste circostanze, ovviamente, sono fondamentali la competenza linguistica e fraseologica del traduttore e la consultazione di materiali di supporto.

Dagli esempi esaminati si evince che considerare la variazione fraseologica in un'ottica cognitiva permette di individuare diversi livelli di concettualizzazioni, dal più generale al più specifico, comprendendo da un lato i reciproci rapporti di dipendenza tra i trasferimenti metaforici contenuti in ogni singola espressione, e dall'altro le possibili motivazioni alla base di ogni variante e l'effetto evocativo provocato scegliendo l'una o l'altra. In una prospettiva traduttiva, poi, un confronto tra le concettualizzazioni comuni nella lingua fonte e quelle usuali nella lingua meta per riferirsi a una stessa realtà, può fornire degli indizi sulla possibilità che l'individuazione dell'UF e del suo valore metaforico e la sua corretta interpretazione da parte di lettori o traduttori non nativi si verifichino: tale possibilità potrebbe essere legata, oltre che a fattori personali, ad aspetti linguistici interni all'espressione e al contesto nel quale essa è inserita, in prima istanza alla tendenza, nell'altra lingua, a comprendere metaforicamente quella data realtà, poi dalla coincidenza di concettualizzazioni generali e, infine, dal grado di corrispondenza dei trasferimenti mentali più specifici.

4.2. Traduzione delle varianti fraseologiche

Si passa ora alla fase conclusiva del processo traduttivo, vale a dire la resa dell'UF nella lingua meta. Essa è legata da una parte all'esito del riconoscimento e dell'interpretazione, e dall'altra all'esistenza o all'assenza, nella lingua di arrivo, di *idiom* basati sugli stessi trasferimenti mentali, con un maggiore o minore sforzo cognitivo da parte del traduttore a seconda del grado di corrispondenza tra i sistemi concettuali disponibili nelle lingue in gioco (cf. Mandelblit 1995). Alcuni studiosi hanno ricondotto la traducibilità degli *idiom* anche alla presenza o meno di riferimenti culturali: Van Den Broeck (1981: 80-81), ad esempio, ne sottolinea il vincolo alla disponibilità, nella lingua meta, di tradizioni analoghe e a eventuali differenze nei codici estetici e morali tra le lingue e le culture a confronto, così come Steen (2014: 12, 23) osserva che il processo traduttivo è meno complesso se si realizza tra lingue vicine dal punto di vista genetico, storico e culturale, mentre è reso più difficoltoso qualora le culture e le lingue coinvolte attribuissero associazioni e valori distinti al dominio sfruttato per concettualizzare una data realtà.

Il concetto di equivalenza, parlando di espressioni metaforiche lessicalizzate, non assume contorni ben definiti in letteratura e varia in base all'ambito di studio: da un punto di vista traduttologico, ad esempio, Rabadán Álvarez (1991: 78, 284) sembra descrivere un'equivalenza di tipo funzionale-relazionale data non solo dalla corrispondenza semantica, ma anche dalla capacità di due espressioni di ricoprire la stessa funzione nel testo fonte e nel testo meta e di relazionarsi con i rispettivi lettori e sistemi culturali in maniera identica, e dal carattere dinamico, in quanto risponde a parametri distinti a seconda della situazione specifica; benché nel campo della linguistica contrastiva dedicata agli *idiom* si tenda invece a considerarla in termini decontestualizzati, senza tener conto della situazione comunicativa, distinguendo tra un'equivalenza totale, se due lingue presentano UF con struttura, valore semantico e contenuto uguali, parziale, se queste differiscono in qualcuno degli aspetti appena menzionati, e nulla, se non coincidono in nessuna delle loro caratteristiche, spesso le scelte traduttive sono invece dettate da fattori funzionali, contestuali e connotativi (cf. Zuluaga 1999: 543, 547-548), come prova il fatto che a volte, nonostante l'avvenuto riconoscimento e la corretta interpretazione di un *idiom*, il traduttore non fornisce quello che sembrerebbe

essere l'equivalente nella lingua meta, perché considerato poco adatto al contesto di arrivo (cf. Corpas 2003: 219-222, in Timofeeva 2012: 413). In virtù di ciò, la traduttologia fraseologica funzionale punta a cercare equivalenze fraseologiche prendendo in esame proprio i contesti d'uso e gli aspetti connotativi, sociolinguistici e culturali delle UF (Timofeeva 2012: 414).

Pur riconoscendo l'importanza delle questioni considerate nell'ambito della traduttologia fraseologica funzionale, l'equivalenza su cui ci si concentrerà in questo frangente è quella cognitiva. Si parte dalla supposizione che i trasferimenti mentali evocati da una determinata UF abbiano un certo peso sul sistema concettuale di chi la sfrutta o di chi la ascolta o legge e che, pertanto, sia importante quanto meno cercare di mantenerli nei limiti del possibile inalterati all'interno del testo meta, conciliando comunque sempre tale obiettivo con le esigenze pragmatico-contestuali: tener conto dei meccanismi cognitivi alla base degli *idiom* nella lingua fonte potrebbe aiutare proprio a trovare una soluzione concettualmente adeguata al testo e alla cultura meta, intendendo la ricerca della massima equivalenza cognitiva come un puntare a raggiungere il maggior livello di specificità dei trasferimenti attivati nel pensiero.

Partendo dal presupposto ideale che, nell'affrontare quest'ultima fase, il traduttore abbia individuato e interpretato correttamente l'UF, si può ipotizzare che, in presenza di varianti cognitive oltre che linguistiche, il grado di equivalenza concettuale con la proposta traduttiva dipenda dal grado di somiglianza di quelli che sono stati presentati come sistemi di *matrioske* caratterizzanti i meccanismi di comprensione delle lingue coinvolte per quella determinata realtà. Se si prendono in esame le varianti italiane analizzate, i passi che un traduttore potrebbe seguire per offrire l'UF concettualmente più vicina nella lingua meta sono: 1) esprimere lo stesso significato mediante trasferimenti concettuali convenzionali; 2) esprimere lo stesso significato mediante il medesimo trasferimento concettuale generale; 3) esprimere lo stesso significato mediante il trasferimento concettuale più specifico possibile. Tale gerarchia, nel confronto tra italiano, tedesco e spagnolo, è già stata illustrata all'interno nel paragrafo 4.1., ma si ritiene utile completare il punto a. relativo ad A.1 con quanto segue:

A) metafora direzionale: concettualizzazione riscontrata solo in spagnolo

A.1) metafora direzionale dentro-fuori:

- a. *essere fuori di testa / essere uscito di testa / non starci con la testa / essere via di testa / essere fuori di capoccia*: nessuna corrispondenza completa
- i. testa concettualizzata in termini di ortaggio o frutto con (IT) *essere fuori di melone/cocomero* e (IT) *essere fuori di zucca/cocuzza*: corrispondenza riscontrata in questa concettualizzazione specifica ma oltre i limiti della concettualizzazione generica con (ES) *estar mal del coco*
- ii. testa concettualizzata in termini di contenitore con (IT) *essere fuori di coccia* e (IT) *essere fuori di brocca*: corrispondenza riscontrata in questa concettualizzazione specifica ma oltre i limiti della concettualizzazione generica con (ES) *estar mal del tarro*, (ES) *estar mal de la olla*, e con (DE) *nicht alle Tassen im Schrank haben*, (DE) *einen Sprung in der Schüssel/Tasse haben* e (DE) *nicht alle Daten im Speicher haben*

Lo scenario presentato mette in luce delle corrispondenze incrociate o parziali: nell'ottica di una ricerca di equivalenza cognitiva, nel momento in cui un traduttore, lavorando sulla combinazione linguistica italiano-spagnolo, si ritrova a dover tradurre espressioni come *essere fuori di zucca* o *essere fuori di brocca*, dopo averle individuate e interpretate adeguatamente, se lo spagnolo dispone di UF per riferirsi alla pazzia, potrebbe mirare a trovare delle soluzioni capaci di ricoprire il maggior numero di trasferimenti concettuali possibile presenti nelle UF di partenza, ovviamente cercando di non alterarne le dimensioni pragmatico-contestuali. Procedendo per gradi, vista la tendenza dello spagnolo a intendere la pazzia attraverso meccanismi di comprensione metaforica che si manifestano in espressioni lessicalizzate, l'accesso a quello che è stato descritto come un sistema di matrioske non è ostacolato. Selezionando la concettualizzazione generale direzionale dentro-fuori, si osserva che ad essa corrisponderebbero UF come *estar alguien fuera de juicio / su acuerdo / sus cabales*, *no estar alguien en sus cabales* o *estar más para allá que para acá*, ma, procedendo sul trasferimento concettuale più specifico relativo all'associazione della testa al regno vegetale o al contenitore, si incontra una barriera, non essendo state identificate espressioni che includano sia la metafora direzionale che quella ontologica. Tuttavia, il traduttore potrebbe scegliere se dare più peso alla metafora generale, proponendo una di quelle direzionali menzionate, oppure a quella più specifica, ad esempio con *estar mal del coco* in un caso ed *estar mal del tarro / de la olla* nell'altro, cercando, al contempo, un equilibrio tra dimensione concettuale ed esigenze pragmatico-contestuali. La scelta potrebbe anche essere motivata dall'eventuale

reinserimento di un termine contenuto nell'UF all'interno del discorso al fine di creare ulteriori immagini. Con il tedesco, invece, l'ostacolo compare appena superata la prima fase: nessuna delle concettualizzazioni riferite alla pazzia individuate, infatti, sembra avere carattere direzionale. Il traduttore, quindi, può scegliere l'UF che ritiene più adeguata per registro, contesti e frequenza d'uso, o, nel caso di *essere fuori di coccia/brocca* mantenere solo l'associazione testa-contenitore mediante *nicht alle Tassen im Schrank haben, einen Sprung in der Schüssel/Tasse haben e nicht alle Daten im Speicher haben*.

5. Conclusioni

Le conclusioni alle quali ha condotto lo studio presentato riguardano essenzialmente due aspetti: la possibilità di trattare la variazione fraseologica in un'ottica cognitiva e l'utilità che un approccio cognitivo alla variazione fraseologica può avere in ambito traduttivo.

È stato scelto il dominio della pazzia per poter identificare un numero consistente di UF italiane tra loro sinonimiche e stabilire se e in che modo sia possibile classificarle concettualmente, esaminandone i reciproci rapporti cognitivi, ed è stata constatata l'esistenza di più livelli di variazione concettuale. Ciò consente di immaginare quest'ultima come un peculiare sistema di matrioske: il grado più generale di variazione riguarda la presenza stessa di più matrioske, ossia di concettualizzazioni distinte per comprendere un solo dominio; i gradi successivi si riferiscono alla concettualizzazione rappresentata da ognuna delle madri, alla facoltà che ciascuna bambolina ha di contenere due o più pezzi di pari grandezza qualora esistessero alternative concettuali per uno stesso livello, e alla possibilità di essi di ospitare altre bambole più piccole; ogni pezzo contenuto in uno più grande costituisce la concettualizzazione sempre più specifica, finché non si raggiunge il seme, cioè il massimo grado di specificità. Il numero di bambole di cui è composta ciascuna matrioska varia a seconda della quantità e della natura concettuale delle UF alternative.

L'esame cognitivo della variazione fraseologica sviluppato per l'italiano è stato poi rapportato all'universo traduttivo, prendendo in considerazione lo spagnolo e il tedesco come ipotetiche lingue meta. In base alle corrispondenze interlinguistiche concettuali individuate, è stato ipotizzato che l'origine cognitiva delle varianti italiane possa influenzare il riconoscimento delle stesse,

la loro corretta interpretazione e lo sforzo mentale messo in campo durante questi due processi: la consuetudine, in tedesco e in spagnolo, di comprendere la pazzia mediante trasferimenti figurati potrebbe indurre il traduttore (ma anche solo il lettore) di madrelingua spagnola o tedesca a riconoscere l'UF e coglierne il significato, proprio perché già abituato a intendere metaforicamente tale dominio; la possibilità di successo aumenta se nella sua lingua esistono concettualizzazioni analoghe a quella più generica alla quale l'UF italiana appartiene e cresce con il loro avvicinarsi al più alto livello di specificità da essa espresso, e si suppone che sia inversamente proporzionale allo sforzo cognitivo richiesto.

Il processo traduttivo e l'impegno mentale necessario nella comprensione del testo fonte e nella produzione del testo meta, pertanto, dipenderanno, oltre che da fattori personali quali la conoscenza fraseologica del traduttore nelle due lingue in gioco, contestuali o linguistici, relativi alla composizione interna dell'UF, dall'affinità concettuale tra lingue fonte e meta nella comprensione del dominio in questione. Questa inizialmente guida verso l'individuazione dell'UF e la sua adeguata interpretazione, e alla fine determina la soluzione traduttiva più consona da un punto di vista cognitivo, la quale dovrebbe mirare a ricoprire il maggior numero di concettualizzazioni rappresentate dall'espressione italiana in base alle possibilità a disposizione nella lingua di arrivo.

Con ciò non si intende sminuire l'importanza delle componenti pragmatiche e contestuali, ma, piuttosto, motivare alla ricerca di un equilibrio tra queste ultime e il mantenimento della natura concettuale originaria.

Riferimenti bibliografici

- ALBALADEJO MARTÍNEZ, Juan Antonio. (2016) "Tratamiento lexicográfico de los fraseologismos nacional-regiolectales austriacos." In: Mogorrón Huerta, Pedro; Analia Cuadrado Rey; María Lucía Navarro Brotons & Iván Martínez Blasco (eds.) 2016. *Fraseología, variación y traducción*. New York/Francoforte: Peter Lang, pp. 175-187.
- ARRIBAS, Nieve. (2006) "Consideraciones metalexiconográficas sobre fraseología y lexicografía italo-españolas." *Quaderni del CIRSIL* 5, pp. 153-184.
- BAKER, Mona. (1992) *In Other Words: A Coursebook on Translation*. New York: Routledge.

- BLACK, Max. (1962) *Models and metaphors: Studies in language and philosophy*. Ithaca: Cornell University Press.
- CACCIARI, Cristina. (2001) *Psicologia del linguaggio*. Bologna: Il Mulino.
- CASTILLO CARBALLO, María Auxiliadora. (1997-1998) “El concepto de unidad fraseológica.” *Revista de Lexicografía* 4, pp. 67-79.
- CORPAS PASTOR, Gloria & Florentina Mena Martínez. (2003) “Aproximación a la variabilidad fraseológica de las lenguas alemana, inglesa y española.” *ELUA. Estudios de Lingüística* 17, pp. 181-201.
- CORPAS PASTOR, Gloria. (1996) *Manual de fraseología española*. Madrid: Gredos.
- CORPAS PASTOR, Gloria. (2001) “Corrientes actuales de la investigación fraseológica en Europa.” *EUSKERA* 46:1, pp. 21-49.
- CORPAS PASTOR, Gloria. (2003) *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. Madrid: Vervuert.
- DOBROVOL'SKIJ, Dmitrij. (2016) “Conceptual structure and systematic variation of Russian idioms.” In: Dal Maso, Elena & Carmen Navarro (eds.) 2016. *Gutta cavat lapidem Indagini fraseologiche*. Mantova: Universitas Studiorum, pp. 27-37.
- GARCÍA PAGE, Mario. (2016) “Variantes léxicas correlativas.” In: Dal Maso, Elena & Carmen Navarro (eds.) 2016. *Gutta cavat lapidem Indagini fraseologiche*. Mantova: Universitas Studiorum, pp. 39-54.
- GOATLY, Andrew. (1997) *The language of metaphors*. London: Routledge.
- GONZÁLEZ REY, María Isabel. (1998) “Estudio de la idiomaticidad en las unidades fraseológicas.” In: Wotjak, Gerd (ed.) 1998. *Estudios de fraseología y fraseografía del español*. Frankfurt am Main/Madrid: Vervuert/Iberoamericana, pp. 57-73.
- KEYSAR, Boaz & Benjamin Bly. (1995) “Intuitions of the transparency of idioms: Can you keep a secret by spilling the beans?” *Journal of Memory and Language* 34, pp. 89-109.
- KÖVECSES, Zoltan. (1986) *Metaphors of anger, pride, and love: a lexical approach to the structure of concepts*. Amsterdam: John Benjamins.
- KÖVECSES, Zoltan. (2002) *Metaphor. A practical Introduction*. Oxford: Oxford University Press.
- KÖVECSES, Zoltan. (2005) *Metaphor in culture: universality and variation*. Cambridge: Cambridge University Press.
- KÖVECSES, Zoltan. (2010a) “A new look at metaphorical creativity in cognitive linguistics.” *Cognitive Linguistics* 21:4, pp. 663-697.

- KÖVECSE, Zoltan. (2010b) "Metaphor and Culture." *Acta Universitatis Sapientiae, Philologica* 2:2, pp. 197-220.
- KÖVECSE, Zoltan. (2010c) "Metaphor, Creativity, and Discourse." *DELTA* 26, pp. 719-738.
- LAKOFF, George & Mark Johnson. (1980) *Metaphors we live by*. Chicago/London: The University of Chicago Press.
- LANGLOTZ, Andreas. (2006) *Idiomatic Creativity. A cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- MANDELBLIT, Nili. (1995) "The cognitive view of metaphor and its implications for translation theory." In: Thelen, Marcel & Barbara Lewandowska-Tomaszczyk (eds.) 1996. *TRANSLATION AND MEANING, Part 3*. Maastricht: Hogeschool Maastricht, School of Translation and Interpreting, pp. 483-495.
- MONTORO DEL ARCO, Esteban Tomás. (2004) "La variación fraseológica y el diccionario." In: Battaner, Paz & DeCesaris, Janet Ann (eds.) 2004. *De lexicografía: actes del I Symposium Internacional de Lexicografía*. Barcelona: Institut Universitari de Lingüística Aplicada, pp. 591-604.
- QUIROGA, Paula. (2006) *Fraseología italo-española: Aspectos de lingüística aplicada y contrastiva*. Granada: Granada Lingvistica.
- RICHARDS, Ivor Armstrong. (1936) *The Philosophy of Rhetoric*. New York/London: Oxford University Press.
- SJØRUP, Annette Camille. (2013) *Cognitive effort in metaphor translation. An eye-tracking and key-logging study*. Copenhagen: LIMAC PhD School. Tesi di dottorato.
- STEEN, Gerard. (2014) "Translating metaphor: What's the problem?" In: Monti, Enrico & Donna Rose Miller (eds.) 2014. *Tradurre figure. Translating figurative language. Quaderni del CeSLiC. Atti di Convegni (3)*. Bologna: Centro di Studi Linguistico-Culturali (CeSLiC), pp. 11-24.
- TIMOFEEVA, Larissa. (2012) "Sobre la traducción fraseológica." *ELUA* 26, pp. 405-432.
- VEGA MORENO, Rosa. (2005) "Idioms, Transparency and Pragmatic Inference." *UCL Working Papers in Linguistics* 17, pp. 389-426.
- ZULUAGA, Alberto. (1980) *Introducción al estudio de las expresiones fijas*. Francoforte/Berna: Peter Lang.
- ZULUAGA, Alberto. (1999) "Traductología y Fraseología." *Paremia* 8:5, pp. 537-549.

NOTA BIOGRAFICA / BIONOTE

SILVIA CATALDO ha conseguito il titolo di dottorato in Traduttologia, Traduzione Professionale e Audiovisiva presso l'Università di Alicante e la laurea magistrale in Traduzione Specializzata presso l'Università di Bologna. Attualmente lavora come docente di scuola secondaria in Italia insegnando spagnolo e tedesco. I suoi principali interessi di ricerca sono legati ai linguaggi metaforici in italiano, spagnolo e tedesco e alla loro analisi in chiave contrastiva e traduttiva.

SILVIA CATALDO holds a PhD in Translation Studies, Professional and Audiovisual Translation from the University of Alicante and a Master's Degree in Specialized Translation from the University of Bologna. She currently works as Spanish and German teacher at secondary schools in Italy. Her main research interests encompass metaphorical languages in Italian, Spanish and German and their contrastive and translational analysis.

Recibido / Received: 30/10/2019
Aceptado / Accepted: 28/02/2020

Para enlazar con este artículo / To link to this article:
<http://dx.doi.org/10.6035/MonTI.2020.ne6.3>

Para citar este artículo / To cite this article:

Koza, Walter & Natalia Rivas Folch. (2020) "Modelización computacional de las locuciones verbales del español chileno a partir de la propuesta de la léxico-gramática." En: 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. 94-120.

MODELIZACIÓN COMPUTACIONAL DE LAS LOCUCIONES VERBALES DEL ESPAÑOL CHILENO A PARTIR DE LA PROPUESTA DE LA LÉXICO-GRAMÁTICA

COMPUTATIONAL MODELIZATION OF VERBAL IDIOMS OF CHILEAN SPANISH BASED-ON LEXICON-GRAMMAR PROPOSAL

WALTER KOZA

walter.koza@pucv.cl

Pontificia Universidad Católica de Valparaíso

Proyecto FONDECyT 1171033

NATALIA RIVAS FOLCH

natalia.rivas.f@gmail.com

Pontificia Universidad Católica de Valparaíso

Proyecto FONDECyT 1171033

Resumen

En el presente trabajo, se propone un estudio formal de las locuciones verbales (LV) pertenecientes al español de Chile con el propósito de diseñar un algoritmo computacional para el análisis automático. Para ello, se describen las LLVV incluidas en el *Diccionario de uso del español de Chile*, según los criterios de la léxico-gramática. Posteriormente, se procede a la modelización computacional recurriendo a NooJ, un programa que cuenta con diversas utilidades para el análisis lingüístico. Este método fue evaluado mediante el reconocimiento automático en un corpus compuesto por textos de prensa chilena. Se obtuvo 96,4% de precisión, 92,23% de cobertura y 94,53%



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

de medida F. Los resultados demuestran que el algoritmo diseñado es adecuado para el análisis automático de las LLVV.

Palabras clave: Locuciones verbales; Español chileno; Análisis automático; léxico-gramática; NooJ.

Abstract

In this work, a formal study for Verbal Idioms (VI) of Chilean Spanish is proposed with the objective of developing a computational algorithm for the automatic analysis of texts. For this, the VIS included in *Diccionario de uso del español de Chile* were described according to the lexicon-grammar theory. Then, a computational modelization was made using NooJ, a program that has several utilities. This method was evaluated with an automatic recognition in a corpus composed by texts from Chilean press. 96.4% of precision, 92.23% of coverage and 94.53% of F-measure were obtained. The results show that the algorithm designed is adequated for the automatic analysis of the VI.

Keywords: Verbal idioms; Chilean Spanish; Automatic Analysis; Lexicon-grammar; NooJ.

1. Introducción

Uno de los problemas más relevantes que ha venido enfrentando la lingüística computacional ha sido establecer un modo efectivo para el análisis de las locuciones (Vietri 2014, Silberztein 2016). Las complejidades que implican estas expresiones para el análisis automático tienen que ver, fundamentalmente, con el grado de fijación, pues, mientras algunas locuciones no permiten separar sus elementos, otras sí lo hacen, como en (1). A su vez, otro inconveniente se deriva de casos en los que es posible modificar uno de sus miembros (2). Y ya en el caso de las locuciones verbales, un desafío lo constituye el hecho de que el verbo puede adquirir condiciones sintácticas que no posee cuando aparece de manera libre (3):

- (1) Juan **agarra para el hueveo** a María. --> Juan **agarra** a María **para el hueveo**.
- (2) Juan **dejó la escoba** --> Juan **dejó la mansa escoba**.
- (3) Juan tiene una casa / *La casa es tenida por Juan. --> Juan **tiene en cuenta** a María. / María **es tenida en cuenta** por Juan.

A tales efectos, en el presente trabajo se propone un análisis de las locuciones verbales (LLVV) pertenecientes al español de Chile con el propósito de diseñar un algoritmo computacional para el análisis automático de textos del lenguaje natural. Con este objetivo, se busca realizar una descripción exhaustiva del comportamiento morfosintáctico que poseen estos elementos tomando como fundamentos teóricos y metodológicos, los planteamientos de la léxico-gramática (Gross 1975; 1984). Para ello, se analizan las LLVV listadas en el *Diccionario de uso del español de Chile* (DUECh) (Academia Chilena de la Lengua 2010), categorizándolas en diversas clases, según su estructura y en relación con la exigencia de argumentos y posibilidades transformacionales.

La fase anteriormente descrita es modelizada computacionalmente. Esta labor se realiza con el software de acceso libre NooJ, desarrollado por Silberstein (2003; 2016). Este programa presenta una estructura basada en la jerarquía de Chomsky-Schützenberger (Chomsky 1957) y cuenta con diversas utilidades para el análisis lingüístico. El trabajo computacional implicó la construcción de un diccionario electrónico de LLVV y gramáticas sintácticas. Este método fue evaluado mediante el reconocimiento automático en un corpus compuesto por textos de prensa chilena, logrando 96,4% de precisión, 92,23% de cobertura y 94,53% de medida F. Estos resultados demuestran que el algoritmo diseñado es adecuado para el análisis automático de las LLVV.

2. Marco teórico

En este apartado se presenta el marco teórico que sustenta la investigación. Se incluyen, por un lado, los trabajos referentes a la naturaleza de la locución y, por otro lado, la propuesta de la léxico-gramática (Gross 1975; 1984). Dentro de esta última se hace mención al trabajo de Vietri (2014), quien lleva a cabo un análisis de las LLVV del italiano, el cual, al igual que en este estudio, implica la creación de un algoritmo computacional en NooJ.

2.1. Respecto al concepto de locución

En el sistema de la lengua es posible encontrar, además de combinaciones libres de palabras, ciertas estructuras fijas que poseen mayor estabilidad. Se trata de estructuras que, según Corpas Pastor (1996: 18):

- i. Constan de al menos dos palabras ortográficas.
- ii. Presentan cierto grado de lexicalización.
- iii. Se caracterizan por la alta frecuencia de coaparición en la lengua.

Estas construcciones son denominadas por la autora como *unidades fraseológicas* (UF) y cuentan con una tradición de estudios en la lingüística. Dentro de este tipo de expresiones, se incluyen las locuciones, entendidas, según la propuesta de Casares (1950) reformulada por Penadés (2012), como “la combinación fija de palabras que funciona como elemento de la oración y cuyo significado no se corresponde con la suma de los significados de sus componentes” (Penadés 2012:25).

El criterio de clasificación de las locuciones más utilizado remite a la categoría gramatical que se identifica según su función sintáctica. De este modo, se establecen locuciones nominales, adjetivales, preposicionales, adverbiales y verbales (Zuluaga 1975; Corpas Pastor 1996; Penadés 2012). Las LLVV contienen un verbo, el cual forma un predicado complejo a partir de la asimilación de un complemento o adjunto. Esto implica que: (i) van a exigir argumentos; (ii) están supeditadas a las restricciones de concordancia mediante la flexión; (iii) pueden designar estados y eventos (Silvagni 2017); (iv) pueden ser modificadas por un adjunto.

Ahora bien, en relación con el ítem (i), para la identificación de argumentos exigidos por las LLVV, es necesario tener en cuenta las restricciones que la locución le impone a sus argumentos. Así, por ejemplo, ‘calentar’ en ‘calentar el asiento’ solo exige un único argumento que debe tener el rasgo +humano, dado que presenta a un complemento directo cristalizado:

- (4) Juan calentó el asiento.
*El sol calentó el asiento.

Asimismo, también es pertinente observar las posibilidades transformacionales de las LLVV, como la voz pasiva, la nominalización, etcétera.

- (5) Juan agarra a María para el hueveo = María fue agarrada para el hueveo por Juan.
Pedro agachó el moño = Pedro tuvo una agachada de moño.

Para una descripción exhaustiva tanto de la naturaleza de los argumentos como de las posibilidades transformacionales de las LLVV, se recurrió al marco de la léxico-gramática, cuyos principales postulados se presentan en la sección siguiente.

2.2 La léxico-gramática

La léxico-gramática (Gross 1975; 1984) se presenta como un método de descripción de las lenguas naturales basado en tres aspectos principales: (i) la sintaxis es indisoluble del lexicón; (ii) la unidad mínima de análisis es la oración simple (a diferencia de palabras o sintagmas), y (iii) el marco debe ofrecer una formalización y un método descriptivo aplicable a cualquier lengua (Elia, Monteleone & Marano 2011). Este modelo plantea un estudio combinado de las reglas sintácticas y preferencias de selección léxica en relación con las posibilidades de transformación que tiene una oración determinada.

Para poder llegar a tal descripción, Gross (1975) propone la elaboración de tablas organizadas en clases, las cuales reúnen elementos léxicos de una categoría gramatical que comparten propiedades (Tolone 2012). Una tabla se presenta sobre la forma de una matriz, la cual contiene: (i) en línea, los elementos de la clase correspondiente; (ii) en columna, las propiedades sintáctico-semánticas (que no son necesariamente aceptadas por todos los miembros de la clase), y (iii) al cruce de una línea y de una columna, el signo + o -, en función de si la entrada léxica descrita por la línea acepta o no la propiedad descrita por la columna. Una propiedad sintáctico-semántica es una información que se refiere directamente a la construcción de base asociada a la clase, a una transformación de la construcción de base o a una construcción adicional. Si una unidad léxica tiene dos sentidos distintos, entonces presentará dos entradas. A modo de ejemplo, en la tabla 1 se muestra un fragmento de una de las tablas de verbos del francés adaptadas por Pivaut (1989), presentada en Tolone (2012).

	N0=: Nhum	N0=: Nhum	N0=: Nnr	<ENT> Ppv	Ppv =: se figé	Ppv =: les figé	Ppv =: Neg	<ENT>V	Neg	N0V	N0 être. Vant	N0V de N0pc	N1=: Nhum	N1=: Nhum	N1=: le fait. Ou P	Ppv =: lui	Ppv =: y	[extrap]	N0idée V Loc N1esprit	<OPT>
+	-	-	-	les	-	+	-	lâcher Advm	-	+	-	-	-	-	-	-	-	-	-	Max les lâche difficilement à Ida
+	-	-	-	<É>	-	-	-	renaitre	-	+	-	-	-	-	-	-	-	-	-	Max renait au bonheur de vivre
+	-	-	-	se	+	-	-	rendre	-	+	-	-	-	-	+	-	-	-	-	Max s'est rendu à mon opinion
+	-	-	-	se	+	-	-	rendre	-	+	-	-	-	-	-	-	-	-	-	Le caporal s'est rendu à l'ennemi
+	-	-	-	<É>	-	-	-	renoncer	-	-	-	-	-	-	-	-	-	-	-	Max renonce à son héritage
+	+	+	+	ne	-	-	+	revenir	+	-	-	-	-	-	-	+	-	-	-	La tête de Luc ne revient pas à Max

Tabla 1. Ejemplo de tabla Léxico-Gramática (Tolone 2012)

Como se puede observar, un verbo como *rendre* ('rendir') va a tener doble entrada según los tipos de objetos que exija (N0=Humano, N1=No Humano: 'Max se rindió a mi opinión'; N0=Humano, N1=Humano: 'El cabo se rindió al enemigo').

Posteriormente, se identifican las transformaciones que admite cada oración ilustrada en la última columna (<OPT>). La noción de transformación es propuesta por Harris (1970) y alude a las diversas posibilidades en las que puede proyectarse una estructura argumental. Esta idea no debe confundirse con la que se plantea en la gramática generativa transformacional, puesto que desde la perspectiva harrisiana, todas las transformaciones tendrían lugar en la estructura superficial. En el ejemplo presentado se puede apreciar que las dos oraciones admiten la nominalización ('Tuvo lugar la rendición de Max ante mi opinión', 'Tuvo lugar la rendición del cabo ante el enemigo').

En relación con las expresiones idiomáticas, Gross (1975) señala que esta metodología puede extenderse a este tipo de expresiones. En relación con ello, Vietri (2014) realiza un análisis de las LLVV del italiano a partir de este aparato teórico-metodológico. Para llevar a cabo sus descripciones, Vietri (2014) parte de la introspección y juicios lingüísticos, corroborando luego estos dos criterios con información proveniente de Internet. Una vez realizada la descripción de las unidades, la autora da cuenta de las posibilidades que introduce la léxico-gramática para poder realizar una formalización de las expresiones idiomáticas descritas. A tales efectos, emplea el software NooJ y crea gramáticas que le permiten detectar las propiedades de las unidades, con las respectivas variaciones que estas admiten.

En este sentido, algunas de las conclusiones a las que llega Vietri (2014) son: (i) las LLVV se rigen por las mismas reglas sintácticas que los verbos predicativos; (ii) respecto de las posibilidades de descomposición de una locución, esto debe verse caso por caso; (iii) la representación semántica de las expresiones idiomáticas se encuentra directamente ligada con su sintaxis; (iv) el uso de la léxico-gramática constituye un marco adecuado para realizar una descripción exhaustiva de unidades como las estudiadas considerando la posterior formalización de reglas para la modelización computacional.

Aquí se toma como base la propuesta de esta autora para las LLVV del español chileno. Al igual que aquella, se realiza una descripción morfosintáctica de estas unidades, incluyendo sus posibilidades combinatorias a través de la estructura argumental que proyectan. Asimismo, se considerarán sus posibilidades transformacionales (nominalización, voz pasiva, etcétera). Posteriormente, se propondrá una formalización y una modelización computacional para la detección automática en textos de lenguaje natural.

3. Metodología

En este apartado se presentan los lineamientos metodológicos que sustentan esta investigación. Se compone de dos secciones, la primera de ellas describe la naturaleza de las LLVV y la segunda, el trabajo computacional.

3.1. Descripción de las LLVV

La descripción se realizó sobre una lista de LLVV extraída del DUECh (Academia Chilena de la Lengua 2010). Este diccionario concentra una serie de entradas léxicas o lemas que remiten a expresiones usadas en el habla de Chile. De acuerdo con Matus, director del diccionario, la obra se propone reflejar el uso corriente, socialmente estabilizado del habla de dicho país (Sáez 2012). En este sentido, primó un criterio lexicográfico en el que se concibió como LV a toda expresión indicada de esa manera en el DUECh, más allá de ciertas omisiones señaladas por Sáez (2012), como el hecho de ser omisiones. Ahora bien, el presente trabajo pretende dar cuenta de un método para describir morfosintácticamente cualquier expresión fija conformada por un verbo más argumentos o adjuntos cristalizados, por lo cual se podría aplicar a nuevas unidades léxicas de este tipo que se detecten o se creen en el futuro.

El primer paso fue la extracción de todas aquellas entradas que pertenecen a la categoría de “locución verbal”, y se recopiló un total de 833 expresiones de este tipo. Posteriormente se las clasificó en dos grandes grupos; el primero incluye a las LLVV que no expresaban negación (‘agachar el moño’) y el segundo, las que sí lo hacen (‘no ver una’).

	Total lemas	Porcentaje
Locuciones Grupo 1	792	95,1 %
Locuciones Grupo 2	41	4,9 %
Locuciones totales (G1 + G2)	833	100 %

Tabla 2. Resumen de locuciones extraídas del DUECh

En la tabla 2 se puede apreciar un resumen del total de locuciones extraídas del DUECh. En el presente trabajo se realizó una descripción del primer grupo, quedando para futuras investigaciones el segundo.

Posterior a la recopilación se llevó a cabo una categorización y descripción de estas estructuras según los criterios que plantea la léxico-gramática (Gross 1975; 1984). A continuación, se detalla el procedimiento.

3.1.1. Análisis distribucional

En una primera instancia se tomó en consideración el modo en el que se estructura el lema en el diccionario, esto es, se observaron los elementos constitutivos de la locución. De este modo, expresiones como ‘dejar la escoba’ y ‘sacarse el pillo’ fueron agrupadas en clases diferentes, dado que una requiere de clítico y la otra no. Sin embargo, al realizar las tablas se pudo observar que ciertas locuciones poseen requerimientos argumentales diferentes al contemplado inicialmente. Así, aquellas locuciones declaradas en el DUECh con pronombre enclítico, como por ejemplo, “meterse en un queso” podían emplearse sin él en casos como (4):

(4) Juan *metió en un queso* a María.

De la misma manera, en locuciones cuyo lema aparece sin el pronombre, como en “sacar en cara”, su uso sí lo exige:

(5) María *te sacó en cara* todo lo que había hecho por ti.

Este análisis arrojó las siguientes clases:

Clase	Ejemplo de lema	Ejemplo en uso	Total	%
Cl 1: (N0)+V+Cl	echar al agua.	María echó al agua a Juan.	405	51,14
Cl 2: (N0)+CL+V+Cl	arrancarse con los tarros.	Juan se arrancó con los tarros.	287	36,24
Cl 3: (N0)+CL+CL+V+Cl	sabérselas por libro.	Ella jura que se las sabe por libro.	6	0,76
Cl 4: (a)+(N0)+CLdat+V	crujirle a alguien.	A Juan le crujió por fin.	6	0,76
Cl 5: (a)+(N0)+CLdat+V+Cl	faltarle a alguien chauchas para el peso.	A María le faltan chauchas para el peso.	35	4,42
Cl 6: (a)+(N0)+se+CLdat+V+Cl	aconchársele a alguien los meados.	A Juan se le aconcharon los meados, por eso no reacciona.	40	5,05
Cl 7: (a)+(N0)+se+CLdat+V	chupársele a alguien.	A María se le chupa ir a hablarle.	3	0,38
Cl 8: (a)+(N0masc)+CLdat+V+Cl	gustarle a un hombre las patitas de chanco.	Está claro que a Juan le gustan las patitas de chanco.	2	0,25
Cl 9: (a)+(N0masc)+se+CLdat+V+Cl	quemársele a un hombre el arroz.	María no sabía que a Juan se le quemaba el arroz.	7	0,88
Cl 10: (a)+(N0fem)+CL+V+Cl	dejar a una mujer el tren.	A María la dejó el tren de tanto esperar a Juan.	1	0,13

Tabla 3. Resumen de categorización de locuciones pertenecientes al primer grupo.

En la tabla anterior, es posible observar las diversas categorías que emergieron a partir de los diversos requerimientos argumentales que poseen las

locuciones. En cada una de las clases se da cuenta de la organización prototípica de la expresión al emplearse. Asimismo, cada categoría contempla una serie de elementos que contiene la locución:

Elemento	Significado
(a)	Introduce el complemento indirecto en la locución.
(N0)	Primer argumento exigido por la locución, que puede tener tanto función de sujeto como de complemento indirecto. Un N0 puede ser tácito y no tener realización fonológica, esta particularidad se indica entre paréntesis.
le/se	Dan cuenta de los pronombres clíticos determinados o definidos por una locución en particular. En estos casos, la expresión únicamente admite estas opciones de clíticos.
CL	Indica que en esa parte de la locución se puede insertar un pronombre clítico dativo o acusativo. Sin embargo, algunas locuciones exigen un tipo particular de pronombre; en esos casos, se indica en la clase como “CLdat”, pues en aquel lugar solo se admiten pronombres dativos.
V	Forma verbal que debe ser conjugada. Se constituye como el núcleo de la locución.
C1	Representa la parte “invariable” de la locución.

Tabla 4. Descripción de los elementos constitutivos por cada clase.

Posteriormente al trabajo de categorización se prosiguió con la confección de tablas léxico-gramática para cada clase. Estas permiten contar con una base de datos de los distintos argumentos de las LLVV. Respecto de la estructura de las tablas, se siguen los lineamientos de Gross (1984) y Vietri (2014). A continuación, se ejemplifica con un fragmento del análisis distribucional de las LLVV de la clase 5.

PREP	ND	PronCl				Verbo	CS	CS					PREP	Ejemplo										
		a	+Anim	-Anim	Se			Dativo	Acusativo	SD	SN	SP			SA	SAd	SV	a	de	con	por	para	N3	
(+)	(+)	-	-	+	-	bajarle	la indiada.	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A Juan le bajó la indiada, después se teimó y se fue.
(+)	(+)	-	-	+	-	caerle	el tejazo.	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Finalmente, a Juan le cayó el tejazo
(+)	(+)	-	-	+	-	caerle	el teja.	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Hasta que le cayó el teja a Juan
(+)	(+)	-	-	+	-	caerle	la chauscha.	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A Juan le cayó la chauscha, se demostró pero ya se dio cuenta
(+)	(+)	-	-	+	-	caerle	la teja.	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A María le cayó la teja y encaráó a Juan
(+)	(+)	-	-	+	-	comerle	las manos.	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	A María le comen las manos por volver a competir
(+)	(+)	-	-	+	-	crecerle	el pelo.	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A Pedro le creció el pelo, ya no pertenece a este lugar
(+)	(+)	-	-	+	-	crujirle	el mate.	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Hasta que a María le crujió el mate y se dio cuenta de cómo eres
(+)	(+)	-	-	+	-	crujirle	la mollera.	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A Juan le crujió la mollera, gracias a eso se solucionó el problema

Tabla 5. Ejemplo tabla léxico-gramática de la clase 5.

Como se puede apreciar, las tablas léxico-gramática poseen forma de matriz, en cuyas líneas se define cada una de las locuciones y en sus columnas se

precisan las diversas propiedades sintáctico-semánticas. A continuación, se sintetiza la cantidad de LLVV que incluye cada clase.

Clase	Total locuciones	Total Entradas
Cl 1: (N0)+V+C1	405	798
Cl 2: (N0)+CL+V+C2	287	459
Cl 3: (N0)+CL+CL+V+C3	6	12
Cl 4: (a)+(N0)+CLdat+V	6	6
Cl 5: (a)+(N0)+CLdat+V+C5	35	36
Cl 6: (a)+(N0)+se+CLdat+V+C6	40	40
Cl 7: (a)+(N0)+se+CLdat+V	3	4
Cl 8: (a)+(N0masc)+CLdat+V+C8	2	2
Cl 9: (a)+(N0masc)+se+CLdat+V+C9	7	7
Cl 10: (a)+(N0fem)+CL+V+C10	1	2

Tabla 6. Comparación cantidad de lemas por clase *versus* cantidad de entradas en tabla léxico-gramática

Puede observarse que las primeras clases poseen una gran diferencia entre la cantidad de lemas y las entradas en el diccionario. Esto porque, si se piensan las clases en un *continuum*, en el que la 1 es la más simple y la 10 la más compleja, se puede vislumbrar que, a medida que se van complejizando las clases, los usos o las propiedades que admiten para su empleo se restringen, lo cual se traduce en una reducción de la cantidad de entradas en las tablas.

3.1.2. Transformaciones

Posteriormente, se llevó a cabo la tabla de transformaciones. En el presente estudio se identificaron trece; sin embargo, se tiene como proyección expandir esta lista. Se resume cada transformación en la siguiente tabla:

Transformación	Significado	Ejemplo
1	Negación con “no”	-dar bola Ella le dio bola = Ella no le dio bola.
2	Negación con “sin”	-hacer bolsa Ella hizo bolsa sus juguetes = Sin que haga bolsa tus juguetes.
3	Voz pasiva	-hacer bolsa María hizo mierda el auto = El auto fue hecho mierda por María.
4	Reemplazo de C1 por clítico	-pegar la PLR María le pegó la PLR= María <i>se la</i> pegó.
5	Reemplazo de A1 por un clítico acusativo	-hacer mierda Juan hizo mierda a María = Juan la hizo mierda
6	Pluralización de los clíticos	-dejarle la cagada María le dejó = María les dejó la cagada.
7	Interrogación	-Dejar la escoba Juan dejó la escoba = ¿ <i>Quién</i> dejó la escoba?
8	Condicional	-Quedar pegado Juan quedó pegado = <i>si</i> Juan quedó pegado, María también.
9	C1 con flexión de número	-Estar cagado Él está cagado = Ellos están cagados.
10	C1 con flexión de género	-hacer huevón Él lo hizo huevón= Él <i>la</i> hizo huevona.
11	C1 con diminutivo	-hacer tuto Ellos hicieron tuto = Ellos hicieron <i>tutito</i> .
12	C con aumentativos	-caer gordo Él le cayó gordo= Él le cayó <i>gordísimo</i> .
13	Inclusión de un modificador de C1	-Quedar la cagada Ayer quedó la cagada = Ayer quedó <i>la mansa</i> cagada.

Tabla 7. Transformaciones de locuciones verbales del español de Chile

En este punto es pertinente aclarar que las transformaciones 11, 12 y 13 podrían corresponderse con lo que se considera desautomatización (Zuluaga 2001; Mena Martínez 2003). Este concepto alude al fenómeno de desviación que hace un hablante de una unidad fraseológica de su forma canónica, como ejemplifica Mena Martínez (2003): ‘dar gato por euro’ en lugar de ‘dar gato por liebre’. Generalmente, esto se realiza por motivos expresivos, humorísticos, etcétera, lo que se correspondería con lo que se considera la actuación, lo cual excede los objetivos del presente trabajo. No obstante, se va a tener en cuenta los tipos de desautomatización descritos en 11, 12 y 13 como transformaciones, puesto que se mantiene el contenido semántico de la LV.

Una vez elaborado el glosario de las transformaciones, se prosiguió con la confección de las tablas. Se presenta un fragmento de las transformaciones de la clase 1.

Verbo	Cl	Ejemplo	Transformaciones												
			T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13
hacer	mierda.	Ellos hicieron mierda todo lo que provenía de sus enemigos	+	+	-	-	+	+	-	-	-	-	-	-	-
hacer	mierda.	La podadora hizo mierda todas las flores	+	+	+	-	+	+	-	-	-	-	-	-	-
hacer	mierda.	Ellos se harán mierda la ropa si siguen peleando	+	+	-	-	+	+	-	-	-	-	-	-	-
hacer	mierda.	Viene el aluvión y los carteles se harán mierda con la tormenta	+	+	-	-	+	+	-	-	-	-	-	-	-
hacer	mierda.	Pamela me hará mierda las uñas	+	+	-	-	+	+	-	-	-	-	-	-	-
hacer	mierda.	Las espinas me harán mierda la ropa	+	+	-	-	+	+	-	-	-	-	-	-	-

Tabla 8. Ejemplo transformaciones locución Clase 1

En la tabla anterior se detalla, en primera instancia, los componentes nucleares de la locución (V+C) y, posteriormente, las transformaciones que admite, lo cual, al igual que en las tablas léxico-gramática, se va indicando con un signo + si la expresión admite determinada modificación o con un – si no es así. Tomando como ejemplo el caso de la tabla 8, se puede señalar que la oración ‘Ellos hicieron mierda todo lo que provenía de sus enemigos’ admite:

Transformación	<OPT>
T1	Ellos no hicieron mierda todo lo que provenía de sus enemigos
T2	Sin que ellos hicieran mierda todo lo que provenía de sus enemigos.
T3	Todo lo que provenía de sus enemigos fue hecho mierda por ellos.
T5	Ellos lo hicieron mierda
T7	¿ Quiénes hicieron mierda todo lo que provenía de sus enemigos?
T8	Si ellos hicieron mierda todo lo que provenía de sus enemigos...

Tabla 9. Ejemplo transformaciones

Como se puede observar, la locución descrita no admite todas las transformaciones, por lo que, por ejemplo, construcciones como (6) resultarán agramaticales.

- (6) a. *Ellos *los hicieron* (T4)
- b. *Ellos *hicieron mierdas* (T9)

Una vez realizados estos procedimientos, que permiten proponer una exhaustiva descripción de las LLVV del español de Chile, se procedió a una modelización computacional en NooJ.

3.2. Trabajo computacional

Para llevar a cabo el procedimiento de detección automática de las LLVV del español de Chile fue necesario formalizar las estructuras que posee cada clase. Se recurrió a NooJ, una herramienta que cuenta con diversas utilidades para el tratamiento de lenguaje natural y que se puede organizar a partir de la jerarquía propuesta por Chomsky y Schützenberger (1963). El programa cuenta con los siguientes recursos:

- diccionarios: listas de palabras con diversos tipos de información lingüística.
- gramáticas morfológicas y derivacionales: modelos de flexión y derivación.
- gramáticas productivas: sistemas regulares o gráficos útiles para el tratamiento de cadenas de caracteres con determinadas propiedades formales.
- gramáticas sintácticas (archivos .nog): sistemas regulares o gráficos útiles para el tratamiento de cadenas de caracteres formadas por dos o más unidades léxicas, generalmente, separadas por espacios en blanco.

Como se ha mencionado, la lista de palabras del diccionario contempla información de diversa índole, según sean los intereses del usuario (morfológica, sintáctica, léxica, semántica, etcétera). A su vez, los lemas pueden asociarse a modelos morfológicos que permiten la flexión automática. Por ejemplo, una palabra como ‘médico’ está listada de la siguiente manera.

médico,N+anim+hum+pm+FLX=NIÑO

Eso significa que ‘médico’ es un nombre (‘N’) que posee los rasgos ‘animado’ (‘+anim’), ‘humano’ (‘+hum’) y ‘profesional de la medicina’ (‘pm’), y tiene una flexión (‘FLX’) correspondiente con el modelo NIÑO, especificado en la gramática morfológica:

NIÑO = <E>/masc+sg | s/masc+pl | a/fem+sg | as/fem+pl;

Con dicha gramática es posible generar automáticamente ‘médico’, ‘médicos’, ‘médica’, ‘médicas’. Este procedimiento resulta mucho más eficaz y económico, debido a que el mismo modelo permite flexionar numerosas palabras (‘médico’, ‘enfermero’, ‘farmacéutico’, etcétera); en el caso de los verbos en español, que tienen una enorme riqueza flexiva, las ventajas de este tipo de gramáticas son indiscutibles (Bonino 2015).

A su vez, también el programa cuenta con la posibilidad de elaborar gramáticas de diversos niveles (dependientes del contexto, de estados finitos, etcétera). Por ejemplo, como se puede apreciar en Silberztein (2016), una gramática sencilla para el SN del inglés puede esquematizarse de la siguiente manera:

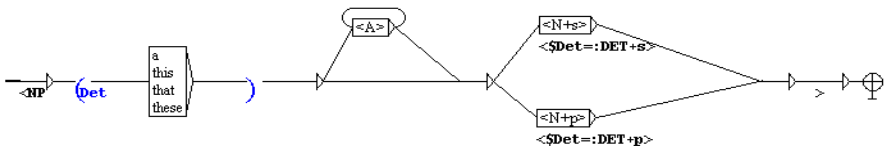


Figura 1. Ejemplo de una gramática sintáctica NooJ para SN del inglés

Como indica la figura 1, la gramática comienza con un estado inicial compuesto por un determinante, indicado entre paréntesis, puede incluir un adjetivo (<A>) y finalizar en el núcleo nominal (<N>). A la vez, se especifica la regla de concordancia de número: si N tiene el rasgo singular (‘+s’), entonces el número del determinante será singular; en cambio, si el N posee el rasgo plural, el determinante tomará ese mismo rasgo. De este modo, cada expresión que contenga la información declarada será etiquetada como NP (Noun Phrase) (<NP ... >). Estas gramáticas pueden operar tanto en el nivel

sintáctico como en el morfológico y permiten reconocer y generar de manera automática.

En esta ocasión se siguieron los siguientes procedimientos: (i) elaboración de gramáticas léxicas; (ii) confección de un diccionario que contemplara todas las propiedades de cada clase de locución.

3.2.1. Elaboración de gramáticas y diccionarios

Cada una de las gramáticas elaboradas presenta como núcleos básicos: (i) el nombre de la clase de locución; (ii) la unidad V, que indica el núcleo verbal de la expresión, y (iii) los elementos que componen la parte “fija” de la expresión. De esta manera, la gramática presenta estructuras como las que se ejemplifican en la tabla 10.

Ejemplo	V	PREPC	CONT	DETC	C
hacer mierda	Hacer	-	-	-	mierda
dejar la escoba	Dejar	-	-	la	escoba
atornillar al revés	Atornillar	-	al	-	revés
caer de cajón	Caer	de	-	-	cajón
agarrar para el chuleteo	Agarrar	para	-	el	chuleteo

Tabla 10. Ejemplo composición de elementos centrales de una locución verbal del español

El número en el elemento C cambiará según cada clase. Por ejemplo, para la clase 1, se denomina C1 y para la clase 2, C2 y así sucesivamente.

Las gramáticas irán variando según sus elementos constitutivos y, para evidenciarlo, se ejemplificará con las gramáticas de clase 1 y clase 2. En la figura 2 se muestra la primera de ellas.

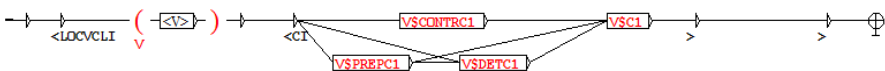


Figura 2. Gramática léxica para detectar locuciones clase 1

En esta figura se pueden observar los elementos antes referidos: al inicio del nodo, el nombre con el que se etiquetará la expresión ‘LOCVLI’ (Locución Verbal de Clase 1); luego, entre paréntesis, la variable V, con la que se detectará el verbo nuclear de la locución y, por último, en CI, las combinatorias de elementos que pueden componer la parte fija de la locución.

La gramática toma la información declarada en el diccionario electrónico, que contiene los lemas declarados de la siguiente manera:

agarrar,V+FLX=AMAR+C1=papa+FXC
 agachar,V+FLX=AMAR+DETC1=el+C1=moño+FXC
 echar,V+FLX=AMAR+CONTRC1=al+C1=agua+FXC
 caer,V+FLX=CAER+PREPC1=de+C1=cajón+FXC

El lema se inicia con el verbo de la locución, seguido de su modelo de flexión (FLX=); posteriormente, se indican los elementos que constituyen la parte invariable. Todo esto debe llevar al final el comando +FXC para indicar que se trata de una expresión fija (*frozen expression*).

Con todo esto, es posible lograr una detección incluso en aquellos casos en los que la locución tenga los elementos separados, como en ‘dejó Juan la mansa escoba’:

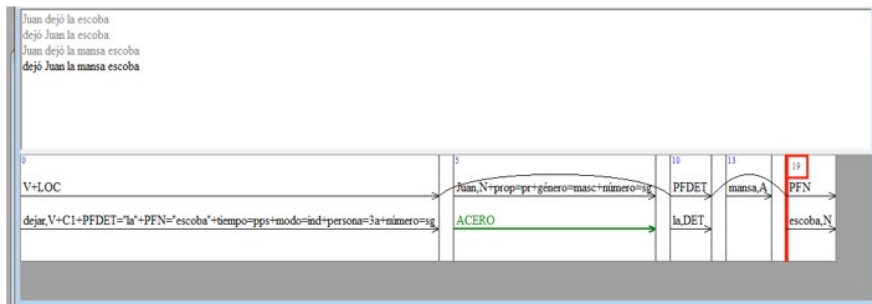


Figura 3. Ejemplo de detección de locución clase 1

La detección presenta “saltos” para unir los elementos de la LV que se encuentran separados.

The screenshot shows the NooJ software interface. The title bar reads "NooJ - [(Modified) Untitled]". The menu bar includes "File", "Edit", "Lab", "Project", "Windows", "Info", and "TEXT". The toolbar shows a minus sign, the number "2", and a plus sign, followed by "/ 3 TUs". A dropdown menu is open, showing "Characters", "Tokens", and "Digrams". A checkbox labeled "Show Text Annotation Structure" is checked. The main text area contains the sentence "Juan agacha el moño" and "Juan se afirma los pantalones". Below the text area is a syntactic tree diagram with grammatical annotations. The tree is structured as follows:

Index	Text	Grammatical Annotation
0	Juan	NPR
5	se	CL
8	afirmar	V+pres+ind+3a+sg
15	los	DETDEF
19	pantalón	N+masc+pl
	los	CL
	pantalones	N+masc+sg

Green arrows indicate the following locutions: "LOCVCLII" spans from index 5 to 8, and "CII" spans from index 15 to 19.

Figura 5. Ejemplo de detección de locución clase 2

Incluso si una locución de clase 1 admite un pronombre clítico antepuesto al verbo, lo que se asemejaría a la gramática de la clase 2, NooJ realiza la detección de manera adecuada sin mezclar la información declarada en las gramáticas. Al igual que en el ejemplo anterior, en la figura 6 se muestra la detección realizada empleando los diccionarios con las respectivas gramáticas de las clases 1 y 2.

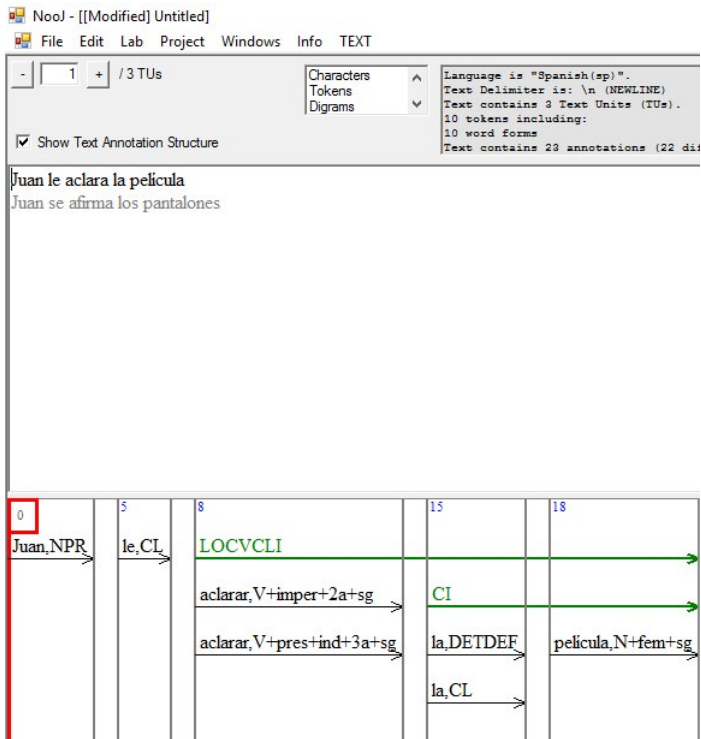


Figura 6. Ejemplo de detección locución Clase 1 que admite pronombre

Esta metodología fue probada en un corpus compuesto por textos extraídos de la prensa chilena. En la sección siguiente, se presentan los resultados obtenidos.

4. Resultados obtenidos

Para llevar a cabo la detección automática se cargaron en NooJ los diccionarios correspondientes a cada una de las clases de locuciones con sus respectivas gramáticas. En la figura 7 se presenta un fragmento del output realizado por software.

Text	Before	Seq.	After
	ganas locas de salir a	mover el esqueleto	y sacudirse las pulgas con
	con mi señora es al	hacer las tareas	. La casa es chica y
	pesca, y me pide sólo	hacer las tareas	. El cuerpo no me da
	películas, pero para eso debe	tirar toda la carne a la parrilla	y ver si la cosa
	conversar, después nos abrazamos, y	quedó la tendalada	. Nos agarramos a besos, nos
	tengo una polola como para	sacar pica	en mi ciudad natal. Hace
	las ideas del encuentro y	dijo upa	. De ahí a comer, unos
	vamos por calor humano, me	dijo upa	, chalupa. Estoy re feliz, jefe
	es que mis compañeros me	agarraron para el palanqueo	, señalándome que sólo quiero acostarme
	reina, la besará y le	sacará pica	a sus ex malos compañeros
	considerando que donde estaba le	atormillaban al revés	. 19 Doc: No sé qué hacer
	llamé tipo una, dijo que	venía de vuelta	y nada... Me la encontré
	que de a poco irán	cachando el mote	. Sin embargo, eso no quiere
	Es que le dije que	duraba menos que un candy	, y usted me señaló que
	feliz, pues me dice que	soy una máquina	sexual. Me halaga todo el
	no quiero parar y temo	dejar la escoba	en casa. César Emperador: No
	y, obvio, sin mina joven. ¿	Cachó el mote	? Ya, ahora enmiende el rumbo
	la cabeza al WC y	tiró la cadena	con ganas. No le creo
	el baño y que lo	hagan bolsa	por pobre. Si la jemu
	Quizás usted en muchos años	valió callampa	en el catre o sólo
	a su familia. El fútbol	vale callampa	frente a un problema de
	cada vez que puede me	saca pica	con ella cuando entra a
	Y que su hermano se	haga un lulo	. Así es la diversidad, mi
	casa de la Ana para	hacer las tareas	, en eso estábamos hace un
	y más al momento de	gritar viva Chile	Pero siempre hubo un abejorro
	la veterana es porque le	calienta la sopa	nomás. Pero puro problema. Es
	Pregúntele cuánto le queda para	gritar viva Chile	, respírele en la oreja y
	Doctor Cariño: Ahora si que	corto las huinchas	. Llevo dos años y medio

Figura 7. Ejemplo de detección de locuciones clase 1

El corpus empleado para poner a prueba la detección automática proviene del periódico nacional *La Cuarta*. De este medio se extrajeron textos de una de sus secciones denominada “Ventanita sentimental”. En este espacio, los lectores envían cartas con diversas inquietudes. Se seleccionó este apartado debido a que esta sección se caracteriza por emplear un registro informal, y por ello se tuvo como supuesto que podrían ser textos ricos en modismos y locuciones verbales como las que se trabajan en esta investigación. En total, el corpus se compone de 33.094 palabras, lo cual equivale a más de cien cartas con sus respectivas respuestas.

Para la obtención de resultados se contabilizaron manualmente las LLVV incluidas en el corpus y se comparó con el output de NooJ. Con ello, se calcularon los porcentajes de precisión, cobertura y medida F. Vale aclarar que, en esta ocasión, no se tuvieron en cuenta aquellas locuciones del corpus que presentaban la estructura verboide más enclítica (‘sabiéndoselas por libro’),

quedando el análisis de estos para etapas futuras de trabajo. En la tabla 11 se listan los porcentajes logrados por el algoritmo.

Resultados	Resumen
Locuciones por clase	103
Detectó bien	95
Detectó mal	3
Detección total	98
Cobertura	92,23%
Precisión	96,94%
Medida F	94,53%

Tabla 11. Resultados detección por total

Puede observarse que los resultados obtenidos fueron positivos. Esto permite comprobar que la formalización realizada a partir del aparato metodológico que entrega la léxico-gramática fue adecuada. Sin embargo, debido a la limitada cantidad de clases que presentaba el corpus, no se pudieron validar los recursos construidos para las ausentes. Esto implica que la compilación de un nuevo corpus que contenga locuciones de todas las clases se constituye como un desafío, necesario para el estudio de este tipo de expresiones.

En relación con los resultados obtenidos por clase, se puede indicar que la mayor cantidad de locuciones se concentran en la clase 1 y 2. Este resultado no llama la atención considerando que más del 50% de los lemas o entradas en el DUECh pertenecen a estas clases, por lo que se constituyen como locuciones diversas y a la vez comunes en el uso del español chileno. En cuanto a los porcentajes obtenidos, se puede observar que los resultados más descendidos se encuentran en la cobertura, lo que implica que para próximos trabajos se requerirá realizar gramáticas en las que se incluyan más propiedades de las locuciones.

Resultados	Clase 1	Clase 2	Clase 4	Clase 5
Locuciones por clase	61	47	3	2
Detectó bien	52	39	3	1
Detectó mal	1	2	0	0
Detección total	53	41	3	1
Cobertura	85,25%	82,98%	100%	50%
Precisión	98,11%	95,12%	100%	100%
Medida F	91,23%	88,64%	100%	66,67%

Tabla 12. Resultados detección por clase

Del mismo modo, se deberán revisar las cualidades declaradas en los diccionarios para detectar aquellas 9 locuciones que quedaron fuera en el caso de la clase 1, 8 en relación con la 2 y 1 en la clase 5. No obstante, se puede señalar que los resultados son satisfactorios como primera aproximación al fenómeno de las LLVV en el español de Chile.

5. Conclusiones

Para concluir, cobra relevancia recordar los objetivos generales propuestos: (i) describir las propiedades morfosintácticas y la estructura argumental que proyectan las LLVV del español de Chile y (ii) detectar LLVV del español de Chile en textos de lenguaje natural. Estos dos grandes propósitos guiaron el quehacer investigativo y, a partir de los resultados evidenciados en el apartado anterior, se puede señalar que se consiguieron ambos. A este respecto, es importante indicar la efectividad del aparato metodológico que aporta la léxico-gramática en trabajos descriptivos como este, pues, por medio de las tablas, se pueden extraer las cualidades de las locuciones y, a su vez, se permite formalizar las estructuras para poder llevar a cabo la modelización computacional.

Así, por un lado, fue posible develar que, en la mayoría de los casos, las locuciones requieren de un argumento N0 +Humano. También se observó que, en relación con el uso de pronombres, un amplio porcentaje se emplea con pronombres dativos ('Juan le da bola a María'). En cuanto a los elementos "fijos", se pudo determinar que, en la mayoría de las clases, estas unidades

pertenecen a sintagmas determinantes, nominales o preposicionales. Y, por último, se dio cuenta de la poca recurrencia al empleo del argumento N1, ya sea introducido o no por preposiciones.

Esta descripción posibilitó la formalización de estas unidades para realizar la posterior detección computacional. Al respecto, se obtuvieron resultados positivos en cuanto a la detección, pues todos los indicadores, ya sea, precisión, cobertura y medida F, estuvieron por encima del 80%. Esto, gracias también a las propiedades que posee el software, ya que la elaboración del diccionario ofrece la opción de introducir un comando especial (+FXC) para el tipo de unidades que se desea detectar.

Por otro lado, se observó que, de las trece transformaciones contempladas en este estudio descriptivo, existen tres que son recurrentes de manera transversal a las clases. Asimismo, cobra relevancia indicar el alto promedio de transformaciones que admiten las locuciones de cada clase, pues eso da cuenta de que estas unidades son ricas en cuanto a propiedades sintáctico-semánticas y morfológicas y, por tanto, pese a que tradicionalmente fueron consideradas como invariables, se debe renovar esta visión.

Ahora bien, a pesar de que, en general, se puede señalar que se lograron los objetivos propuestos, algunos aspectos que quedaron fuera del alcance de esta investigación deberán ser abordados en un futuro. Al respecto, queda como proyecto la elaboración de tablas léxico-gramática para las locuciones pertenecientes al segundo grupo, esto es, las que en su entrada del diccionario poseen el adverbio de negación “no”.

Asimismo, también se plantea para un trabajo posterior, la elaboración de reglas gramaticales que permitan la detección de locuciones que se emplean con uno o dos pronombres enclíticos en la unidad verbal (‘hacérselo pebre’). Del mismo modo, cobra relevancia, como ya se mencionó anteriormente, tanto la inclusión de más transformaciones a la tabla, como la formalización de estas para enriquecer el procedimiento de detección. Para ello, se podría requerir de una mayor descripción de los elementos constitutivos de las locuciones, lo que se traduce en una descripción más detallada de las propiedades que estas unidades poseen.

En síntesis, el trabajo con vistas al futuro, en relación con el aspecto descriptivo, se organiza en torno a los siguientes ejes: (i) profundizar la descripción lingüística realizada en las tablas léxico-gramática para mejorar la

detección automática; (ii) extender las posibilidades transformacionales que poseen las LLVV del español de Chile, y (iii) elaborar tablas para locuciones cuyo lema en el DUECh contenga el adverbio de negación “no”. En cuanto al trabajo computacional, se pretende: (a) ampliar el diccionario electrónico de las LLVV; (b) enriquecer las gramáticas informáticas, y (c) elaborar gramáticas para la generación automática de oraciones que contengan una LV como predicado. Asimismo, también se incluye dentro de las tareas que se deberán realizar la compilación de un corpus mayor para poder probar la totalidad de los recursos elaborados.

Referencias bibliográficas

- ACADEMIA CHILENA DE LA LENGUA. (2010) *Diccionario de uso del español de Chile*. Santiago: MN Editorial Ltda.
- BONINO, Rodolfo. (2015) “Una propuesta para el tratamiento de los enclíticos en NooJ.” *Infosur Revista* 7, pp. 31-40.
- CHOMSKY, Noam. (1957) *Syntactic structures*. Berlín: Mouton & Co.
- CORPAS PASTOR, Gloria. (1996) *Manual de fraseología española*. Madrid: Gredos.
- GROSS, Maurice. (1975) *Méthodes en syntaxe*. París: Hermann.
- GROSS, Maurice. (1984) “Lexicon-grammar and the syntactic analysis of French.” En: Association for Computational Linguistics (1984). *Proceedings of the 10th International Conference on Computational Linguistics and 22nd anual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, pp. 275-282.
- HARRIS, Zelig. (1970) *Papers in structural and transformational linguistics*. Dordrecht: Springer.
- MENA MARTÍNEZ, Florentina. (2003) “En torno al concepto de desautomatización fraseológica: aspectos básicos.” *Tonos Revista electrónica de estudios filológicos* 5. Versión electrónica: <<https://www.um.es/tonosdigital/znum5/estudios/H-Edesautomatizacion.htm>>
- ELIA, Annibale; Mario Monteleone & Federica Marano. (2011) “From the concept of transformation in Harris and Chomsky to the Lexique-grammaire of Maurice Gross.” En: Kasevich, Vadim; Yuri Kleiner & Patrick Sériot (eds.) (2011). *History of Linguistics*. Amsterdam: John Benjamins, pp- 76-82.
- PENADÉS, Inmaculada. (2012) *Gramática y semántica de las locuciones*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.

- PIVAUT, Laurent. (1989) *Verbes supports et vocabulaire technique: sport, musique et activités intellectuels*. Paris: Universidad de París. Tesis doctoral inédita.
- SÁEZ, Leopoldo. (2012) “El léxico del dialecto chileno: *Diccionario de uso del español de Chile DUECh*.” *Estudios filológicos* 49, pp. 137-155.
- SILBERZTEIN, Max. (2003) *NooJ Manual*. Versión electrónica: <<http://www.NooJ4nlp.net/>>
- SILBERZTEIN, Max. (2016) *Formalizing Natural Languages: The NooJ Approach*. Londres: Wiley Eds.
- SILVAGNI, Federico. (2017) *Entre estados y eventos. Un estudio del aspecto interno del español*. Barcelona: Universidad Autónoma de Barcelona. Tesis doctoral inédita.
- TOLONE, Elsa. (2012) “Conversión de las tablas del léxico-gramática del francés en el léxico LGLex”. Ponencia presentada en el *2nd Argentinian Workshop on Natural Language Processing (WNLP'11)*. Córdoba: Universidad Nacional de Córdoba.
- VIETRI, Simonetta. (2014) *Idiomatic Constructions in Italian. A Lexicon-Grammar Approach*. Amsterdam Netherlands: John Benjamins BV.
- ZULUAGA, Alberto. (1975). “La fijación fraseológica.” *Thesaurus* 30, pp. 235-288.
- ZULUAGA, Alberto. (2001) “Análisis y traducción de unidades fraseológicas des-automatizadas.” *PhinN* 16, pp. 67-83.

NOTA BIOGRÁFICA / BIONOTE

WALTER KOZA es Doctor en Humanidades y Artes con Mención en Lingüística, egresado de la Universidad Nacional de Rosario (Argentina). Actualmente, es Profesor Auxiliar de la Pontificia Universidad Católica de Valparaíso (Chile). Sus áreas de investigación son la lingüística computacional, los estudios del léxico y la gramática del español.

WALTER KOZA is PhD in Humanities and Arts with mention in Linguistics graduated of Universidad Nacional de Rosario (Argentina). Nowadays is Assistan Professor at Pontificia Universidad Católica de Valparaíso (Chile). His research areas are computational linguistic, the lexicology and Spanish grammar.

NATALIA RIVAS es Profesora de Castellano y comunicación y Licenciada en Lengua y Literatura Hispánica, egresada de la Pontificia Universidad Católica de Valparaíso (Chile).

NATALIA RIVAS is Teacher of Spanish and Comunication and Bachelor in Language and Hispanic Literature, graduated at Pontificia Universidad Católica de Valparaíso (Chile).

Recibido / Received: 31/10/2019
Aceptado / Accepted: 22/03/2020

Para enlazar con este artículo / To link to this article:
<http://dx.doi.org/10.6035/MonTI.2020.ne6.4>

Para citar este artículo / To cite this article:

Zuccalà, Daniele & María Bernarda Espejo Olaya. (2020) “Estudio de caso de la fraseología empleada en Colombia: análisis crítico de un corpus *ad hoc* de locuciones.” En: 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. 121-153.

ESTUDIO DE CASO DE LA FRASEOLOGÍA EMPLEADA EN COLOMBIA: ANÁLISIS CRÍTICO DE UN CORPUS AD HOC DE LOCUCIONES

CASE STUDY OF THE PHRASEOLOGY USED IN COLOMBIA: CRITICAL ANALYSIS OF AN AD HOC CORPUS OF MULTI WORD UNITS

DANIELE ZUCCALÀ

daniele.zuccala@uma.es

Universidad de Málaga / Università degli Studi di Torino

MARÍA BERNARDA ESPEJO OLAYA

maria.espejo@caroycuervo.gov.co

Instituto Caro y Cuervo

Resumen

El objetivo del presente artículo es cubrir un vacío al someter a estudio el universo fraseológico del español hablado en Colombia, con particular atención a la esfera de las locuciones. El diseño metodológico del estudio se ha dividido en dos fases: en primer lugar, se ha compilado un corpus *ad hoc* de más de 1400 unidades fraseológicas, extraídas principalmente del *Diccionario de colombianismos* y del *Diccionario de americanismos*; en segundo lugar, se ha analizado el contenido fraseológico de estas unidades siguiendo criterios cuantitativos y cualitativos. Para el análisis del corpus se tuvo en cuenta la función oracional desempeñada en el discurso, el campo semántico al que pertenece el principal constituyente léxico de la unidad fraseológica (Sánchez y Aguirre 1992), la situación comunicativa, el registro (Penadés Martínez 2015), la



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

intención del hablante, las variantes (Corpas Pastor 1996; Mellado Blanco 2004) y su presencia en el *Diccionario de la Lengua Española*.

Palabras clave: Fraseología; Unidad Fraseológica; Locuciones; Colombia; Variación diafásica.

Abstract

The aim of this study is to fill a gap by studying the phraseological universe of the Spanish spoken in Colombia, with particular attention to the sphere of locutions. The methodological design of the study has been divided into two phases: firstly, an *ad hoc* corpus of more than 1400 phraseological units has been compiled, taken mainly from the *Diccionario de colombianismos* and the *Diccionario de americanismos*; secondly, the phraseological content of these units has been analysed according to quantitative and qualitative criteria. The analysis of the corpus took into account the sentence functions performed in the discourse, the semantic field to which the main lexical constituent of the phraseological unit belongs (Sánchez and Aguirre 1992), the register (Penadés Martínez, 2015), the speaker's intentionality, the variants (Corpas Pastor 1996; Mellado Blanco 2004), and its presence in the *Diccionario de la Lengua Española*.

Keywords: Phraseology; Multi Word Units; MWU; Colombia; Diaphasic Variation.

1. Introducción

Urge una aproximación pormenorizada hacia la fraseología empleada en Colombia, puesto que, a pesar de la riqueza fraseológica demostrada por el *Diccionario de fraseología usada en Colombia* (Espejo Olaya 2015), no ha sido estudiada de manera estructurada, como sí ha ocurrido con el español peninsular. Por ello, el objetivo principal de la presente investigación es analizar la fraseología empleada en la actualidad en el país latinoamericano¹, con particular atención a la esfera de las locuciones. El telón de fondo del estudio está representado por las aproximaciones a la fraseología española y,

1. Cabe destacar que el presente estudio no analiza la “fraseología colombiana”, es decir, aquellas unidades fraseológicas empleadas de manera exclusiva en el país latinoamericano, sino que aborda la “fraseología empleada en Colombia”, es decir, la fraseología empleada en el país, incluidas aquellas UF que sus hablantes comparten con otros países de habla hispánica.

en particular, pone sus raíces en la clasificación y las aportaciones de Corpas Pastor (1996, 2002, 2003, 2013).

En primer lugar, se examinará el marco teórico entendiendo la fraseología como una ciencia autónoma que se enraíza en dos áreas de conocimiento: por un lado, la terminología y, por el otro, la lexicografía. Al tratarse de un estudio empírico en el que se analiza un amplio fichero fraseológico, se ha optado por estudiar las unidades fraseológicas siguiendo un método puramente terminológico. Por ello, se ha recurrido a la Organization for Standardization (ISO) y, en particular, a la ISO 704 Terminology work - Principle and methods y a la ISO 1087 Terminology work - Vocabulary, estándares fundamentales para respetar el rigor terminológico y, por ende, fraseológico. Por otra parte, el estudio se nutre de la lexicografía a través de las aproximaciones sobre disponibilidad léxica iniciadas en los años sesenta por Georges Gougenheim (1964, 1967) y su grupo de investigación, así como de las propuestas de Sánchez Lobato y Aguirre Beltrán, que en su obra *Léxico fundamental del español* (1992) definen quince campos semánticos para aglutinar el léxico.

Dos fases complementarias forman el diseño metodológico de la presente investigación. Esta fórmula ha permitido explotar un corpus *ad hoc* compuesto por más de 1400 unidades fraseológicas (UF) extraídas de fuentes primarias, entre las que destacan el *Diccionario de colombianismos* (Instituto Caro y Cuervo, 2018) y el *Diccionario de americanismos* (Asociación de Academias de la Lengua Española, 2010). La aplicación de un método riguroso y los pilares teóricos sobre los que se ha construido el estudio han permitido explotar los resultados parciales de una ficha fraseológica.

2. Marco teórico

La fraseología es un área líquida de la lingüística, parafraseando a Zygmunt Bauman. El sociólogo y humanista polaco empleaba el concepto de modernidad líquida (1999) para definir el actual momento histórico, en el que las realidades sólidas se han desvanecido, dejando espacio a un mundo más precario, provisional y, a menudo, difícil de encasillar. Algo parecido ocurre con la fraseología: en comparación con el léxico simple, las propiedades intrínsecas de la fraseología hacen que huya de etiquetas y cuadros preestablecidos, escabulléndose de los encasillamientos.

Por esta misma razón, las unidades fraseológicas han despertado el interés de lexicógrafos y terminólogos que, atraídos por sus propiedades, han intentado con más o menos éxito organizar esta área de conocimiento en clasificaciones de diferente naturaleza. A continuación, se analizarán las clasificaciones que el presente estudio ha empleado, con particular atención a la que Corpas Pastor elaboró en 1996.

2.1. Clasificación de las unidades fraseológicas

Desde hace siglos se registran unidades fraseológicas en el ámbito hispánico mediante obras lexicográficas, gramáticas y otros tipos de textos, principalmente literarios (Sevilla, 2012). Sin embargo, es a mediados del siglo XX cuando Julio Casares, miembro de la Real Academia Española, marca un hito en los estudios fraseológicos y paremiológicos con su *Introducción a la lexicografía moderna* (1950). Su pionero trabajo despierta la atención de numerosos lexicógrafos, entre los cuales se encuentra Alberto Zuluaga, quien publica su *Introducción al estudio de las unidades fijas* (1980), que, si por un lado bebe de las intuiciones de Casares, por otro añade algunas novedades relevantes para futuros estudios fraseológicos.

Sin embargo, habrá que esperar hasta los años noventa para observar el verdadero auge de la fraseología española. Desde entonces se han presentado numerosas clasificaciones de la misma, algunas de las cuales han tenido amplia aceptación entre la comunidad científica. Debido al reducido espacio del que se dispone, se ofrece a continuación solamente una pincelada de las principales taxonomías aportadas a la fraseología española².

En 1996 Gloria Corpas Pastor publica su *Manual de fraseología española*, consiguiendo, por primera vez, reunir toda la fraseología en una taxonomía holística que sigue siendo considerada un clásico de referencia en los estudios fraseológicos y es todavía empleada para clasificar el amplio espectro del universo fraseológico (Barani 2012; Velázquez Puerto 2015; Martín Salcedo

2. A falta de espacio, se mencionan otras aportaciones que, a pesar de no ofrecer una clasificación holística del universo fraseológico, son relevantes para definir el marco teórico de la presente investigación: Carmen Mellado Blanco (2004, 2008, 2015), Julia Sevilla Muñoz (1993, 1997, 2013, 2015), Mena y Pamies Bertrán (2002), Castillo Carballo (2017), Mogorrón Huerta (2009, 2015) y Zamora Muñoz (2003).

2017). Asimismo, dicha clasificación ha sido aplicada con éxito a otros idiomas, como el catalán y el valenciano, por Josep Guia (1999) y por Sancho Cremades (1999); al gallego, en el centro Ramón Piñeiro; y se realizó también una aproximación a la lengua italiana llevada a cabo por Jorge Jesús Leiva Rojo (2005).

Es precisamente la clasificación de Corpas Pastor la que define el marco teórico del presente estudio. La autora divide el universo fraseológico en tres esferas. La primera está constituida por las colocaciones, que la autora considera unidades fraseológicas (UF) fijadas en la norma, que no constituyen actos de habla completos y que presentan aquella “propiedad de las lenguas por la que los hablantes tienden a producir ciertas combinaciones de palabras entre una gran cantidad de combinaciones teóricamente posibles” (Corpas Pastor 1996: 82).

Las locuciones componen la segunda esfera. Se trata de unidades fijadas en el sistema que tampoco son actos de habla y que están caracterizadas por su “institucionalidad, su estabilidad sintáctico-semántica y su función denominativa” (1996: 89), características que las distinguen de las combinaciones libres de palabras. Como toda unidad fraseológica, las locuciones están compuestas por al menos dos términos que no muestran unión ortográfica; de otra manera se trataría simplemente de palabras compuestas.

La tercera y última esfera a la que hace referencia Corpas Pastor en su manual está formada por los enunciados fraseológicos, unidades fraseológicas que, a diferencia de las colocaciones y las locuciones, representan actos de habla por sí mismos y presentan fijación interna y externa. La autora incluye dos subgrupos en los enunciados fraseológicos: las paremias y las fórmulas rutinarias; e insiste en que la separación entre una y otra no es neta, y en que incluso resulta difícil a veces distinguirlas de las locuciones.

Otra taxonomía que se destaca en este breve recorrido es la que Leonor Ruiz Gurillo (1997) presenta en sus *Aspectos de fraseología teórica española*, obra fundamental, entre otras razones, porque se articula en torno a los conceptos de centro y periferia dentro del sistema de la lengua propuesto por la Nueva escuela de Praga³. En particular, la autora señala que las zonas periféricas de la lengua representan un área dinámica más propensa al cambio, y esta

3. Para más información sobre la denominada escuela de Praga, véase Vachek (1967).

transformación puede realizarse en direcciones divergentes, ya que puede producirse una atracción de los elementos periféricos hacia el centro del sistema o, por el contrario, pueden alejarse hasta desaparecer. Ruiz Gurillo considera que las clases nucleares, es decir, las que presentan mayor fijación e idiomática, no poseen una variación significativa, como, en cambio, sucede con las clases periféricas, que reflejan una gran riqueza léxica. Establece, por lo tanto, una escala gradual que va de las locuciones totalmente fijas e idiomáticas hacia aquellas con un alto grado de idiomática (entre ellas, las combinaciones semifijas). En este modelo, la progresión de continuidad viene establecida fundamentalmente por la fijación y la idiomática: cuanto mayor sea la idiomática de la UF, mayor será su fijación y, por lo tanto, mayor el acercamiento al núcleo.

Finalmente, la última clasificación que se presenta es la que García-Page expone en su *Introducción a la fraseología española. Estudio de las locuciones* (2008). El autor se adscribe a la visión estrecha de la fraseología, es decir, considera que las locuciones son las únicas UF que entrarían con todo derecho en el universo fraseológico, puesto que las colocaciones pertenecen a un fenómeno propio de la sintaxis y las paremias y los refranes, en línea con Casares (1950) y Zuluaga (1980), son objeto de estudio de la paremiología, por lo que tampoco tendrían un lugar en los estudios de fraseología. En este sentido, su taxonomía se ciñe a las locuciones, que clasifica en ocho categorías según su función oracional, a saber: locuciones nominales, pronominales, adjetivales, adverbiales, prepositivas, conjuntivas, verbales y oracionales.

2.2. La fraseología en Colombia

Los estudios de fraseología en Colombia tomaron impulso a partir de los trabajos de Zuluaga Ospina, que ha venido ocupándose del análisis de las expresiones fijas, tanto a nivel teórico como práctico, en trabajos como *Estudios generativo-transformativistas de las expresiones idiomáticas* (1975a) y *La fijación fraseológica* (1975b). Posteriormente, publica su tesis doctoral *Introducción al estudio de las expresiones fijas* (1980), en la que ofrece un análisis y clasificación de las expresiones fijas o idiomáticas que hace hincapié en las locuciones.

Francisco Octavio Zuluaga Gómez continuó con la línea de trabajo trazada por Zuluaga Ospina. Su tesis de doctorado *Análisis pragmalingüístico de las unidades fraseológicas en español con atención especial a los refranes* (2003) está dedicada al análisis pragmático de dichos, locuciones, frases hechas, frases proverbiales, refranes, fórmulas de cortesía, entre otras unidades, desde una perspectiva que integra la sintaxis, la semántica y la pragmática, enfatizando este último aspecto, que según el autor no ha recibido un análisis pormenorizado. En palabras de Zuluaga Gómez:

“El uso frecuente de todo tipo de expresiones fijas en el lenguaje cotidiano y su presencia en la obra literaria, en el discurso periodístico y en otro tipo de discursos, permite afirmar que las UFF estén lejos de ser un fenómeno marginal o de uso restringido a determinados niveles de lengua” (2003: 1).

En 2004, el autor publica “Locuciones, dichos y refranes sobre el lenguaje: unidades fraseológicas fijas e interacción verbal”. En dicho artículo, Zuluaga Gómez plantea que las locuciones y refranes sobre el lenguaje se pueden considerar como fórmulas metalingüísticas en lenguaje corriente (fomelcos) que funcionan como instrucciones pragmáticas de la lengua. Según la función específica que desempeñan en la práctica discursiva y en la interacción, presenta una clasificación en tres categorías: descriptivas, instructivas y regulativas. Con el objetivo de ilustrar este planteamiento analiza algunas muestras siguiendo un modelo de entrada lexical en el que se explica la estructura semántica y la función pragmática de las expresiones, y concluye que las fomelcos son un registro del conocimiento que tiene el hablante de la racionalidad que rige las relaciones de uso del lenguaje.

Finalmente, en una reciente publicación editada por el Instituto Caro y Cuervo y realizada en coautoría con Viviana Díaz Orozco y Lirian Astrid Ciro, que lleva por título *Dichos, refranes y locuciones en las novelas de Gabriel García Márquez* (2018), el autor estudia la fraseología en la narrativa del escritor colombiano. Dicho trabajo es de particular interés porque, según los autores,

“los estudios de fraseología aplicados a obras literarias conforman un campo de investigación fecundo y están prácticamente sin explorar en el caso de obras literarias colombianas”.⁴

4. Para más información véase: <https://www.caroycuervo.gov.co/Noticias/dichos-refranes-y-locuciones-en-las-novelas-de-gabriel-garcia-marquez>.

De otra parte, cabe mencionar el *Lexicón de fraseología del español de Colombia* (1996), de Siervo Mora Monroy, investigador del Instituto Caro y Cuervo. En él, el autor elabora una revisión de la fraseología de los cinco principales diccionarios de colombianismos existentes en ese momento con el objetivo de explorar los usos actuales de la fraseología empleada en Colombia.

Otro trabajo de interés para presentar los estudios de fraseología colombiana es el *Estudio gramatical de las locuciones verbales con doble pronombre clítico*, de Orduña López (2011). En él, el investigador describe y clasifica desde una perspectiva gramatical un conjunto de locuciones verbales que en su estructura presentan dos pronombres clíticos. A través de dichas locuciones verbales, recogidas del corpus lexicográfico del macroproyecto de investigación *Diccionario descriptivo del español del Valle de Aburrá*, el autor analiza la estructura de este tipo de unidades léxicas en relación con diversos aspectos gramaticales (forma y función de los clíticos, subcategoría del verbo, argumentos y adjuntos del verbo y de la locución).

Entre las principales conclusiones, Orduña López identifica la necesidad de armonizar el concepto de locución y propone restringirlo a las unidades fraseológicas que cumplen una función sintáctica dentro de la oración. Por otra parte, deja abierto el debate sobre el carácter fraseológico de las locuciones formadas por un verbo y dos pronombres clíticos, mientras que, en cuanto a los clíticos no reflexivos, destaca la predominancia del clítico dativo, al que le sigue el acusativo femenino en su forma singular. Finalmente, subraya que las locuciones con doble clítico presentan una gran variedad estructural en relación con los argumentos que aparecen ligados tanto al verbo de la locución, nivel interno de análisis, como a la locución *per se*, nivel externo de análisis.

Arboleda (2012) publica “La intensificación de las locuciones de la prensa escrita de Medellín: una aproximación pragmática”. En ella, presenta la descripción de las locuciones y su función pragmática dentro del discurso periodístico de la prensa escrita de Medellín. Partiendo de algunos fundamentos teóricos en el ámbito sintáctico y semántico analiza las locuciones desde el punto de vista pragmático teniendo en cuenta la intensificación como una estrategia comunicativa de producción.

Por otra parte, en 2012 Adriana María Ortiz Correa publica “Las locuciones y sus variantes en el Diccionario descriptivo del Valle de Aburrá”, y

observa que el fenómeno lingüístico de la variación es propio de algunas locuciones. En particular, observa que se producen cambios o alteraciones de carácter morfológico, gramatical, sintáctico o léxico en una gran cantidad de locuciones registradas en el *Diccionario Descriptivo del Valle de Aburrá* (DEVA). El concepto de variación cuestiona el carácter fijo propio de las locuciones y plantea la discusión sobre la posibilidad constante de creación por parte de los hablantes y la interpretación que ellos mismos hacen de sus significados en relación con los contextos discursivos de emisión.

En este sentido, la investigadora concluye que las locuciones pertenecen a la norma lingüística social y no se explican a nivel del sistema, ya que son fijadas por el uso que hacen los hablantes. Las variaciones son necesarias porque son adaptaciones que hacen los hablantes a las necesidades del discurso y pueden producir diferentes efectos discursivos de significado, como proporcionar expresividad a un enunciado, lograr la comprensión del mensaje por parte del receptor, aclarar una situación, etc.

Por otra parte, Mira (2011) publica “Función pragmática de las unidades fraseológicas en Rosario Tijeras”, en el que presenta un análisis pragmático desde el punto de vista de la teoría de la cortesía de un corpus de unidades fraseológicas extraídas de una novela con alto contenido de lenguaje coloquial. Se analizan las unidades en contexto contrastándolas con los conceptos teóricos relacionados con las estrategias de cortesía utilizadas, las máximas conversacionales observadas y los actos de habla presentes en los enunciados. El resultado del análisis permite concluir, en primer lugar, que los principios universales de la cortesía se mantienen en las interacciones conversacionales analizadas y, en segundo lugar, que las estrategias de cortesía varían y están estrechamente condicionadas por el contexto social y situacional de las interacciones, ya que el tipo de UF utilizadas se consideraría descortés en otro contexto, pero, en el que refleja la novela, estas contribuyen al vínculo de identidad y pertenencia grupal, de tal manera que los miembros del grupo no las perciben como descorteses.

Villegas (2012) publica “Unidades fraseológicas en Aire de tango de Manuel Mejía Vallejo”. En este ensayo se encuentran dichos, refranes y otras unidades fraseológicas propias de la cultura antioqueña de la época. Se resalta el estudio de las unidades fraseológicas fijas presentes en la obra, las cuales se cotejaron con tres diccionarios de fraseologismos, y a la vez se verifica

su actualidad a través de una encuesta realizada a 50 personas de distintas edades. Algunas de las UF reunidas y clasificadas no aparecen en los diccionarios, pero los hablantes las usan en la actualidad, y otras han caído en desuso. Asimismo, se constata la permanencia en el habla urbana de la tradición popular del campo, expresada en el conocimiento y el uso de UF en el intercambio verbal de los hablantes.

Lopera Medina (2017) publica “Las unidades fraseológicas en los cumplidos en hablantes universitarios de Medellín, Colombia”, ensayo en el que analiza la fraseología en los cumplidos. La muestra se recogió mediante cuestionario aplicado a la población universitaria de Medellín. Los cuestionarios presentaban situaciones simuladas para que los informantes produjeran y respondieran cumplidos. Para el análisis se verificaron las UF en el *Diccionario de Parlache*, el *Diccionario Fraseológico Documentado del Español Actual* y el DRAE. Además, se utilizó el programa lingüístico computacional Cratilo para buscar frecuencias y arrojar información estadística de algunas de las UF. Los resultados indican que las locuciones adverbiales y la fórmula rutinaria *muchas gracias* aparecen con mayor frecuencia en los cumplidos y que las UF se presentan más en las respuestas que en la producción de los cumplidos.

Finalmente, con el objetivo de ofrecer un amplio estado de la cuestión con respecto a los estudios realizados sobre la fraseología empleada en Colombia se presenta el *Diccionario de fraseología usada en Colombia* (2015), aproximación fraseográfica elaborada a partir de la recolección de datos de algunos estudiantes de la Facultad de Educación de la Universidad de La Salle. Esta recopilación se presenta como un esbozo de diccionario en línea en el que se ofrecen más de 300 unidades fraseológicas empleadas en el país.

3. Fraseología y terminología

La fraseología es una disciplina estrechamente relacionada con la terminología y la lexicografía, puesto que su aplicación no se limita exclusivamente a las lenguas especializadas, sino que incluye también la lengua común, es decir, la que se utiliza en la vida cotidiana. Es por tanto indispensable, en primer lugar, disponer de sólidas nociones en terminología para poder elaborar una metodología cónsona con el estudio.

Silvia Padel y Diane Nolet (2001: xvii), en su *Manual de terminología*, definen la terminología como la “disciplina lingüística dedicada al estudio científico de los conceptos y términos utilizados en las lenguas de especialidad”. Por ende, es una ciencia que forma parte de la lingüística aplicada y que comprende el trabajo en lexicografía especializada, traducción, redacción y enseñanza de idiomas. Dicha disciplina está estandarizada por la International Organization for Standardization (ISO), y una de las publicaciones principales al respecto es la ISO 704:2000, Terminology work. Principle and methods. En este documento, la organización internacional proporciona una estructuración y estandarización de los elementos esenciales para el trabajo cuantitativo en terminología (ISO 704 2000: vi). Este protocolo

establishes and harmonizes the basic principle and methods for preparing and compiling terminologies both inside and outside the framework of standardization [...] is applicable to terminology work in scientific, technological, industrial, administrative, and other fields of knowledge (ISO 704 2000:1).

Dado su rigor académico, la elaboración de la presente metodología y, en particular, la estructuración del fichero fraseológico ha seguido las pautas proporcionadas por la ISO como cimientos para la construcción del fichero empleado en la anotación de las unidades fraseológicas usadas en Colombia y que se define a continuación.

4. Metodología

El diseño metodológico elaborado para el presente estudio incluye dos pasos que conllevan al desarrollo de una investigación exploratoria con rasgos descriptivos y enfoque cualitativo. A saber:

1. Recopilación de más de 1400 unidades fraseológicas empleadas en el español de Colombia.
2. Anotación de las unidades a través de un fichero fraseológico monolingüe.

A continuación, se describen pormenorizadamente las dos fases del presente diseño metodológico, así como la elaboración del fichero o expediente fraseológico monolingüe, es decir, la consignación de los datos recabados de la muestra para informar sobre sus características (Padel y Nolet: 2001).

4. 1. Creación del corpus objeto de estudio

El corpus *ad hoc* que se conformó está compuesto por más de 1400 locuciones. Para la recolección de los datos se recurrió al portal de Internet del *Diccionario de americanismos* (Asociación de Academias de la Lengua Española, 2010) y se usaron los archivos digitales del *Diccionario de colombianismos* (Instituto Caro y Cuervo 2015). Asimismo, se utilizaron datos para el corpus de algunos periódicos digitales como *El Tiempo* y *El Espectador*, principalmente.

4.2. Anotación de las unidades en el fichero fraseológico

La compilación de un fichero fraseológico exige rigor en el método de consignación de las unidades recabadas de un corpus. Dada la conexión formal previamente definida entre fraseología y terminología, el presente estudio toma como modelo las pautas establecidas por la International Standard Organization, en particular en dos estándares internacionales: Terminology work - Principles and methods (ISO 704:2000, segunda edición) y Terminology work - Vocabulary (ISO 1087:2000, primera edición). Dichos estándares han guiado la compilación del fichero fraseológico para examinar la fraseología colombiana. En la anotación de las UF se han respetado las siguientes pautas ofrecidas por la ISO 704:

1. Las unidades fraseológicas se anotan en minúscula.
2. Los sustantivos en número singular, excepto si su forma canónica es en plural, y en género masculino.
3. No se anotan los artículos.
4. Los verbos en su forma infinitiva.
5. Se respeta el orden natural de aparición de los componentes que conforman la UF.
6. Se emplea la correcta ortografía.

La anotación de las locuciones en el fichero fraseológico ha adoptado el principio uninocional, criterio fundamental de los estudios terminológicos según el cual a cada entrada corresponde un solo concepto. De este modo, las variantes fraseológicas no son integradas en la misma entrada, sino que disponen de una entrada propia. Las entradas han sido registradas según su disposición alfabética (ISO 1087-a 2000: 13), es decir, en función del orden de

la letra del alfabeto con la que empieza la unidad. La herramienta elegida para elaborar el fichero fue Microsoft Excel, herramienta ampliamente empleada en terminología al ser más versátil que otras y personalizable a varios niveles.

Tras definir las pautas empleadas para compilar el fichero fraseológico, se detalla a continuación la composición del mismo y los parámetros estudiados en cada unidad fraseológica. El fichero fraseológico está compuesto por diez columnas: en primer lugar, la columna ‘Unidad fraseológica’, en la que se han anotado las UF siguiendo las pautas previamente descritas. La segunda columna, ‘Definición’⁵, proporciona una explicación concisa y clara del concepto (Pavel y Noet 2011: 26) de la unidad analizada junto con su referencia. Con el objetivo de proporcionar una definición adecuada a cada unidad, se ha recurrido en primer lugar al *Diccionario de la Lengua Española (DLE)* y, cuando este no cumplía con los requisitos de concisión y claridad o no contenía la unidad, se ha elaborado una definición propia. La tercera columna es el ‘Contexto’; en ella se proporciona un contexto eficaz y se anota su referencia. Con respecto a la transcripción de los contextos, se ha recurrido al manual del usuario de IATE2⁶, disponible *online* en la página web de la base de datos:

A context (i.e. a short quotation from a reliable source that illustrates the use of a term) is always useful, particularly if your entry does not have a definition. The use of specific formatting (e.g. using quotation marks or not, putting the relevant term in bold, etc.) should be agreed upon at interinstitutional level by each language community. It is fine to use a context containing an inflected or plural form, or a context where a term composed of more than one word is split in the running text. (IATE2 User Handbook 2019: 55-56)

La cuarta columna, ‘Función oracional’, describe el tipo de unidad fraseológica en relación con la taxonomía propuesta por Corpas Pastor (1996) y especifica el nivel gramatical que la locución desempeña en la lengua (verbal, nominal, adverbial, adjetival, prepositiva, conjuntiva o clausal). La quinta

5. De acuerdo con Pavel y Noet, se entiende por definición la “fórmula lexicográfica que explica el concepto designado por un término” (2001: 124).

6. IATE (Terminología Interactiva para Europa) nace en 1999 y, en la actualidad, es la mayor base de datos terminológica del mundo, con más de ocho millones de términos en las 24 lenguas oficiales de la UE. IATE es coordinado por el Centro de Traducción de los Órganos de la Unión Europea. Para más información véase: <<https://iate.europa.eu/home>> y <<http://cdt.europa.eu/es/iate-es>>.

columna identifica el ‘Campo semántico’ al que pertenece el constituyente léxico principal que compone la unidad. En este sentido, se ha optado por agrupar las unidades en función de uno de sus componentes léxicos en lugar de priorizar el significado holístico de la unidad⁷. Para ello se ha recurrido a la clasificación lexicográfica elaborada por Sánchez Lobato y Aguirre Beltrán en el *Léxico fundamental del español* (1992), en la que los autores proponen quince temas o nociones para agrupar el léxico o, en este caso, las unidades fraseológicas, en función de los siguientes campos asociativos:

1. Identificaciones personales.⁸
2. Circunstancias personales: lugar de residencia.
3. Profesión, trabajo, creencias.
4. El paisaje, la fauna.
5. Condiciones atmosféricas.
6. Entretenimientos, aficiones.
7. Actividades artísticas, la prensa.
8. Aficiones: deportes y toros.
9. Los viajes, el alojamiento y el transporte.
10. La salud, la enfermedad y la figura humana.
11. La educación y la enseñanza.
12. La alimentación, la comida y la bebida.
13. El comercio, la industria, el vestido.
14. La moneda, el correo, la sociedad y el estado.
15. Pesos y medidas: los números.

Cabe destacar que el presente trabajo se nutre de las aportaciones sobre disponibilidad léxica elaboradas por Georges Gougenheim, Paul Rivenc, René Michéa y Aurélien Sauvageo (1964). En este sentido, se ha demostrado la

7. Este procedimiento es considerado como una alternativa a la agrupación por significado unitario de las unidades y es avalado por Mario García-Page en *VI Congresso Internazionale di Fraseologia e Paremiologia Phrasis Fraseologia e paremiologia: Modelli, strumenti e prospettive* (2019, actas en prensa).

8. Este campo semántico recoge los vocablos correspondientes a: el nombre, la persona y la edad, el estado civil, la familia, las actividades intelectuales y los sentimientos

dificultad de poder encajar todo el léxico y, por ende, la fraseología, en un número reducido de campos semánticos.⁹

El sexto y el séptimo parámetro analizados por cada UF siguen las prescripciones de Penadés Martínez (2015: 265) y corresponden respectivamente a la ‘Situación comunicativa’ de las unidades fraseológicas, que la autora distingue entre formal e informal, y al ‘Registro’, neutro y vulgar. Para ello, se ha recurrido a la 23ª edición del *Diccionario de la Lengua Española* (DLE) y, en particular, se aplicaron las marcas lexicográficas propuestas por el diccionario¹⁰ a las prescripciones de Penadés Martínez (2015). En este sentido, se contrastaron las unidades extraídas con el corpus del DLE con el objetivo de asignar un parámetro objetivo a la situación comunicativa y al registro de las unidades¹¹. En el caso de que el diccionario no presentase ninguna marca lexicográfica para una UF, los autores optaron por asignar dichos valores de acuerdo con su experiencia y su conocimiento nativo de la variedad colombiana del español.

La octava característica que se estudió en cada locución ha sido definida como ‘Presencia en el DLE’; en este sentido, se ha anotado si la UF está recopilada en la 23ª edición del *Diccionario de la Lengua Española* (DLE), fruto de la colaboración de las veintidós corporaciones integradas en la Asociación de Academias de la Lengua Española (ASALE), cuya edición electrónica recoge las modificaciones aprobadas por todas las academias en 2018.

La novena característica estudiada en cada UF corresponde a las ‘Variantes’. Para la definición e identificación de las variantes se ha seguido a Zuluaga, quien afirma que “son parcialmente idénticas en su estructura y en sus componentes” (1980: 109), así como a Mellado Blanco, para quien “las variantes fraseológicas se reconocen porque el número de elementos que no varían en ellas supera al de elementos que varían” (2004: 91). Por otra parte, se ha

9. Para profundizar en la disponibilidad léxica véase: Zuccalà (2014).

10. Véanse los epígrafes 2.4. y 5.3.2 del DLE, “Advertencias para el uso del diccionario”: https://www.rae.es/diccionario-de-la-lengua-espanola/que-contiene/item-numero-2#_Toc85519289.

11. Se estableció la siguiente correspondencia: el registro vulgar y el neutro de nuestro corpus corresponden respectivamente a las marcas “vulg.” (vulgar) y “coloq.” (coloquial) del DLE; la situación comunicativa formal y la informal corresponden respectivamente a “cult.” (culto) y “jerg.” (jergal) del DLE.

decidido emplear la clasificación de Corpas Pastor para identificar y distinguir las que son variantes de las variaciones¹². La autora (1996: 28-30) divide las variantes en dos, a saber:

1. Variantes estructurales: variaciones concernientes al uso de preposiciones, artículos, número y orden de constituyentes, formas acortadas de los constituyentes o número gramatical de los mismos, que no cambian la organización interna de las unidades fraseológicas (*irse de/a pico pardos*).
2. Sinónimos estructurales: se distinguen por la congruencia estructural completa o la identidad parcial del componente léxico (*poner a uno las peras a cuatro/ocho*).

Para ejemplificar tomando una muestra del corpus objeto de estudio, las unidades *meter la mano al fuego* y *meter la mano a la candela* serán variantes, en cuanto presentan “congruencia estructural completa” así como “identidad parcial del componente léxico” (Corpas Pastor 1996: 28) En cambio, no serán consideradas variantes las locuciones *comer como un pajarito* y *vivir del aire*, dos unidades diferentes que expresan el mismo concepto: comer poco.

Como se ha mencionado anteriormente, la inserción de las variantes sigue también el principio uninocional, según el cual a cada entrada corresponde un solo concepto. En este sentido, las variantes de una locución han sido anotadas, por un lado, como entradas independientes y, por otro, añadidas en esta columna de manera especular: si una entrada dispone de dos variantes, cada variante será a su vez una entrada *per se* y cada una de ellas tendrá a su vez dos variantes. Finalmente, la última columna está dedicada a los ‘Comentarios’ que los autores de esta publicación han considerado relevantes en el análisis de las UF

12. Según la autora, no son variantes sino variaciones: 1. Las variaciones por derivación (*ser un culo/culillo de mal asiento*); 2. Las variaciones por transformación (*metedura de pata a partir de meter la pata*); 3. Las variaciones por antónimos estructurales (*cabeza arriba y cabeza abajo*); 4. Las variaciones diatópicas (como la expresión española *hacer dedo* y la hispanoamericana *ir al dedo*); 5. Las variaciones diafásicas como *tirar dedo* (plano neutro) e *ir al dedo* (plano informal); 6. Las UF con casillas vacías como *¡a tu (su, vuestra, nuestra, etc.) salud!*; 7. Las modificaciones creativas, personales de cada hablante y directamente proporcionales al grado de fijación de las mismas (Corpas Pastor 1996: 29).

Para finalizar el proceso de transferencia de las UF se han respetado los pasos de la verificación de términos estandarizados por ISO 1087-2:2000, Terminology work - Vocabulary - Part 2: Computer applications, en materia de Gestión de la terminología y Control de calidad. Dichos pasos se realizan con el objetivo de detectar errores *a posteriori* en la base de datos elaborada y se resumen a continuación:

1. Double-entry check: “Data validation to determine whether the same data entry has been stored more than once” (ISO 1087-2), es decir, verificar que el mismo término no se haya registrado dos o más veces. Asimismo, se ha empleado la herramienta de extracción terminológica SynchroTerm para identificar eventuales dobles entradas y eliminar toda repetición.
2. Consistency check: “Data validation to check whether interdependent entries comply with predefined conditions” (ISO 1087-2). En este sentido, se ha verificado que no hubiese referencias cruzadas entre las diferentes entradas en la base de datos elaborada.
3. Spelling check: “Data validation to determine that all words comply with predefined spelling rules” (ISO 1087-2). En la transcripción manual, aunque también en la automática, son frecuentes las erratas. Además, es preferible utilizar un único sistema ortográfico de acuerdo con los cánones sugeridos por la Academia de la lengua objeto de estudio. Aun así, pueden darse casos de localismos o regionalismos que tienen que ser marcados como variantes y destinarse a una entrada específica.
4. Completeness check: “Data validation to see whether data are presented where required” (ISO 1087-2). Debido a las características de la presente investigación, no ha sido necesario definir ninguna de las categorías que caracterizan el concepto (ej. definición, género, número, contexto, fuente, etc.). Por ello no se ha adoptado este criterio de evaluación, pues es innecesario evaluar tanto la información gramatical como la información semántica de las entradas.
5. Format check: “Data validation to determine whether data conform to a specific format” (ISO 1087-2). Asegurarse de que los datos sean

transferidos en la columna correcta de la base de datos es fundamental para el desarrollo de cualquier investigación.

6. Plausibility check: “Data validation to determine whether data conform to a specified criteria (ISO 1087-2). En la redacción de una base terminológica común es necesario confirmar que las categorías gramaticales (género, número, etc.), las definiciones, las fuentes y todo el aparato descriptivo siguen las reglas que se han establecido para la constitución del modelo terminológico empleado.

5. Resultados

A continuación, se presentan los principales resultados de la investigación desde una perspectiva sintáctica y semántica. En particular se detallan los resultados obtenidos con respecto a las locuciones nominales, adjetivas, adverbiales, clausales y verbales. Asimismo, con respecto a esta última clase, se definen aparte las locuciones verbales que se conforman con el verbo ‘dar’. Finalmente, se describen las locuciones que componen el corpus desde un punto de vista semántico en función de los 15 campos semánticos utilizados.

5.1. Locuciones nominales

Por lo general, estas locuciones tienen las siguientes estructuras: S + Prep + S y S + Adj. En la base de datos encontramos las siguientes: *arroz con mango* (lío, problema, en la Costa Atlántica), *burro con plata* (persona ignorante, pero con dinero), *callejón sin salida* (conflicto sin solución), *golpe de ala* (*colq.* mal olor en las axilas), *golpe de gracia* (rematar), *lágrimas de cocodrilo* (simular una pena inexistente), *mano dura* (severidad en el mando o en el trato personal), *policía acostado* (*colq.* resalto, reductor de velocidad).

5.2. Locuciones adjetivas

La estructura de este tipo de locuciones, por lo general es la siguiente: Adj (participio) + Prep + S, como *duro de corazón* (que nada le compadece ni lastima, insensible), *chapado a la antigua* (anticuado). Por otra parte, dentro de este grupo destacan las locuciones que indican comparaciones estereotipadas y que están constituidas por la estructura Adj + como +S: *fuerte como un roble*, *fuerte como un toro* (muy fuerte), o Adv + Adj + SN: *más suave que la*

seda (muy suave), *más cansón que novia fea*. Por otra parte, se observan unidades formadas por SP compuestos por la preposición *de* que funciona como núcleo sintáctico y su sintagma nominal o determinante correspondiente que actúa como un elemento clausal: *de medio pelo* (de baja calidad), *de ambiente* (sociable), *de perros* (de ínfima clase), *de pelo en pecho* (vigoroso, varonil), etc.

En las locuciones adjetivas se encuentran estructuras como las siguientes: Prep + N, aunque también hay estructuras conformadas por Prep + Art + N, Prep + N + Adj, Prep + Adj. Veamos algunos ejemplos: *a la francesa*, *de ataque*, *de buenas*, *de malas*, *de quinta* (de mala calidad), *de un cacho* (a punto de terminar), *en carne viva* (sin piel), *en blanco y negro* (que no reproduce los colores), *en la olla* (en malas condiciones), *común y corriente* (usual, habitual), *ligero de cascos* (irreflexivo), *sin igual* (sin par), *sin compromiso* (libre), *sin oficio* (sin trabajo, ocioso), etc.

5.3. Locuciones adverbiales

En el corpus analizado se ha observado cómo las locuciones adverbiales portan distintos valores referenciales (Corpas Pastor 1996). En este sentido, predominan las de modo, *a ciegas* (ciegamente), *a pie* (caminando), *a escondidas* (en secreto) y, en menor cantidad, las de tiempo, *al instante*, *a primera hora*, *a deshoras*, y de lugar, *al lado*, *a lo lejos*, *por detrás*, *de cerca*.

Desde un prisma sintáctico, las locuciones adverbiales que conforman el corpus presentan la siguiente estructura: Prep + N, Prep + Art + N, Prep + Art + Adj, Prep + N + Prep + N: *en picada* (hacia abajo y a gran velocidad) *a conciencia* (con empeño), *a la brava* (bruscamente), *a la guachapanda* (descuidadamente), *a la lata* (en gran cantidad), *a lo desgualeado* (en gran cantidad, sin medida), *a morir* (mucho), *a palo seco* (sin anestesia), *a precio de huevo* (muy barato), *a vuelo de pájaro* (rápidamente, superficialmente), *de golpe* (súbitamente), *de papayita* (colq. oportunamente), *de un solo tacazo* (de una vez), etc.

Hay que destacar que un número significativo de locuciones adverbiales denotan rapidez o velocidad y su estructura comienza por las preposiciones *a* o *en*: *a toda* (con gran velocidad), *a toda prisa*, *a las carreras*, *a toda mierda* (vulg.), *a toda mecha* (inf.), *a marchas forzadas*, *a toda máquina*, *a toda prisa*,

a vuelo de pájaro (rápidamente) al galope, al trote, en bombas (inf. rápido), en par patadas (inf.), en un abrir y cerrar de ojos, en un dos por tres, en un ya.

Finalmente, se presenta a continuación una muestra de locuciones adverbiales de uso frecuente extraídas del corpus: a borbotones, a brazo partido, a cada instante, a cada momento, a cada rato, a calzón quitado, a cántaros, a carcajadas, a carta cabal, a ciegas, a ciencia cierta, a conciencia, a crédito, a cuerpo de rey, a destajo, a destiempo, a diestra y siniestra, a discreción, a disgusto, a distancia, a domicilio, a duras penas, a empujones, a escondidas, a fin de cuentas, a flor de piel, a flote, a fondo, a fuego lento, a gatas, a golpes, a gran escala, a grandes rasgos, a granel, a grito herido, a gusto, a hombros, a hurtadillas, a intervalos, a la antigua, a la brava, a la buena de (mi) Dios, a la deriva, a la fuerza, a la lata, a la intemperie, a la izquierda, a la larga, a la ligera, a la moda, a la parrilla, a la perfección, a la plancha, a la postre, a la sombra, a la(s) carrera(s), a las buenas, a las claras, a las malas, a lo grande, a lo lejos, a lo loco, a lo mejor, a lo moderno, a lo pobre, a lo salvaje, a lo sumo, a mano, a mano armada, a manos llenas, a mansalva, a media noche, a media voz, a menudo, a montones, a morir, a muerte, a ojo, a pata, a patadas, a pedir de boca, a pequeña escala, a pesar de los pesares, a pico de botella, a pique, a plomo, precio de huevo, a presión, a pulso, a punta de, a puntapiés, a quemarropa, a rastras, a ratos, a raudales, a regañadientes, a rienda suelta, a sabiendas, a simple vista, a sorbos, a sueldo, a todas luces, a todo full, a tontas y a locas, a (los) trancazos, a trechos, a tuta, a veces, a voluntad, a voz en cuello, al ajillo, al azar, al clima, al contado, al dedillo, al descubierto, al detal, al fin y al cabo, al natural, al ojmetro, al pelo, al peluche, al pie de la letra, al por mayor, al rape, al rato, al rompe, al unísono, al vapor, al vuelo, con bombos y platillos.

5.4. Locuciones verbales

Las locuciones verbales constituyen la clase más más numerosas en el corpus analizado. En general, presentan una estructura de tipo V + SN: *gastar pólvora en gallinazos* (esforzarse por algo sin importancia), *alzar el codo* (colq. consumir bebidas alcohólicas), *ganar terreno* (conseguir ventaja en algo), *pagar el pato* (sufrir un castigo no merecido), *perder los estribos* (encolerizarse, obrar fuera de razón por ira o enojo).

Asimismo, se registran locuciones verbales conformadas por V + SN que expresa modo, lugar, tiempo, etc., como *salir con un chorro de babas* (no cumplir las expectativas generadas). Mención aparte merecen las locuciones que se conforman con los verbos *dar* en cuanto que presentan características definitorias de esta clase.

Cabe destacar que un número representativo de locuciones verbales se constituye en negativo de manera cristalizada. De estas, el constituyente verbal que predomina es el verbo 'tener' seguidos por, en términos de frecuencia, los verbos 'ser', 'dar', 'caber' y 'poder': *no caber duda* (con toda seguridad), *no caber en la cabeza algo* (no ser comprensible), *no caber en sí de gozo* (no caber de contento), *no caber ni un tinto* (estar un lugar o un recipiente lleno), *no casarse con nadie* (conservar la independencia de opinión o actitud), *no cocinarse en dos aguas* (no ser joven), *no dar abasto* (no bastar, no rendir lo suficiente), *no dar el brazo a torcer* (no ceder; no dejarse convencer), *no dar marcha atrás* (no retroceder), *no dar pie con bola* (equivocarse frecuentemente), *no dar puntada sin dedal* (hacer algo de manera calculada, en busca del propio beneficio), *no dar una* (no acertar en nada), *no decir ni mu* (permanecer en silencio), *no decir ni pío* (no decir palabra alguna), *no dejar lugar a dudas* (ser claro), *no dejar ni a sol ni a sombra* (perseguir de manera inoportuna a una persona, a todas horas y en todo sitio), *no dejar títere con cabeza* (desacreditar duramente, hablando o escribiendo, a un cierto número de personas sin excluir a ninguna), *no despegar los labios* (no hablar nada), *no dirigir la palabra a alguien* (no hablarle), *no entender ni papa* (nada), *no estar el palo para hacer cucharas* (estar en mala situación económica), *no estar en sus cabales* (estar fuera de juicio), *no haber manera* (no poder realizar algo pese a intentarlo insistentemente), *no haber más remedio* (resignarse ante algo que no tiene alternativa), *no matar ni una mosca* (ser inofensivo), *no pegar ojo* (no dormir), *no perder de vista*, *no perder ripio* (estar atento a una conversación), *no perderse ni la corrida de un catre* (ser curioso, preguntar todo), *no poder ver ni en pintura* (no soportar, odiar), *no poder* (no soportar), *no probar bocado* (no comer nada), *no quitarle los ojos de encima* (no dejar de mirar a alguien o algo), *no saber ni papa* (ignorar), *no saber ni pío* (ignorar), *no ser manco* (ser hábil), *no ser moco de pavo* (ser difícil de conseguir, tener valor), *no ser ni chicha ni limonada* (ser algo indefinido o impreciso), *no soltar prenda* (no decir alguien nada que pueda comprometerlo), *no tener arte ni parte* (ser un asunto totalmente

ajeno a alguien), *no tener desperdicio* (irónico, ser provechoso), *no tener dos dedos de frente* (carecer del más mínimo sentido común o de inteligencia), *no tener la película completa* (no tener toda la información), *no tener más remedio* (no tener alternativa), *no tener ni cinco* (no tener nada de dinero), *no tener ni idea* (no estar enterado), *no tener ni pies ni cabeza* (no tener sentido), *no tener nombre* (ser muy injusto), *no tener parangón* (ser único, no poderse comparar), *no tener para la yuca* (colq. no tener dinero), *no tener pelos en la lengua* (decir sin vergüenza o temor lo que piensa o siente), *no tener pierde* (contar con el éxito asegurado), *no tener precio* (ser muy valioso), *no tener sangre en las venas* (tener un carácter frío e impasible), *no tener un pelo de tonto* (ser una persona difícil de engañar con mentiras), *no tragar entero* (analizar con detalle los hechos), *no tragar a alguien* (sentir marcada antipatía hacia una persona), *no ver más allá de sus narices* (ver poco; no ser sagaz), *ser incapaz de matar una mosca* (ser inofensivo).

5.6. Locuciones verbales con el verbo 'dar'

Hay un número representativo de locuciones formadas con el verbo 'dar' que tienen diversas funciones y que presentan de manera frecuente la estructura dar + SN: *dar pasto* (ventaja), *dar pedal* (incitar), *dar lora* (sermonear), *dar lata* (fastidiar), *dar en el clavo* (acertar), *dar papaya* (exponerse, arriesgarse), *dar sopa y seco* (demostrar superioridad o suficiencia).

Desde un prisma semántico, se observa que los sustantivos que presentan entre sus significados el de 'golpe' son acompañados por el verbo 'dar' como constituyente verbal: *dar un cocotazo*, *dar garrote*, *dar leña*, *dar por la cabeza*, *dar por la nuca*, *dar palo*, *dar una garrotera*, *dar una tunda*, *dar un coñazo*, *dar un garrotazo*, *dar vara*, *dar un trancazo* (dar o recibir un golpe violento).

5.7. Campos semánticos

Al estudiar las 1400 locuciones en función del campo semántico (CS) al que pertenece el principal constituyente léxico de la unidad fraseológica, se observa que los más productivos son los siguientes: CS4 El paisaje, la fauna; CS10 La salud, la enfermedad y la figura humana; CS12 La alimentación, la comida y la bebida; CS13 El comercio, la industria, el vestido. En este sentido, los hablantes colombianos emplean más locuciones cuyo componente léxico

principal está asociado al cuerpo humano, los animales, la edad, estado civil, familia, sentimientos y la ropa.

Las locuciones semánticamente cercanas a la fauna presentan connotaciones socialmente consideradas como negativas en cuanto matizan defectos o actitudes personales como la ira, la fuerza bruta, la astucia, la lentitud, etc. Dichas locuciones son utilizadas en situaciones comunicativas informales en las que la intención del hablante es manifestar dichos sentimientos. En cambio, pocas locuciones destacan atributos positivos: *ser una hormiguita* (ser ahorrador y laborioso), *tener ojos de águila, de lince* (ser buen observador).

Asimismo, se ha observado que un número significativo de locuciones que pertenecen al CS10 de la salud, enfermedad y figura humana son utilizadas en situaciones comunicativas informales y expresan ayuda, interés por algo o alguien, esfuerzo, intromisión y diversas actitudes en las relaciones interpersonales.

A continuación, se presenta una muestra de locuciones que componen cada campo semántico:

CS1. *capar clase* (jerg. faltar a clase), *botar corriente* (pensar mucho), *cumplir quince años* (cumplir quince años), *llorar lágrimas de sangre* (llorar con dolor y arrepentimiento), *matar de hambre* (dar poco de comer), *morir(se) de hambre* (tener un hambre irresistible), *no cocinarse en dos aguas* (no ser joven), *peinar a alguien* (ganarle a alguien en una discusión con argumentos), *poner apodos* (poner nombre a una persona, tomado de sus defectos corporales o de alguna otra circunstancia), *poner sobrenombre* (apodar), *ponerse arrozado* (enojarse), *quedarse para vestir santos* (soltero), *quemarse las pestañas* (estudiar mucho), *ser hijo de papi y mami* (ser adinerado), *ser un tronco* (ser poco hábil realizando una actividad), *tener celos*.

CS2. *cortar los servicios* (suspender los servicios públicos por falta de pago), *echar rulo* (jerg. chismosear), *estar para recoger con cuchara* (agotado), *parar antena* (inf. escuchar), *ser cuchilla* (jerg.juv. ser estricto), *ser lámpara* (jerg.juv. alguien que quiere ser el centro de atención), *ser boleta* (jerg.juv. persona que se hace notar por su mal gusto).

CS3. *ganarse el pan* (sustentarse con el producto de su trabajo), *ganarse la papa* (inf. trabajar para sustentarse), *ganarse la vida* (trabajar), *matarse trabajando* (trabajar duro y mucho), *patear la lonchera* (inf. arriesgar el trabajo), *ponerse las pilas* (disponerse a emprender algo con energía y resolución).

CS4. *a vuelo de pájaro* (rápidamente, superficialmente), *colgar un mico* (artículo, párrafo o término que aparece en un texto parlamentario en el último momento), *como pez en el agua* (con comodidad o soltura), *creerse muy gallito* (furioso), *dormir como una foca/morsa* (profundamente), *estar como un tigre* (furioso), *estar como un titi* (furioso), *estar como una ballena/vaca* (gorda), *estar como una tatacoa* (furioso), *estar con el mico al hombro* (de mal humor), *estar de malas pulgas* (de mal humor), *hablar como cotorra* (mucho), *hacer conejo* (irse de un establecimiento sin pagar), *oler a mico/ tigre* (oler mal a causa del sudor), *pagar el pato* (sufrir un castigo no merecido), *parecer una jirafa* (alta), *puyar el burro* (*inf.* Apurarse, darse prisa, Costa Atlántica) *ser abeja* (astuto), *ser burro* (bruto; ignorante), *ser gallina* (*inf.* cobarde), *ser una hormiga, -ita* (trabajador), *ser una tortuga* (lento), *ser un burro* (*inf.* ser fuerte), *ser la oveja negra* (persona que, en una familia o colectividad poco numerosa, difiere desfavorablemente de las demás), *ser picaflor* (enamorado), *ser (un) sapo* (delator, chismoso), *ser un águila* (tener gran perspicacia y viveza, sobre todo en las cuestiones prácticas), *ser un pulpo* (acaparador; persona que toca a otra repetidamente con fines sexuales), *ser un zorro* (astuto), *ser un lince* (sagaz), *tener ojos de águila/ de lince* (tener vista aguda).

CS5. *Caerle la gota fría* (tener mala suerte), *cogerle la noche/tarde* (hacérsele tarde), *dar la ventolera* (entrar de repente ganas de hacer algo), *estar lloviendo hasta maridos* (*inf.* llover fuerte), *hacer su agosto* (hacer negocio, lucrarse, aprovechando ocasión oportuna para ello), *llover sobre mojado* (repetirse algo innecesario o enojoso).

CS6. *hacer show* (llamar la atención y causar asombro o escándalo), *hacer teatro* (llamar la atención causando asombro o escándalo), *hacerse un video/ una película* (*jerg. juv.* inventar historias para engañar sobre algo o alguien), *poner las cartas sobre la mesa* (poner de manifiesto un propósito u opinión que se guardaba oculto), *salírsele un gallo* (nota falsa y chillona que emite quien canta o habla), *ser un hombre orquesta* (realizar distintas actividades a la vez, sin perder el control y con destreza), *tocar el violín* (acompañar una persona a una pareja a salir a la calle).

CS7. *dárselas de artista* (ser gracioso), *mojar prensa* (salir alguien en la noticia de un periódico), *robar pantalla* (salir reiteradamente en una grabación para TV sin ser necesario).

CS8. *capotear la crisis* (eludir una dificultad), *capotear la tormenta* (eludir con habilidad), *capotear un problema* (eludir), *coger el toro por los cuernos* (afrontar los problemas), *dar pedal a alguien* (instigar), *echar los perros* (cortejar).

CS9. *acumular millas* (en algunas aerolíneas, acumular puntos los viajeros frecuentes, que pueden ofrecer una variedad de beneficios), *ir viento en popa* (con buena suerte, dicha o prosperidad), *irse a pique* (malograrse, perderse).

CS10. *a pedir de boca* (tal como se deseaba), *abrirle los ojos a alguien* (hacer que alguien se dé cuenta de algo que hasta el momento ignoraba), *agachar la cabeza* (obedecer sumisamente), *andar agachado* (encorvado; conforme ante una injusticia), *caer parado* (tener suerte en circunstancias adversas), *con pelos y señales* (detalladamente), *costar un ojo de la cara* (ser excesivo su precio), *dar la cara* (responder de los propios actos y afrontar sus consecuencias), *dejar con un palmo de narices* (incumplir una promesa a alguien, privándolo de lo que esperaba conseguir), *echar un pulso* (pulsear), *echar una mano* (ayudar), *en un abrir y cerrar de ojos* (en un instante, rápidamente), *estar alguien como quiere* (guapo), *estar con el culo al aire* (vulg. estar en una situación económica difícil), *estar pasando aceite* (persona mayor que está enferma), *hablar hasta por los codos* (mucho), *hacer de tripas corazón* (esforzarse para disimular el miedo, dominarse, sobreponerse en las adversidades), *irse de culo* (sorprenderse), *lamer el culo* (vulg. adular), *levantarse con el pie derecho* (empezar bien el día), *meter la pata* (hacer o decir algo inoportuno o equivocado), *meter las de caminar* (equivocarse), *morderse la lengua* (hablar de forma prudente sin expresar lo que realmente se siente), *no dar el brazo a torcer* (no ceder, no dejarse convencer), *no decir esta boca es mía* (permanecer en silencio), *no dolerle una muela* (tener buena salud; no tener problemas económicos), *no pegar el ojo* (no dormir nada), *no levantar cabeza* (estar muy atareado), *no mover un dedo* (no preocuparse ni esforzarse por nada ni nadie), *no tener ni pies ni cabeza* (no tener sentido), *patear el arpa* (poner en riesgo el empleo), *poner el dedo en la llaga* (conocer y señalar el verdadero origen de un mal, el punto difícil de una cuestión, aquello que más afecta a la persona de quien se habla), *poner el pecho* (afrontar las adversidades o peligros con valentía), *poner/prestar el hombro para llorar* (ser solidario), *ponerse con los pelos de punta* (sentir miedo o pavor), *quemarse las pestañas* (estudiar mucho), *romperse el culo* (realizar un trabajo rudo), *romperse la espalda* (realizar un

trabajo pesado), *sacarle el culo alguien* (vulg. evadir a una persona), *ser una teta* (vulg., estúpido) *ser uña y mugre* (tener una amistad estrecha), *tener a alguien entre ceja y ceja* (mirarlo con prevención desfavorable), *tener a alguien pasando aceite* (tener enamorado a alguien), *tener cerebro de gallina* (ser desmemoriado, olvidadizo), *tener la lengua larga* (murmurar), *valer un culo* (no tener importancia).

CS11. *tener palabra* (cumplir lo prometido), *adelantar cuaderno* (ponerse al día, generalmente con un buen amigo, en asuntos personales y de terceros), *perder el año* (no hacer las cosas bien y sufrir las consecuencias).

CS12. *comer callado* (jerg. guardar silencio por conveniencia), *cuidar la cuchara* (inf. cuidar el empleo), *tener malos tragos* (ponerse agresivo o molesto cuando alguien consume bebidas alcohólicas), *ser mala copa* (igual significado que la anterior), *salir más caro el caldo que los huevos* (a veces sale más caro el medio por el cual vamos a obtener algo, que el beneficio que realmente nos da), *estar rojo como un tomate* (sonrojado), *ser zanahorio* (ingenuo), *ser lenteja* (lento), *estar como un fideo* (muy delgado), *estar como un hielo* (frío), *estar como una uva/como un lulo* (muy bien, físicamente; estar bien de salud), *parecer un bagre* (feo), *pedir cacao* (rogar, suplicar), *ponerse trucha* (jerg. estar atento), *estar frito* (estar mal), *comer a la carta* (a elección del consumidor), *ser buena papa* (inf. buena persona), *ser buena muela* (inf. ser de buen apetito), *ser harina de otro costal* (no tener relación un asunto con otro).

CS13. *apretarse el cinturón* (reducir los gastos por escasez de medios), *chuparle las medias a alguien* (adular a alguien), *estar de moda*, *hacerse el de las gafas* (hacerse el desentendido), *ponerse los pantalones* (imponer su autoridad), *alzarse la bata* (salir a divertirse), *meterse en camisa de once varas* (inmiscuirse en lo que no le incumbe o no le importa), *ponerse la camiseta* (comprometerse a hacer algo), *ponerse en los zapatos del otro* (comprender sus problemas y necesidades), *pasar de moda* (perder actualidad), *ser una corbata* (tener un empleo bien remunerado con poco trabajo), *tener pantalones* (tener carácter), *tirar la casa por la ventana* (gastar dinero exageradamente con motivo de una celebración).

CS14. *tomar cartas en el asunto* (intervenir).

CS15. *no valer ni cinco* (nada), *en un dos por tres* (rápido).

6. Conclusiones

El análisis de más de 1400 unidades empleadas en Colombia ha permitido demostrar la riqueza fraseológica del país objeto de estudio. De hecho, la presencia de variantes y sinónimos estructurales (Corpas Pastor, 1996) observada en el fichero fraseológico se considera como un ulterior indicador de la riqueza fraseológica de dicho país.

El análisis sintáctico ha dejado patente que las locuciones más abundantes en el corpus objeto de estudio son de tipo verbal y presentan estructura de V + SN; las locuciones nominales más frecuentes son de tipo N + Adj. y N + Prep.+ N, mientras que las locuciones adjetivas más frecuentes se conforman por Adj. (participio) + Prep.+ N. Finalmente, las locuciones adverbiales más abundantes son las de modo y de tiempo y su estructura mayoritaria se refleja en Prep.+ N; Prep. + Art + N; Prep. + Art + Adj. y Prep. + N + Prep. + N.

Desde un prisma semántico, el análisis del corpus estudiado ha permitido corroborar que, agrupando las unidades fraseológicas en función del principal constituyente léxico, los campos semánticos más productivos son los relacionados con el cuerpo humano, los animales y el vestuario. Las locuciones del CS 4. El paisaje y la fauna vehicular, además, principalmente, conceptos considerados socialmente como negativos: *estar gordo*, *estar furioso*, *ser bruto*, etc. Estas locuciones son utilizadas en situaciones comunicativas informales en las que la intención del hablante es manifestar sentimientos de ira o mal humor.

Asimismo, las locuciones que pertenecen al CS 10. Salud, enfermedad y figura humana son utilizadas principalmente en situaciones comunicativas informales y expresan ayuda, interés por algo o alguien, esfuerzo, intromisión y diversas actitudes en las relaciones interpersonales, como se puede observar en *a pedir de boca* (tal como se deseaba), *a espaldas de alguien* (aprovechando su ausencia, sin que se entere), *tener a alguien entre ceja y ceja* (mirarlo con prevención desfavorable), *tener a alguien pasando aceite* (tener enamorado a alguien). Por último, ciertas prendas del vestuario están relacionadas con la economía, el carácter o la actitud de las personas: *apretarse el cinturón* (reducir los gastos por escasez de medios), *alzarse la bata* (salir a divertirse), *amarrarse alguien los pantalones* (comportarse con coraje y decisión), *ponerse la camiseta* (esforzarse para defender los intereses de una institución o de

cualquier emprendimiento conjunto), *ser una corbata* (tener un empleo de poco esfuerzo y buena remuneración).

Cabe mencionar que el trabajo presentado es un estudio piloto que pretende cubrir el vacío en los estudios de la fraseología empleada en Colombia y que, para futuras líneas de investigación, será necesario ampliar el corpus para extraer y analizar un mayor número de locuciones. Asimismo, será necesario ceñir el estudio a las locuciones puramente colombianas, ya que las locuciones estudiadas, o parte de ellas, no son exclusivas de la variante colombiana del español, sino que son compartidas con otros países de habla hispánica. Finalmente, en futuros estudios será necesario analizar la fraseología usada en Colombia, no solo desde una perspectiva sintáctica, semántica y diafásica como se ha hecho en el presente estudio, sino también analizarlas desde un prisma diatópico, diastrático y diacrónico, aspectos que no se tuvieron en cuenta en este trabajo.

Referencias bibliográficas

- ARBOLEDA GRANDA, Diego León. (2012) “La intensificación de las locuciones de la prensa escrita de Medellín: una aproximación pragmática.” *Lingüística y literatura*, 62, pp. 63-85.
- BARANI, Nazia. (2012) *Aspectos de la utilización de las paremias en el diario El País: hacia el desarrollo de materiales didácticos para la enseñanza del español a hablantes de persa*. Salamanca. Tesis doctoral inédita.
- CASARES, Julio. (1950) *Introducción a la lexicografía moderna*. Madrid: CSIC.
- CASTILLO CARBALLO, María Auxiliadora. (2017): “La producción fraseográfica en su historia”, *Estudios de Lingüística del Español*. 38, pp. 85-106.
- CORPAS PASTOR, Gloria. (1996) *Manual de fraseología española*. Madrid: Gredos.
- CORPAS PASTOR, Gloria & Morvay Károly. (2002) “Los estudios de Fraseología y Fraseografía en la Península Ibérica. (Breve presentación y orientación bibliográfica).” *Anales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae. Sectio Lingüística*, 25, pp. 165-182.
- CORPAS PASTOR, Gloria. (2003) *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. Madrid: Iberoamericana.
- CORPAS PASTOR, Gloria. (2013) “Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas.” En: Olza, Inés (ed.) 2013. *Fraseografía pragmática*, Berlín: Frank and Timme, pp. 335-373.

- ESPEJO OLAYA, M^a Bernarda. (2015) *Diccionario de fraseología usada en Colombia*. Versión electrónica: <<https://lexicografiaunisalle.wordpress.com/>>.
- GARCÍA-PAGE SÁNCHEZ, Mario. (2008) *Introducción a la fraseología española. Estudio de las locuciones*. Barcelona: Anthopos Editorial.
- GOUGENHEIM, George; René Michea; Paul Rivenc & Aurélien Sauvageot. (1964) *Lelaboration du français fondamental (1er degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris : Didier.
- GOUGENHEIM, George. (1977) *Dictionnaire fondamental de la langue française*. París: Didier Nouvelle.
- INSTITUTO CARO Y CUERVO. (2018) *Diccionario de colombianismos*. Bogotá: Legis S.A.
- IÑESTA MENA, Eva María & Antonio Pamies Bertrán. (2002) *Fraseología y metáfora: aspectos tipológicos y cognitivos*. Granada: Método Ediciones.
- LOPERA MEDINA, Sergio Alonso. (2017) “Las unidades fraseológicas en los cumplidos en hablantes universitarios de Medellín, Colombia.” *Tonos digital* 33. Versión electrónica: <<https://digitum.um.es/digitum/bitstream/10201/54052/1/Las%20unidades.pdf>>.
- MARTÍN SALCEDO, Javier. (2017) *Fraseología española en uso ¡Si tú lo dices! ¡Venga! Ya verás como sí*. Madrid: Consejería de Educación de la Embajada de España.
- MELLADO BLANCO, Carmen. (2004) *Fraseologismos somáticos del alemán. Un estudio léxico-semántico*. Berlín: Peter Lang.
- MELLADO BLANCO, Carmen (eds.) (2008) *Colocaciones y fraseología en los diccionarios*. Fráncfort: Peter Lang.
- MELLADO BLANCO, Carmen. (2015) “Parámetros específicos de equivalencia en las unidades fraseológicas del español y el alemán.” *Revista de Filología de la Universidad de La Laguna*, 33, pp. 153-174.
- MIRA ÁLVAREZ, Germán Darío. (2011) “Función pragmática de las unidades fraseológicas en Rosario Tijeras.” *Revista Virtual Universidad Católica del Norte*, 33, pp. 44-59.
- MOGORRÓN HUERTA, Pedro. (2009) “Del origen paremiológico de algunas locuciones.” *Paremia*, 18, pp. 65-76.
- MOGORRÓN HUERTA, Pedro & Fernando Navarro Domínguez. (2015) *Fraseología, Didáctica y Traducción*. Fráncfort: Peter Lang GmbH.
- MORA MONROY, Siervo. (1996) *Lexicón de fraseología del español de Colombia*. Bogotá: Instituto Caro y Cuervo.

- ORDUÑA LÓPEZ, José Luís. (2011) “Estudio gramatical de las locuciones verbales con doble pronombre clítico estudio.” *Revista de Lingüística Teórica y Aplicada*, 49, 2, pp. 87-110.
- ORTIZ CORREA, Adriana María. (2012) “Las locuciones y sus variantes en el Diccionario descriptivo del Valle de Aburrá.” *Lingüística y Literatura*, 62, pp. 87-104.
- PAVEL, Silvia & Diane Nolet. (2001) *Manual de terminología*. Gatineau: Servicio de traducción del Gobierno de Canadá.
- PENADÉS MARTÍNEZ, Inmaculada. (2015) “La enseñanza de la fraseología vinculada a los contenidos de los manuales de ELE.” En: Mogorrón Huerta, Pedro & Fernando Navarro Domínguez (eds.) 2015. *Fraseología, Traducción y Didáctica*. Frankfurt am Main: Peter Lang, pp. 241-260.
- QUIROGA, Paula. (2006) *Fraseología italo-española. Aspectos de lingüística aplicada y contrastiva*. Granada: Granada Lingvistica.
- REAL ACADEMIA ESPAÑOLA. (2001) *Diccionario de la lengua española (23ª ed.)*. Madrid: Espasa.
- RUIZ GURILLO, Leonor. (1997) *Aspectos de fraseología teórica española*. Valencia: Universidad de Valencia.
- SÁNCHEZ LOBATO, Jesús & Blanca Aguirre Beltrán. (1992) *Léxico fundamental del español*. Madrid: Sociedad General Española de Librería SGEL.
- SEVILLA MUÑOZ, Julia. (1993) “Las paremias españolas: clasificación, definición y correspondencia francesa.” *Paremia*, 2, pp. 15-20.
- SEVILLA MUÑOZ, Julia. (1997) “La fraseología y la paremiología en los últimos decenios.” *Linred: Lingüística en la Red*, pp. 1-29.
- SEVILLA, MUÑOZ, Julia. (2013) “Las paremias y su clasificación.” *Paremia*, 22, pp. 105-114.
- SEVILLA MUÑOZ, Julia (2015) “Las TIC en la clase de FLE para la enseñanza de refranes y locuciones.” En: *Fraseología, Didáctica y Traducción*. Fráncfort: Peter Lang GmbH, pp. 261-272.
- TRANSLATION CENTRE FOR THE BODIES OF THE EUROPEAN UNION. (2019) *IATE2 User Handbook, 2019*, Luxemburgo: Translation Centre for the bodies of the European Union. Versión electrónica: <<https://iate.europa.eu/assets/handbook.pdf>>.
- VILLEGAS, María Cristina. (2012) “Unidades fraseológicas en Aire de tango de Manuel Mejía Vallejo.” *Lingüística y Literatura*, 62, pp. 211-225.

- VACHEK, John. (1967) "The Non-Static Aspect of the Synchronically Studied Phonological System." *Phonologic der Gegenwart, Wiener Slavistisches Jahrbuch, Ergänzungsband VI*. Graz: Bohlaus Nachf.
- VARIOS AUTORES (Asociación de Academias de la Lengua Española). (2010) *Diccionario de americanismos*. Madrid: Santillana.
- VARIOS AUTORES (Grupo de Investigación en Educación, Sociedad y Paz). (2010) *Diccionario descriptivo del español del Valle de Aburrá (DEVA)*. Medellín, Antioquia: Universidad de Antioquia.
- VARIOS AUTORES (International Standard Organization). (2000) *ISO 1087-2 Terminology work - Vocabulary - Part 2: Computer applications*. Ginebra: International Standard Organization.
- VARIOS AUTORES (International Standard Organization). (2000) *ISO 704 Terminology work. Principle and methods*. Ginebra: International Standard Organization.
- VELÁZQUEZ PUERTO, Karen. (2015) *Análisis fraseológico contrastivo español-alemán y su aplicación a ELE: Una propuesta didáctica para la enseñanza-aprendizaje de las unidades fraseológicas desde una perspectiva cognitiva*. Madrid: Universidad Complutense de Madrid. Tesis doctoral inédita.
- ZAMORA MUÑOZ, Antonio Pablo. (2003) "Esquemas sintácticos fraseológicos pragmáticos." En: Vera Luján, Agustín; Ramón Almela Pérez & José María Jiménez Cano (eds.) (2003): *Homenaje al Profesor Estanislao Ramón Trives*. Murcia: Universidad de Murcia, pp. 825-836.
- ZUCCALÀ, Daniele. (2014) "Disponibilidad léxica del E/LE: el caso brasileño. Estudio piloto realizado con estudiantes de últimos cursos de Traducción en la UNED." *Actas del V congreso Internacional FIAPE*. Cuenca: REDELE, pp. 154-179.
- ZULUAGA GÓMEZ, Francisco Octavio. (2003) *Análisis pragmlingüístico de las unidades fraseológicas en español con atención especial a los refranes*. Ámsterdam: Universidad de Ámsterdam.
- ZULUAGA GÓMEZ, Francisco Octavio. (2004) "Locuciones, dichos y refranes sobre el lenguaje: unidades fraseológicas fijas e interacción verbal." *Forma y Función*, Bogotá: Universidad Nacional.
- ZULUAGA GÓMEZ, Francisco Octavio; Viviana Díaz Orozco & Lirian Astrid Ciro. (2018) *Dichos, refranes y locuciones en las novelas de Gabriel García Márquez*. Bogotá: Instituto Caro y Cuervo.

- ZULUAGA OSPINA, Alberto. (1975a) "Estudios generativo-transformativistas de las expresiones idiomáticas." *Thesavrus Boletín del Instituto Caro y Cuervo*, 30, 1, pp. 1-48.
- ZULUAGA OSPINA, Alberto. (1975b) "La fijación fraseológica." *Thesavrus Boletín del Instituto Caro y Cuervo* 30, 2, pp. 223- 247.
- ZULUAGA OSPINA, Alberto. (1980) *Introducción al estudio de las expresiones fijas*. Frankfurt: Peter Lang.
- ZULUAGA OSPINA, Alberto. (1992) "Spanische Phraseologie." En: Holtus, Günter (ed.) 1992. *Lexikon der Romanistischen Linguistik*. Tübingen: Niemeyer, VI, 1, pp. 125-131.

NOTA BIOGRÁFICA / BIONOTE

DANIELE ZUCCALÀ es doctorando en Lingüística, Traducción y Literatura por la Universidad de Málaga (UMA) en cotutela con la Università degli Studi di Torino (UNITO), Máster en Traducción Editorial (UMA) y Máster en Traducción y Nuevas tecnologías (ISTRAD). Actualmente trabaja en el IATE Support and Development Team en el Centro de Traducción de los Órganos de la Unión Europea (Luxemburgo). Ha realizado diferentes estancias de investigación y es autor de distintos artículos en los que estudia el léxico y la fraseología del italiano y del español. Sus actuales líneas de investigación son la fraseología y su relación con la didáctica.

DANIELE ZUCCALÀ is PhD in Linguistic, Translation and Literature from the Universidad de Málaga (UMA) in co-tutelle with the Università degli Studi di Torino (UNITO), Master degree's in Editorial Translation (UMA) and in Translation and New Technologies (ISTRAD). He is currently working in the IATE Support and Development Team at the Translation Centre for the Bodies of the European Union (Luxembourg). He has carried out different research stays and is the author of several articles in which he studies Italian and Spanish lexicon and multi word units. His current lines of research are phraseology and its relationship with didactics.

MARÍA BERNARDA ESPEJO OLAYA es Profesora Investigadora en Lengua y Literatura Española, título otorgado por la Agencia Española de Cooperación Internacional, AECI. Forma parte del grupo de investigación de Lingüística

del Instituto Caro y Cuervo, en la línea de investigación Estudios sobre el español de Colombia. Es docente en la Licenciatura en Español y Lenguas Extranjeras de la Universidad de La Salle. Sus actuales líneas de investigación son la lexicografía y la fraseología en el español de Colombia.

MARÍA BERNARDA ESPEJO OLAYA es Research Professor in Spanish Language and Literature at the Spanish Agency for International Cooperation, AECI. She is part of the Linguistics research group of the Caro y Cuervo Institute, in the line of research: Studies on Spanish in Colombia. She teaches in the Bachelor's Degree in Spanish and Foreign Languages of the University of La Salle. Her current lines of research are lexicography and phraseology in Spanish in Colombia.

Recibido / Received: 08/11/2019
Aceptado / Accepted: 12/02/2020

Para enlazar con este artículo / To link to this article:
<http://dx.doi.org/10.6035/MonTI.2020.ne6.5>

Para citar este artículo / To cite this article:

Hidalgo-Ternero, Carlos Manuel. (2020) "Google Translate vs. DeepL: analysing neural machine translation performance under the challenge of phraseological variation." 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. 154-177.

GOOGLE TRANSLATE VS. DEEPL: ANALYSING NEURAL MACHINE TRANSLATION PERFORMANCE UNDER THE CHALLENGE OF PHRASEOLOGICAL VARIATION¹

CARLOS MANUEL HIDALGO-TERNERO
cmhidalgo@uma.es
University of Malaga (Spain)

Abstract

The present research analyses the performance of two free open-source neural machine translation (NMT) systems —Google Translate and DeepL— in the (ES>EN) translation of somatisms such as *tomar el pelo* and *meter la pata*, their nominal variants (*tomadura/tomada de pelo* and *metedura/metida de pata*), and other lower-frequency variants such as *meter la pata hasta el corvejón*, *meter la gamba* and *metedura/metida de gamba*. The machine translation outcomes will be contrasted and classified depending on whether these idioms are presented in their continuous or discontinuous form (Anastasiou 2010), i.e., whether different n-grams split the idiomatic sequence (or not), which may pose some difficulties for their automatic detection and translation.

1. Acknowledgements: This research has been funded by a FPU (Formación de Profesorado Universitario) contract, ref. FPU16/02032, awarded to Carlos Manuel Hidalgo Ternero by the Ministerio de Educación, Ciencia y Deporte (Spanish Ministry of Education, Science and Sport). Furthermore, this publication has been carried out within the framework of the thematic network TRAJUTEC and the teaching network "TACTRAD: aplicación de nuevas herramientas TAC para la enseñanza-aprendizaje de la traducción especializada" (ref. 719/2018) coordinated by the University of Malaga, as well as the research projects Interpreta 2.0 (PIE 17-015, type B), VIP (FFI2016-75831-P) and TERMITUR (HUM2754).



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

Overall, the insights gained from this study will prove useful in determining for which of the different scenarios either Google Translate or DeepL delivers a better performance under the challenge of phraseological variation and discontinuity.

Keywords: Neural machine translation; Phraseological variation; Discontinuity; Somatismos.

Resumen

La presente investigación tiene por objetivo analizar el rendimiento de dos sistemas de traducción automática neuronal (TAN) —Google Translate y DeepL— en la traducción (ES>EN) de somatismos tales como *tomar el pelo* y *meter la pata*, sus variantes nominales (*tomadura/tomada de pelo* y *metedura/metida de pata*) y otras variantes con menor frecuencia de aparición tales como *meter la pata hasta el corvejón*, *meter la gamba* y *metedura/metida de gamba*. Los resultados de ambos sistemas se contrastarán y clasificarán según los diferentes somatismos se presenten en su forma continua o discontinua (Anastasiou 2010), es decir, dependiendo de si distintos n-gramas dividen la secuencia idiomática (o no), lo que puede presentar serias dificultades para su detección y traducción automáticas. En general, los resultados obtenidos permitirán dilucidar en cuál de los diferentes escenarios Google Translate o DeepL ofrece un mejor rendimiento ante el desafío de la variación y discontinuidad fraseológica.

Palabras clave: Traducción automática neuronal; Variación fraseológica; Discontinuidad; Somatismos.

1. Introduction

The not so paradoxical instability of fixed expressions poses serious challenges for Natural Language Processing in general, and Machine Translation in particular. In this regard, one of the fertile grounds for debate in computational linguistics is how to approach these fixed expressions, admitting complete fixedness is the exception.

In this context, Sag, Baldwin et al. (2002) have thoroughly elaborated on the two opposing (and yet complementary) approaches in the computational treatment of multiword expressions (MWE). On the one hand, a *fully compositional account*, which may experience no problem with lexical variation, but it does with idiomaticity (for example, interpreting the “joke” semantics in *pull someone’s leg*) and overgeneration (for example, in creating

unregistered instances such as **the leg was pulled*). On the other hand, the *word-with-spaces* approach suffers from a flexibility problem as it does not properly capture phraseological and inflectional variation, and it also impedes lexical proliferation (e.g. light-verb constructions such as *make a payment*, *make a deposit*, etc. would have to be listed individually, which entails an important lack of generality and prediction) (Sag, Baldwin et al. 2002: 2-3). This approach does not consider either that MWEs do not always consist of adjacent tokens (e.g. *take my piece of advice into account*), which hence poses further serious difficulties for the automatic detection of these expressions (Rohanian, Taslimipoor et al. 2019).

Against this background, it will first be necessary to analyse the somatisms (i.e., idioms containing terms that refer to human or animal body parts) under study in order to determine to which category it is possible to relate them. These somatisms include *tomar el pelo* (literally, ‘to take someone’s hair’; figuratively, ‘to fool someone’), and *meter la pata* (lit., ‘to put the leg in’; fig., ‘to mess things up’), and their variants. In the case of *tomar el pelo*, only the nominal variants, *tomadura de pelo* and *tomada de pelo*, retrieve a significant number of concordances in the analysed corpora. Regarding other possible variants such as *tomar el cabello* or *tomar los pelos*, a very scarce number of concordances of them as idioms has been recalled: 5 in the case of *tomar el cabello* and 3 for *tomar los pelos*. The latter ones can therefore be considered as *anti-collocations*, i.e., “potential lexical variants of a given institutionalized phrase which are observed with zero or markedly low frequency” (Pearce, in Sag, Baldwin et al. 2002: 7).

Concerning the idiom *meter la pata*, it was possible to detect a wider range of variants: not only *variants through syntagmatic relation* (Koike 2007) such as *metedura de pata* and *metida de pata*, but also through insertion, *meter la pata hasta el corvejón*, and *through paradigmatic relation* (Koike 2007), in a relation of co-hyponymy, such as *meter la gamba*, which also holds two nominal variants, *metedura de gamba* and *metida de gamba*. These latter variants are diaphasically marked as informal, and their appearance usually connotes irony or humour, as it was possible to observe in the concordances retrieved from the analysed corpora.

Regarding the idioms *tomar el pelo* and *meter la pata*, and their variants, the different retrieved concordances reveal that both can undergo verb inflection (*tomo, tomas, tomó... el pelo*) but they are not subject to noun inflection (**tomar los pelos, *meter las patas*). In these sequences, *pelo* and *pata* are only modified by the definite article (**tomar un pelo, *tomar otro pelo, *meter una pata...*). Therefore, both sequences can be considered as *semi-fixed* MWEs as they can undergo both morphosyntactic and internal alterations from their canonical form, unlike other MWEs such as *by and large, in vitro* or *in a nutshell*, which are hence designated as *fixed expressions* (Sag, Baldwin et al. 2002: 4).

As *semi-fixed* MWEs, both somatisms (*tomar el pelo* and *meter la pata*) and their variants (*tomadura de pelo, metedura de pata, meter la gamba...*) present a certain degree of syntactic flexibility. The different concordances reveal that adverb phrases such as *más* ('more'), *tanto* ('so much') or *un poco* ('a bit'), noun phrases such as *otra vez* ('again') or the subject, among others, consistently occur within (and hence split) the sequences in the verb-phrase variants of the idioms (*tomar el pelo, meter la pata* and *meter la gamba*). For the noun-phrase variants, the sequence is regularly split by different adjective modifiers such as *total* ('total'), *descomunal* ('enormous') or *constante* ('constant') in the case of *tomadura/tomada de pelo*, and *total, grave* ('serious'), or *histórica* ('historic') in the case of *metedura/metida de pata*. Therefore, whereas it would not be possible to capture the full range of lexical variants of these idioms if being treated as *words-with-spaces*, a *fully compositional account* would overgenerate anti-collocations such as *tomar los pelos* or *tomar los cabellos*, all of which thus necessitates employing a hybrid approach (Sag, Baldwin et al. 2002).

Throughout this research we will thus distinguish between continuous or discontinuous forms of the idiom (Anastasiou 2010) depending on whether or not the sequence is split by those different n-grams. One of the main objectives of our research will hence be to determine in which cases and to what extent discontinuity poses some challenges for the automatic detection and translation of the idioms. The ultimate objective will be to analyse and contrast the performance of two neural machine translation (NMT) systems, Google Translate vs. DeepL, when accurately detecting and translating the continuous and discontinuous forms of the somatisms *tomar el pelo* and *meter la pata*, their nominal variants (*tomadura de pelo, tomada de pelo, metedura de*

pata, and *metida de pata*), as well as other variants such as *meter la pata hasta el corvejón*, *meter la gamba*, *metedura de gamba*, and *metida de gamba*, in order to determine which of these NMT systems delivers a better performance under the challenge of phraseological variation and discontinuity.

The overall structure of this paper takes the form of five sections. Once the somatisms under study have been introduced in section 1, they will be further analysed in section 2, shifting the focus onto their diatopic variation. Section 3 begins by describing the methodology of the present research, and section 4 exposes the results retrieved from both NMT systems when facing the challenge of phraseological variation and discontinuity. In this regard, the results will be contrasted and classified depending on whether these idioms are presented in their continuous or discontinuous form (Anastasiou 2010), i.e., whether different n-grams split the idiomatic sequence (or not). In this line, section 5 analyses the repercussions of the retrieved results for the NMT systems and, finally, the conclusions of the present research will be displayed in section 6.

2. Diatopic variation of the somatisms under study

In order to exhibit the diatopic variation of the somatisms under study, Corpus Pastor's (2015 & 2018) corpus-based research protocol will be implemented. In this regard, the somatisms have been retrieved and analysed from the monolingual corpora esTenTen (general Spanish), eseuTenTen (European Spanish), and esamTenTen (Latin American Spanish). The latter includes subcorpora corresponding to the 18 American national varieties. These corpora, belonging to the TenTen family, are available through Sketch Engine, a language corpus management and query system with 500 corpora in more than ninety different languages. Comprising over 17.5 billion words, esTenTen corpus (including eseuTenTen and esamTenTen) is compiled from texts retrieved from the Internet. These corpora are lemmatised and part-of-speech tagged (Kilgarriff, Baisa et al. 2003).

The different somatisms will be displayed in Table 1 along with their raw/normalised frequencies for each of the 19 Spanish national varieties (identified by their country code top-level domain: .ar, .bo, .cl, .co, .es, etc.). Whereas the first column, representing the *raw frequency*, refers to the total

number of occurrences of a specific somatism, the second one (in bold) stands for the *normalised frequencies*, presented as percentage scores per million tokens. Analysing normalised frequencies to a common base is essential when contrasting results retrieved from corpora of dissimilar sizes (Corpas Pastor 2018).

	<i>Tomar el pelo</i>		<i>Tomadura de pelo</i>		<i>Tomada de pelo</i>		<i>Meter la pata</i>		<i>Metedura de pata</i>		<i>Metida de pata</i>		<i>Meter la gamba</i>		<i>Metedura de gamba</i>		<i>Metida de gamba</i>		<i>Meter la pata hasta el corvejón</i>	
.AR	3775	1.13	388	0.12	1322	0.4	2651	0.79	159	0.05	710	0.21	80	0.02	1	>0.01	10	>0.01	1	>0.01
.BO	29	0.45	13	0.2	3	0.05	40	0.62	3	0.05	41	0.64	-	-	-	-	-	-	-	-
.CL	576	0.49	234	0.2	63	0.05	587	0.5	55	0.05	582	0.5	-	-	-	-	1	>0.01	-	-
.CO	157	0.32	27	0.05	7	0.01	99	0.2	17	0.03	134	0.27	-	-	-	-	-	-	-	-
.CR	27	0.43	14	0.22	2	0.03	33	0.52	4	0.06	33	0.52	-	-	-	-	-	-	-	-
.CU	125	0.45	69	0.25	-	-	141	0.5	74	0.26	24	0.09	-	-	-	-	-	-	-	-
.DO	54	0.92	13	0.22	-	-	141	2.39	12	0.2	58	0.99	1	0.02	-	-	-	-	-	-
.EC	36	0.41	45	0.51	1	0.01	43	0.49	16	0.18	12	0.14	-	-	-	-	-	-	-	-
.SV	9	0.25	7	0.19	-	-	18	0.49	6	0.16	22	0.6	-	-	-	-	-	-	-	-
.GT	22	0.61	4	0.11	1	0.03	63	1.75	2	0.06	57	1.58	-	-	-	-	-	-	-	-
.HN	6	0.57	4	0.38	-	-	-	-	3	0.28	-	-	-	-	-	-	-	-	-	-
.MX	1240	0.63	455	0.23	488	0.25	1054	0.53	101	0.05	459	0.23	3	>0.01	-	-	-	-	1	>0.01
.NI	32	0.46	5	0.07	1	0.01	40	0.57	5	0.07	49	0.7	-	-	-	-	-	-	-	-
.PA	13	0.62	4	0.19	-	-	45	2.14	3	0.14	38	1.81	-	-	-	-	-	-	-	-
.PY	28	0.4	15	0.22	1	0.01	41	0.59	4	0.06	22	0.32	-	-	-	-	-	-	-	-
.PE	264	0.78	111	0.33	7	0.02	642	1.9	21	0.06	482	1.43	-	-	-	-	-	-	6	0.02
.UY	496	2.35	168	0.8	180	0.85	221	1.05	9	0.04	87	0.41	2	0.01	-	-	-	-	-	-
.VE	57	0.19	27	0.09	2	0.01	129	0.44	28	0.1	77	0.26	1	>0.01	-	-	3	0.01	-	-
.ES	4429	1.89	3356	1.43	29	0.01	4523	1.93	1979	0.84	220	0.09	276	0.12	17	0.01	9	>0.01	41	0.02

Table 1: Raw and normalised frequencies of the somatisms under study in the 19 national varieties of *esTenTen*

The results in Table 1 exhibit distinctly different frequency scores for the somatisms across national varieties. A significant number of occurrences for the verbal variants (*tomar el pelo* and *meter la pata*) has been detected in the 19 national varieties, albeit with different frequency scores: whereas *tomar el pelo* displays a higher normalised frequency in the dialectal areas of Uruguay (2.35), Spain (1.89) and Argentina (1.13), *meter la pata* is more frequent in the varieties of Dominican Republic (2.39), Panama (2.14), Spain (1.93) and Peru (1.9).

Regarding the nominal variants, quite revealing results have been retrieved on the frequency scores of *tomadura de pelo* vs. *tomada de pelo*, and *metedura de pata* vs. *metida de pata*. While, in the first dichotomy, *tomadura de pelo* appears as the most frequent variant in most national varieties, except for Argentina (0.12 vs. 0.4), Mexico (0.23 vs. 0.25), and Uruguay (0.8 vs. 0.85);

in the second opposition, *metida de pata* is the variant with the higher normalised frequency across national varieties, with the exception of Cuba (0.26 vs. 0.09), Ecuador (0.18 vs. 0.14) and Spain (0.84 vs. 0.09).

With respect to the somatism *meter la gamba*, the results permit to observe that its occurrence is predominant in European Spanish (0.12), distantly followed by the national variety of Argentina (0.02), Dominican Republic (0.02) and Uruguay (0.01). Its nominal variants, *metedura de gamba* and *metida de gamba*, exhibit low frequency scores in the few varieties presenting occurrences for these sequences: 0.01 in European Spanish in the case of *metedura de gamba* and 0.01 in the national variety of Venezuela for *metida de gamba*. Finally, regarding the variant through insertion *meter la pata hasta el corvejón*, only significant occurrences have been detected in the national varieties of Spain and Peru (0.02 each). In the case of Peru, 3 out of the 6 retrieved occurrences presented instances of another variant of this somatism, *meter la pata hasta los corvejones*.

Once the diatopic variation of the somatisms under study has been analysed, one of the objectives of the research will be to examine whether, in all the cases, there is a direct correspondence between a higher frequency score of a given somatism across national varieties and a better performance of the NMT system, or whether other factors enter the scene, such as some specific national varieties being privileged over others in the translation of these idioms. The final results will also shed some light on these NMTs' performance regarding rare and out-of-vocabulary (OOV) words, as is the case of the lower-frequency variants *meter la pata hasta el corvejón*, *meter la gamba*, and their nominal variants *metedura de gamba* and *metida de gamba*.

3. Methodology

Following a corpus-based methodology, the cases under study have also been retrieved from the corpus esTenTen through Sketch Engine. In this context, it is necessary to take into account that this corpus is created from a wide compilation of texts collected from the web and hence it includes linguistically valuable text corpora from heterogeneous sources ranging from NMT canonical parallel training data to non-canonical user-generated content

(UGC), with the acknowledged challenges that UGC's pervasive source-text error, noise and out-of-vocabulary tokens, inter alia, still pose to even the most robust NMT systems (Sperber, Niehues et al. 2017; Belinkov & Bisk 2018; Anastasopoulos 2019; Lohar, Popović et al. 2019). This heterogeneity in terms of text sources, types and diatopic varieties has however been maintained in the retrieval and analysis process in order to avoid sampling bias, which would otherwise originate from exclusively selecting a sole text typology for the somatisms under study. Since the main objective of the present research is to determine these NMT systems' performance under the challenge of phraseological variation and discontinuity, the results have been manually assessed following a reference-based MT evaluation with several possible target-text candidates for each of the somatisms under study in their continuous and discontinuous form for the different contexts. In this way, those results presenting morphological, syntactic, and/or orthographic dissimilarities or containing either source-text or translation errors (omissions, additions, etc.) which affected other elements in the sentences were unified within the same category depending on the NMTs' performance in the translation of the analysed source-text somatisms under analogous phenomena (continuous or discontinuous form split by a subject, object, adjective modifier, etc.)

In this context, in order to retrieve the discontinuous forms of the somatisms under study, Sketch Engine's CQL (Corpus Query Language) code has been employed. CQL is a special query code to search for complex lexico-grammatical patterns which would otherwise be laborious to retrieve through the simple query system. In order to exclude as much noise as possible, the following CQL schemas have been employed:

Sequence	CQL schemas
“Tomar [1-2 n-grams] el pelo”	[lemma="tomar"][] {1,2} [word="el"] [word="pelo"]
“Tomadura/tomada [1-2 n-grams] de pelo”	[lemma="tomadura tomada"][] {1,2} [word="de"] [word="pelo"]
“Meter [1-2 n-grams] la pata”	[lemma="meter"][] {1,2} [word="la"] [word="pata"]
“Metedura/metida [1-2 n-grams] de pata”	[lemma="metedura metida"][] {1,2} [word="de"] [word="pata"]

Table 2: CQL schemas for the discontinuous form of the somatisms

Whereas the CQL code *lemma* finds examples of all inflected forms of the introduced token (e.g. *tomar*, *tomó*, *tomaría*, *tomaste* in the case of *tomar*), the code *word* retrieves the exact form of the token, without inflection. In the somatisms *tomar el pelo* and *meter la pata*, and their nominal variants (*tomadura de pelo*, *tomada de pelo*, *metedura de pata*, and *metida de pata*) only the verbs *tomar* and *meter*, and the nouns *tomadura/tomada* and *metedura/metida*, respectively, allow inflection. That is the reason why, in order to refine the concordances retrieved in esTenTen, the CQL code *lemma* has only been employed for those elements in the idiom allowing inflection, and the code *word* has been selected for the rest.

In order to search for the different n-grams that may split the sequence, the CQL codes “[] {1,2}” have regularly been used, which would retrieve instances of unigrams or bigrams within the sequence. In two specific cases the number of n-grams has been augmented to “[] {1,4}”: when searching for indirect objects within the sequence *tomar el pelo*, and when seeking several different n-grams dividing the verb-phrase variants of the idioms *tomar el pelo* and *meter la pata*, all of which can contain a number of n-grams higher than 2.

The retrieved concordances, containing discontinuous somatisms, have been stored and classified depending on which n-gram split the sequence. Two free open-source NMT systems —Google Translate and DeepL— have then been employed to translate these somatisms in their discontinuous scenarios. These sequences have subsequently been manually altered into their continuous form in order to evaluate and contrast the performance of both

NMT systems against both the continuous and discontinuous display of the somatisms in identical contexts, with the purpose of securing that a dissimilar performance in the NMT systems was chiefly caused by phraseological variation and discontinuity, discarding other possible factors. Against such a background, the different translation outcomes will be presented in the following section.

4. Results

A total of 380 cases has been analysed, considering the continuous and discontinuous forms of the idioms *tomar el pelo*, *meter la pata* and their nominal variants (*tomadura de pelo*, *tomada de pelo*, *metedura de pata*, and *metida de pata*) split by different modifiers (*tanto*, *un poco*, *otra vez*, etc.), as well as other variants such as *meter la pata hasta el corvejón*, *meter la gamba*, *metedura de gamba* and *metida de gamba*.

In the case of *tomar el pelo*, 140 cases have been examined (70 for the continuous and 70 for the discontinuous form of the sequence). The itemised results for *tomar el pelo* (Chart 1) display notable differences between both NMT systems' performance. Whereas the inclusion of modifiers such as *más* (9 vs. 10)² or the indirect object (6 in both) only presented minor dissimilarities between both NMT systems, DeepL delivered a better performance in those cases where the sequence was split by other modifiers such as *tanto* (6 vs. 10), *otra vez* (7 vs. 10), and the subject (7 vs. 9). In two other scenarios, it was Google Translate that offered more suitable equivalents: when the sequence was divided by *un poco* (6 vs. 4) and more significantly when several different n-grams were included within the sequence (10 vs. 3). An example of these NMTs' performance in the translation of the continuous and discontinuous form of *tomar el pelo* will be illustrated in Table 3.

Itemised NMT results for *tomar el pelo*

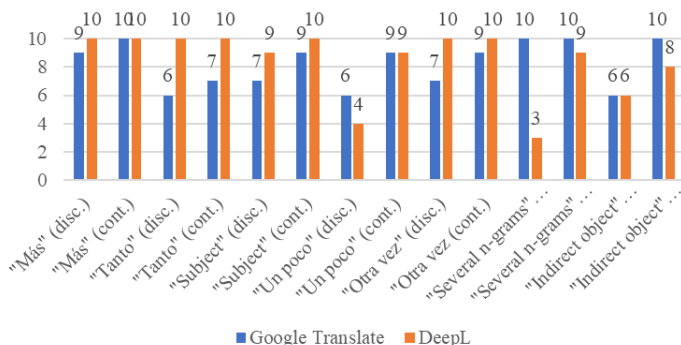


Chart 1: Itemised NMT results for *tomar el pelo*

	KWIC extracts	Comments
ST [ES] Disc. form	En 1969, cuando una periodista le preguntó si este "¿no?" provenía de su herencia santiagueña (junto con las empanadas), ella respondió: "pare un poco, ya no lo uso más". Al respecto, decía: "me tomaron tanto el pelo en todas partes que lo tuve que dejar" (Mactas 1969: 55).	
TT [EN] Disc. form with DeepL	In 1969, when a journalist asked her if this "no" came from her Santiago heritage (along with the empanadas), she replied: "stop a little, I don't use it anymore". In this respect, she said: "I was teased so much everywhere that I had to give it up" (Mactas 1969: 55).	The ST idiom has been detected and an appropriate TT equivalent has been provided.
TT [EN] Disc. form with Google Translate	In 1969, when a journalist asked if this "right?" It came from her Santiago heritage (along with the empanadas), she replied: "stop a little, I don't use it anymore". In this regard, he said: "They took my hair so much everywhere I had to leave it" (Mactas 1969: 55).	A literal and inadequate translation of the ST idiom has been provided: "they took my hair so much..." instead of "I was teased so much..."
ST [ES] Cont. form	En 1969, cuando una periodista le preguntó si este "¿no?" provenía de su herencia santiagueña (junto con las empanadas), ella respondió: "pare un poco, ya no lo uso más". Al respecto, decía: "me tomaron el pelo tanto en todas partes que lo tuve que dejar" (Mactas 1969: 55).	
TT [EN] Cont. form with DeepL	In 1969, when a journalist asked her if this "no" came from her Santiago heritage (along with the empanadas), she replied: "stop a little, I don't use it anymore". She said: "I was teased so much everywhere that I had to give it up" (Mactas 1969: 55).	The ST idiom has been detected and an appropriate TT equivalent has been provided (analogously to the discontinuous scenario).
TT [EN] Cont. form with Google Translate	In 1969, when a journalist asked if this "right?" It came from her Santiago heritage (along with the empanadas), she replied: "Stop a little, I don't use it anymore". In this regard, he said: "They teased me so much everywhere that I had to leave it" (Mactas 1969: 55).	In contrast to the discontinuous scenario with Google Translate, the ST idiom has now been detected and an appropriate TT equivalent has been provided.

Table 3: KWIC extracts with the translation of *tomar el pelo*

However, when analysing the total NMT results (Chart 2), only slight differences have been observed in the performance of Google Translate versus DeepL when adequately detecting and establishing equivalences for the idiom: 51 (73%) vs. 52 (74%) appropriate equivalents, respectively, for the discontinuous form, and 64 (92%) vs. 67 (96%) adequate equivalents for the continuous form of the sequence.

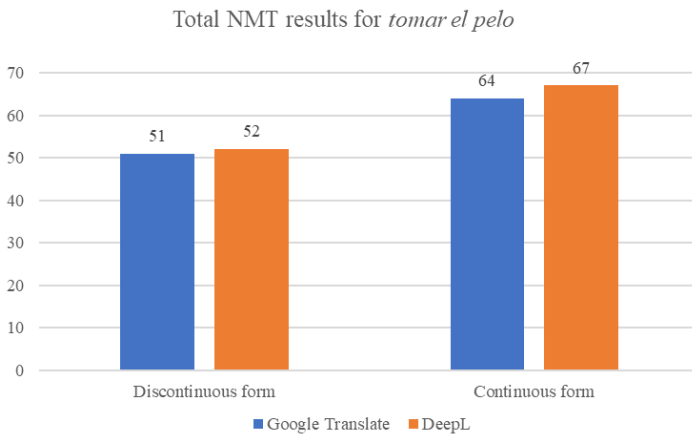
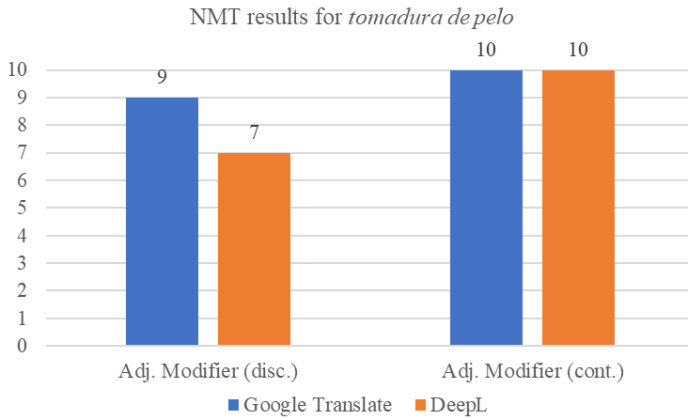
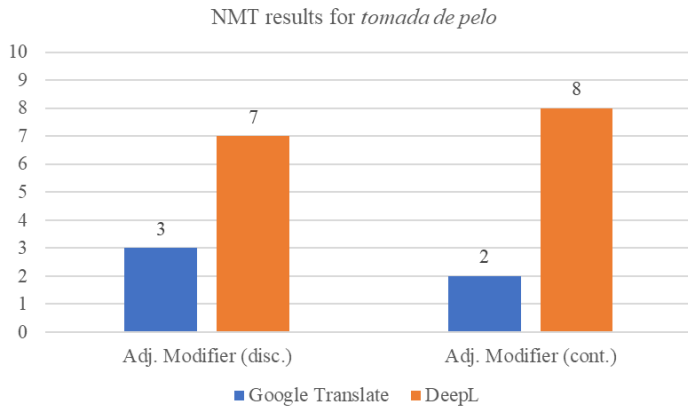


Chart 2: Total NMT results for *tomar el pelo*

When examining the results of the nominal variant *tomadura de pelo* (Chart 3), it can be observed that Google Translate provided more appropriate equivalents in the discontinuous form, when the sequence was split by an adjective modifier (9 vs. 7). Both NMT systems offered appropriate equivalents in all those concordances with the continuous form of *tomadura de pelo* (10 out of 10). Regarding the somatism *tomada de pelo* (Chart 4), while DeepL exhibited similar results, the performance of Google Translate dramatically diminished both in the discontinuous (3 vs. 7) and the continuous (2 vs. 8) form of the sequence.

Chart 3: NMT results for *tomadura de pelo*Chart 4: NMT results for *tomada de pelo*

In the case of the idiom *meter la pata*, 120 cases have been analysed (60 for the continuous and 60 for the discontinuous form of the sequence). Concordances with the same modifiers as in the case of *tomar el pelo* were translated and examined, except for the absence of indirect object in the scenario of *meter la pata*, since it does not take the dative. The itemised

NMT results (Chart 5) exhibit, overall, a quasi-analogous outcome for both NMT systems. Google Translate and DeepL delivered a similar performance in those instances in which the sequence was split by modifiers such as *más* (9 in both), *tanto* (10), the subject or *un poco* (9 vs. 10), and *otra vez* (8 vs. 9). However, there was a specific case in which more dramatic differences could be observed: when several different n-grams divided the sequence (as those in the examples of Table 4), Google Translate achieved an accuracy rate of 5 out of 10, while DeepL's performance obtained a score of 9 out of 10.

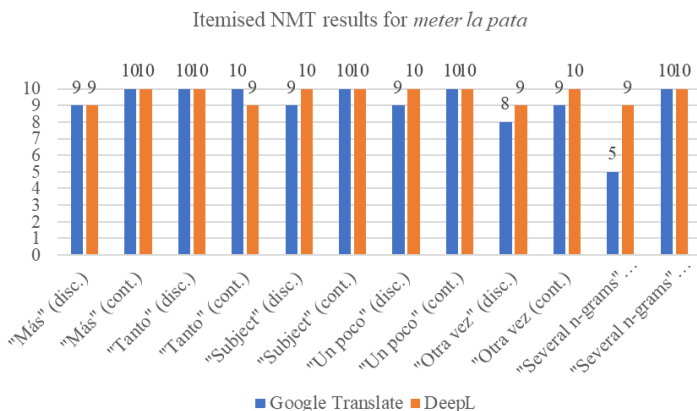


Chart 5: Itemised NMT results for *meter la pata*

	KWIC extracts	Comments
ST [ES] Disc. form	Una vez que hayamos hecho bien el <u>trimming</u> lo que toca ahora es el secado y curado de nuestras plantas de marihuana para obtener el aroma y fragancia adecuada, no podemos <u>meter</u> aquí tampoco la pata.	
TT [EN] Disc. form in DeepL	Once we have done the trimming well, it is now time to dry and cure our marijuana plants to obtain the right aroma and fragrance.	The ST sentence containing the discontinuous idiom has been omitted in the TT.
TT [EN] Disc. form in Google Translate	Once we have done the trimming well, what it touches now is the drying and curing of our marijuana plants to obtain the right aroma and fragrance, we can <u>not put the leg here either</u> .	A literal and inadequate translation of the ST discontinuous somatism has been provided in the TT.
ST [ES] Cont. Form	Una vez que hayamos hecho bien el <u>trimming</u> lo que toca ahora es el secado y curado de nuestras plantas de marihuana para obtener el aroma y fragancia adecuada, no podemos <u>meter</u> la pata aquí tampoco.	
TT [EN] Cont. form in DeepL	Once we've done the trimming right, it's time to dry and cure our marijuana plants to get the right aroma and fragrance, we can't screw up here either.	The ST continuous idiom has been detected and an appropriate equivalent has been provided.
TT [EN] Cont. form in Google Translate	Once we have done the trimming well, what it touches now is the drying and curing of our marijuana plants to obtain the right aroma and fragrance, we cannot screw up here either.	The ST continuous idiom has been detected and an appropriate equivalent has been provided.

Table 4: KWIC extracts with the translation of *meter la pata*

The global results for the sequence (Chart 6) also show that, analogously to *tomar el pelo*, no notable differences in the overall performance of both NMT systems have been detected: whereas Google Translate offered an accuracy rate of 83% (50 out of 60 cases) for the discontinuous form of the sequence, and of 98% (59/60 cases) for the continuous one, DeepL slightly outperformed Google Translate in the discontinuous case up to 95% (57/60 cases) but it showed a similar accuracy rate to Google Translate for the continuous form of the sequence (98%, i.e., 59/60 cases).

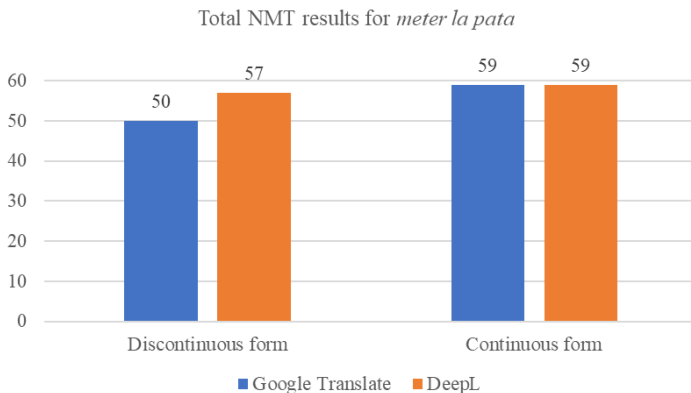


Chart 6: Total NMT results for *meter la pata*

Furthermore, as can be observed in Chart 7, Google Translate and DeepL also delivered a similar performance for the nominal variant *metedura de pata*: 7 vs. 8 for the discontinuous form and 8 in both for the continuous one. That was not the case of the variant *metida de pata* (Chart 8), for which Google Translate’s performance drastically declined both in the discontinuous (1 vs. 7) and the continuous form (3 vs. 8).

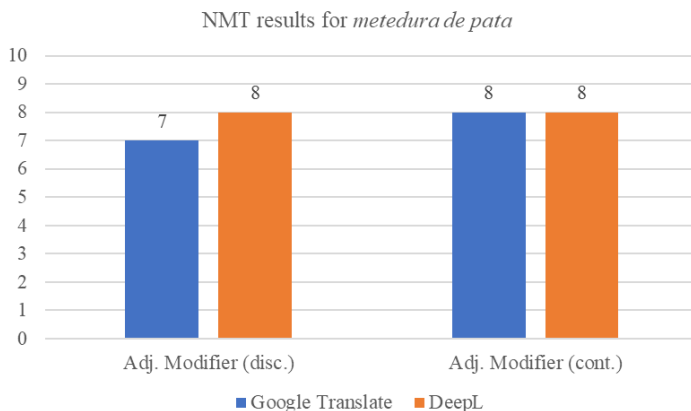


Chart 7: NMT results for *metedura de pata*

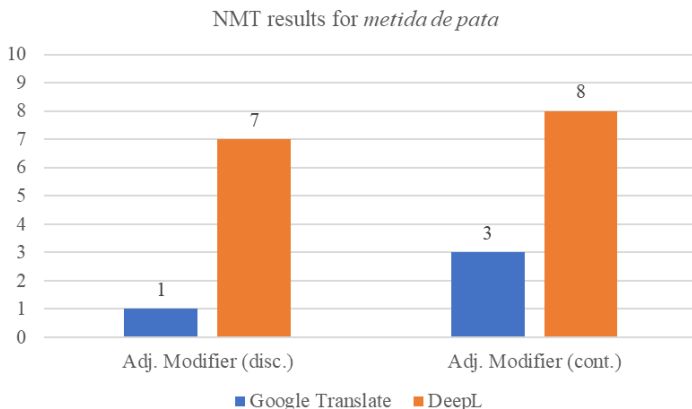


Chart 8: NMT results for *metida de pata*

Finally, concerning the lower-frequency variants *meter la pata hasta el corvejón*, *meter la gamba*, *metedura de gamba* and *metida de gamba*, neither Google Translate nor DeepL could adequately detect and translate these idioms, which led to a literal (and, hence, inappropriate) interpretation and translation of them in a total of 40 analysed cases.

Once the total of 380 cases has been classified and exposed, the global results will be displayed. As can be observed in Chart 9 and 10, overall, DeepL slightly outperformed Google Translate in the task of finding appropriate equivalents for the somatisms *tomar el pelo* and *meter la pata*, and their nominal variants *tomadura de pelo*, *tomada de pelo*, *metedura de pata*, and *metida de pata*. For the continuous form of these idioms, while Google Translate offered an accuracy rate of 86% (146 out of 170), DeepL achieved a score of 94% (160/170). In the discontinuous scenario, Google Translate was newly surpassed by DeepL: 71% (121/170) vs. 81% (138/170). The addition of both continuous and discontinuous results hence shows a minor contrast between both NMTs' performance: 79% (267/340) with Google Translate and 88% (298/340) with DeepL. Finally, when including all the remaining variants (*meter la pata hasta el corvejón*, *meter la gamba*, *metedura de gamba*, and *metida de gamba*) the global results exhibit an accuracy rate of 70% (267/380) for Google Translate and 78% (298/380) for DeepL.

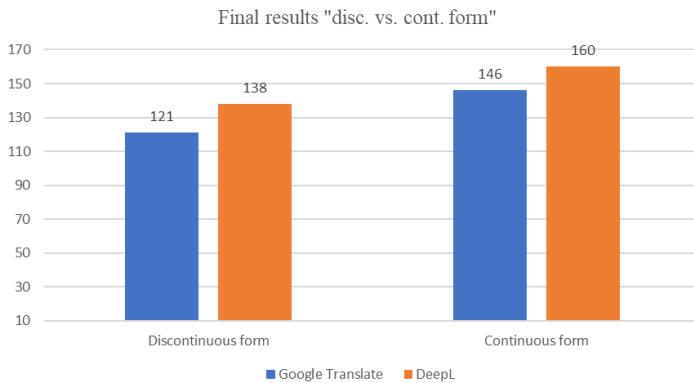


Chart 9: Final results “disc. vs. cont. form”

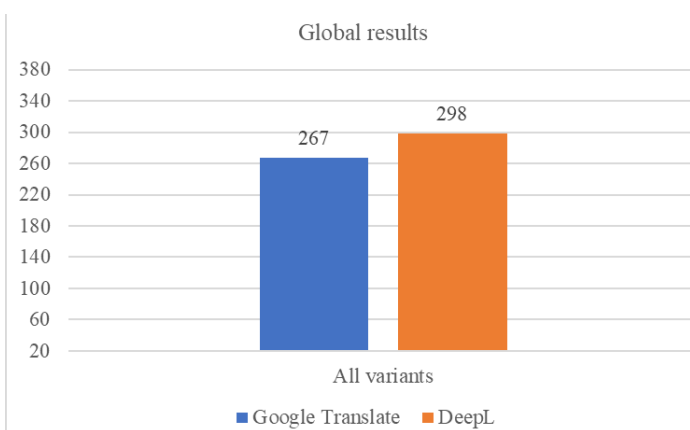


Chart 10: Global results

5. Analysis of the results

The overall results show that both Google Translate and DeepL delivered a high-quality performance in the detection and establishment of equivalences for the idioms *tomar el pelo* and *meter la pata*, and their respective nominal variants *tomadura de pelo*, and *metedura de pata*, both in their continuous and discontinuous forms, with a final global accuracy rate of 86% with Google Translate and 89% with DeepL for these somatismos.

Nevertheless, that was not the case of the nominal variants *tomada de pelo* and *metida de pata*, for which Google Translate's performance dramatically declined to 20% in the continuous and 30% in the discontinuous form of the sequence *tomada de pelo*, and to 30% in the continuous and 10% in the discontinuous form of *metida de pata*. In the case of *tomada de pelo*, despite presenting higher normalised frequencies in widely spoken varieties such as those of Argentina (0.12 vs. 0.4), Mexico (0.23 vs. 0.25), and Uruguay (0.8 vs. 0.85) and occurrences in most of the remaining national varieties, Google Translate's accuracy rate decreased 80% in the continuous and 60% in the discontinuous form when compared to the variant *tomadura de pelo*, for which European Spanish presents the highest comparative normalised frequency (1.43 vs. 0.01). Even more revealing is the case of *metida de pata* since, despite presenting a higher normalised frequency in most of the national varieties, it posed considerably more serious difficulties for Google Translate than *metedura de pata*, with a comparative decline of 50% in the continuous and 60% in the discontinuous form; even if *metedura de pata* is exclusively predominant in the dialectal areas of Cuba (0.26 vs. 0.09), Ecuador (0.18 vs. 0.14) and, mainly, Spain (0.84 vs. 0.09).

Therefore, the inclusion of the diatopically-marked variants *tomada de pelo* and *metida de pata* in the analysis resulted in a considerable decrease in the global NMTs' accuracy rate from 89% to 88% with DeepL and, more dramatically, from 86% to 79% with Google Translate. These results are in line with previous studies on phraseological variation (Corpas Pastor 2015 & 2018; Valencia Giraldo & Corpas Pastor 2019) which corroborate the overall tendency towards European Spanish to the detriment of American national varieties in the translation of phraseology. Furthermore, the results also highlight the pressing need for ameliorating NMT training data to encompass the diatopic variation of a given language as well as the necessity of searching alternative approaches for low-resourced language varieties (Honnet, Popescu-Belis et al. 2017; Costa-Jussà, Zampieri et al. 2018; Farhan, Talafha et al. 2020).

In addition to these obstacles, neither Google Translate nor DeepL could adequately detect and translate other lower-frequency variants such as *meter la pata hasta el corvejón* and *meter la gamba* (or the nominal variants *metedura de gamba* and *metida de gamba*). That is the reason why the final global results

(those containing all the variants) showed a further dramatic decline in the overall NMTs' performance from 79% to 70% for Google Translate, and from 88% to 78% for DeepL, all of which emphasises the additional necessity to optimise the NMTs' performance regarding rare and out-of-vocabulary words (Sennrich, Haddow et al. 2015; Gülçehre, Ahn et al. 2016; Liu, May et al. 2018)

It is also possible to detect in the results a decline in the NMTs' performance from the continuous to the discontinuous form of the idioms (from 86% to 71% in the case of Google Translate, and from 94% to 81% with DeepL). As can be observed in the itemised results, in the case of the verbal variants (*tomar el pelo* and *meter la pata*) only the insertion of common 1-gram modifiers such as *más* or the subject, and the 2-gram modifier *otra vez*, did not pose serious problems for any of the NMT systems. However, in the case of *tomar el pelo* the 1-gram modifier *tanto* presented difficulties for Google Translate, and so did the inclusion of several different n-grams for DeepL. Other modifiers such as *un poco* or the indirect object caused problems for both NMT systems. In the scenario of *meter la pata*, none of the modifiers posed an obstacle, except for the inclusion of several different n-grams, in this case only for Google Translate. As for the nominal variants (*tomadura de pelo* and *metedura de pata*), none of the adjective modifiers presented, overall, any serious difficulties for the NMT systems. Consequently, the itemised results indicate that the machine translation outcomes largely depend on the interplay of the idiom in question, the modifiers within the discontinuous form of the sequence, and the NMT system itself.

6. Conclusion

The present research has aimed to contribute to the growing body of literature (Rikters & Bojar 2017; Wang, Tu et al. 2017; Huang, Wang et al. 2018, inter alia) that is bringing into focus the treatment of MWEs by current neural machine translation systems as well as new avenues for its optimisation. In this context, the NMT systems Google Translate and DeepL have proved to deliver an overall high-quality performance in the detection and establishment of equivalences for idioms with high normalised frequencies such as *tomar el pelo* and *meter la pata*, and their nominal variants *tomadura de pelo*

and *metedura de pata*, both in their continuous and discontinuous forms, with a final average of 86% vs. 89%, respectively.

In spite of their high accuracy rate in the translation of these idiomatic sequences, both NMTs still encountered serious difficulties to adequately detect and translate other lower-frequency phraseological variants of the idioms (such as *meter la pata hasta el corvejón*, *meter la gamba*, and its nominal variants *metedura de gamba* and *metida de gamba*) or, in the case of Google Translate, other variants with higher frequency scores in widely spoken Latin-American national varieties, as was the case of the diatopically-marked variants *tomada de pelo* and *metida de pata*. These results hence concur with the need for further research to improve on NMTs' performance not only for rare and out-of-vocabulary words but also when facing the challenge of diatopic variation within a given language, which could not be properly seized by standard one-size-fits-all models.

References

- ANASTASIOU, Dimitra. (2010) *Idiom Treatment Experiments in Machine Translation*. Newcastle upon Tyne: Cambridge Scholars Publishing. ISBN-13: 978-1-4438-2515-3 ISBN-10: 1-4438-2515-8
- ANASTASOPOULOS, Antonios. (2019) "An analysis of source-side grammatical errors in NMT." *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, Florence, Italy: Association for Computational Linguistics, pp. 213–223.
- BELINKOV, Yonatan & Yonatan Bisk. (2018) "Synthetic and natural noise both break neural machine translation." *CoRR* abs/1711.02173
- CORPAS PASTOR, Gloria. (2015) "Translating English Verbal Collocations into Spanish: on Distribution and other Relevant Differences related to Diatopic Variation." *Lingvisticae Investigationes Special Issue 'Spanish Phraseology. Varieties and variations'*, 38:2, pp. 229–262. ISSN: 03784169 / E-ISSN: 15699927. DOI 10.1075/li.38.2.03cor.
- CORPAS PASTOR, Gloria. (2018) "Laughing one's head off in Spanish subtitles: a corpus-based study on diatopic variation and its consequences for translation." In: Mogorrón Huerta, Pedro & Albadalejo-Martínez, Antonio. (eds.) 2018. *Fraseología, Diatopía y Traducción, Colección "IVITRA Research in*

- Linguistics and Literature*". Amsterdam: John Benjamins, pp. 54-106. ISBN: 9789027202253.
- COSTA-JUSSÀ, Marta R.; Marcos Zampieri & Santanu Pal. (2018) "A neural approach to language variety translation." *Proceedings of the fifth workshop on NLP for similar languages, varieties and dialects* (VarDial 2018), pp. 275–282.
- FARHAN, W.; Bashar Talafha; Analle Abuammar; Ruba Jaikat; Mahmoud Al-Ayyoub; Ahmad Bisher Tarakji & Anas Toma. (2020) "Unsupervised dialectal neural machine translation." *Information Processing & Management*, 57:3. ISSN 0306-4573. DOI: <https://doi.org/10.1016/j.ipm.2019.102181>
- GÜLÇEHRE, Çaglar; Sungjin Ahn; Ramesh Nallapati; Bowen Zhou & Yoshua Bengio. (2016) "Pointing the unknown words." *CoRR*. abs/1603.08148
- HONNET, Pierre-Edouard; Andrei Popescu-Belis; Claudiu Musat & Michael Baeriswyl. (2017) "Machine translation of low-resource spoken dialects: Strategies for normalizing swiss german." *CoRR*. abs/1710.11035.
- HUANG, Po-Sen; Chong Wang; Sitao Huang; Denny Zhou, & Li Deng. (2018) "Towards neural phrase-based machine translation." *ICLR*.
- KILGARRIFF, Adam; Vít Baisa; Jan Bušta; Miloš Jakubiček; Vojtěch Kovvář; Jan Michelfeit; Pavel Rychlý & Vít Suchomel. (2003) "The Sketch Engine." <<https://www.sketchengine.eu>>
- KOIKE, Kazumi. (2007) "Relaciones paradigmáticas y sintagmáticas de las locuciones verbales en español." In: Cuartero Otal, Juan & Emsel, Martina (eds.), *Vernetzungen Bedeutung in Wort, Satz und Text. Festschrift für Gerd Wotjak zum 65. Geburtstag*. Frankfurt am Main: Peter Lang, pp. 263-275.
- LIU, Nelson F.; Jonathan May; Michael Pust & Kevin Knight. (2018) "Augmenting statistical machine translation with subword translation of out-of-vocabulary words." *arXiv preprint arXiv:1808.05700*.
- LOHAR, Pintu; Maja Popović; Haithem Alfi & Andy Way. (2019) "A systematic comparison between SMT and NMT on translating user-generated content." *20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2019)*. La Rochelle, France.
- RIKTERS, Matīss & Ondřej Bojar. (2017) "Paying attention to multi-word expressions in neural machine translation." *arXiv preprint arXiv:1710.06313*
- ROHANIAN, Omid; Shiva Taslimipoor; Samaneh Kouchaki; Le An Ha & Ruslan Mitkov. (2019) "Bridging the Gap: Attending to Discontinuity in Identification of Multiword Expressions." *Proceedings of the 2019 Conference of the North*

- American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1, pp. 2692 - 2698.
- SAG, Ivan A.; Timothy Baldwin; Francis Bond; Ann Copestake & Dan Flickinger. (2002) "Multiword Expressions: A Pain in the Neck for NLP." In: Alexander Gelbukh (ed.) 2002. *Computational Linguistics and Intelligent Text Processing. CICLing 2002. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, pp. 1–15.
- SENNRICH, Rico; Barry Haddow & Alexandra Birch. (2015) "Neural machine translation of rare words with subword units." *arXiv preprint arXiv:1508.07909*.
- SPERBER, Matthias; Jan Niehues & Alex Waibel. (2017) "Toward robust neural machine translation for noisy input sequences." *International Workshop on Spoken Language Translation (IWSLT)*. Tokyo, Japan.
- VALENCIA GIRALDO, M. Victoria, & Gloria Corpas Pastor. (2019) "The Portrait of Dorian Gray: A Corpus-Based Analysis of Translated Verb + Noun (Object) Collocations in Peninsular and Colombian Spanish." In: Gloria Corpas Pastor & Ruslan Mitkov (eds.) 2019. *Computational and Corpus-Based Phraseology. Europhras 2019, LNAI 11755*. Cham: Springer Nature Switzerland AG. DOI: https://doi.org/10.1007/978-3-030-30135-4_13
- WANG, Xing; Zhaopeng Tu; Deyi Xiong & Min Zhang. (2017) "Translating phrases in neural machine translation." *EMNLP 2017*, pp. 1421–1431.

NOTA BIOGRÁFICA / BIONOTE

CARLOS MANUEL HIDALGO TERNERO es graduado en Traducción e Interpretación (ES<>EN/DE) por la Universidad de Málaga (2015), con un Máster Universitario en Enseñanza del Español como Lengua Extranjera y de Otras Lenguas Modernas, por la Universidad de Sevilla (2016) y un Máster en Profesorado de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanza de Idiomas (especialidad lengua inglesa) por la UMA (2017). En la actualidad es Personal Investigador en Formación (PIF) en la UMA con un contrato FPU, ref. FPU16/02032, concedido por el Ministerio de Educación, Ciencia y Deporte. En este contexto, lleva a cabo su labor investigadora en varios proyectos tales como Interpreta 2.0 (PIE 17-015, type B), VIP (FFI2016-75831-P) y Termitur (HUM2754), así como ha sido miembro del Comité Organizador del Congreso Europhras 2019.

CARLOS MANUEL HIDALGO TERNERO holds a Bachelor's Degree in Translation and Interpreting (ES<>EN/DE), from the University of Malaga (2015), as well as a Master's Degree in Teaching Spanish as a Foreign Language and Other Modern Languages, from the University of Seville (2016), and a Master's Degree in Secondary Education and Language Teaching (English), from the University of Malaga (2017). He's currently a trainee researcher at the University of Malaga thanks to a FPU (Formación de Profesorado Universitario) contract, ref. FPU16/02032, awarded by the Ministerio de Educación, Ciencia y Deporte (Spanish Ministry of Education, Science and Sport). In this context, he carries out his research work in different projects such as Interpreta 2.0 (PIE 17-015, type B), VIP (FFI2016-75831-P) and TERMITUR (HUM2754), and he has also been a member of the Organising Committee for the international conference EUROPHRAS 2019.

Recibido / Received: 10/11/2019
Aceptado / Accepted: 05/02/2020

Para enlazar con este artículo / To link to this article:
<http://dx.doi.org/10.6035/MonTI.2020.ne6.6>

Para citar este artículo / To cite this article:

Valencia Giraldo, M.^a Victoria. (2020) "Standardisation in translated language: A study of collocations in Peninsular and Colombian Spanish translations of *The Picture of Dorian Gray*." 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. 178-209.

STANDARDISATION IN TRANSLATED LANGUAGE: A STUDY OF COLLOCATIONS IN PENINSULAR AND COLOMBIAN SPANISH TRANSLATIONS OF *THE PICTURE OF DORIAN GRAY*¹

M.^a VICTORIA VALENCIA GIRALDO
victoriavalencia@usal.es
University of Salamanca

Abstract

The law of growing standardisation (Toury 1995) appears to be particularly at play in diatopy, and more specifically in the case of transnational languages. Some studies have revealed the tendency to standardise the diatopic varieties of Spanish in translated language (Corpas Pastor 2015a, 2015b, 2017, 2018). However, to our knowledge, no work has studied this tendency in the Spanish translations of a literary work. This paper focuses on verb + noun (object) collocations of Spanish translations of *The Picture of Dorian Gray* by Oscar Wilde. Two different varieties have been chosen (Peninsular and Colombian Spanish). The techniques used to translate this type of collocations in both Spanish translations will be analysed. Further, the diatopic distribution of these collocations will be studied by means of large corpora. Based on the results, it is argued that the Colombian Spanish translation is actually closer to general or standard Spanish than to the variety of this country.

Keywords: Corpus-based translation studies; Diatopic variation; Collocations; Law of growing standardisation; Translation universals.

1. Acknowledgements: The author of this paper is supported by Banco Santander–Universidad de Salamanca International PhD Scholarship.



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

Resumen

La ley de estandarización de Toury (1995) juega un papel importante en la diatopía y, particularmente, en el caso de las lenguas transnacionales. Se ha evidenciado una tendencia a estandarizar las variedades diatópicas del español en la lengua traducida (Corpas Pastor 2015a, 2015b, 2017, 2018). Sin embargo, hasta donde conocemos, no se han realizado estudios sobre esta tendencia en traducciones al español de obras literarias. Este artículo se centra en la traducción al español de las colocaciones verbo-nominales con función de objeto directo de *El retrato de Dorian Gray* de Oscar Wilde, en una traducción al español peninsular y en otra al español de Colombia. Se analizan las técnicas utilizadas para traducir este tipo de colocaciones y se estudia la distribución diatópica de estas mediante corpus. Con base en los resultados, se plantea que la traducción al español de Colombia se acerca más al español general que a la variedad de este país.

Palabras clave: Estudios de traducción basados en corpus; Variación diatópica; Colocaciones; *Law of growing standardisation*; Universales de traducción.

1. Introduction

Corpus-based Translation Studies have promoted research on the features of translated language, by focusing on the process and product of translation, from a descriptive perspective. Some of these features or regularities have been proposed by Toury (1995) under the term of laws of translation, namely the law of growing standardisation and the law of interference. According to Toury's law of growing standardisation, "in translation, source-text textemes tend to be converted into target-language (or target-culture) repertoemes" (1995: 303). In other words, the complicated textual relations established in the source text are replaced in the target text by more common ones in the target discursive-textual community. Thus, translated texts tend to be more conventional and standardised than non-translated texts. Another related notion was introduced by Baker (1993, 1996) under the umbrella term 'translation universals' in order to encompass several common features of translated texts (simplification, explicitation, normalisation, convergence and transference). Based on Toury's law and the universal of normalisation, some research has been carried out on the tendency to standardise the diatopic varieties of Spanish in translated language by means of collocations and other phraseological units (Corpas Pastor 2015a, 2015b, 2017, 2018). In general, it can be

inferred from those studies that the richness of the different Spanish national varieties is not reflected in translated language, nor in the existing bilingual dictionaries. However, to the best of our knowledge, no study has already determined whether the translation of collocations in a literary work reflects a specific Spanish variety or whether it is closer to general or standard Spanish.²

In this paper, we argue that the translation of verb + noun (object) collocations found in the first two chapters of *The Picture of Dorian Gray* by Oscar Wilde (1890/1985) in a translation into Colombian Spanish appears closer to general or standard Spanish than to the Spanish variety of Colombia. In order to demonstrate this, the techniques used to translate this type of collocations in a translation into Peninsular Spanish³ (Wilde 1890/1999) and in another one into Colombian Spanish (Wilde 1890/1994) will be analysed. Moreover, the diatopic distribution of these Spanish collocations will be studied by means of large corpora. Section 2 addresses the definition of collocations, the diatopic variation they exhibit and their rendering into Spanish. Section 3 gives a cursory look at Baker's (1996) concept of translation universals, particularly normalisation, and Toury's law of growing standardisation, in order to explain the reasons why we endorse Toury's law as a convenient theoretical background, as well as previous studies on standardisation of diatopic varieties of Spanish in translation. Then, the choice of corpora is accounted for in Section 4. The main analysis and discussion of results is presented in Section 5, followed by the conclusions (Section 6) and the appendix.

2. Translation of Collocations and Diatopic Varieties of Spanish

The term collocation is understood as the property of language by which speakers tend to produce certain combinations of words among a large number of possible combinations (see Haensch, Wolf, Ettinger *et al.* 1982: 251). From a linguistic point of view, collocations are phraseological units which are free syntagms generated from rules, although, at the same time, they show different ranges of combinatorial restriction conditioned by use

2. General or standard Spanish is understood here as 'global' or 'unified' Spanish spoken/written all over the Spanish-speaking world (see Paffey, 2012).

3. The term Peninsular Spanish will be used to refer to Spanish from Spain, including the islands, Canary and Balearic Islands, Ceuta and Melilla.

(Corpas Pastor 1996: 53). A full discussion on the notion of collocation is beyond the scope of this work. We will present the main features we took into consideration to conduct this study.

Collocations are considered composite units and conventionally restricted structures, in which both collocates have a different semantic status (Hausmann 1989). For example, in *to make a decision* and *tomar una decisión*, nouns constitute the autosemantic bases, while *to make* and *tomar* are collocates, i.e., the meanings of these depend entirely on their respective bases.

As far as verb + noun (object) collocations are concerned, verbs (as collocates) in these combinations exhibit a variety of ranges of collocability: from practically unlimited to practically fixed with intermediate categories (Corpas Pastor 1996: 68-69). In the first type, we find collocations which share collocates and a base of the same semantic field, for instance *desempeñar un cargo*, *una función* or *un papel*, in Spanish. On the opposite extreme of the spectrum, we find practically fixed or very restricted collocations, such as *conciliar el sueño*, *acariciar una idea*, *grind one's teeth*, or *shrug one's shoulders*, whose collocates exhibit a figurative meaning or have a narrow or specific meaning (Aisenstadt 1979 in Howarth 1996: 39). Among the intermediate cases (*asumir una responsabilidad*, for example), a homogeneous group is formed by a delexical verb (or, rather, a verb used in a delexical sense, Howarth 1996), almost grammaticalised, and a noun, generally deverbal, which provides the fundamental meaning. Delexical verbs are highly polysemous verbs, such as *dar*, *tomar*, *hacer*, or *poner*; and *to take* and *to make*, etc., in English.

Collocations express typical relations between components, as can be observed in noun + verb and verb + noun collocations (Koike 2001). *Tocar la guitarra* and *rasguear la guitarra* are collocations, due to the typical relation between its components (Koike 2001). On the contrary, *limpiar la guitarra* and *guardar la guitarra* are not considered collocations, because the noun *guitarra* can only express a typical relation as a musical instrument (Koike 2001).

By repeating a combination of words, speakers recognise it as familiar and use it as if it were a prefabricated fragment (Corpas Pastor 2001). One way to determine whether a certain sequence is sufficiently frequent or salient to be considered conventionalised, habitual or typical is to count its occurrences in a corpus, as suggested by Manning and Schütze (1999). A large number of occurrences is enough to consider that a collocation is frequent, and therefore

entrenched in the mind of speakers. Nevertheless, a significant collocation could be considered typical and cognitively salient (entrenched), even if it might not strike as particularly frequent (Corpas Pastor 2015b: 6).

As Koike (2001) points out, lexical units that form a collocation do not necessarily have to be followed one after the other in a text. In some cases, one or more words appear in the middle of them. Some authors (Sinclair 1966; Jones and Sinclair 1974) have proposed the concept of *collocational span* of four words before and after the keyword or node, but there are collocations in which constituent elements are separated from each other by more than four words (Koike 2001).

Collocations tend to be semantically transparent, but not always predictable (Corpas Pastor 2015b, 2017). Unlike idiomatic phraseological units (idioms), collocations generally do not pose comprehension problems, but mainly in production, especially in second language learning, languages for academic/specific purposes, translation, and interpretation (Corpas Pastor 2016, 2017).

Collocations are not only peculiar to a particular language, but they also reveal variation in their collocational patterns when it comes to a particular text or literary style (Hardy 2007), an author style (Hoover 2003), domain-specific genres and registers (Biber and Conrad 1999, Gledhill 2000, Williams 2002, Copras Pastor 2015a, etc.). They are also useful to characterise language change over time (see Hilper 2006), and diatopic variation (see also Copras Pastor 2015b).

The translation of collocations into the target language is often characterised by the anisomorphism of languages when it comes to choosing collocates (Corpas Pastor 2015b, 2017). While bases are usually translated literally, collocates vary enormously. For example, *to pay homage* cannot be translated into Spanish as **pagar homenaje*, but, generally, as *rendir homenaje*. In general, the translator should try to make the translation of collocations sound as natural as possible in the target language, except, of course, when the function of a collocation is precisely to generate surprise or to draw the attention of the recipients, as frequently occurs in poetry, advertising and humour (Baker 2011).

In the case of transnational languages, such as English or Spanish, the rendering becomes even more complicated, since it is necessary to choose

the language variety into which to translate (Corpas Pastor 2015b). Let us illustrate this again with the collocation *to pay homage*. Spanish offers a wide range of collocates on a scale of formality and grammaticalisation, the most frequent of which are: *dar/hacer un homenaje* (informal), *rendir/tributar/brindar un homenaje* (formal). However, this list is further restricted if diatopic variation is taken into consideration. For instance, *dar y brindar* are not plausible verbal collocates for *homenaje* in Nicaraguan Spanish; in Mexican and Argentinean Spanish *dar* is not used, but *hacer* as delexical verb with *homenaje*. Dominican Spanish only allows collocations with *rendir* and *tributar*; in Honduran Spanish the only option is *rendir*, etc., (Corpas Pastor 2015b).

Translating diatopy is not a small or trivial task, and bilingual dictionaries are not of much help either, as they tend to maintain a simplified and conventional approach to the translation of collocations.⁴ In fact, there is currently no Spanish-English bilingual collocations dictionary (Corpas Pastor 2016, 2017), and in those where equivalents of collocations are included, e.g. Collins, Larousse and Oxford, *rendir homenaje* is offered as the only translation equivalent of *to pay homage* (Corpas Pastor 2015b), which does not at all reflect the richness of the Spanish language. However, according to Corpas Pastor (2015b), the varieties of transnational languages, like Spanish, should be taken into consideration in order to increase the fluency and naturalness of translated texts.

The tension between standard Spanish and transnational Spanish (pluri-centric) has consequences for the type of Spanish (transnational variety) used in translation. For commercial reasons, the translation industry has adopted the presumed neutral Spanish, as opposed to the Spanish located for a specific market (diatopic variety) (Corpas Pastor 2018). But this variety, far from being neutral, is loaded with a positive sociocultural value and prestige, in opposition to other diatopically marked varieties (Corpas Pastor 2018).

4. The term *simplification/simplified* is used in this paper following Lanstyak's & Heltai's (2012: 106) definition: it "can be defined as non-use of complex, marginal, marked or less consolidated vocabulary items or grammatical forms".

3. Features of translated language: The tendency to standardise language

For decades now, many scholars within descriptive translation studies have identified and hypothesised general features of translational behaviour exhibited by translators across time and cultures. In this regard, corpus-based translation studies have promoted empirical research on the search for regularities in translational behaviour combining qualitative and quantitative analysis. The examination of high and low-level linguistic aspects from huge corpora containing texts from different periods of time, places and genres have allowed researchers to confirm or disconfirm the supposedly universal features of translated texts in comparison with non-translated texts.

Yet this issue is far from uncontentious. First, there is still no consensus as to terminology to refer to tendencies of translational behaviour. Some scholars opt to use the term hypothesis, namely the explicitation hypothesis (Blum-Kulka 1986), and the simplification hypothesis (Laviosa-Braithwaite 1996); Toury (1995) proposes his laws of translation, namely the law of growing standardisation and the law of interference; others prefer to talk about universals (Baker 1993, 1996); Chesterman (2004a, 2004b) even distinguishes between S-universals and T-universals.

Particularly, the tendency to standardise/normalise translated language in comparison to non-translated language has received several terms: normalisation, conventionalisation, conservatism (Baker 1993, 1996) and standardisation (Toury 1995). Let us define this tendency, which is central to this paper, and then illustrate it with some studies carried out by different scholars. Baker (1996: 183) defines normalisation/conservatism as follows: “a tendency to exaggerate features of the target language and to conform to its typical patterns”. Normalisation is commonly observed in the use of typical grammatical structures, punctuation and collocational patterns or clichés.

As regards Toury’s law of growing standardisation, it can be synthesised as follows: “In translation, source-text textemes tend to be converted into target-language (or target-culture) repertoremes”. A *repertoire* is understood as “the range of *choices* which makes functions realisable through real products and practices” (Even-Zohar 1990: 40-43). And a *repertoreme* is any sign which forms part of a repertoire. When a repertoreme is retrieved from the repertoire and put into use in a particular text, it becomes part of a unique

network of internal relations in that individual act/text. These relations provide the retrieved repertoire with *ad hoc* textual functions, whereby it is rendered a texteme (Toury 1995: 304). Based on that account, Toury (1995) restates his law: “In translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by a target repertoire”. Unique aspects in a source text are more likely to be transformed into institutionalised aspects in the target text, therefore translated texts show greater standardisation.

In fact, “there seems to be a discernible correlation between the degree of flexibility (or rigidity) with which the law is adhered to in a particular (sub) culture and the position assigned in that culture to translation [...]” (Toury 1995: 306). Therefore, “the more peripheral this status, the more translation will accommodate itself to established models and repertoires” (Toury 1995: 306-307). Standardisation, for example, may be observed in a lesser degree within a culture where translated literature is essential for the development of its own literary system (Even-Zohar 1990, Toury 1995) and, conversely, the degree of standardisation may be greater within a prestigious culture/language.

Now let us provide some examples where we can observe this tendency in translated language. May (1997) identified patterns of normalisation in the French and Russian translations of the modernist fictional works of Virginia Woolf and William Faulkner. The prose of these authors is characterised by the experimental use of punctuation and syntactic structure – conjunctions and parataxis –, which produces visual effects and highlights the interplay of different textual voices. However, these idiosyncratic structures of the authors are lost in translations when adapted to the conventions of the target language, for instance, when finishing incomplete sentences or when separating them if intertwined.

Besides grammar and punctuation, there are other aspects in which standardisation is observed, like lexical-combinatory patterns (May 1997). In this regard, it is worth mentioning Øverås’ (1998) research. This author finds in a bidirectional English-Norwegian parallel corpus that the most creative collocations are translated into more stable and common collocations, which are completely institutionalised in the target language. For instance, the Norwegian free collocation *nyfødt blankt* (literally ‘newborn blank’) is

translated into English as the restricted collocation *pristinely white*. Øverås terms this phenomenon *neutralisation* and explains it by means of Toury's law of growing standardisation.

As far as standardisation of regional varieties in translated language is concerned, Englund Dimitrova (1997: 62) analyses the way dialectal markers in *A Time on Earth* by Swedish writer Vilhem Moberg were translated into Russian and English. She notices that both English and Russian translators “do not employ in the translation any linguistic markers that can be identified as belonging to a specific dialect, nor to some larger regional variety”, but they rather translate the characters' speech into a marked colloquial language. This finding supports the idea that, since translated language tends to be more standardised and conventional, regional or diatopic varieties are underrepresented in comparison with non-translated language.

By conducting some corpus-based studies and the survey of certain bilingual dictionaries, Corpas Pastor (2015a, 2015b, 2017, 2018) has confirmed the fact that translated Spanish does not reflect the actual richness of the diatopic varieties of Spanish (see also Section 2). In her 2018 study, she analyses the translation into Spanish of the English idiom *to laugh one's head off*. This study presents the enormous diatopic variety of expressions in transnational Spanish (more exactly, 74 synonymous idioms) to express this same meaning, extracted from several (sub)corpora of the TenTen⁵ family: esTenTen (general, standard Spanish), esEuTenTen (Peninsular and Latin American Spanish varieties) and EsAmTenTen (Latin American national varieties). By contrast, the author provides the reduced number of equivalents (only 4) used in the Spanish subtitling of English-language films, which were analysed in the parallel subcorpus of Spanish-English subtitles: The Open Subtitles Corpus (2011 version). The results reveal a strong trend towards the standardisation of the idiomaticity and diatopic variety in film subtitling.

The works reviewed above examine standardisation from different perspectives. On one hand, May's (1997), Øverås' (1998) and Englund Dimitrova's (1997) works deal with standardisation of translated texts in comparison with their source texts. On the other hand, Corpas Pastor's (2015a, 2015b, 2017,

5. Available through Sketch Engine: <https://www.sketchengine.co.uk/>. (Subscription required).

2018) investigations highlight standardisation of translated texts with regard to non-translated texts in the target language (which is Spanish in these cases). This distinction of viewpoints when studying features of translated language was first drawn and termed by Chesterman (2004a, 2004b, 2010), namely S-universals and T-universals: “an S-universal formulates a generalisation about a difference between translations and source texts”, whereas “a T-universal claims something about typical differences between translations and non-translations in the target language” (Chesterman 2010: 40). In the specific case of standardisation, we could talk about S-standardisation and T-standardisation. Certainly, this paper delves into T-standardisation of Spanish.

3.1. Why is Toury’s concept of law of growing standardisation more useful than Baker’s universal of normalisation/conservatism?

Most studies on the features of translated language have used Baker’s term *universals*. Yet this term entails a number of problems, which are intended to be outlined very briefly in this section. First of all, it is worth noting that the definition of each universal feature (simplification, explicitation, normalisation, convergence and transference) is actually vague, and the boundaries between the categories used to describe them are not clear-cut, neither are the relations between them (see Pym 2008, for a detailed criticism on Baker’s categories of translation universals). For instance, Baker (1996) illustrates the universals of simplification and normalisation with the same example of experimental uses of punctuation and the way they are normalised in translated texts, which leads one to wonder what the actual difference between the two is.

As defined by Baker, the concept of translation universals implies that those tendencies are the result of the specific constraints of the process of translation, which make them unique and inherent to the process of translation only. Yet, some authors claim that these features of translated texts, as opposed to non-translated texts, are likely derived from specificities and constraints of language in general, rather than from translated language in particular (Halverson 2003, House 2008, Lanstyák and Heltai 2012, Szymor 2017, 2018). Szymor (2017: 34) states that “the universal or general tendencies

in linguistic behaviour of translators will not result from the constraints of the translation process but rather from the constraints that bilingualism and human cognition place on the said linguistic behaviour”. In addition to bilingualism, we believe that interculturalism is at play too. For instance, we may observe standardisation of linguistic varieties when speakers of different diatopic varieties of any transnational language, such as Spanish or English, try to communicate effectively to each other. In this attempt, they may be inclined to use lexical and phraseological units that are comprehensible for users of other varieties they are communicating with at a specific moment. Furthermore, standardisation can also be a strategy that speakers apply to express themselves in more formal communicative situations within their own culture and language.

Additionally, the use of the term *universal* is also problematic because, by definition, it is not suitable for the explanation of regularities of translated language. Since most studies on this issue have adopted the concept of universals, they have barely focused on accounting for the conditions that favour the recurrence of those features. Speaking of translation universals implies the assumption that the translation event is unchangeable and that it occurs within the same socio-cultural and historical context, and under the same conditions and variables. Obviously, this is not the case. There are numerous variables behind each translation event which inevitably exert an influence on the translational behaviour. Among those variables we find cognitive aspects (see Halverson 2003, Halverson 2007), cross-linguistic, or socio-cultural issues, namely: commercial interests of the translation industry (Corpas Pastor, 2016), editing processes (Bisiada 2017, 2019), the use of translation memories (Pym 2008) and bilingual dictionaries (at least in the case of English-Spanish) (Corpas Pastor 2016, 2017), risk-averse translator performance (Pym 2008, 2015). Toury (1995) claims that each factor can be influenced by the presence of another, that is enhanced, offset or mitigated. Consequently, he rather uses the term laws or regularities, instead of universals, because these concepts allow for the existence of exceptions and nuances.

Due to the problems with the term translation universals, and the advantages that Toury's laws present – as we outline below –, we endorse Toury's concept of law of growing standardisation in the development of this study

and advocate employing laws as a convenient theoretical background for describing and explaining general features that spread beyond a certain culture/language, text-type or period of time. Laws of translation offer the possibility to formulate conditioned statements that show the relationship between multifarious factors involved in the recurrence of a given translational behaviour. This concept embraces all kinds of factors, besides cognitive ones, to exert an influence on recurring translational behaviour. Taking into consideration the sociocultural nature of translation as activity, it would be absurd to dismiss the influence of the sociocultural environment; as the concept of translation universals, by definition, does. Laws of translation can be formulated in the form of restricted hypotheses, that is, by adding conditions under which they may or may not occur (or in a lesser or greater degree) and establishing relations between them, difficult though as it may appear. It is worth highlighting that laws can be further modified or enriched as more research is conducted. Laws are certainly more flexible and useful categories to work with than universals. In addition, it is worth mentioning that the idea of formulating conditioned statements about patterns of translational behaviour, encourages scholars to conduct empirical studies on the plausible factors involved in those behaviours. This indeed fosters a better understanding of the translation activity. Finally, an interesting feature of laws of translation is that they have a predictive quality, as Toury (1995, 2004) has emphasised. By establishing proper relations between variables, it is possible to predict future behaviour in translation products.

Considering that the case study which is presented below is part of an ongoing investigation, no in-depth analysis of the variables involved in the translation event has been performed, nor all of them have been identified yet.

4. Choice of Corpora

There are some diachronic corpora of Spanish documents on the web nowadays, such as *Corpus del Español* (CE)⁶, *Corpus de Documentos Españoles Anteriores a 1800* (CODEA)⁷, *Corpus Biblia Medieval* (BM)⁸, *Corpus Hispánico*

6. Available online through: <http://www.corpusdelespanol.org/>.

7. Available online through: <http://corpuscodea.es/>.

8. Available online through: <http://corpus.bibliamedieval.es/>.

y *Americano en la Red: Textos Antiguos* (CHARTA)⁹, *Corpus Diacrónico del Español* (CORDE) and *Corpus Diacrónico y Diatópico del Español Americano* (CORDIAM). For this study, the following diachronic corpora have been selected: the *Corpus Diacrónico del Español* (CORDE)¹⁰ (Diachronic Corpus of Spanish), which contains American Spanish varieties and Peninsular Spanish (available through the Spanish Royal Academy); and the *Corpus Diacrónico y Diatópico del Español de America* (CORDIAM)¹¹ (Diachronic and Diatopic Corpus of American Spanish), which is accessible through the Mexican Academy of the Language.

CORDE was created by the Royal Spanish Academy in the late 1990s. It was the first historical corpus of Spanish. It is a corpus of texts from all the places where Spanish is spoken (23 countries): Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, USA, Spain, Philippines, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Portugal, Puerto Rico, Dominican Republic, Uruguay, Venezuela. Due to its historical perspective, CORDE allocates a total of 74% to Peninsular Spanish and 26% to the remaining 22 varieties. Additionally, there are records of Spanish from all periods, from its beginnings to 1974. CORDE currently contains 250 million words (approximately) from texts pertaining to various subjects (fiction, didactics, science and technology, press, religion, history, law and legal). The CORDE interface allows narrowing searches to specific places, periods of time, and domains.

The *Corpus Diacrónico y Diatópico del Español de América* (CORDIAM) is divided into three subcorpora: CORDIAM-Documents, which contains chronicle, administrative, legal texts and letters; CORDIAM-Literature, where narrative, poetry, prose (satire, dialogues, first manifestations of modern essay, etc.) and theatre are found; and CORDIAM-Press, which is constituted by documents written exclusively in American Spanish from the 19 Latin American countries, besides Southern and Western United States, Jamaica, Haiti and Guyana (former territories of the Spanish crown). CORDIAM-Press

9. Available online through: <http://www.corpuscharta.es/>.

10. *Corpus Diacrónico del Español* (CORDE): <http://corpus.rae.es/cordenet.html>.

11. *Corpus Diacrónico y Diatópico del Español de America* (CORDIAM): <http://www.cordiam.org/>.

contains only texts from the eighteenth and nineteenth centuries because the press exists in America from the eighteenth century. CORDIAM covers texts from four centuries, the oldest dating from 1494 and the most recent from 1905. It is a relatively small corpus containing 4,475,759 words in 6,058 texts (4,178 documents, 15 literature and 1,865 newspapers).

In this study, the searches in CORDIAM were not restricted to any particular type of text or genre, because the portion of literary texts in CORDIAM is too small. However, they were narrowed to the nineteenth century (chronological criterion).

We have chosen to work with CORDE and CORDIAM for the following reasons: both corpora offer users the possibility of performing query searches by using geographical and chronological filters; both contain a wide variety of genres and a large number of words.

The searches for verb + noun (object) collocations in CORDE have been restricted geographically for obtaining individual results from Colombia, Spain and also from general Spanish. Due to the fact that *The Picture of Dorian Gray* was published in 1890, the chronological restriction here has been established between 1850 and 1950, because we consider this to be a reasonable period of time to observe the use of the language of that time, which must also be reflected in the target texts.

5. Rendering of Collocations

Collocations were extracted manually from the source text (ST), as well as their corresponding translation equivalents from the two selected target texts, one from Colombia (Wilde 1890/1994) and the other one from Spain (Wilde 1890/1999). The translation techniques used in the rendering of collocations into Spanish were identified and analysed in the selected texts, based on the classification of techniques proposed by Hurtado Albir (2001).

First, all verb + noun (object) combinations of the first two chapters of the ST and their corresponding translation equivalents in the target texts (TT) were extracted manually. Then some were eliminated until a list of 60 collocations was obtained. The criteria to discard some combinations of this type were typicality and frequency of use, hence the selected collocations express typical relations between their components and present high frequency of

use. Among the combinations that were discarded are to *cause excitement*, *preach the importance*, *return to the ideal*, etc. In cases where there was doubt, some dictionaries and resources were consulted, namely, the monolingual dictionary *Oxford Collocations* (2002), the corpus EnTenTen13 (see footnote 3) (Word Sketch through *Sketch Engine*), which indicates collocations or word combinations for any lemma queried, the *Diccionario Combinatorio Práctico del Español Contemporáneo*, the EsTenTen11 (see also footnote 3) and the diachronic corpora CORDE and CORDIAM. This was done with caution, since these resources do not take into account the diachrony of language, except for CORDE and CORDIAM.

5.1. Translation Techniques

It can be clearly observed that in the Colombian Spanish target text (hereafter CO_SP_TT) (see Table 1) a great variety of techniques were applied in the translation of verb + noun (object) collocations, while in the Peninsular Spanish target text (hereafter PEN_SP_TT) fewer techniques were used (see Table 2). Established equivalence was the preferred option in both target texts.

Table 1. Translation techniques identified in the rendering of collocations in the CO_SP_TT

TECHNIQUE	CO_SP_TT	PERCENTAGE
Established equivalent	47	77%
Linguistic compression	6	10 %
Modulation	2	3 %
Transposition	2	3 %
Literal translation	1	2 %
Discursive creation	1	2 %
Linguistic amplification	1	2 %
Reduction	1	2 %

77% of collocations were translated using the established equivalent technique in the CO_SP_TT. In this category, we group those collocations that were translated as other collocations recognised (by usage or by the dictionary)

in the target language (TL), and which, therefore, constitute the established equivalents, as we can observe in Table 2:

Table 2. Examples of established equivalents in the CO_SP_TT

	ST	CO_SP_TT
(1)	To catch the gleam	Ver el brillo
(2)	To elevate one's eyebrows	Arquear las cejas
(3)	To close one's eyes	Cerrar los ojos
(4)	To tell the truth	Decir la verdad
(5)	To exhibit a picture	Exponer un retrato Exhibir un retrato
(6)	To make a success	Tener éxito
(7)	To play the piano	Tocar el piano
(8)	To make a difference	Establecer una diferencia
(9)	To pay attention	Prestar atención
(10)	Disclose a mystery	Descubrir el misterio

In the vast majority of cases, the base was translated literally and the collocate did vary. In just one case the base was changed in the collocation, as in the translation of *to bear the burden* as *soportar el peso*.

Table 3. Translation techniques identified in the rendering of collocations in the PEN_SP_TT

TECHNIQUE	PEN_SP_TT	PERCENTAGE
Established equivalent	53	89 %
Linguistic compression	5	8 %
Literal translation	1	2 %
Reduction	1	2 %

Similarly, the established equivalent was the most frequently used technique in the translation of collocations in the PEN_SP_TT (see Table 3 and 4). For example, the collocations *to make a success* translated as *tener éxito* or *to play the piano* as *tocar el piano* (see also Table 2) in both target texts. It would not

make any sense to translate literally *to make a success* as *hacer un éxito*, hence it must be translated using the established equivalent in Spanish: *tener éxito*. Other examples are:

Table 4. Examples of established equivalents in the PEN_SP_TT

	ST	PEN_SP_TT
(1)	To produce an effect	Producir un efecto
(2)	To convey the sense	Transmitir la sensación
(3)	To elevate one's eyebrows	Alzar las cejas
(4)	To tell the truth	Decir la verdad
(5)	To exhibit a picture	Exponer un retrato
(6)	To make a contrast	Producir un contraste
(7)	To develop an interest	Fomentar interés
(8)	To paint a portrait	Pintar el retrato
(9)	To gain a reputation	Ganar una reputación
(10)	Break one's promises	Faltar a las promesas

As for the literal translation technique, this was applied in only one instance in both target texts. Many collocations in the English language are transparent, not only in terms of comprehension but also of production, which permits a word-for-word translation into Spanish, although this does not always imply that the resulting translation is a typical and frequent collocation in the TL, that is, the established equivalent. For instance, the collocation *to think (...) thoughts* translated as *pensar (...) ideas* in the PEN_SP_TT and *pensar con pensamientos* in the CO_SP_TT.

On the other hand, the linguistic compression technique was applied in the translation of collocations as single verbs in the TL, instead of as other collocations, resulting in a synthesis of linguistic elements. Most of these cases occur with collocations formed by delexical, almost grammaticalised verbs, and a generally deverbal noun, which provides the main semantic load (Corpas Pastor 1996: 69). Some collocations that illustrate this are:

Table 5. Examples of linguistic compression in both target texts

	ST	CO_SP_TT	PEN_SP_TT
(1)	to have no appreciation	No apreciar	No apreciar
(2)	to make no answer	Sin responder	No contestar
(3)	to make a difference	-	Distinguir
(4)	To ring the bell	Llamar	Llamar

Likewise, as an example of the use of this technique, we include collocations that were translated as more general verbs, like (4) *to ring the bell* and *llamar* (in both TT).

Other techniques were used less frequently in the translations analysed. In the CO_SP_TT, for instance, a change of point of view (modulation) can be observed in the translation of *to keep one's promises* as *faltar a sus promesas* and *to break one's promises* as *no cumplir sus promesas*. Additionally, a change of grammatical category is noticed in the translation of *to cure the soul* as *remedio para el alma* (transposition). In this same TT, a linguistic amplification is employed in the translation of the collocations *to frame [a portrait]* as *poner en marco [el retrato]*; and an ephemeral equivalence (discursive creation) is established for the collocation *to convey the sense* translated as *plasmar la velocidad*, which is not a frequent collocation in Spanish. Also, the reduction technique is clearly used in the omission of the collocations *to have a romance* in the PEN_SP_TT and *to give a definition* in the CO_SP_TT.

Not all collocations from the source language were translated as other collocations in the target language. Some of them were translated as free combinations of verb + noun, ephemeral equivalents (discursive creation), as verbs that contain all the semantic load, or they were simply omitted. Most of these instances were observed in the CO_SP_TT and others, such as the discursive creation, were pinpointed exclusively in this TT.

The English collocations found in the ST were consulted in some popular bilingual online dictionaries too, namely the *Oxford Dictionary*¹², *Collins*

12. *Oxford English-Spanish Dictionary* <<https://es.oxforddictionaries.com/>> (Accessed: 01-04-19).

*Dictionary*¹³, *Cambridge Dictionary*¹⁴, and the *Linguee Dictionary*¹⁵, with the aim of observing the repertoire of translation (English-Spanish) options they offer for collocations. Undoubtedly, the treatment of collocations in bilingual dictionaries is now extremely poor compared to that in monolingual dictionaries (Corpas Pastor 2017), and their approach to the translation of collocations is usually very simplified. In general, the *Linguee dictionary* is the one that provides the highest number of possible equivalents for collocations, for example:

Table 6. Equivalents for *to produce an effect* in bilingual dictionaries

To produce an effect	<i>Linguee</i>	<i>Collins Dictionary</i>
	(1) Causar un efecto	(1) Producir un efecto
	(2) Tener un efecto	(2) Surtir un efecto
	(3) Generar un efecto	
	(4) Producir un efecto	

Producir un efecto (4) is the one that has more instances. Conversely, in the other two bilingual dictionaries analysed no translation equivalent is found for this same collocation. Another example is *to yield to (a) temptation*.

Table 7. Equivalents for *to yield to (a) temptation* in bilingual dictionaries

To yield to (a) temptation	<i>Linguee</i>	<i>Oxford dictionary</i>	<i>Collins</i>
	(1) Rendirse a la tentación	(1) Ceder a la tentación	(1) Ceder a la tentación
	(2) Caer en la tentación		
	(3) Sucumbir a la tentación		

13. *Collins English-Spanish Dictionary* <<https://www.collinsdictionary.com/>> (Accessed: 01-04-19).

14. *Cambridge English-Spanish Dictionary* <<http://dictionary.cambridge.org/es/diccionario/ingles-espanol/>> (last accessed: 01-04-19).

15. *Linguee English-Spanish Dictionary* <<http://www.linguee.es/>> (last accessed: 01-04-19).

The English collocation *to exhibit the picture/portrait* is included in the *Collins* and *Cambridge dictionaries* with the verb *exponer* as the only translation equivalent, albeit in some Latin American countries the verb *exhibir* is a more common collocate for the nouns *cuadro/pintura*. Regarding the collocation *to pay attention*, its only translation equivalent found in these dictionaries is *prestar atención*, despite the fact that the verb *poner* is also a very frequent collocate used in American Spanish. It is evident that bilingual dictionaries do not reflect the diatopic variation of Spanish, but rather favour general Spanish as regards the translation of verb + noun (object) collocations. These results support those obtained by Corpas Pastor (2015b, 2017). In the following section, the results of the search for collocations in CORDE and CORDIAM are presented.

5.2. Searching Collocations in CORDE and CORDIAM

After identifying and analysing the techniques used in the translation of collocations in chapters 1 and 2 of *The Picture of Dorian Gray*, it was necessary to discard those that are not considered collocations in the target language, in order to carry out the analysis in the Spanish corpora (CORDE and CORDIAM), whose results are presented below and in Appendix 1¹⁶.

A total of 50 collocations were extracted from the PEN_SP_TT, out of which 23 (46%) were also found in the CO_SP_TT, which means that both target texts share a high percentage of collocations. With respect to the results yielded by the large corpora, 44 (88%) collocations out of those 50 also occurred in general Spanish (GEN_SP) from CORDE. Forty-seven (94%) of these collocations appeared in Peninsular Spanish (PEN_SP) from CORDE and 3 (6%) were not found in CORDE or CORDIAM.

On the other hand, in CORDIAM only 9 (18%) collocations out of the total 50 were found in the GEN_SP and only 3 (6%) collocations in the Colombian Spanish (CO_SP) from CORDIAM: *producir un efecto*, *decir la verdad* and *prestar atención*.

A total of 48 collocations were found in the CO_SP_TT, out of which 23 (47.91%) were also found in the PEN_SP_TT. As for large corpora, 46

16. See Appendix 1 for an account of the number of occurrences of some collocations in CORDE and CORDIAM.

(95.83%) collocations out of those 48 appeared in the GEN_SP from CORDE, and 29 (95.83%) out of these also occurred in the CO_SP from CORDE, 2 (4.16%) were not found in CORDE or CORDIAM, and all of these collocations occurred in the PEN_SP (100%) from CORDE. As we can observe, a high percentage of the collocations that occur in the Colombian Spanish target text are used in Peninsular Spanish and in many other diatopic varieties.

Regarding CORDIAM, only 11 (22.91%) collocations out of 48 were found in the GEN_SP. Only 2 (4.16%) collocations occurred in the CO_SP. Let us summarise this information in Table 8 and 9:

Table 8. Summary of results retrieved from CORDE

PEN_SP_TT (50 collocations)	Percentage	CO_SP_TT (48 collocations)	Percentage
44 collocations occurred in the GEN_SP.	88%	46 collocations appeared in the GEN_SP.	95.83%
47 collocations appeared in the PEN_SP.	94%	29 also occurred in the CO_SP.	60.41%
		All of these collocations occurred in the PEN_SP.	100%

Table 9. Summary of results retrieved from CORDIAM

PEN_SP_TT (50 collocations)	Percentage	CO_SP_TT (48 collocations)	Percentage
9 collocations were found in the GEN_SP.	18%	11 collocations were found in the GEN_SP.	22.91%
3 collocations occurred in the CO_SP.	6%	2 collocations occurred in the CO_SP.	4.16%

The results retrieved from CORDIAM (see Table 9 and Appendix 1) were very poor, mainly those concerning Colombian Spanish. This is due to the small size of CORDIAM and the fact that the queries were restricted to the 19th century, which further reduced the chances of finding occurrences in this corpus.

Furthermore, the larger proportion of Peninsular Spanish words contained in CORDE (see Table 8 and Appendix 1) is justified by the fact that this is a historical corpus of Spanish from all the countries where Spanish has been spoken, thus the documents from Spain outnumber those from elsewhere.

The data reveals that the translation of collocations in the CO_SP_TT is closer to general Spanish than to Colombian Spanish, inasmuch as this TT and the PEN_SP_TT share a high percentage of collocations. Those collocations that were found only in the CO_SP_TT occurred in the corpora in other varieties too, especially in Peninsular Spanish, which indicates that none of these collocations are diatopically restricted to Colombian Spanish. Some collocations, despite appearing in the CO_SP_TT, were not found in the Colombian Spanish from CORDE or CORDIAM, albeit they were retrieved from other national varieties of CORDE. For example:

1. *Crear un efecto*
2. *Arquear las cejas*
3. *Exhibir el retrato*
4. *Sentir placer*
5. *Ceder a la tentación*

However, no occurrences of the collocations *crear (un) efecto* and *ganar (un) empuje*, extracted from the CO_SP_TT, were observed in any of the corpora, suggesting perhaps a very low frequency of use of such collocations in Spanish between 1850–1950. In the case of the collocation *crear (un) efecto*, as in the vast majority, it was necessary to check each of the concordances and discard some of them for not complying with the syntactic function we expected to find in the corpora.

It is also noteworthy that there were no instances in CORDE or CORDIAM of four collocations extracted from the PEN_SP_TT:

1. *Captar el efecto*
2. *Despilfarrar el oro*
3. *Captar el significado*
4. *Enmarcar un retrato*

This may be due to several factors: first, these collocations are neither very frequent nor very typical in the language; second, generally collocations

allow the variation of at least one of their components without affecting the meaning of the others, for instance *despilfarrar el oro*, where the noun *oro* could be replaced by *dinero/fortuna/recursos*, etc. *Despilfarrar dinero* does have occurrences in CORDE, although not in Colombian Spanish. As for the absence of the collocations *captar (el) significado* and *enmarcar (un) retrato* from both corpora, it could be attributed to a rare use of them in the period of time consulted.

The established equivalent was the most used technique in the translation of collocations in both target texts, which results in more conventional and standard translations, and less creative and experimental target language. Although a greater number of techniques were applied in the Colombian Spanish target text besides the established equivalent, this TT reflects a general Spanish. In other words, it does not exhibit specific features of the Colombian Spanish variety, as regards collocations, but, on the contrary, it is more similar to general Spanish.

On the other hand, the bilingual dictionaries consulted are characterised by maintaining this tendency to standardise translated language, listing only one or two translation equivalents for certain collocations. These dictionaries are indeed far from reflecting the reality of the diatopic variety of a transnational language like Spanish. In most cases, the collocations that we looked up in these dictionaries were not even listed. This is evidence of standardisation in translated language and, at the same time, can be a contributing factor for maintaining standardisation in translated language since dictionaries are one of the translators' essential tools. Other important tools are translation memories and machine translation. According to Pym (2008: 325), the use of translation memories reinforces standardisation because they offer consistency in terminology and phraseology. By the same token, we consider that corpus-based machine translation tools (statistical machine translation and Neural machine translation) also reinforce standardisation because, broadly speaking, they rely on huge parallel corpora, which results in more usual and standard translated language and less creative and unusual one.

Additionally, Pym (2008, 2015) argues that standardisation is a strategy used by translators to reduce communicative risk. Translators tend to select more frequent and typical collocations that are recognised and accepted by most of the Spanish-speaking community as the prestigious variety, rather

than select a diatopically marked variety that might sound odd and incorrect to the majority. Furthermore, standardisation practices may also take place during the editing process carried out by publishing houses, as Bush (1998:129) puts it: “In the English, Spanish and Portuguese-speaking worlds, for example, there will be issues of different dialects and editors who will only accept their variety of standard [...]. The editor’s reading, however, need not simply be a threatening and standardizing project”.

Today, research by means of diachronic and diatopic Spanish corpora faces numerous limitations. In this study, we had to deal with the fact that CORDE and CORDIAM are neither lemmatised nor grammatically annotated, which implies a slower data extraction process. Therefore, it was necessary to check each of the results yielded by the corpora to verify that these actually reflected the linguistic phenomenon we intended to analyse. To illustrate this, let us give the example of the query *mezcl* dist/5 color**, which yielded instances containing adjectives like *mezclado*, nouns such as *mezcla*, and verb + noun (subject) collocations, but these were ruled out as being beyond the scope of this study.

The total size of the subcorpus of Colombia, or that of any particular national variety, is not specified in either of the corpora. For this reason, it was not possible to establish the normalised frequency of each collocation. This, in turn, prevented us from comparing objectively the frequency of use of the collocations in different countries (subcorpora), in order to determine whether each collocation was more typical of and more frequent in one national variety or another – which is what was intended, generally speaking, in this research. Another disadvantage of CORDE is that in order to be able to check the concordances resulting from a given query, the maximum number of these cannot exceed 2,000. Consequently, it is necessary to formulate query searches as specific as possible so that all results are shown.

As far as CORDIAM is concerned, its size is much smaller than that of CORDE. Therefore, when query searches in CORDIAM are narrowed using geographical and chronological criteria, the results are very poor, or even zero. Furthermore, this corpus is composed mainly of legal, administrative, private documents and chronicles. Only a small percentage of the total corpus is made up of literary texts, which largely conditions the results of queries.

6. Conclusions

Generally, translated language tends to exhibit more features of standardisation than non-translated language. In the case of translated Spanish, we observe a tendency to avoid the use of diatopically marked phraseological units and, on the contrary, favour those that correspond to general or standard Spanish. In this paper, we argue that the rendering of verb + noun (object) collocations into Colombian Spanish of the first two chapters of *The Picture of Dorian Gray* by Oscar Wilde is indeed closer to general Spanish than to the Colombian Spanish variety. The data reveal that the most frequently applied technique in the rendering of collocations in both target texts was the established equivalent. The Colombian Spanish target text shares 23 out of the 48 collocations (47.91%) with the peninsular Spanish target text. Secondly, the results of the analysis of the diatopic distribution of the collocations in CORDE show that 46 out of the 48 collocations (95.83%) found in the Colombian Spanish target text are also used in general Spanish and 100% of these in Peninsular Spanish. Moreover, no diatopically restricted collocations to Colombian Spanish variety were found.

The evidence presented in this study also reveals that bilingual dictionaries (English-Spanish) show a general and simplified approach to translation since they usually list only one or two translation equivalents, or even none, for certain collocations. This may indicate that the reality of diatopic varieties of Spanish, as a transnational language, is not taken into consideration. In conclusion, these findings support the starting hypothesis of this study, and they contribute to the research on standardisation of national varieties of Spanish in translated texts, insofar as they provide evidence on translated literature from English into Spanish. This could be explained by many factors, including, but not limited to, commercial interests of the translation industry, the use of translation memories, machine translation tools, and bilingual dictionaries (English-Spanish), risk-averse translator performance, editing process in both countries, among others. Undoubtedly, more thorough research on the variables that affect translational behaviour, and specifically the tendency towards standardisation (publishing-related, socio-cultural, and cognitive variables) is badly required to be able not only to describe this

translational behaviour, but also to explain it more accurately by means of Toury's law of growing standardisation.

However, a note of caution is in order at this point. The limitations of this study could have affected the distributional results obtained from CORDE and CORDIAM as to frequency and diatopy. Moreover, these conclusions are also limited by the fact that this study only deals with two translated texts. Further studies of this nature for Spanish and other types of phraseological units are badly needed in order to reveal 'true' regularities in translated language. A first step will be to revise and expand the existing diachronic and diatopic Spanish corpora, by size increase and annotation enrichment (particularly at the morphosyntactic level).

Appendix 1

TT	COLLOCATION	GEN_SP		CO_SP		PEN_SP
		CORDE	CORDIAM	COR-DE	CORDIAM	CORDE
PEN_SP_	Vislumbrar (el) resplandor	2	-	-	-	2
CO_SP_	Ver (el) brillo	16	-	-	-	10
Both	Soportar (el) peso	69	-	2	-	54
PEN_SP_	Producir (un) efecto	486	12	5	1	251
CO_SP_	Crear (un) efecto	-	-	-	-	-
PEN_SP_	Transmitir (la) sensación	2	-	-	-	1
Both	Cerrar (los) ojos	1022	2	27	-	738
CO_SP_	Arquear (las) cejas	51	-	-	-	48
PEN_SP_	Alzar (las) cejas	11	-	-	-	10
PEN_SP_	Ganar (una) reputación	7	-	-	-	4
CO_SP_	Adquirir fama	72	4	2	-	57
Ambos	Exponer (el) retrato	7	-	-	-	6

PEN_SP_	Exponer (el) cuadro	19	-	-	-	19
CO_SP_	Exhibir (el) retrato	7	-	-	-	4
CO_SP_	Menear (la) cabeza	139	-	2	-	97
Both	Decir (la) verdad	575	104	10	3	450
Both	Contar historias	163	-	1	-	130
Both	Responder (a la) pregunta	199	2	4	-	143
Both	Tener éxito	502	1	5	-	364
Both	Tocar (el) piano	199	-	-	-	126
CO_SP_	Establecer (una) diferencia	75	-	4	-	61
PEN_SP_	Presentar (una) idea	13	1	-	-	9
CO_SP_	Exponer (una) idea	28	-	1	-	26
PEN_SP_	Hallar placer	23	-	-	-	19
CO_SP_	Sentir pacer	63	-	-	-	34

References

- ACADEMIA MEXICANA DE LA LENGUA, *Corpus Diacrónico y Diatópico del Español de América (CORDIAM)*. Accessed 20 Apr 2019.
- BAKER, Mona. (1993) “Corpus Linguistics and Translation Studies: Implications and Applications”. In: Baker, Mona; Gill Francis & Elena Tognini-Bonelli (eds.) 1993. *Text and Technology: in Honor of John Sinclair*. Amsterdam: John Benjamins, pp. 233–250.
- BAKER, Mona. (1996) “Corpus-Based Translation Studies: The Challenges That Lie Ahead”. In: Somers, Harold (ed.) 1996. *Terminology, LSP and Translation*. Amsterdam: John Benjamins, pp. 175–186.
- BAKER, Mona. (2011) *In Other Words: A Coursebook on Translation*. London: Routledge.
- BIBER, Douglas; Conrad, Susan. (2009) *Register, Genre, and Style. Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University Press.

- BISIADA, Mario. (2017) "Universals of Editing and Translation." In: Hansen-Schirra, Silvia; Oliver Czulo & Sascha Hofmann (eds.) 2017. *Empirical Modelling of Translation and Interpreting*. Berlin: Language Science Press, pp. 241–275.
- BISIADA, Mario. (2019) "Translated Language or Edited Language? A Study of Passive Constructions in Translation Manuscripts and Their Published Versions." *Across Languages and Cultures* 20:1, pp. 35–56.
- BLUM-KULKA, Shoshana. (1986). "Shifts of Cohesion and Coherence in Translation". In House, Juliane & Shoshana Blum-Kulka (eds.) 1986. *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, pp. 17-36.
- BOSQUE, Ignacio (2004): *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: Ediciones SM.
- BUSH, Peter. (1998) "Literary translation practices". In: Baker, Mona (ed.) 1998. *Routledge Encyclopedia of Translation Studies*. London: Routledge, pp. 127–130.
- CHESTERMAN, Andrew (2004a). "Beyond the particular". In: Mauranen, Anna & Pekka Kujamäki (eds.) 2004. *Translation universals: Do they exist?* Amsterdam: John Benjamins, pp. 33–49.
- CHESTERMAN, Andrew (2004b). "Hypotheses about Translation Universals". In: Gyde Hansen, Kirsten Malmkjaer & Daniel Gile (eds) 2004. *Claims, Changes and Challenges in Translation Studies*, Amsterdam: John Benjamins, pp. 1-13.
- CHESTERMAN, Andrew. (2010) "Why study translation universals?" *Kiasm. Acta Translatologica Helsingiensia (ATH)* 1, pp. 38–48.
- CORPAS PASTOR, Gloria. (1996) *Manual de Fraseología Española*. Madrid: Gredos.
- CORPAS PASTOR, Gloria. (2001) "En torno al concepto de colocación". *Euskera* 1, p. 89–108.
- CORPAS PASTOR, Gloria. (2015a) "Register-specific collocational constructions in English and Spanish: a usage-based approach". *Journal of Social Sciences*, 11:3, pp.139–151.
- CORPAS PASTOR, Gloria. (2015b) "Translating English verbal collocations into Spanish: on distribution and other relevant differences related to diatopic variation". *Lingvisticæ Investigationes*, 38:2, pp. 229–262.
- CORPAS PASTOR, Gloria. (2016) "Collocations dictionaries for English and Spanish: the state of the art". In: Orlandi, Adriana & Laura Giacomini (eds.)

2016. *Defining collocations for lexicographic purposes: from linguistic theory to lexicographic practice*. Frankfurt: Peter Lang, pp. 173–208.
- CORPAS PASTOR, Gloria. (2017) “Collocations in e-bilingual dictionaries: from underlying theoretical assumptions to practical lexicography and translation issues”. In: Torner Castells, Sergi & Elisenda Bernal Gallen (eds.) 2017. *Collocations and Other Lexical Combinations in Spanish. Theoretical and Applied Approaches*. London: Routledge, pp. 139–160.
- CORPAS PASTOR, Gloria. (2018) “Laughing one’s head off in Spanish subtitles: a corpus-based study on diatopic variation and its consequences for translation”. In: Mogorrón, Pedro & Juan Antonio Albaladejo Martínez (eds.) 2018. *Phraseology, Diatopy and Translation*, Amsterdam: John Benjamins, pp. 54–106.
- ENGLUND DIMITROVA, Birgitta. (1997) “Translation of Dialect in Fictional Prose-Vilhelm Moberg in Russian and English as a Case in Point.” *Norm, variation and change in Language (= Stockholm Studies in Modern Philology)* 11, pp. 49-65.
- EVEN-ZOHAR, Itamar. (1990). “Polysystem Theory.” *Poetics Today* 11:1, pp. 9–26.
- GLEDHILL, Christopher. (2000) *Collocations in Science Writing*. Tübingen: Gunter Narr.
- HAENSCH, Günther; Larry Wolf; Stephan Ettinger & Ray Werner. (1982) *La Lexicografía. De La Lingüística Teórica a La Lexicografía Práctica*. Madrid: Gredos.
- HALVERSON, Sandra L. (2003). “The Cognitive Basis of Translation Universals.” *Target* 15:2, pp. 197–241.
- HALVERSON, Sandra L. (2007) “A Cognitive Linguistic Approach to Translation Shifts.” *Belgian Journal of Linguistics* 21, pp. 105–121.
- HARDY, Donald E. (2007) *The Body in Flannery O’Connor’s Fiction: Computational Technique and Linguistic Voice*. Columbia: University of South Carolina.
- HAUSMANN, Franz Josef. (1989) “Le Dictionnaire de Collocations”. In: Hausmann, Franz Josef; Oskar Reichmann; Herbert Ernst Wiegand & Ladislav Zgusta (eds.) 1989. *Wörterbücher. Dictionaries. Dictionnaires. Eininternationals Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie Internationale de Lexicographie*. Berlin: Walter de Gruyter, pp. 1000-1019.
- HILPER, Martin. (2006) “Distinctive collexeme analysis and diachrony”. *Corpus Linguistics and Linguistic Theory* 2:2, pp. 243–257.

- HOOVER, David. (2003) "Frequent collocations and authorial style". *Literary and Linguistic Computing* 18:3, pp. 261–286.
- HOUSE, Juliane. (2008) "Beyond Intervention: Universals in Translation?" *Trans-Kom* 1:1: pp. 6–19. Electronic version: <http://www.trans-kom.eu/bd01nr01/trans-kom_01_01_02_House_Beyond_Intervention.20080707.pdf>
- HOWARTH, Peter Andrew. (1996) *Phraseology in English Academic Writing. Some Implications for Language Learning and Dictionary Making*. Tübingen: Max Niemeyer.
- HURTADO ALBIR, Amparo. (2001) *Traducción y Traductología: Introducción a la Traductología*. Madrid: Cátedra.
- JONES, Susan & John M. Sinclair. (1974) "English lexical collocations. A study in computational linguistics". *Cahiers de Lexicology* 24, pp. 15–61.
- KOIKE, Kazumi. (2001) *Colocaciones léxicas en el español actual: estudio formal y léxico-semántico*. Alcalá de Henares: Universidad de Alcalá/Takushoku University.
- LANSTYÁK, István & Pál Heltai. (2012) "Universals in Language Contact and Translation." *Across Languages and Cultures* 13:1, pp. 99–121.
- LAVIOSA-BRAITHWAITE, Sara. (1996) *The English Comparable Corpus (ECC): A Resource and Methodology for the Empirical Study of Translation*. PhD Thesis, Manchester: UMIST.
- MANNING, Christopher D. & Hinrich Schütze. (1999) *Foundations of Statistical Natural Language Processing*. Cambridge: The MIT Press.
- MAY, Rachel. (1997) "Sensible Elocution." *The Translator* 3:1, pp. 1–20.
- OXFORD COLLOCATIONS DICTIONARY FOR STUDENTS OF ENGLISH. (2002) OXFORD: OXFORD UNIVERSITY PRESS.
- ØVERÅS, Linn. (1998) "In Search of the Third Code: An Investigation of Norms in Literary Translation." *Meta: Journal Des Traducteurs* 43:4, pp. 557-570. Electronic version: <<https://doi.org/10.7202/003775ar>>
- PAFFEY, Darren. (2012) *Language Ideologies and the Globalisation of 'Standard' Spanish*. London: Bloomsbury.
- PYM, Anthony. (2008) "On Toury's laws of how translators translate". In: Pym, Anthony; Miriam Shlesinger & Daniel Simeoni (eds.) 2008. *Beyond Descriptive Translation Studies*. Amsterdam: John Benjamins, pp. 311–328.
- PYM, Anthony. (2015) "Translating as Risk Management." *Journal of Pragmatics* 85, pp. 67–80.
- REAL ACADEMIA ESPAÑOLA: *Data Bank (CORDE)*. *Corpus Diacrónico del Español*. Electronic versión: <<http://www.rae.es>>

- SINCLAIR, John. (1966) "Beginning the study of lexis". In: Bazell, Charles Ernest; John Catford; Michael Halliday & Robert Robins (eds.) 1966. *In Memory of J.R. Firth*. London: Longman, pp. 410–430.
- SZYMOR, Nina. (2017) *Translation Universals: A Usage-Based Approach*. PhD thesis. Sheffield: University of Sheffield. Electronic version: <<http://etheses.whit-rose.ac.uk/18431/>>
- SZYMOR, Nina. (2018) "Translation: universals or cognition. A usage-based perspective", *Target* 30:1, pp. 53–86.
- TOURY, Gideon. (1995) *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins, 2012 (2nd. ed.)
- TOURY, Gideon. (2004) "Probabilistic Explanations in Translation Studies: Welcome as They Are, Would They Qualify as Universals?" In: Mauranen, Anna & Pekka Kujamäki (eds.) 2004. *Translation Universals. Do They Exist?* Amsterdam: John Benjamins, pp. 15-32.
- WILDE, Oscar. (1890) *The Picture of Dorian Gray*. London: Penguin Classics, 1985.
- WILDE, Oscar. (1994) *El Retrato de Dorian Gray*. Translated by anonymous. Bogotá: Panamericana.
- WILDE, Oscar. (1999) *El retrato de Dorian Gray*. Translated by J. López Muñoz. Madrid: El Mundo.
- WILLIAMS, Geoffrey. (2002) "In search of representativity in specialised corpora – categorisation through collocation". *International Journal of Corpus Linguistics* 7:1, pp. 43-64.

BIONOTE / NOTA BIOGRÁFICA

MARÍA VICTORIA VALENCIA GIRALDO holds a degree in Foreign Languages (English and French) from the University of Valle (Colombia). She completed her master's degree in Translation at the University of Málaga with funding provided by the AUIP (the Iberoamerican Association of Postgraduate Universities). She is currently undertaking PhD studies in Translation at the University of Salamanca, supported by Banco Santander–Universidad de Salamanca International PhD Scholarship. Her doctoral thesis deals with the study of the tendency to standardise diatopic varieties of Spanish in literary translation from an empirical and descriptive perspective. She is a member of the editorial board of the journal CLINA from the Department of Translation

and Interpreting, University of Salamanca. M.^a Victoria has worked as an English teacher at the University of Valle and as a freelance translator of academic texts.

MARÍA VICTORIA VALENCIA GIRALDO es licenciada en Lenguas Extranjeras (inglés y francés) de la Universidad del Valle (Colombia). Tiene un máster en traducción de la Universidad de Málaga, el cual fue financiado gracias a una beca de la AUIP (Asociación Universitaria Iberoamericana de Posgrado). Actualmente está llevando a cabo estudios de doctorado en traducción en la Universidad de Salamanca, con la financiación de la beca de doctorado del Banco Santander –Universidad de Salamanca. Su tesis doctoral se centra en la investigación de la tendencia a la estandarización de las variedades dia-tópicas del español en la traducción literaria, desde una perspectiva empírica y descriptiva de la traducción. Es miembro del consejo de redacción de la revista CLINA del Departamento de Traducción e Interpretación de la Universidad de Salamanca. M.^a Victoria ha trabajado como profesora de inglés en la Universidad del Valle y se ha desempeñado como traductora de textos académicos.

Recibido / Received: 11/11/2019

Aceptado / Accepted: 16/03/2020

Para enlazar con este artículo / To link to this article:

<http://dx.doi.org/10.6035/MonTI.2020.ne6.7>

Para citar este artículo / To cite this article:

León Araúz, Pilar & Melania Cabezas García. (2020) "Term and translation variation of multi-word terms." 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. 210-247.

TERM AND TRANSLATION VARIATION OF MULTIWORD TERMS

LA VARIACIÓN TERMINOLÓGICA EN LAS TRADUCCIONES DE LOS TÉRMINOS COMPUESTOS

PILAR LEÓN ARAÚZ

pleon@ugr.es

Universidad de Granada

MELANIA CABEZAS GARCÍA

melaniacabezas@ugr.es

Universidad de Granada

Abstract

Phraseology is central to specialized language. In scientific and technical communication, multiword terms (MWTs) (e.g. *volatile organic compound*) are the most frequent type of phraseological units. Rendering them into another language is not an easy task due to their cognitive complexity, the proliferation of different forms, and their unsystematic representation in terminographic resources. This often results in a broad spectrum of translations for MWTs, leading to higher term variation as a result of their composition by two or more constituents. In this study we carried out a quantitative and qualitative analysis of English term variants of MWTs from the environmental domain and their translations into Spanish. The focus was on translation variation and its occurrence in different linguistic resources.



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

Keywords: Multiword term; Term variation; Translation; Corpus linguistics; Terminographic resource.

Resumen

La fraseología es una parte fundamental del lenguaje especializado. En la comunicación científico-técnica, los términos compuestos (p. ej. *compuesto orgánico volátil*) son las unidades fraseológicas más habituales. Trasladarlos a otra lengua presenta dificultades debido a su complejidad cognitiva, la proliferación de diferentes formas y su representación inconsistente en los recursos terminográficos. Ello suele derivar en un amplio abanico de traducciones para estos términos compuestos, que de por sí son proclives a la variación terminológica debido a su formación por dos o más elementos. En este trabajo realizamos un análisis cuantitativo y cualitativo de las variantes terminológicas de términos compuestos en inglés pertenecientes al ámbito medioambiental y de sus traducciones al español. En concreto, dedicamos especial atención a la variación en la traducción y su cobertura en diferentes recursos lingüísticos.

Palabras clave: Término compuesto; Variación terminológica; Traducción; Lingüística de corpus; Recurso terminográfico.

1. Introduction

Phraseology has been extensively studied in both general language (Firth 1957; Sinclair 1991; Mel'čuk et al. 1995; Corpas Pastor 1996, 2000; Mogorrón-Huerta 2010; *inter alia*) and specialized discourse (Picht 1991; Meyer & Mackintosh 1996; *inter alia*). The guiding principle of phraseology is that language is not only composed of individual words, but also of larger structures that function as a whole (Sinclair 1991). In their broadest sense, phraseological units are lexical units formed by two or more words that frequently co-occur and exhibit a variable degree of lexicalization, syntactic and semantic stability, and a possible idiomatic nature (Gläser 1988; Corpas Pastor 2000). There are different types of phraseological units (e.g. idioms, collocations, proverbs, etc.), whose consideration as such depends on the scope of the approach and how they comply with these features.

Specialized phraseological units have been defined as phraseological units conveying a specialized meaning which are frequently used in a scientific or technical domain, contain at least one term, and exhibit a certain degree of

lexicalization (Bevilacqua 2004: 28), e.g. *to generate power*. Specialized phraseological units are fundamental since they give semantic precision to terms, play a central role in conceptual systems, and allow expression according to the conventions of the different specialized domains. However, they have been less studied than their homologues in general language.

Multiword terms (MWTs) are the main type of phraseological unit in specialized discourse (Meyer & Mackintosh 1996; Ramisch 2015). In MWTs, two or more lexemes converge to form a new unit of meaning (e.g. *shrouded wind turbine*). These lexemes can join and form a graphic compound (*sacacorchos*) or keep the space in between (*pez espada*). Although the designation of *compound* is often restricted to the first type (Corpas Pastor 1996), we consider both types as compounds, even though our study focuses on separated or syntagmatic compounds.

Some authors refuse to consider MWTs as phraseological units (Zuluaga 1975; García-Page 2008). However, we agree with those including them in the group (Benson et al. 1986; Pawley 2001; Ramisch 2015; *inter alia*), because they share the defining features of phraseological units: the formation by two¹ or more elements, the frequent co-occurrence, the functioning as a whole, and a certain degree of lexicalization. They differ from idioms, the undeniable example of phraseological unit according to narrow approaches (García-Page 2008), because they are less lexicalized and idiomatic (thus, more transparent), and they convey concepts.

Precisely because these combinations are less lexicalized than other phraseological units, they are especially inclined to term variation or the coexistence of different denominations. For example, the Spanish idiom *a tenor de lo establecido* is a very lexicalized structure that does not seem to admit term variation. On the contrary, MWTs such as *contaminación por ozono* show different term variants, such as *contaminación por ozono troposférico* or *contaminación fotoquímica*. This happens because MWTs allow lexical expansion by combining concepts in different ways that are feasible in a particular

1. One of the defining features of phraseological units is their formation by at least two elements. However, these can often converge in a graphic compound (e.g. *rompeolas*). This does not mean that they lose their phraseological nature (in fact, there are studies, such as Oltra Ripoll's [2018], which include monolexical units in their phraseological analysis). Nevertheless, these are not the focus of our study.

domain (Picht 1991). For this reason, they can be more prone to variation than general phraseological units or other fixed, specialized combinations.

Furthermore, knowledge transfer and market expansion highlight the need for specialized translation, which implies dealing with a vast quantity of MWTs. Rendering them into other languages is not exactly easy, in part due to their characteristics, which make them cognitively and structurally complex, as well as their unsystematic treatment in terminological resources. This results in a wide variety of translation solutions, some more adequate than others in each context, which evidence a high degree of term variation when translating MWTs.

We carried out a quantitative and qualitative analysis of the Spanish translations of a set of English MWTs and their variants, with a focus on translation variation. Our goals were: (1) to define a typology of term and translation variants of MWTs; and (2) to investigate term variation of Spanish MWTs in different contexts (translation, bilingual or multilingual lexicography, and original production in Spanish). Our results showed that MWTs exhibit an enormous degree of term variation of different characteristics, which is particularly present in translation scenarios (i.e. parallel corpora).

The rest of this article is organized as follows. Section 2 explains term variation and its presence in specialized discourse. Section 3 focuses on variation in translation contexts. Section 4 presents the materials and methods of the study, and Section 5 describes the results obtained. Particularly, in Section 5.1 we present a typology of term variation found in translation, and in Section 5.2 translation variation is compared across different resources. Finally, Section 6 summarizes the conclusions of this study and future research lines.

2. Term variation in terminology

Variation is an essential feature of all languages, even in specialized discourse. It can be conceptual, when it affects meaning, or denominative, as explored in this study, when different designations are used to name the same concept (e.g. *wildfire* and *rural fire*). Not only are monolexical units variable, but also phraseological units, in particular MWTs.

However, in specialized discourse, both phraseology and variation were obscured for a long time, namely because the General Theory of Terminology

(TGT, Wüster 1968) downplayed problematic aspects such as context, phraseology, and variation. In fact, whilst aware of the existence of these phenomena in specialized discourse, phraseology and variation were perceived as obstacles for effective expert communication. The General Theory of Terminology took thus a prescriptive approach in which a term was said to allude to only one concept, and a concept was named by only one term. The richness resulting from variation was thus artificially ignored for the sake of precision, even though variation often emerges for the same reason, since new ways of conveying meaning are constantly sought for.

Variation did not become a focus until the advent of the new theories of terminology, which formulated communicative and cognitive approaches, and acknowledged the variable nature of both terms and concepts (Cabré 1993; Temmerman 2000; Freixa 2006; León Araúz 2017). Although to a lesser extent compared to general language (Freixa 2006; Sanz Vicente 2011), specialized discourse exhibits a considerable degree of variation, which has been discovered by means of corpora (Fernández-Silva 2018).

MWTs are largely used to illustrate specialized term variation. These terms are formed by two or more elements: a head and one or several modifiers. The head often indicates the category to which the concept belongs, whereas the modifier often indicates the criterion for subdivision of the category (Bowker 1998: 487). Depending on the nature of the head (e.g. object, property, process, etc.), the modifier can specify different types of features (e.g. purpose, location, method, material, etc.) (*ibid*: 488). Although some MWTs are fixed, it is evident that most of them are very often prone to variation (e.g. *air/atmospheric/air-borne/airborne pollutant*), as has been explained in different studies (Bowker 1998; Fernández-Silva & Kerremans 2011; Daille 2017; Giacomini 2018; Gledhill & Pecman 2018; Cabezas García & Chambó in press; *inter alia*).

Initially, the use of one variant or the other might seem arbitrary. However, as shown by Rogers (1997: 219), there are certain systematic patterns of variation that need to be explained. In the same line, Bowker & Hawkins (2006) affirm that variation cannot be attributed to the carelessness of subject field experts, but rather to their desire for precision and the carefulness invested in their choice of expression. On the one hand, variation sometimes happens with a specific purpose (Bowker 1998; Fernández-Silva et al. 2009; Kerremans

2017; Freixa & Fernández-Silva 2017; Gledhill & Pecman 2018), and, on the other, it can also reveal the novelty of concepts (i.e. neologisms) (Cabré 1993; Picton 2011).

As stated by Candel Mora & Carrió Pastor (2012), discovering the causes or types of variation is important for both theoretical and practical reasons. From a theoretical perspective, it reflects the mental processes involved in the selection of one term over another. On a practical level, this information is helpful for terminologists or translators in production tasks, since they need to know when to use one variant over another and the reasons why it is the best choice in a particular context.

Traditionally, the reasons for variation have been user-based (resulting in temporal, geographic, or social variation) or usage-based (i.e. field, tenor, and channel) (Gregory & Carroll 1978). Nevertheless, additional reasons can be involved in the variation of term denominations. As Freixa (2006) states, causes for term variation can be (1) dialectal; (2) functional; (3) discursive; (4) interlinguistic; and (5) cognitive. Several of these causes can also cooccur.

Dialectal reasons are based on the geographical, chronological, or social origin of speakers. Functional reasons are related to field, tenor, and channel. Discursive reasons can lead to stylistic and rhetoric changes. For instance, Collet (2003) argues that MWTs contribute to text cohesion by molding their shape in different contexts. Freixa & Fernández-Silva (2017), and Fernández-Silva (2018) state that intratextual term variation, especially in the case of MWTs, facilitates cohesion thanks to its repetitive nature. Along these lines, Fernández-Silva (2016) points out that MWT variants that emphasize different perspectives of the same concept contribute to knowledge construction. Accordingly, Gledhill & Pecman (2018) argue that English N+of+N MWT variants (e.g. *release of plumes*) usually convey knowledge rich contexts, while N+N MWTs (e.g. *plume release*) introduce new information.

Reasons for term variation can also be interlinguistic, when two or more languages are in contact and influence each other; and cognitive, which are based on the different conceptualizations of reality or user motivations. Cognitive variants are thus the natural reflection of multidimensionality, that is, the nature of those concepts that can be organized according to different facets or dimensions (Bowker 1998). As Meyer & Mackintosh (1996) argue, MWTs are the ideal scenario for multidimensionality since different facets

can be shown in the modifiers (e.g. *photochemical smog* [cause], *summer smog* [time], *ozone smog* [agent]). Term and concept variation, far from being unrelated, are thus the consequence of the convergent influence of multidimensionality, context and dynamism in specialized domains (León Araúz 2017).

Apart from the causes of variation, it is also important to reflect on its consequences. According to Fernández-Silva et al. (2009), term variation can have no cognitive consequences when there is only a change in the form but not in the meaning (e.g. *marine product* and *sea product*). Alternatively, term variation can have cognitive effects, when there is a shift in perception along with the change in form (e.g. *sea product* and *fishing product*). In this line, some classifications based on the semantic distance of term variants have emerged, such as the ones presented in Aguado de Cea & Montiel Ponsoda (2012) and Fernández-Silva (2018). Both studies distinguish three main groups: variants with (1) minimum, (2) medium, and (3) maximum semantic distance.

In the first group, the semantic content does not change, that is, terms are conceptually equivalent. For Aguado de Cea & Montiel Ponsoda (2012), this set includes synonyms, such as graphical and orthographical variants (*localization*, *localisation*), inflectional variants (*cat*, *cats*), and morphosyntactic variants (*nitrogen fixation*, *fixation of nitrogen*). Fernández-Silva (2018) adds morphological variants (*ozone-depletion potential*, *ozone-depleting potential*) and specifies that, in MWTs, synonymy can affect just one of the constituents (*organic matter pollution*, *organic matter contamination*).

For Aguado de Cea & Montiel Ponsoda (2012), variants with a medium semantic distance are partial synonyms or terminological units that highlight different aspects of the same concept, such as stylistic or connotative variants (*man*, *bloke*), diachronic variants (*tuberculosis*, *phthisis*), dialectal variants (*gasoline*, *petrol*), pragmatic or register variants (*headache*, *cephalalgia*), and explanatory variants (*immigration law*, *law for regulating and controlling immigration*). Fernández-Silva (2018) studies medium semantic distance in MWTs where the conceptual change is reflected in the modifiers. They can be subject to reductions (*motor vehicle emission*, *vehicle emission*), additions or deletions of non-defining characteristics (*non-point pollution*, *non-point source pollution*), or the use of a different defining feature (*summer smog*,

Los Angeles smog). Again, the implication of some of these characteristics is directly linked to multidimensionality and cognitive variants.

Finally, variants that entail a maximum semantic distance are terminological units that highlight different features of the same concept which belong to different conceptualizations, or variants that refer to two conceptually related concepts (Aguado de Cea & Montiel Ponsoda 2012; Fernández-Silva 2018). In the specific case of MWTs, they involve changes in the MWT head or even in both constituents. These variants are the most cognitively complex since the semantic category of the concept is altered. This change can be just an emphasis of a specific perspective of the conceptualization (*exhaust emission, exhaust pollution*). Alternatively, a hypernymic category (*doubly fed induction generator, DFIG machine*) or, on the contrary, a hyponymic category (*high level ozone, stratospheric ozone*) can be employed. It is also possible to replace both the head and modifier with totally different categories (*oil pollution, discharged hydrocarbon*), which is the most semantically distant type of variation.

Evidently, term variation can acquire a wide range of forms, as suggested in the examples above. The following list shows the classification proposed in Faber & León Araúz (2016: 12-13)², which encompasses different proposals found in the literature and offers a thorough picture of term variation types, specifying whether semantics or communicative situations are affected:

- (A) Orthographic variants that are not influenced by geographic origin and do not alter semantics or the communicative situation, e.g. *groundwater, ground water*.
- (B) Diatopic variants:
 - (i) Orthographic variants that do not modify semantics, e.g. *fecal, faecal*.
 - (ii) Dialectal variants, which can alter semantics if cultural factors are involved, e.g. *gasoline, petrol*.
 - (iii) Culture-specific variants, which affect semantics and the communicative situation, e.g. *dry lake, sabkha*.

2. Some of the examples have been changed with variants found in this study.

- (iv) Calques, which can modify semantics and the communicative situation, e.g. *environmentally hazardous substance* > *sustancia ambientalmente peligrosa*, *sustancia peligrosa para el medio ambiente*.
- (v) Borrowings, which can alter semantics and the communicative situation and can be adapted or not, e.g. *smog* > *smog*, *esmog*
- (C) Short form variants, which have an effect on the communicative situation:
 - (i) Abbreviation, e.g. *greenhouse gas*, *GHG*.
 - (ii) Acronym, e.g. *laser*, *Light Amplification by Stimulated Emission of Radiation*.
- (D) Diaphasic variants:
 - (i) Scientific variants, which influence the communicative situation:
 - Scientific names, e.g. *Dracaena draco*, *drago*.
 - Expert neutral variants, e.g. *Ocellaris clownfish*, *Amphiprion ocellaris*.
 - Jargon, e.g. *lap-appy*, *laparoscopic appendectomy*.
 - Formulas, e.g. *carbon dioxide*, *CO₂*.
 - Symbols, e.g. €, *euro*.
 - (ii) Informal variants, which affect the communicative situation and can also influence semantics:
 - Lay user variants, e.g. *Dragon tree*, *drago*.
 - Colloquial variants, e.g. *motor vehicle pollution*, *car pollution*.
 - Generic variants, e.g. *pollution*, *contamination*.
 - (iii) Domain-specific variants, which can modify semantics or the communicative situation if the domains have different term preferences, e.g. *mud* and *sludge* represent the same concept except that the former is used in Geology and the latter in Water Treatment.
- (E) Cognitive variants, which are usually MWTs:
 - (i) Dimensional variants, which alter semantics because they express different dimensions of the same concept, e.g. *esmog fotoquímico* [*photochemical smog*], *niebla tóxica estival* [*summer smog*].
 - (ii) Intentional variants, which have an effect on the semantics and/or the communicative situation, since they are used depending on the receiver's reaction sought, e.g. *climate change*, *climate emergency*.
- (F) Metonymic variants, which affect semantics by alluding to a part or material of the concept, e.g. *accidental water pollution*, *accidental marine pollution*.

- (G) Diachronic variants, e.g. *anhídrido carbónico* [*carbonic anhydride*], *dióxido de carbono* [*carbon dioxide*].
- (H) Non-recommended variants, e.g. since *mental retardation* now has negative connotations, it has been substituted by *intellectual disability*.
- (I) Morphosyntactic variants, which do not usually influence semantics but depend on the communicative situation, as well as on term preferences and collocations, e.g. *contaminación acústica* [*acoustic pollution*], *contaminación de ruido* [*noise pollution*].

The nature and scope of variants are very diverse and can have different consequences in communication. Nevertheless, terms can activate more than one variant type, which might make term choice more difficult. For example, *H₂O* and/or *water* can be domain-based variants since the first one is more frequently used in Chemistry and Water Treatment domains than in Oceanography, for example. However, their use also depends on the communicative situation (i.e. formal or informal). On the contrary, the same type of variant can be expressed by more than one term. Diaphasic variants, in particular, form a continuum from more formal to informal (e.g. *thermal low pressure system*, *thermal low*, *thermal trough*, and *heat low*) (Faber & León Araúz 2016).

3. Term variation and translation

Term variation has also been explored in interlinguistic contexts. The notion of ‘equivalence’ becomes thus crucial since terminologists and translators often follow different criteria. While terminologists usually understand equivalence at the term level, because their goal is the inclusion of terms in terminographic resources, translators look for correspondence at the sentence or text level. That is why they search for the functional equivalence of their translations (Reiss & Vermeer 1984; Nord 1997) instead of a direct term correspondence, which would be the case for terminologists. In this line, Gerzymisch-Arbogast (2008) highlights the fact that translation is a text-based rather than a corpus-based activity, and Kerremans & Temmerman (2016) state that translators are not always guided by the principle (often adopted for structuring multilingual terminology databases) that a term in the source language should be rendered as a direct (literal) equivalent in the

target text. Therefore, terminological equivalence does not always correspond to translation equivalence (Kerremans & Temmerman 2016: 59).

Equivalence is thus a broader concept for translators than for terminologists and allows for a wide spectrum of translation mechanisms. Consequently, equivalents at the sentence or text level can be those reproducing the same function or effect than the source text (Reiss & Vermeer 1984; Nord 1997), rather than just those conveying the same concept. For this reason, the use of hypernyms or other variants reflecting different conceptualizations (e.g. as a result of multidimensionality) can be justified in a translation equivalence context.

Along these lines, understanding and finding equivalence for phraseological units can be problematic for translators. In the case of MWTs, comprehension problems may be related to the identification of the term as a single unit of understanding, its internal dependencies, or the semantics of its formants. In turn, production-related problems include the order in which its formants should be translated, the prepositions that should accompany them, or the distinction among different variants should they be found in any resource. The well-known notions of *translation problem*, *error*, *strategy*, and *technique* become thus central. A phraseological unit poses a *translation problem* when it “puts up resistance to being translated” (Oltra Ripoll 2018: 102). This can lead to *translation errors* or the inadequate resolution of a problem (Hurtado 2001: 279). Therefore, the analysis of term variation in translation, as the one carried out in this study, must consider the *strategies* or mechanisms employed to solve translation problems, and, specifically, the *techniques* or procedures to find the solution to a particular translation problem (e.g. omission, addition, modulation, paraphrase, etc.). As for the identification of translation equivalences, parallel corpora have traditionally been used (Déjean & Gaussier 2002; Daille & Morin 2005). These are collections of original texts and their aligned translations, which facilitate equivalent identification. However, they are scarce, particularly in specialized languages and in language pairs in which English is not involved, and the influence of the source language is obvious. These corpora are useful for error analysis or the identification of the different translation options, but complicate the study of idiomatic language uses. In consequence, comparable corpora are increasingly present, which consist in collections of texts of

the same type and domain, originally written in every language. These are more easily obtained and allow for the study of real language uses, although equivalent identification is evidently more complicated.

One of the studies devoted to term variation in translation is Fernández-Silva et al.'s (2009), who investigate, among other aspects, the role of the cultural system by means of term variants in French and Galician. Kerremans (2010, 2016) also studies term variation in specialized translation, with a focus on the reflection of the English source language variants in the target languages (Dutch and French). Along these lines, Fernández-Silva & Kerremans (2011) study cognitive term variants, and affirm that source language variants in Galician are reflected in the English target texts. Miyata & Kageura (2016) argue that translated texts (from Japanese into English) show a higher density of term variants as a result of the different translations possibilities. This finding was also confirmed in a previous study by Sanz Vicente (2011), which focused on the translation of English MWTs into Spanish and observed the higher coexistence of term variants in the target language. Accordingly, Jiménez-Crespo & Tercedor-Sánchez (2017) explore term variation in translated (English > Spanish) and non-translated texts (Spanish), paying particular attention to register, terminologization, explicitation, and term variation in translated documents. Consciously or unconsciously, these studies often use MWT examples to illustrate term variation in translations.

As some of these research studies suggest, MWT variation can influence their translations, where a higher degree of variation is usually found. Every language has different degrees of variation depending on their own linguistic characteristics. For instance, English is prone to graphical variation due to the use of hyphens, among other factors, whereas Spanish is more prone to morphosyntactic variation due to its compounding rules. For this reason, ascertaining term variants in the source language is relevant with a view to facilitating equivalent identification. In this sense, Rogers (1997) underlines that translators are frequently obliged to make decisions not only regarding synonymy within the source text and target text, but also regarding the cross-linguistic relations between these synonyms.

The representation of the different translation possibilities in terminographic resources becomes thus central in descriptive settings (Kerremans 2010), contrary to what has traditionally been done. These resources often

describe just a small portion of variants (if at all), which are represented in an unsystematic way. As for MWTs, those formed by more than two constituents are rarely included (Giacomini 2018). Users are often confronted with a lack of information on how term variation arises and which selection criteria to choose. Evidently, they need to know when to use each variant as well as its conceptual and communicative implications, since this will affect the receiver's interpretation of the message. Otherwise, translators can actually over-standardize, thus creating consistency in places where the use of variants was deliberate and well-reasoned (Bowker & Hawkins 2006: 80). Translators and terminologists must thus learn to recognize the patterns that lie beneath variation so that they do not inappropriately standardize terminology (*ibid*: 101). However, while refusing the wüsterian principle of univocity, we believe that the proliferation of term variants in target texts is often puzzling and overwhelming, partly due to an unsystematic treatment of MWTs, since translators do not always consider them as a single unit of understanding. Consequently, besides describing different types of variants, which is undoubtedly important, the added value of a linguistic resource lies in the enhancement of those data with additional information, such as semantic, pragmatic, and usage aspects (Faber & León Araúz 2016; Giacomini 2018), which improves a sound use of variation in texts.

4. Materials and methods

According to Kerremans & Temmerman (2016: 45), corpora allow us to study the textual and linguistic features of translations, taking into account different contextual parameters that have an impact on translation choices and, ultimately, on the translation product. For this reason, different corpora were used for MWT extraction and term variant identification.

Since one of our goals was to explore MWT variation patterns in translation situations, the OPUS2 English corpus (Tiedemann 2012) was first used to extract a set of MWTs, which would then be compared with their equivalents in the OPUS2 Spanish parallel corpus (Tiedemann 2012). The OPUS2 English corpus is an open source parallel corpus that can be accessed at Sketch Engine (<https://www.sketchengine.eu/>) (Kilgarriff et al. 2014) and encompasses 40 languages. It is organized in subcorpora, such as the European Central

Bank (ECB), the European Parliament Proceedings (EUROPARL), and the Translated UN documents (MultiUN). Aware of the scarcity of specialized parallel corpora, we selected this parallel corpus, which includes both general and specialized corpora, and decided to focus on specialized terms of general interest. This is the case of POLLUTION, a specialized concept that is present in everyday communication due to the increasing climate awareness.

Therefore, starting from the term *pollution*, a conceptual analysis was carried out in the corpus, which allowed us to identify pollution-related concepts, such as *emission*, *ozone*, *gas*, *pollutant*, *substance*, *contamination*, *smog*, *air*, etc. These terms were then used as MWT heads in CQL (Corpus Query Language) queries in Sketch Engine, which allow to search for specific morphosyntactic patterns, such as MWTs premodified by different elements (Table 1):

[tag="N.* JJ.* RB.* VFN.* VVG.*"]{1,}[lemma="pollution"][tag!="N.* JJ.*"]

Table 1: CQL query to extract MWTs whose head is *pollution*.

The CQL expression in Table 1 elicits premodified MWTs (the most frequent structure of these terms), such as *heavy metal pollution* or *indoor air pollution*. It searches for the lemma *pollution* ([lemma="pollution"]) (or *emission*, *ozone*, *gas*, *pollutant*, etc., in the following queries) preceded by nouns, adjectives, adverbs, past participles, or present participles ([tag="N.*|JJ.*|RB.*|VFN.*|VVG.*"]) appearing one or more times ({1,}). On the right of the head *pollution*, a restriction is included in order to exclude nouns or adjectives ([tag!="N.*|JJ.*"]), which could indicate that *pollution* is not the MWT head but takes part in a longer term.

This query was repeated for the other possible heads (*emission*, *ozone*, *gas*, etc.), and the 234 most frequent MWTs were selected. Several of these MWTs were term variants, and were thus grouped by concept. Finally, we obtained a set of 98 pollution-related concepts.

To identify the Spanish variants of the English MWTs, different resources were used. Firstly, with a view to investigating MWT variants in translation contexts, we used three parallel corpora, which allowed an easy identification of variants thanks to their alignment. Although, ideally, parallel corpora of specialized texts on the environment in English and Spanish would be the

best option, due to their scarcity we used parallel corpora encompassing both general and specialized discourse in a wide variety of domains.

These parallel corpora were the OPUS2 English-Spanish corpus (Tiedemann 2012), which was presented above in its monolingual version in English; the EurLex English-Spanish corpus (Vaisa et al. 2016), a multilingual corpus in all the official languages of the European Union that includes texts in the EUR-Lex database and is available in Sketch Engine; and Linguee (<https://www.linguee.es/>), an online corpus of aligned translations in different languages, such as English and Spanish, which includes general as well as specialized texts. Even though Linguee does not allow specific CQL queries and shows just a summary of the possible translations, it complemented the alignment mismatches that were often found in OPUS2 and EurLex. As a consequence, it allowed us to collect more different translations.

Secondly, in order to compare term variants found in translation contexts with those present in bilingual or multilingual lexicographic scenarios, Spanish equivalents of English MWTs were also looked up in two terminological databases: TERMIUM Plus and IATE. TERMIUM Plus (<https://www.btb.termiumplus.gc.ca/>) is a terminological database developed by the Government of Canada, which describes millions of concepts from specialized domains in English, French, and Spanish. Additionally, IATE (<https://iate.europa.eu/>) is the EU's terminology database and includes terms from a wide range of specialized domains in the official languages of the European Union. The entries consulted in these resources also allowed us to expand the collection of English source terms, since many of their entries contain synonyms. Therefore, a new set of terms was researched in the parallel corpora in order to expand the collection of Spanish translation variants. The final set of terms (a total of 277) ranged from two-word terms (e.g. *oil pollution*) to six-word terms (e.g. *aggregate anthropogenic carbon dioxide equivalent emissions*).

Finally, with a view to analyzing term variants in a context of original production in Spanish, we used a Spanish comparable corpus of specialized texts on the environment. The corpus was compiled by the LexiCon research group of the University of Granada while building EcoLexicon (<https://ecolexicon.ugr.es>, León Araúz, Reimerink & Faber 2019), a terminological knowledge base on the environment, and consists of approximately 10 million words. Since the size of the corpus cannot compete with the size of parallel ones,

this was compensated by the use of Google Scholar (<https://scholar.google.com/>) as a second comparable corpora.

Evidently, Google Scholar is not a comparable corpus strictly speaking. Moreover, it does not allow for flexible searches, such as lemmatized or CQL queries. Searches were time-consuming, since many variants needed to be looked up several times taking into account the inflectional rules of Spanish (i.e. number and gender). However, this was useful to obtain more results and measure the frequency of all variants found in the previous resources. We decided to only retain those terms that occurred a minimum of 10 times. At least in terms of frequency, this corpus should be more reliable due to its size. We agree with Bowker & Hawkins (2006), who state that the Web represents a huge and easily accessible body of linguistic data. Grefenstette (2002: 207) supports this claim by arguing that the size of the Web compensates for its “dirtiness”: “the correct form is always orders of magnitude more frequent than the erroneous form (...). The Web is dirty but the signal (correct forms and correct usage) is so strong noise can easily be ignored”. In this case, restricting the queries to Google Scholar ensures specialized language since Google Scholar is limited to research works.

It should be noted that the sequence of resources presented (i.e. parallel corpora, terminographic resources, and comparable corpora) was not random. We started with resources that provide direct access to interlinguistic variants (i.e. parallel corpora and terminographic databases), and then the last step was querying the comparable corpora, which required specific strategies to findequivalences the search process of which was more complex.

Our equivalence identification strategy in the comparable corpora involved the following queries. The terms found in parallel corpora (e.g. *contaminante atmosférico*, *contaminante aéreo* [*atmospheric pollutant*]) were literally searched for to confirm their presence in the comparable corpora. It should be noted that some of the variants obtained from parallel corpora were not queried in the comparable corpora, since they could bias the results. This was the case of hypernyms used as term variants. For instance, although *contaminación acústica* [*acoustic pollution*] and *ruido* [*noise*] are used as variants in the parallel corpora, searching for *ruido* [*noise*] in the comparable corpora would elicit additional meanings that would spoil the results. The

same was true of polysemic acronyms in Google Scholar (e.g. *GEI* [*GHG*]) and *ad hoc* variants that could not convey exactly the same meaning (e.g. *daños medioambientales* [*environmental damages*] as a term variant of *contaminación medioambiental* [*environmental pollution*]).

Additionally, the MWT heads and modifiers found in the parallel corpora were used with a span in between, as in the CQL query shown in Table 2. The span was set in 5 elements in order to allow for different possibilities without being too broad, since concepts in an MWT don't usually have a wider distance, as found in previous studies. However, in larger MWTs, such as those including participles or relative sentences, a higher span was used. Since CQL queries are not possible in Google Scholar, the * wildcard was employed to indicate the span. This allowed us to obtain new possibilities of extended MWTs, such as *contaminantes liberados a la atmósfera* or *contaminantes vertidos a la atmósfera* [*pollutants released into the air*].

[lemma="contaminante"][]{}0,5[lemma="aire atmósfera atmosférico aéreo"]

Table 2: CQL query for eliciting MWTs of the type *contaminante * aire/atmósfera/atmosférico/aéreo*.

Other queries in the specialized corpus of EcoLexicon included searching for the head and sorting the results by right so as to distinguish the different modifiers accompanying the head easily. The same strategy was applied to modifiers, sorting by left to discover the different possible heads. This type of queries is more time consuming, that is why they were only used in the EcoLexicon corpus, given the corpus size and the restriction possibilities, but not in Google Scholar. Figure 1 shows some of the modifiers that follow *sustancia* [*substance*] in the EcoLexicon Spanish corpus: *sustancia contaminante* [*polluting substance*], *sustancia explosiva* [*explosive substance*], *sustancia nociva* [*harmful substance*], *sustancia peligrosa* [*dangerous substance*], *sustancia que agota el ozono* [*ozone-depleting substance*], *sustancia que agota la capa de ozono* [*ozone-depleting substance*], its abbreviation *SAO* [*ODS*], *sustancia química* [*chemical substance*], and *sustancia tóxica* [*toxic substance*].

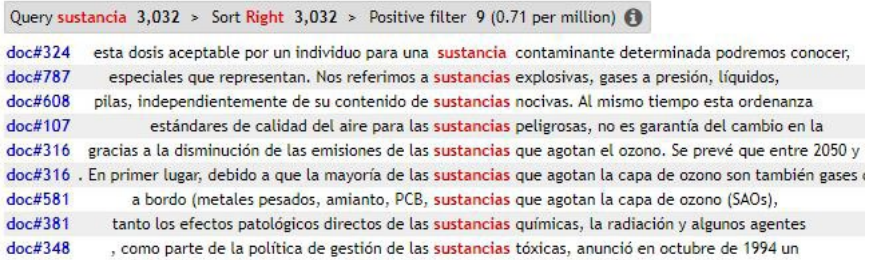


Figure 1: Sample of the modifiers of *sustancia* in the query *sustancia* + sort by right.

Evidently, concordance lines can also reveal knowledge-rich contexts that provide new term variants, as exemplified in the abbreviation SAO (*sustancia que agota la capa de ozono*) [ODS, *ozone-depleting substance*]. The following example also shows the usefulness of knowledge-rich contexts in variant identification, based on explicit knowledge patterns (*o*, the disjunctive conjunction): “Hemos de distinguir entre *contaminación natural* o *endógena* y *contaminación antrópica* o *exógena*”. As can be seen, *contaminación natural* [natural pollution] and *contaminación endógena* [endogenous pollution] are term variants and antonyms of *contaminación antrópica* [anthropogenic pollution] and *contaminación exógena* [exogenous pollution], another variant that had not been elicited in parallel corpora or terminological resources. Consequently, these strategies revealed new variants and complemented the queries in the parallel corpora and the terminological databases. In total, 1428 Spanish term variants were retrieved. Table 3 shows the English terms naming the concept OZONE POLLUTION and the Spanish variants found in the different resources.

OZONE POLLUTION	
English term variants	
ozone pollution tropospheric ozone pollution lower atmospheric ozone pollution surface ozone pollution	pollution by ozone photochemical pollution pollution caused by tropospheric ozone ground-level ozone pollution
Spanish translation variants	
OPUS2 Spanish	EurLex Spanish
contaminación por ozono contaminación fotoquímica ozono troposférico contaminante	contaminación por ozono contaminación por el ozono contaminación fotoquímica contaminación del ozono de la troposfera contaminación por ozono troposférico contaminación provocada por el ozono contaminación troposférica por ozono contaminación por ozono en la baja atmósfera
Linguee	Google Scholar
contaminación por ozono contaminación del ozono contaminación por ozono troposférico contaminación fotoquímica contaminación de ozono a nivel del suelo contaminación por ozono de la baja atmósfera contaminación del ozono a nivel del piso contaminación por ozono en la baja atmósfera	contaminación de ozono a nivel del suelo contaminación del ozono contaminación fotoquímica contaminación por el ozono contaminación por ozono contaminación por ozono en la baja atmósfera contaminación por ozono troposférico
EcoLexicon	IATE and Termium
contaminación por ozono contaminación fotoquímica	contaminación por ozono

Table 3: English variants for OZONE POLLUTION and Spanish variants found in the different resources.

5. Results and discussion

5.1. Characterizing translation variation

The analyzed concepts show a high degree of variation in Spanish, ranging from 2 term variants (e.g. *sectorial emission*) to more than 46 (e.g. *aircraft noise*). Few of them appear to be highly lexicalized, but among those who are, what seems to be a trend is the fact that having an acronym makes them more stable, such as *compuesto orgánico volátil (COV)* [*volatile organic compound, VOC*]. On the contrary, codifying a causal relation was found to make MWTs more prone to variation, since they usually present multiple periphrastic structures making the semantics of the concept explicit. One of these examples is *anthropogenic emissions*, which can be rendered as *emisiones antropogénicas*, but also as *emisiones procedentes de fuentes humanas*, *emisiones generadas por el hombre*, *emisiones causadas por el hombre*, *emisiones producidas por los humanos* or *emisiones provocadas por el hombre*, among other variants.

It is also worth noting that the more specific the concepts are, the more stable their designations were found to be, even if their hypernyms show many more variants. For instance, it is striking to see that while *contaminación atmosférica transfronteriza* [*transboundary air pollution*] can be rephrased as *contaminación transfronteriza del aire*, its hyponym *contaminación atmosférica transfronteriza a larga distancia* [*long-range transboundary air pollution*] does not follow the same pattern to be rephrased as *contaminación transfronteriza del aire a larga distancia* or *contaminación transfronteriza a larga distancia del aire* (not found in any of the resources). Thus, depending on the bracketing structure (i.e. internal dependencies in the MWT), the distance between the different elements of the MWT seems to present a limited span.

In what follows variation found in the different resources has been parametrized in a typology specifically conceived to characterize translation correspondences. Although many of the variants described in Faber & León Araúz (2016) occurred (e.g. orthographic variants, dialectal variants, calques, adapted and non-adapted borrowings, short form variants, etc.), most of them were morphosyntactic and cognitive, which calls for an expansion of this variation categories in translation scenarios. Partially based on traditional

classifications of translation strategies³ (Hurtado 2001; Newmark 1981, *inter alia*) and Daille's typology for French term variants (Daille 2003), our typology is divided in five main groups (omissions, structural shifts, transpositions, expansions and inaccuracies), each of them including other strategies, such as hyponymy, synonymy, paraphrases and modulation.

1. Omissions

a. Omission of articles

- e.g. *total de las emisiones agregadas de GEI*, *total de emisiones agregadas de GEI*
- e.g. *contaminación de las aguas subterráneas*, *contaminación de aguas subterráneas*

b. Omission of formants

i. Omission of modifiers

1. Omission implying activation of hypernym

- e.g. *contaminación atmosférica transfronteriza*, *contaminación atmosférica*
- e.g. *contaminación marina por petróleo*, *contaminación marina*, *contaminación por petróleo*

2. Omission not implying activation of hypernym

- e.g. *contaminación de fuentes difusas*, *contaminación difusa*
- e.g. *emisión de gases de efecto invernadero*, *emisión de gases invernadero*

ii. Omission of the head (sometimes implying transposition)

- e.g. *emisión de gases de efecto invernadero*, *gases de efecto invernadero*
- e.g. *sustancia tóxica*, *tóxico*

2. Structural shifts

a. Shift of prepositions

i. Prepositions expressing cause

3. The scope of traditional classifications is usually wider and not necessarily term-oriented. Furthermore, strategies are always described with regards to a source text, whereas in this study comparisons are established between different term variants within the same language.

- e.g. *contaminación por metales pesados, contaminación con metales pesados, contaminación de metales pesados*
- ii. Prepositions expressing location
 - e.g. *ozono en el nivel del suelo, ozono al nivel del suelo, ozono del nivel del suelo*
- iii. Prepositions expressing affected entity
 - e.g. *contaminación del agua, contaminación en el agua*
- b. Permutations
 - i. Permutation of modifiers
 - 1. Permutation with no meaning change
 - e.g. *emisiones antropogénicas de gases de efecto invernadero, emisiones de gases de efecto invernadero antropogénicas*
 - e.g. *contaminación marina accidental, contaminación accidental marina*
 - 2. Permutation with meaning change
 - e.g. *reducción certificada de emisiones, reducción de emisiones certificadas* (See Inaccuracies)
 - ii. Permutation of head and modifiers
 - e.g. *contaminación de fuentes no localizadas, fuentes de contaminación no localizadas*
 - e.g. *emisiones totales de GEI, total de emisiones de GEI*
 - *contaminante gaseoso, gas contaminante*
 - *contaminante atmosférico, aire contaminado*
- c. Shift of noun within modifier
 - i. Shift by synonym
 - e.g. *emisión de los tubos de escape de los automóviles, emisión de los tubos de escape de los coches*
 - ii. Shift by hypernym
 - e.g. *contaminación por metales pesados, contaminación por materiales pesados*
 - iii. Shift by metonym
 - e.g. *sustancia destructora de la capa de ozono, sustancia destructora del ozono*
 - iv. Shift by modulation (conceptual dimension)
 - e.g. *smog de verano* (time), *smog de Los Ángeles* (location)

- d. Shift of noun within head
- i. Shift by synonym
 - e.g. *aumento del nivel del mar*, *subida del nivel del mar*
 - ii. Shift by modulation
 1. Shift by near-synonym
 - e.g. *destrucción de la capa de ozono*, *empobrecimiento de la capa de ozono*
 2. Shift by metonym
 - e.g. *contaminación de las aguas subterráneas*, *contaminantes de las aguas subterráneas*
 3. Shift by conceptual dimension
 - e.g. *contaminación acústica*, *molestias acústicas*
- e. Shift of both modifiers and head (modulation)
- e.g. *contaminación causada por el ozono*, *polución urbana*
 - e.g. *smog fotoquímico*, *niebla tóxica estival*
 - e.g. *contaminación marítima por hidrocarburos*, *marea negra*
- f. Shift of modifying adjectives
- e.g. *contaminación no localizada*, *contaminación dispersa*
 - e.g. *emisiones antropógenas de gases de efecto invernadero*, *emisiones antropogénicas de gases de efecto invernadero*
 - *contaminación marina*, *contaminación marítima*
3. Transposition
- a. Transposition of adjective by periphrasis
 - e.g. *cambio climático antropógeno*, *cambio climático producido por el hombre*
 - e.g. *contaminación lumínica*, *contaminación producida por la iluminación*
 - e.g. *ozono troposférico*, *ozono en la baja atmósfera*
 - b. Transposition of adjective by “of + noun” (transposition)
 - e.g. *contaminante atmosférico*, *contaminante del aire*
 - e.g. *ozono superficial*, *ozono de superficie*
 - c. Transposition by inflection of one of the formants
 - e.g. *contaminación atmosférica transfronteriza a larga distancia*, *contaminación atmosférica transfronteriza a largas distancias*
 - e.g. *contaminación del agua*, *contaminación de las aguas*

4. Expansions

a. Making conceptual information explicit

i. Agent

- e.g. *contaminación de origen terrestre, contaminación por efluentes provenientes de tierra*
- e.g. *contaminación del ozono, contaminación provocada por el ozono*

ii. Patient

- e.g. *contaminación de origen terrestre, contaminación marina de origen terrestre*

b. Adding a second variant

- e.g. *ozono troposférico, ozono troposférico (al nivel del suelo)*
- e.g. *contaminación de fuente no puntual, contaminación de fuente no puntual (difusa)*
- e.g. *contaminación a larga distancia, contaminación a larga distancia o transfronteriza*

5. Inaccuracies (some of them can be considered translation errors [a-d] and some others as cognitive or intentional variants [e]).

a. Inaccuracies related to the semantics of one of the formants

- e.g. *contaminación atmosférica transfronteriza a larga distancia, contaminación atmosférica transfronteriza prolongada*
- e.g. *contaminación por nitratos, contaminación por nitrógeno*
- e.g. *contaminación por fuente no localizada, contaminación que no viene de fuente*

b. Inaccuracies related to the semantic relation between the formants

i. Confusion origin-patient

- e.g. *contaminación del terreno, contaminación de origen terrestre*

ii. Confusion patient-agent

- e.g. *contaminación en materias orgánicas, contaminación por materia orgánica*

c. Inaccuracies related to bracketing (i.e. internal dependence analysis)

- e.g. *contaminación del aire urbano, contaminación urbana del aire*

- e.g. *gases de efecto invernadero antropogénico, gases antropogénicos con efecto invernadero*
- e.g. *gases de efecto invernadero causados por la actividad humana, gases de efecto invernadero causado por la actividad humana*
- e.g. *calidad exterior del aire, calidad del aire exterior*
- d. Inaccuracies due to style and redundancy
 - e.g. *contaminación del aire al aire libre*
 - e.g. *emisión fugitiva fugaz*
 - e.g. *marea negra por petróleo*
 - e.g. *contaminación por contaminantes orgánicos*
- e. Inaccuracies due to ad hoc translations (modulation)
 - e.g. *desastre ecológico marino, contaminación marina*
 - e.g. *paisajes degradados, contaminación visual*
 - e.g. *exceso de ruido, contaminación acústica*
 - e.g. *envenenamiento industrial, contaminación industrial*

As can be inferred from the classification above there are several structural shifts that also convey a difference in meaning (i.e. inaccuracies, modulations and certain omissions). Cognitive variants occur on the changes mostly affecting nouns, whether in the modifier or in the head, but especially in the latter. As for term opacity, different structures convey more transparent meanings thanks to explicitation. For instance, the preposition *por* is more specific than *de* for making causal relations explicit (e.g. *contaminación por petróleo, contaminación de petróleo*), since *de* is naturally more ambiguous in Spanish.

The most frequent types of variation found in our study were: (1) the omission of articles; (2) the changes in modifiers (reflecting structural or semantic modifications); and (3) the introduction of periphrastic structures, often through participles such as *causado, producido, provocado, inducido, originado, ocasionado*, etc. (e.g. *cambio climático producido/provocado por el hombre*), and relative clauses followed by verbs such as *provocar, causar, contribuir, originar, and producir* (e.g. *emisión de gases que provocan el efecto invernadero*). Inaccuracies, while not that frequent, are worth mentioning,

since they are especially due to the incorrect interpretation of the semantics and internal dependencies of MWTs.

Quite often, several of these types coincide within the same set of variants conveying the same concept, as in *contaminación transfronteriza a gran distancia* (where only structural changes and synonyms apply) (Table 4) or *smog fotoquímico* (where cognitive variants stand out) (Table 5):

contaminación atmosférica transfronteriza a larga distancia (adjective changes)
contaminación atmosférica transfronteriza de larga distancia (preposition changes)
contaminación atmosférica transfronteriza de largo alcance (noun changes)
contaminación atmosférica transfronteriza a largas distancias (number changes)
contaminación atmosférica internacional de largo alcance (adjective changes)
contaminación atmosférica transfronteriza a gran distancia (adjective changes)

Table 4: Spanish variants for *contaminación transfronteriza a gran distancia*.

<i>bruma de verano</i>	<i>niebla fotoquímica</i>
<i>bruma fotoquímica</i>	<i>niebla sucia</i>
<i>bruma industrial de verano</i>	<i>niebla tóxica estival</i>
<i>bruma industrial estival</i>	<i>niebla urbana de verano</i>
<i>contaminación causada por el ozono</i>	<i>nubes de ozono</i>
<i>contaminación de verano</i>	<i>ozono en el aire ambiente</i>
<i>contaminación estival</i>	<i>ozono fotoquímico</i>
<i>contaminación estival por ozono</i>	<i>ozono troposférico</i>
<i>contaminación fotoquímica</i>	<i>polución urbana</i>
<i>contaminación por ozono</i>	<i>smog de Los Ángeles</i>
<i>contaminación veraniega</i>	<i>smog de verano</i>
<i>esmog de ozono</i>	<i>smog estival</i>
<i>esmog de verano</i>	<i>smog fotoquímico</i>
<i>esmog fotoquímico</i>	<i>smog fotoquímico oxidante</i>
<i>neblina estival</i>	<i>smog relacionado con el ozono</i>
<i>neblumo fotoquímico</i>	<i>smog tipo Los Ángeles</i>

Table 5: Spanish variants for *smog fotoquímico*.

Whereas, in Table 4, the variants for *contaminación transfronteriza a gran distancia* usually imply structural changes and synonyms (e.g. *gran* > *larga*; *distancia* > *alcance*; *a larga distancia* > *de larga distancia*; *a larga distancia* > *a largas distancias*, etc.), term variants in Table 5 are quite different. In this

case, cognitive variants are noticeable, which result from the different conceptualizations and modulation of the same concept. For instance, different dimensions are highlighted in the modifiers, whether they are adjectives or nouns. Some of them point to the time when this type of pollution usually occurs (*bruma de verano*, *contaminación de verano*, *contaminación estival*, *esmog de verano* [summer smog]). Others show the city where it was first described (*smog de Los Ángeles*, *smog tipo Los Ángeles* [Los Angeles smog]). Variants can also introduce the agent producing this pollution (*contaminación por ozono*, *esmog de ozono*, *nubes de ozono*, *ozono en el aire ambiente*), or even the process that causes it, the chemical reaction of ozone and light (*bruma fotoquímica*, *contaminación fotoquímica*, *esmog fotoquímico*, *neblumo fotoquímico*). Heads also show cognitive variation, which can also entail changes in the general conceptual category, as in *esmog de ozono* and *ozono fotoquímico* or less complex and less accurate conceptualizations, such as *bruma de verano*, *niebla sucia*, and *polución de verano*. Additional structural aspects are also observed in Table 5, such as the use of the adapted borrowing *esmog* and its non-adapted variant *smog*. SMOG FOTOQUÍMICO is thus a clear example of the richness of term variation.

Therefore, the representation of term variation in terminological resources should be adapted to the different types and consequences of variants. A possible way of covering structural variants (e.g. morphological or morpho-syntactic variants) would be by grouping and ranking them by frequency. However, frequency alone cannot be used as the sole criterion of classification, since other motivations, such as stylistic, cognitive or functional aspects, can be involved in term selection and should also be somehow represented. Consequently, for cognitive variants, frequency as well as the semantic content emphasized by the term should be made explicit. In turn, morphological and morphosyntactic variants should highlight their differences contrastively (as in Table 4). Besides, when necessary, specific markers on term formation devices should also be included (e.g. calque, borrowing, acronym, etc.). However, these issues are not the scope of this study and will be investigated in future research.

5.2. Comparing translation variation across different resources

In line with the findings reported by Sanz Vicente (2011) and Miyata & Kageura (2016), parallel corpora show more translation variants than comparable corpora. Based on our results, 68 out of the 98 concepts were designated with more different terms in parallel corpora, whereas 30 showed more term variants in comparable corpora. In those cases where the comparable corpora provided more variants, there was a slight difference in the number of possible term choices. In contrast, when the parallel corpora showed more variants, the difference amounted to up to 24 additional translations for the same concept. The proliferation of variants in parallel corpora could confirm the hypothesis that translators do not always consider MWTs as a whole unit of understanding but rather as different chunks of independent strings.

In parallel corpora, the use of Spanish variants often depends on the choice made in the source language. For example, *anthropogenic climate change* is usually translated as *cambio climático antropógeno* or *cambio climático antropogénico*, whereas when the same concept is referred to as *human-induced climate change*, most frequent Spanish variants include *cambio climático inducido por el hombre* or *cambio climático provocado por las actividades humanas*. A similar observation can be made regarding synthetic or longer variants in source terms. *Emissions of greenhouse gases* gives rise to more periphrastic choices in Spanish than *greenhouse gas emissions* (e.g. *producción de gases que contribuyen a la creación del efecto invernadero*). However, not only does it occur with regards to synthetic variants, but also in relation to register and transparency. When the source text contains *atmospheric pollution*, it is usually translated as *contaminación atmosférica*, whereas when *air pollution* is chosen, both *contaminación atmosférica* and *contaminación del aire* emerge in Spanish texts. *Contaminación de la atmósfera*, a much less frequent variant in Spanish, only arises when *atmospheric pollution* does in English texts, which would be an example of how parallel corpora do not always show most idiomatic forms of specialized terms. On the contrary, there are also cases where the use of English variants does not give rise to clear-cut cross-linguistic equivalent variants, as in *ozono troposférico*, which is not only used as the literal equivalent of *tropospheric ozone* but also as that of *ground-level ozone* in approximately half the cases where *ground-level ozone* is used in English texts.

Looking at the quantitative results⁴ in more detail, it is worth noting that among the seven resources (Figure 2), OPUS and Google Scholar were those where more translation variants were found (597 and 758 respectively), followed by Linguee (480) (even though not all results are displayed), EurLex (444) and EcoLexicon (450), and finally Termium Plus (105) and IATE (118), which is not surprising since terminological resources usually reflect up to two or three variants.

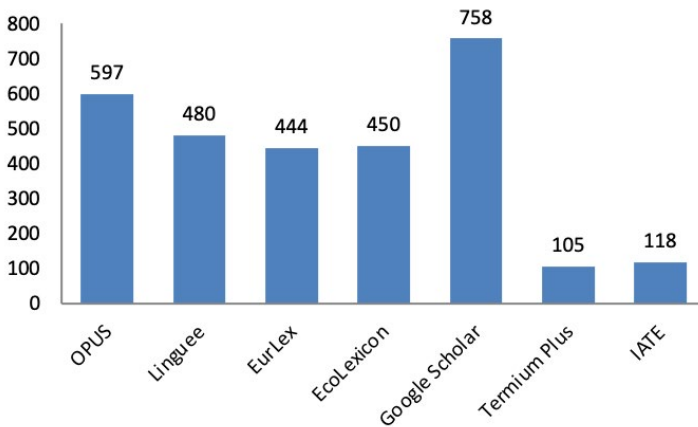


Figure 2: Number of translation variants found in each resource.

Evidently, this does not mean that OPUS and Google Scholar are the best resources to find more reliable results, since reliability would rather depend on the amount of resources containing the same term choices. In this sense, not all translation variants were found in all seven resources; only 25 Spanish terms (pertaining to 25 concepts) were confirmed by all. A staggering 73% of the terms were found only in one of the resources (434 terms) or two of the resources (598 terms) (Figure 3). This means that only 25 concepts have a clearly preferred variant and that most translation variants are not highly lexicalized.

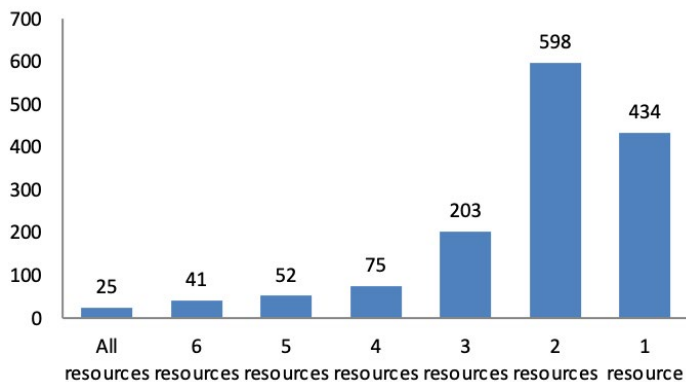


Figure 3: Number of translation variants found in all, several or one of the resources.

The results of analyzing the amount of terms that were found only in one of the resources confirm that parallel corpora contain a higher amount of infrequent or *ad hoc* translation variants (Figure 4).

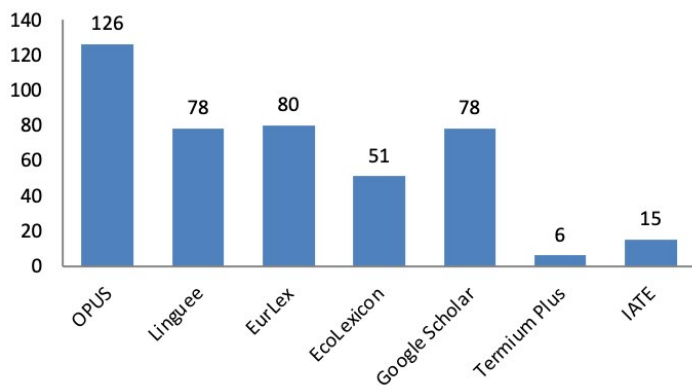


Figure 4: Number of translation variants found only once in each resource.

What is striking is the fact that, although few in number, terminological resources also present very rare translation variants, such as *bióxido de carbono equivalente* (Termium Plus) or *contaminante corpuscular* (IATE). Furthermore, frequency is not always taken into account in these term bases. For example,

Termium Plus only includes the variant *cambio climático mundial* as an equivalent of *global climate change*, although *cambio climático global* exhibits a much higher frequency since the last forty decades, as shown in the Google Ngram Viewer (Figure 5).

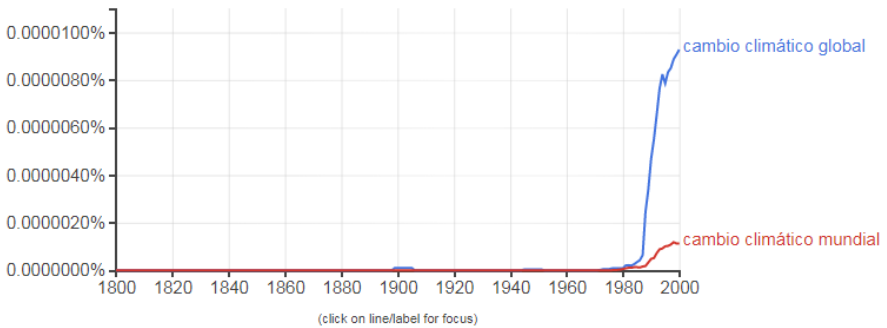


Figure 5: Frequency of *cambio climático global* and *cambio climático mundial* in the Google Ngram Viewer.

6. Conclusions

This study delves into English term variants of MWTs from the environmental domain and their translations into Spanish. The focus was on translation variation and its occurrence in different linguistic resources (i.e. parallel corpora, terminographic resources, and comparable corpora). A typology specifically conceived to characterize translation correspondences of MWTs has been proposed based on omissions, structural shifts, transpositions, expansions and inaccuracies. Additionally, parallel corpora have been found to include more translation variants than other linguistic resources. This proliferation could confirm the hypothesis that translators do not always consider MWTs as a whole unit of understanding but rather as different chunks of independent strings. However, the greater number of translation variants in parallel corpora can also respond to the general texts included in these corpora, even if it was expected that our set of MWTs did not appear very frequently in general texts, given their specialization. Since, to the best of our knowledge, parallel corpora on environmental science are scarce or even non-existent,

an investigation of term variation in this type of corpus should be carried out in the future. Our results highlight the need for an accurate term variation representation in terminographic resources, such as terminological knowledge bases (TKBs). Since they are conceived for different user types and specialized knowledge is not only conveyed in expert-to-expert scenarios, we believe that term variants should be extensively covered in TKBs. However, a balance between thoroughness and over-information must be sought for, because entries presenting 46 different variants (as was found in some of the concepts studied) could be counter-productive. Variation information should thus cover different types of variants and usage-related information, such as frequency and semantic or pragmatic indications. As a future line of research, we plan to implement the findings of this study in the TKB EcoLexicon, where new modules will be offered to compare the use, frequency, and meaning of term variants.

References

- AGUADO DE CEA, Guadalupe & Elena Montiel Ponsoda. (2012) "Term variants in ontologies." In: *XXX Congreso Internacional de AESLA 2012*. Lleida: AESLA, pp. 436-443.
- BENSON, Morton; Evelyn Benson & Robert F. Ilson. (1986) *Lexicographic Description of English*. Studies in Language Companion Series 14. Amsterdam, Philadelphia: John Benjamins.
- BEVILACQUA, Cleci Regina. (2004) *Unidades fraseológicas eventivas: descripción y reglas de formación en el ámbito de la energía solar*. Barcelona: Universitat Pompeu Fabra. PhD thesis.
- BOWKER, Lynne. (1998) "Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study." *Meta* 43:4, pp. 631-651.
- BOWKER, Lynne & Shane Hawkins. (2006) "Variation in the organization of medical terms. Exploring some motivations for term choice." *Terminology* 12:1, pp. 79-110.
- CABEZAS GARCÍA, Melania & Santiago Chambó. (in press) "Multi-word term variation: Prepositional and adjectival complex nominals in Spanish." *RESLA (Revista Española de Lingüística Aplicada)*.
- CABRÉ, María Teresa. (1993) *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Antártida, Empúries.

- CANDEL MORA, Miguel Ángel & María Luisa Carrió Pastor. (2012) "Corpus analysis: a pragmatic perspective on term variation." *RESLA (Revista Española de Lingüística Aplicada)* 2012, pp. 33-50.
- COLLET, Tanja. (2003) "A two-level grammar of the reduction processes of French complex terms in discourse." *Terminology* 9:1, pp. 1-27.
- CORPAS PASTOR, Gloria. (1996) *Manual de fraseología española*. Madrid: Gredos.
- CORPAS PASTOR, Gloria. (2000) "Fraseología y traducción." In: Salvador, Vicent & Adolf Piquer. (eds.) 2000. *El discurs prefabricat. Estudis de fraseologia teòrica i aplicada*. Estudis Filològics 4. Castelló de la Plana: Servei de Publicacions de la Universitat Jaume I, pp. 49-62.
- DAILLE, Béatrice. (2003) "Conceptual structuring through term variations." Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment. Sapporo, Japan: ACL, pp. 9-16.
- DAILLE, Béatrice. (2017) *Term Variation in Specialised Corpora: Characterisation, Automatic Discovery and Applications*. Amsterdam: John Benjamins.
- DAILLE, Béatrice & Emmanuel Morin. (2005) "French-English Terminology Extraction from Comparable Corpora." In: Dale, Robert; Kam-Fai Wong; Jian Su & Oi Yee Kwong. (eds.) 2005. *IJCNLP 2005*. Lecture Notes in Computer Science 3651. Berlin, Heidelberg: Springer, pp. 707-718.
- DÉJEAN, Hervé & Éric Gaussier. (2002) "Une nouvelle approche à l'extraction de lexiques bilingues à partir de corpus comparables." *Lexicometrica* 2002, pp. 1-22.
- FABER, Pamela & Pilar León Araúz. (2016) "Specialized knowledge representation and the parameterization of context." *Frontiers in Psychology* 7:196, pp. 1-20.
- FERNÁNDEZ-SILVA, Sabela. (2016) "The Cognitive and Rhetorical Role of Term Variation and its Contribution to Knowledge Construction in Research Articles." *Terminology* 22:1, pp. 52-79.
- FERNÁNDEZ-SILVA, Sabela. (2018) "The cognitive and communicative functions of term variation in research articles: a comparative study in Psychology and Geology." *Applied Linguistics* 2018, pp. 1-23.
- FERNÁNDEZ-SILVA, Sabela; Judit Freixa & María Teresa Cabré. (2009) "The multiple motivation in the denomination of concepts." *Journal of Terminology Science and Research* 20, pp. 1-24.
- FERNÁNDEZ-SILVA, Sabela & Koen Kerremans. (2011) "Terminological variation in source texts and translations: a pilot study." *Meta* 56:2, pp. 318-335.

- FIRTH, John Rupert. (1957) *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- FREIXA, Judit. (2006) "Causes of Denominative Variation in Terminology: A typology proposal." *Terminology* 12:1, pp. 51-77.
- FREIXA, Judit & Sabela Fernández-Silva. (2017) "Terminological variation and the unsaturability of concepts." In: Drouin, Patrick; Aline Francœur; John Humbley & Aurélie Picton. (eds.) 2017. *Multiple Perspectives on Terminological Variation*. Amsterdam: John Benjamins, pp. 155-181.
- GARCÍA-PAGE, Mario. (2008) *Introducción a la fraseología española*. Barcelona: Anthropos.
- GERZYMISCH-ARBOGAST, Heidrun. (2008) "Fundamentals of LSP Translation." *Mutra* 2, pp. 7-55.
- GIACOMINI, Laura. (2018) "Frame-based Lexicography: Presenting Multiword Terms in a Technical E-dictionary." In: *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts*. Ljubljana: EURALEX, pp. 309-318.
- GLÄSER, Rosemarie. (1988) "The grading of idiomaticity as a presupposition for a taxonomy of idioms." In: Hüllen, Werner & Rainer Schulze. (eds.) 1988. *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics*. Tübingen: Max Niemeyer, pp. 264-279.
- GLEDHILL, Christopher & Mojca Pecman. (2018) "On Alternating Pre-Modified and Post-Modified Nominals Such As Aspirin Synthesis Versus Synthesis of Aspirin: Rhetorical and Cognitive Packing in English Science Writing." *Fachsprache* 40:1-2, pp. 26-48.
- GREFENSTETTE, Gregory. (2002) "Multilingual corpus-based extraction and the Very Large Lexicon." In: Borin, Lars. (ed.) 2002. *Parallel corpora, parallel worlds*. Amsterdam, New York: Rodopi, pp. 135-149.
- GREGORY, Michael & Susanne Carroll. (1978) *Language and Situation: Language Varieties and their Social Contexts*. London: Routledge.
- HURTADO ALBIR, Amparo. (2001) *Traducción y Traductología: Introducción a la traductología*. Madrid: Cátedra.
- JIMÉNEZ-CRESPO, Miguel Ángel & Maribel Tercedor-Sánchez. (2017) "Lexical variation, register and explicitation in medical translation: A comparable corpus study of medical terminology in US websites translated into Spanish." *Translation and Interpreting Studies* 12:3, pp. 405-426.

- KERREMANS, Koen. (2010) "A comparative study of terminological variation in specialised translation." In: Heine, Carmen & Jan Engberg. (eds.) 2010. *Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009*. Aarhus: Aarhus University, pp. 1-14.
- KERREMANS, Koen. (2016) "Variation in the translation of terms: corpus-driven terminology research." In: Lewandowska-Tomaszczyk, Barbara; Marcel Thelen; Gys-Walt van Egdom; Dirk Verbeeck & Łukasz Bogucki. (eds.) 2016. *Translation and Meaning*. Frankfurt am Main: Peter Lang, pp. 217-227.
- KERREMANS, Koen. (2017) "Towards a resource of semantically and contextually structured term variants and their translations." In: Drouin, Patrick; Aline Francœur; John Humbley & Aurélie Picton. (eds.) 2017. *Multiple Perspectives on Terminological Variation*. Amsterdam: John Benjamins, pp. 83-108.
- KERREMANS, Koen & Rita Temmerman. (2016) "Finding the (un)expected? Quantitative and qualitative comparisons of term variants and their translations in a parallel corpus of EU texts." In: Corpas Pastor, Gloria & Miriam Seghiri. (eds.) 2016. *Corpus-based Approaches to Translation and Interpreting: from theory to applications*. Frankfurt am Main: Peter Lang, pp. 43-63.
- KILGARRIFF, Adam; Vít Baisa; Jan Bušta; Miloš Jakubíček; Vojtěch Kovář; Jan Michelfeit; Pavel Rychlý & Vít Suchomel. (2014) "The Sketch Engine: ten years on." *Lexicography* 1:1, pp. 7-36.
- LEÓN ARAÚZ, Pilar; Arianne Reimerink & Pamela Faber. (2019) "EcoLexicon and by-products: integrating and reusing terminological resources." *Terminology* 25:2, pp. 222-258.
- LEÓN ARAÚZ, Pilar. (2017) "Term and Concept Variation in Specialized Knowledge Dynamics." In: Drouin, Patrick; Aline Francœur; John Humbley & Aurélie Picton. (eds.) 2017. *Multiple Perspectives on Terminological Variation*. Amsterdam: John Benjamins, pp. 213-258.
- MEL'ČUK, Igor; André Clas & Alain Polguère. (1995) *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.
- MEYER, Ingrid & Kristen Mackintosh. (1996) "Refining the terminographer's concept-analysis methods: How can phraseology help?" *Terminology* 3:1, pp. 1-26.
- MIYATA, Rei & Kyo Kageura. (2016) "Constructing and Evaluating Controlled Bilingual Terminologies." In: *Proceedings of the Fifth International Workshop on Computational Terminology (Computerm2016)*. Osaka: ACL, pp. 83-93.

- MOGORRÓN-HUERTA, Pedro. (2010) "Analyse du figement et de ses possibles variations dans les constructions verbales espagnoles." *Linguisticae Investigationes* XXXIII:1, pp. 86-152.
- NEWMARK Peter. (1981) *Approaches to Translation*. Oxford: Pergamon.
- NORD, Christiane. (1997) *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Manchester: St. Jerome Publishing.
- OLTRA RIPOLL, María Dolores. (2018) "La fraseologia com a tret estilístic en novel·les originals en anglés i les seues traduccions al català." *Caplletra* 65, pp. 95-124.
- PAWLEY, Andrew. (2001) "Phraseology. Linguistics and the Dictionary." *International Journal of Lexicography* 14:2, pp. 122-134.
- PICHT, Heribert. (1991) "Fraseología LSP desde el Punto de Vista Terminológico." *Sendebär* 2, pp. 91-105 (translated from German by Marián Hens Córdoba).
- PICTON, Aurélie. (2011) "Picturing short-period diachronic phenomena in specialised corpora: A textual terminology description of the dynamics of knowledge in space technologies." *Terminology* 17:1, pp. 134-156.
- RAMISCH, Carlos. (2015) *Multiword Expressions Acquisition: A Generic and Open Framework*. Cham: Springer.
- REISS, Katharina & Hans Josef Vermeer. (1984) *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer.
- ROGERS, Margaret. (1997) "Synonymy and equivalence in special-language texts. A case study in German and English texts on Genetic Engineering." In: Trosborg, Anna (ed.) 1997. *Text Typology and Translation*. Amsterdam, Philadelphia: John Benjamins, pp. 217-245.
- SANZ VICENTE, Lara. (2011) *Análisis contrastivo de la terminología de la teledetección. La traducción de compuestos sintagmáticos nominales del inglés al español*. Salamanca: Universidad de Salamanca. PhD thesis.
- SINCLAIR, John. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- TEMMERMAN, Rita. (2000) *Towards New Ways of Terminology Description: The sociocognitive approach*. Amsterdam, Philadelphia: John Benjamins.
- TIEDEMANN, Jörg. (2012) "Parallel Data, Tools and Interfaces in OPUS." In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: ELRA, pp. 2214-2218.
- VAISA, Vít; Jan Michelfeit; Marek Medved & Miloš Jakubiček. (2016) "European Union Language Resources in Sketch Engine." In: *Proceedings of the Tenth*

- International Conference on Language Resources and Evaluation (LREC'16)*.
Portorož: ELRA, pp. 2799-2803.
- WÜSTER, Eugen. (1968) *The Machine Tool. An Interlingual Dictionary of Basic Concepts*. London: Technical Press.
- ZULUAGA, Alberto. (1975) "La fijación fraseológica." *Thesaurus* XXX:2, pp. 225-248.

BIONOTE / NOTA BIOGRÁFICA

PILAR LEÓN ARAÚZ is a Senior Lecturer at the Faculty of Translation and Interpreting in Granada, where she teaches courses in the fields of Terminology and CAT tools. She received her PhD in Translation and Interpreting from the University of Granada in 2009. She also holds a BA in Applied Foreign Languages from the Université de Provence (France) and a BA in Applied Modern Languages from the University of Northumbria (UK). Her major fields of research include Terminology, Knowledge Representation and Extraction and Corpus Linguistics. She has published papers in international journals such as *Terminology*, *International Journal of Lexicography* and *Frontiers*, and serves on the scientific boards of international journals and conferences.

PILAR LEÓN ARAÚZ es Profesora Titular en la Facultad de Traducción e Interpretación de la Universidad de Granada, donde imparte asignaturas relacionadas con la Terminología y las herramientas de traducción asistida por ordenador. En 2009 obtuvo el Doctorado en Traducción e Interpretación por la Universidad de Granada. Es Licenciada en Lenguas Extranjeras Aplicadas por la Universidad de Provence (Francia) y en Lenguas Modernas Aplicadas por la Universidad de Northumbria (Reino Unido). Sus principales áreas de investigación son la terminología, la representación y extracción del conocimiento, y la lingüística de corpus, sobre las cuales ha publicado artículos en revistas internacionales como *Terminology*, *International Journal of Lexicography* y *Frontiers*. Además, forma parte del comité científico de revistas y congresos internacionales.

MELANIA CABEZAS GARCÍA holds a degree in Translation and Interpreting from the University Pablo Olavide and a Master's degree in Professional Translation from the University of Granada, where she teaches courses in the Bachelor's

Degree in Translation and Interpreting. She received her PhD in Translation and Interpreting from the University of Granada in 2019. She is a member of the LexiCon research group and was awarded a research fellowship by the Spanish Ministry of Education to write her PhD dissertation on the formation, translation, and representation of English and Spanish multiword terms. Her research interests are Terminology, Corpus Linguistics, and Specialized Translation. She has published papers in international journals on Linguistics, Terminology, and Specialized Translation, and serves on the scientific boards of international journals and conferences, such as *LREC* and the *International Journal of Lexicography*.

MELANIA CABEZAS GARCÍA es Graduada en Traducción e Interpretación por la Universidad Pablo de Olavide y Máster en Traducción Profesional por la Universidad de Granada, donde imparte docencia en el Grado en Traducción e Interpretación. Obtuvo el Doctorado en Traducción e Interpretación por la Universidad de Granada en 2019. Perteneció al grupo de investigación LexiCon y contó con una beca de investigación predoctoral otorgada por el Ministerio de Educación de España para realizar su tesis doctoral, en la que estudió la formación, traducción y representación de los términos compuestos en inglés y español. Sus áreas de investigación son la terminología, la lingüística de corpus y la traducción especializada. Ha publicado artículos en revistas internacionales sobre lingüística, terminología y traducción especializada, y forma parte del comité científico de revistas y congresos internacionales, como *LREC* o *International Journal of Lexicography*.

Recibido / Received: 30/10/2019

Aceptado / Accepted: 21/03/2020

Para enlazar con este artículo / To link to this article:

<http://dx.doi.org/10.6035/MonTI.2020.ne6.8>

Para citar este artículo / To cite this article:

López Ruiz, María del Carmen. (2020) "La recepción y traducción de unidades fraseológicas en el discurso político. Análisis de un corpus *ad hoc* (EN, FR > ES)." En: 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. 248-286.

LA RECEPCIÓN Y TRADUCCIÓN DE UNIDADES FRASEOLÓGICAS EN EL DISCURSO POLÍTICO. ANÁLISIS DE UN CORPUS *AD HOC* (EN, FR > ES)¹

THE RECEPTION AND TRANSLATION OF PHRASEOLOGICAL UNITS IN POLITICAL SPEECHES. ANALYSIS OF AN *AD HOC* CORPUS (EN, FR > ES)

MARÍA DEL CARMEN LÓPEZ RUIZ

maricarmen.lopezruiz@uco.es

Universidad de Córdoba (España)

Resumen

El objetivo de este trabajo es investigar las UFS más habituales en los discursos políticos partiendo de las premisas que se han abordado hasta el momento en este lenguaje de especialidad en relación con la fraseología (Coseriu, 1987; Van Dijk, 1996; Navarro, 2002; Fuentes *et al.*, 2002). Adoptando una visión traductológica, mostraremos las UFS halladas en un corpus de trabajo compuesto por discursos políticos en inglés y en francés con su correspondiente traducción hacia el español. La intención es clasificar las UFS para obtener resultados cuantitativos en lo referente a la fraseología del discurso político, tanto en el corpus de discursos en inglés como en el de francés, para así poder obtener una visión más completa de los diferentes tipos de UFS que se hallan en el discurso político y sus porcentajes correspondientes. El resultado de nuestra

-
1. La presente contribución forma parte de la tesis doctoral que lleva a cabo actualmente la autora en el marco del contrato para la Formación del Profesorado Universitario (FPU) en la Universidad de Córdoba. Dicha investigación está respaldada económicamente por el Ministerio de Universidades de España.



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

investigación nos permitirá elaborar un glosario trilingüe cuya finalidad es servir de referencia durante el proceso de documentación de otros especialistas en el tema.

Palabras clave: Traducción; Fraseología; Discurso Político; Traducción Especializada; Unidades Fraseológicas.

Abstract

This paper examines the most frequent PUs in political speeches, considering previous works in phraseological studies with regards to this specialized language by authors such as Coseriu, 1987; Van Dijk, 1996; Navarro, 2002 or Fuentes et al., 2002. We intend to display from a traductological point of view all the different PUs found in our corpus of work, composed by English and French political speeches, as well as their corresponding translations into Spanish. Our goal is to classify such PUs in order to obtain quantitative results regarding phraseology in our corpus, and thus attain a broader view of the types of PUs that are commonly used in political speeches and their corresponding percentage frequency. Results will let us bring out a trilingual glossary in our attempt to help other specialists throughout their documentation process when translating.

Keywords: Translation; Phraseology; Political Discourse; Specialized Translation; Phraseological Units.

1. Fraseología y discurso político. Estado de la cuestión

El estudio de la fraseología se ha abordado desde diferentes prismas teóricos. Tanto es así que no sorprende encontrar contribuciones enfocadas desde la perspectiva lingüística (Coseriu 1987; Corpas Pastor 1997; Zuluaga 1997; Martínez Marín 1996; Navarro 2002; García Page Sánchez 2008), la didáctica (García Yelo 2005; García Romero & Sevilla Muñoz 2005; Higuera García 2009; Serrano Lucas 2010; Brazo Millán 2018) o la contrastiva (Lozano 1992; Gutiérrez Díez 1995; Sevilla Muñoz 1997; Corpas Pastor 2003; Radulescu 2006; Timofeeva 2012; Mogorrón Huerta 2014). Sin embargo, lo que interesa para los objetivos de esta contribución es conocer cuál es el estado de la cuestión en lo que concierne a las UFS y el discurso político.

Como sugieren los estudios dedicados a este último, cada vez hay más interés en abordar este terreno por parte tanto de investigadores/as como

de grupos de investigación. Lo que destaca la mayor parte de las publicaciones que abordan la fraseología del discurso político es la capacidad que tienen estas unidades para aportar énfasis a la eficacia comunicativa del discurso. Así lo prueban autores como Van Dijk 1996; Tannen 1999; Fernández Lagunilla 1999a; Navarro 2002; Fabbri *et al.* 2002; Nuñez Cabezas *et al.* 2002; Fuentes *et al.* 2002; Arce Castillo 2006; Arellano *et al.* 2013 o Messina Fajardo 2016.

Si consideramos las conclusiones a las que llegan todos estos autores, podríamos decir que las diferentes UFS se insertan a lo largo del discurso político para cumplir diferentes funciones. Aunque a veces se ayudan de otros recursos estilísticos y retóricos, tales como la metáfora, las UFS pueden transmitir un contenido completo contextualizado en el discurso en que se insertan. Sin embargo, las UFS que se hallan en un discurso político no solo aparecen en ese lenguaje de especialidad; uno de los rasgos distintivos de esta categoría radica precisamente en la posibilidad de ubicarse en distintos contextos y discursos. Esta posibilidad estaría ligada a otro rasgo definitorio de las UFS en el discurso político: la utilización de UFS ayuda al político/a a conectar con el auditorio, pues en muchas ocasiones permite que el mensaje llegue de un modo mucho más accesible, lo cual fomenta la relación orador/auditorio, lo que potencia, a su vez, la función apelativa. Así, el marco del discurso y la comunicación entre el/la dirigente político/a y el auditorio se centran en la base argumental en la que se fundamenta el discurso, propiciando la relación y conexión de las ideas transmitidas entre emisor y receptor. Como Messina Fajardo (2016: 125) establece:

Los políticos recurren a los fraseologismos, o a porciones de refranes disociados de su contexto o trastocados. Las estructuras lingüísticas, al ser memorizadas y al pertenecer a la colectividad son reconocidas por todos sin necesidad de reproducir el texto completo. Esta es una técnica muy empleada también entre los clásicos de la literatura (Cervantes, Rojas, Lope, Quevedo). Con esta técnica, el discurso no solo gana rapidez, eficacia, sino [que] se establece una relación de convivencia y complicidad cultural con el interlocutor.

A raíz de esta afirmación, podemos comprobar que la multitud de funciones que las UFS pueden llegar a representar son ingentes. A propósito de lo establecido por Navarro (2002: 202-206), las UFS cumplen cuatro funciones

diferentes: función lingüístico-textual, función fraseológica (citando a Zuluaga 1997: 631), función connotativa y, finalmente, función icónica. Así, se comprueba que las UFS persiguen sus propias finalidades dentro de los lenguajes de especialidad. Aunque las UFS pueden estudiarse bien como ejemplos de uso de la lengua general, bien como representaciones específicas de los lenguajes de especialidad, está claro que en lo que concierne al lenguaje político dichas funciones están bien delimitadas, puesto que básicamente cumplen con los mismos objetivos que persigue de principio a fin un discurso político: convencer mediante argumentos a su auditorio. En ese sentido, las UFS se utilizan en el lenguaje político como medios para facilitar la comprensión de los enunciados más complejos y para hacer llegar de un modo más cercano la idea transmitida al receptor. Lo que se consigue es que el receptor del mensaje capte a la perfección lo que el político/a intenta transmitir y, en el caso de las UFS más connotativas, se potencia una mayor conexión con el político/a dentro del hilo argumental perseguido en su discurso, gracias al empleo de UFS contextualizadas según cualidades diastráticas o diatópicas, lo que acerca al político/a a su destinatario.

2. Expresiones y unidades fraseológicas y su presencia en el discurso político

Para resumir brevemente la clasificación formal establecida por Corpas Pastor (1997: 51), de modo que con un simple vistazo podamos contextualizar el estado de la cuestión, proponemos el esquema recapitulativo que se ofrece a continuación (véase Figura I). En este punto, debemos decir que, si bien ha habido otras investigaciones posteriores que detallan algunos de los conceptos nombrados a continuación y establecen nuevas dicotomías y clasificaciones posibles (por ejemplo, la de Sevilla Muñoz & Crida Álvarez 2013), decidimos abogar por las de esta autora por tratarse de la nomenclatura más utilizada y recurrente por la mayor parte de los expertos en el tema. En el cuadro recapitulativo posterior también se encontrará una clasificación que alude a unos criterios de división de carácter semántico, para lo cual hemos recurrido a las aportaciones de Luque Durán & Manjón Pozas (2002: 2) y Montoro del Arco (2007: 131). Así, surgiría una clasificación semántica de

las UFS que posibilitaría la existencia de dos subtipos: las UFS ontológicas y las culturales. En el caso de las ontológicas, hacen referencia a estructuras y unidades de naturaleza universal, ya que son extrapolables a todas las culturas; por su parte, las culturales se refieren a rasgos prototípicos de una sociedad o cultura en particular, y estarían en conexión con una de las funciones que comentamos con anterioridad (concretamente, la connotativa, en los términos de Navarro, *op. cit.*). Así lo explican los mismos autores (Durán & Manjón Pozas, *ut supra*):

Como conclusión provisional, puede afirmarse que la tipología fraseológica tendría un esquema de partida en el cual existiría un planteamiento ontológico opuesto a un planteamiento cultural. El planteamiento ontológico postula que una parte de las metáforas no son más que maneras de hablar sobre un determinado fenómeno basándose en asociaciones ontológicas (fisiológicas, etc.) y, por tanto, tendrán un carácter universal. [...] Frente a las metáforas y a la fraseología de planteamiento ontológico existen las metáforas y la fraseología cultural, que se basan en algunas realidades peculiares de un pueblo y de una cultura, como es el caso de los toros en español, el béisbol en el mundo americano, el sumo en japonés, etc., aunque se podían citar centenares de ejemplos. Sin embargo, ontología y cultura son sólo los dos polos en la creación de fraseología metafórica. En realidad, al estudiar la fraseología de diferentes lenguas, encontramos muchos más hechos y problemas que hay que tener en cuenta y dilucidar antes de poder sentar las bases científicas de una tipología fraseológica.

Precisamente, en torno a las unidades fraseológicas culturales, Montoro del Arco (*ibidem*) explica que “La fraseología de las lenguas es uno de los medios a través de los cuales la memoria colectiva da cuenta de las cualidades y connotaciones socioculturales que se han ido asociando durante siglos [...]”. Por ende, el estudio de este tipo de unidades (culturales) nos permite dar cuenta de que, en el contexto de esta investigación, una buena parte del legado lingüístico que destaca entre las UFS presentes en el discurso político es un fiel reflejo de la cultura y de la memoria social e histórica de una comunidad lingüística.

- | |
|---|
| <p>1. Unidades fraseológicas (UFS)</p> <p>a) Clasificación formal (Corpas Pastor, 1997)</p> <ul style="list-style-type: none"> - Locuciones (<i>op. cit.</i>, 91): - Locuciones nominales - Locuciones adjetivas - Locuciones adverbiales - Locuciones verbales - Locuciones prepositivas - Locuciones conjuntivas - Locuciones clausales <p>b) Colocaciones (<i>op. cit.</i>, 66):</p> <ul style="list-style-type: none"> - Sustantivo + verbo - Verbo + sustantivo - Sustantivo + adjetivo/sustantivo - Sustantivo + preposición + sustantivo - Verbo + adverbio - Adverbio + adjetivo <p>c) Enunciados fraseológicos:</p> <ul style="list-style-type: none"> - Paremias (<i>op. cit.</i>, 147): <ul style="list-style-type: none"> • Enunciados de valor específico • Citas • Refranes - Fórmulas rutinarias (<i>op. cit.</i>, 174) <ul style="list-style-type: none"> • Fórmulas discursivas • Fórmulas psicosociales <p>2. Clasificación semántica (Durán y Manjón Pozas, 2002: 2; Montoro del Arco, 2007: 131-132)</p> <p>a) Ontológicas</p> <p>b) Culturales</p> |
|---|

Figura I. Cuadro resumen recapitulativo de la clasificación formal y semántica de unidades fraseológicas (Corpas Pastor, 1997; Durán y Manjón, 2002; Montoro del Arco, 2007)

3. Objetivos de nuestra investigación y metodología de trabajo

Nuestro trabajo ha consistido en recopilar un corpus de veinte discursos políticos en inglés y otros veinte discursos políticos en francés que cuentan con sus respectivas traducciones oficiales hacia el español y hacia el inglés o francés (dependiendo de la lengua del TO). Los textos que componen el corpus son paralelos, lo cual implica que, además de contar con la versión original (en inglés o en francés, hallados en internet, periódicos, páginas web especializadas [previa comprobación con el original], traducciones oficiales

en las páginas de la Embajada o de la UNESCO, etc.), también se ha dispuesto de la versión traducida correspondiente (en medios digitales, generalmente en la red de internet, en páginas especializadas en discurso político o directamente en medios de comunicación)². El criterio que hemos seguido para seleccionar un discurso u otro de un mismo/a dirigente ha sido el de la temática del propio discurso, seleccionando especialmente discursos políticos institucionales, bien de índole nacional o internacional, ya que será en estos donde se registre una mayor frecuencia de terminología especializada, lo que implica, también, una mayor frecuencia de UFS de corte político. Para garantizar que las muestras eran similares en lo que concierne al tamaño, hemos utilizado discursos de una extensión similar, que podría equipararse con unos 10-15 minutos de duración en la oralidad, aproximadamente, como se ha comprobado en aquellos discursos de los que se ha encontrado una versión televisada a través del canal de internet *Youtube*. Además, solo se han contemplado discursos completos, no fragmentos ni noticias.

Dicho lo anterior, a continuación se indica la metodología de trabajo seguida, la cual ha sido dividida en tres fases:

- a. Una primera fase, mediante la que se ha procedido a seleccionar cada una de las UFS halladas en los textos en inglés y en francés. Posteriormente, se compiló información relativa a cada una de estas unidades, utilizando para ello la ficha fraseológica que se mostrará posteriormente (véase Figura IV). Para la realización de dicha ficha, se utilizaron, en esta primera fase, obras lexicográficas en las que se ha basado por completo la fase de documentación de este trabajo, y que detallaremos ahora. El objetivo de esta primera fase consiste en compilar todas las UFS de los textos origen para buscar posteriormente una UF equivalente, teniendo en cuenta el contexto en lengua meta³ (inglés/francés y español). Los resultados de esta primera fase nos han permitido elaborar un glosario con los datos obtenidos hasta el momento con nuestras propias propuestas de traducción, los cuales

2. Véase anexo final para más datos.

3. Se encontrarán, en este trabajo, alusiones al término “lengua meta” a través de la sigla “LM”. De modo análogo, la sigla “LO” remitirá al término “lengua origen”; “TO”, a “texto origen” y, finalmente, “TM” a “texto meta”.

se mostrarán en la parte final de este trabajo, así como conocer con gráficos porcentuales un conjunto de datos relativos al uso de las UFS en inglés y en francés en los discursos políticos que componen nuestro corpus. Esta contribución se centrará exclusivamente en esta primera fase.

- b. Una segunda fase, mediante la que se ha comprobado en las traducciones de los TM si la traducción (oficial o en prensa) ofrece o no un equivalente formal y semántico apto para transmitir el mismo sentido que la UF del TO. Tras esto, se ha evaluado la pertinencia o inadecuación de dicha UF del TM a partir de los datos de nuestras fichas fraseológicas y de nuestro glosario, barajando también la posibilidad de la retroalimentación entre las diversas fuentes. Una vez valoradas todas las UFS, se dispondrán los resultados de forma porcentual igualmente, y se añadirán a nuestro glosario las UFS que se hayan traducido de un modo alternativo a nuestra propuesta, siempre que estén correctamente trasladadas con base en el contexto.
- c. La fase final del proyecto se concibe a través de una herramienta lexicográfica que nos permitirá disponer de todo el glosario junto a las fichas fraseológicas en un portal web.

Inciendo en la primera de estas tres fases, que es a la que nos dedicamos en este artículo, indicaremos que, en primer lugar, se seleccionaron las UFS que contenía cada discurso y se valoraron las dificultades de traducción asociadas al contexto específico en el que se proveía cada una de ellas. Tras un proceso de documentación extensivo, se buscaron decisiones traductológicas convenientes (lingüística y traductológicamente hablando) para representar la realidad a la que aludía la UF origen, respetando, en todo caso, el propósito fundamental que perseguía el orador del texto origen. Con posterioridad, se elaboró una ficha fraseológica para cada una de las UFS del corpus de trabajo, que ejemplificaremos posteriormente tomando como referencia el modelo presentado por Serrano Lucas (2010: 202). El contenido de estas fichas es crucial para este trabajo, ya que esta información será la que aparecerá al utilizar el motor de búsqueda del portal web donde aparecerán todas estas unidades para su consulta, una vez terminado el proyecto en su totalidad.

Las fuentes empleadas a lo largo del proceso de documentación para proceder con nuestras propuestas de traducción han sido las siguientes:

Fuentes documentales en inglés
<ul style="list-style-type: none"> – Cambridge Advanced Learner’s Dictionary. En papel – Cambridge Phrasal Verbs Dictionary. En papel – Collins Cobuild English Grammar. (2011). Bishopbriggs (Glasgow): HarperCollins – Diccionario Cambridge en línea – Diccionario Longman en línea – Diccionario Merriam-Webster en línea – Diccionario Oxford en línea – Diccionario Oxford Inglés/Francés, Francés/Inglés (en papel) – Diccionarios y glosarios de la Unión Europea (en línea): Lexicool – Eurovoc: Tesoro multilingüe de la Unión Europea – Gran Diccionario Oxford Inglés/Español, Español/Inglés (en papel) – IATE: European Union Terminology Database. En línea – Longman Dictionary of Contemporary English. (2005). Harlow: Pearson Education Limited – Macmillan English Dictionary for Advanced Learners. (2007). Oxford: Macmillan – Oxford Collocations Dictionary for students of English. En papel – Oxford Idioms Dictionary for learners of English. En papel – Oxford Phrasal Verbs for students of English. En papel – Rigdom, J. (2018). <i>Handbook of English Idioms, Parables and Colloquial Sayings</i>. WordsRUs PhraseBooks, 1. Versión en línea – The Free Dictionary. En línea – Unbridged Dictionary Spanish/English, English/Spanish Larousse. En papel
Fuentes documentales en francés
<ul style="list-style-type: none"> – ABC Thesaurus (en línea) – Cassagne, J.M et L. Nisset (2000). <i>101 French Idioms</i>. Pennsylvania: McGraw-Hill – Diccionario Larousse en línea – Diccionario Oxford Inglés/Francés, Francés/Inglés (en papel) – Diccionario Reverso en línea – Gran Diccionario Larousse Español/Francés, Francés/Español (en papel) – Gran Dictionnaire Anglais/Français, Français/Anglais Larousse (2017). En papel – IATE: European Union Terminology Database. En línea – Lexilogos (en línea) – P. de Rieux, A. (2006): <i>Dictionnaire des idiomes anglais et américains</i>. Paris: Publibook

Fuentes documentales en español
<ul style="list-style-type: none">– Carbonell Basset, D. (1995). <i>Diccionario fraseológico inglés/castellano, castellano/inglés</i>. Barcelona: Ediciones del Serbal– Carbonell Basset, D. (2004). <i>Diccionario inglés y español de modismos</i>. Barcelona: Ediciones del Serbal– Carbonell Basset, D. (2005). <i>Nuevo diccionario de dichos y refranes actuales, inglés y castellano</i>. Barcelona: Ediciones del Serbal– Diccionario de la Real Academia Española (RAE, versión en línea). Disponible en este enlace– Diccionario del Español Jurídico (DEJ) en línea– Gran Diccionario Larousse de la Lengua Española– Gran Diccionario Larousse Francés/Español, Español/Francés– Gran Diccionario Oxford Inglés/Español, Español/Inglés (en papel)– Seco, M. (2012). <i>Nuevo diccionario de dudas y dificultades de la lengua</i>. Madrid: Espasa

Figura II. Recapitulación de fuentes documentales para las propuestas de traducción en lengua meta

Tras resolver estos esquemas, elaboramos un glosario trilingüe centrado en discursos políticos, el cual será presentado en las páginas siguientes con el ánimo de facilitar las tareas de documentación en este lenguaje de especialidad a los traductores/as que se encuentren ante un reto traductológico que encaje dentro de las dimensiones de este trabajo.

En consonancia con todo lo anterior, los objetivos que se persiguen en este trabajo (tomando como referencia la primera fase a la que nos dedicaremos) son los que siguen:

- Compilar un corpus de trabajo compuesto por discursos políticos en inglés y francés para identificar todas las UFS dispuestas a lo largo de los mismos.
- Identificar todas las UFS halladas en el corpus en inglés y en el corpus en francés, con el fin de cuantificar el número estimado de UFS por pares de idiomas. Para llevar a cabo dicha identificación, se han seguido de cerca los criterios de clasificación de las obras lexicográficas mencionadas con anterioridad, además de nuestros propios conocimientos gramaticales en español y en lenguas extranjeras

(inglés/francés) para proceder a identificarlas. No obstante, la ayuda de algunos programas específicos, tales como la herramienta *Antconc*, han servido de gran ayuda para identificar este tipo de unidades, con el objetivo de seleccionar aquellas cuya visibilidad en la totalidad del corpus fuera realmente representativa.

- Estudiar cada UF en el contexto específico de emisión en la lengua y cultura origen para comprender el significado adquirido en dicho contexto.
- Categorizar cada una de las UFS localizadas en el corpus de trabajo según la disposición de Corpas Pastor, Durán, Manjón Pozas & Montoro del Arco (*op. cit.*), siguiendo el modelo de ficha fraseológica propuesto, con el fin de conocer cuál es el tipo de UFS más recurrente en los discursos políticos por cada par de lenguas.

4. Traducir fraseología: las unidades fraseológicas del discurso político en inglés y en francés y su recepción en español en nuestro estudio

4.1. Corpus de trabajo

En la elaboración del corpus se han seleccionado discursos de dirigentes políticos de diferentes organismos internacionales hasta naciones en particular. En este trabajo, las variantes que hemos introducido como criterios de búsqueda no han estado ligadas al estudio de los políticos/as *per se*, por lo que el único patrón diferenciador ha sido la lengua emitida, y no otros criterios como la ideología de cada orador/a político que pronuncia el discurso, la variable sexo, la variable edad, etc. Aunque estas aportan muchos datos a nuestro estudio, deberán quedarse relegadas a una futura investigación.

Siendo así, el corpus compilado adquirirá la forma siguiente:

Número discurso	Discursos en inglés (TO: EN)	Discursos en francés (TO: FR)
1	Abraham Lincoln	Charles de Gaulle
2	Barack Obama	Dominique de Villepin
3	Bill Clinton	Édouard Philippe
4	Boris Johnson	Emmanuel Macron
5	Donald Trump	François Fillon
6	Dwight Eisenhower	François Hollande
7	Franklin D. Roosevelt	François Legault
8	Geoffrey Howe	François Mitterrand
9	George W. Bush	Georges Pompidou
10	Harold Macmillan	Gautier Mignot
11	John F. Kennedy	Jacques Chirac
12	Margaret Thatcher	Jean Charest
13	Mark Drakeford	Jean-Yves Le Dirian
14	Martin Luther King	Laurent Fabius
15	Nelson Mandela	Léon Blum
16	Nikita Khrushchev	Marine Le Pen
17	Robert F. Kennedy	Nicolas Sarkozy
18	Ronald Reagan	Philippe Couillard
19	Theresa May	Ueli Maurer
20	Winston Churchill	Valéry Giscard d'Estaing

Figura III. Detalle del corpus de trabajo compilado atendiendo al autor (ordenado por orden alfabético)

4.2. Muestra del procedimiento de análisis

Tras compilar el corpus, la primera fase de trabajo consistía en identificar todas las UFS halladas a lo largo de los discursos. Después, había que contextualizar dicha UF en el contexto del acto comunicativo en el que se encuadraba, con el fin de apreciar correctamente el sentido en la lengua origen. Posteriormente, se llevó a cabo el proceso de documentación para localizar un equivalente en lengua meta, y por último, se procedió a realizar un proceso

de revisión para comprobar si la traducción propuesta cumplía o no con las mismas expectativas del TO en lo referente a los fines de la traducción.

En las líneas siguientes, se ha incorporado una pequeña muestra esquemática donde señalaremos los aspectos a los que nos hemos referido anteriormente sobre las fases seguidas para llevar a cabo nuestra investigación. Dadas las constricciones espaciales, no nos será posible valorar una muestra por cada discurso que compone el corpus, así como tampoco un análisis extensivo de la totalidad de UFS del discurso, motivo por el cual tan solo esbozaremos aquí una muestra, escogida al azar para simplemente ejemplificar el modo de proceder en el análisis:

- **Identificación del discurso:**
 - **Nombre del emisor:** Barack H. Obama
 - **Contexto del discurso:** Discurso en El Cairo
 - **Fecha del discurso:** 4 de junio de 2009
- **Muestra:**

(TO) ENGLISH

“But I am convinced that in order to move forward, we must say openly to each other the things we hold in our hearts and that too often are said only behind closed doors. There must be a sustained effort to listen to each other, to learn from each other, to respect one another, and to seek common ground. As the Holy Qur’an tells us, ‘Be conscious of God and speak always the truth.’ That is what I will try to do today, to speak the truth as best I can, humbled by the task before us, and firm in my belief that the interests we share as human beings are far more powerful than the forces that drive us apart”.

- **Valoraciones:** En primer lugar, se han identificado las UFS del discurso halladas en el fragmento. Como podemos ver, no todas pertenecen al mismo nivel, si contemplamos el cuadro recapitulativo de Corpas Pastor (*ut supra*); así, nos encontramos con colocaciones, como por ejemplo *move forward*, *human beings* o *drive us apart*, y también con paremias, del tipo de *Be conscious of God and speak always the truth*, además de locuciones de diversa índole, como por ejemplo *in order to*. Tras considerar el contenido semántico y pragmático en el que se inserta este fragmento, llevaríamos a cabo una fase de comprensión semántica de lo que dicha UF quiere comunicar al lector desde el punto de vista del plano ilocutivo, al lector. Por ejemplo, la expresión *Be*

conscious of God and speak always the truth se ubicaría en un contexto en el que el político/a quiere utilizar una paremia para ilustrar el mensaje que está transmitiendo al auditorio (escucharse mutuamente, haciendo alusión especialmente a las naciones reunidas en la cumbre en la que se encuentra), por lo que se ayuda de esta UF estratégicamente seleccionada, ya que, existiendo otras tantas UFS de significado semejante, escoge una que procede del libro sagrado para el auditorio que tiene delante, puesto que se encuentra en El Cairo: El Corán. Esto añade significado inferido al mensaje transmitido. Por ende, desde el punto de vista semántico, nos encontraríamos frente a una UF de carácter cultural. Esto incluye una puntualización para el traductor: dicha expresión tiene una traducción acuñada, que ya procede de un libro que ha sido traducido a muchas lenguas. Esta UF en concreto se ha traducido en bastantes versiones del libro original por “Tengan conciencia de Dios y digan siempre la verdad” en español, y por “Sois conscients de Dieu et dis toujours la vérité” en francés, traducción que se corresponde con la versión más habitual en este libro sagrado para cada una de estas lenguas, respectivamente, y que incluye, además, una nueva UF: “tener conciencia de” y “être conscient de”. En este último ejemplo, sería la versión en francés la más próxima formalmente a la versión en inglés, ya que respeta la estructura *verbo ser + adjetivo + preposición de*, mientras que en español se procede a una transposición (en términos de Hurtado, 2001) mediante la que se utiliza el sustantivo de la misma familia léxica a la que pertenece el adjetivo correspondiente: *conciencia*. En cuanto a las funciones que llevaría a cabo la UF que nos ocupa, diremos que cumple con las funciones de Navarro (*ut supra*): la función fraseológica, en tanto que este fragmento está plagado de UFS utilizadas para transmitir mejor el mensaje al auditorio, receptor del mensaje comunicado; la connotativa, en tanto que se sirve de expresiones comunes al ámbito cultural del receptor del mensaje; y, finalmente, icónica, en tanto que dicho pasaje es un ejemplo característico de lenguaje metafórico y de la función poética, utilizada para ilustrar el argumento argüido a lo largo del fragmento.

– Ficha fraseológica (EN/FR/ES):

Unidad fraseológica en LO (EN) <i>Behind closed doors</i>	
Fuente del TO	TO: EN. Discurso 2
Definición	'If something happens <i>behind closed doors</i> , it is hidden or kept secret from public view'
Fuente de la definición	Diccionario Cambridge en línea. Disponible en este enlace
Tipo de UF	Locución adverbial
Ejemplo de uso de la UF en LO Fuente	[...] we must say openly to each other the things we hold in our hearts and that too often are said only <i>behind closed doors</i> TO: EN. Discurso 2
Observaciones	
<ul style="list-style-type: none"> – Tema/Campo semántico: Negociaciones privadas; secretismo, privacidad – Variantes en LO: Behind shut doors – Sinónimos: In privacy; be shrouded in secrecy/mystery; take sb to one side; to keep something just between you and me; in camera; in chambers – Antónimos: Openly; Publicly; Familiarly; Commonly – Registro: Formal e informal. En este contexto, formal – Pertenencia a lenguajes de especialidad: Jurisdicción, legalidad, discurso político, dadas las relaciones legales, jurídicas e institucionales y los encuentros y reuniones políticas secretas a las que el público no puede tener acceso y de las que, en ocasiones, ni siquiera se le informa para mantener así el anonimato – Grado de creatividad metafórica: Medio: existencia de metáfora, pero con fácil interpretación cognitiva. Esta UF hace referencia a la imagen de una <i>puerta cerrada</i> para indicar, a través de dicha metáfora, que se trata de algo a lo que no se puede tener acceso de forma fácil. Solo los que pueden “atravesar esa puerta” son los que tendrán acceso a la información privilegiada – Fuentes: Diccionarios Longman, Cambridge y Oxford (versión en línea). 	

Unidad fraseológica en lengua meta (ES) <i>Detrás de puertas cerradas</i>	
Fuente del término meta	Gran Diccionario Oxford Español/Inglés, Inglés/Español (2008) (versión en papel)
Definición	‘Dicho de un juicio o de una vista: Que se desarrolla, por motivos de honestidad, orden público y otros análogos, en presencia solo de las partes, sus representantes y sus defensas’
Fuente de la definición	RAE en línea. Disponible en este enlace
Ejemplo de uso de la UF en LM Fuente	El proceso en materia de transparencia y acceso del público a los documentos durante mucho tiempo se ha desarrollado detrás de puertas cerradas herméticamente. Debate del Parlamento Europeo. Enlace
Observaciones	
<ul style="list-style-type: none"> – Tipo de correspondencia: Locución adverbial. La traducción se mantiene, tanto en la forma como en el sentido. Correspondencia formal y conceptual – Tema/Campo semántico: Legal, jurídico, político – Variantes en LO: A puerta cerrada; a puertas cerradas – Sinónimos: Secretamente; subrepticamente; de forma secreta; en privado; de forma confidencial; a solas; en secreto; extraoficialmente; reservadamente – Antónimos: Abiertamente; públicamente; manifiestamente; oficialmente – Registro: Formal e informal. En este contexto, formal – Pertenencia a lenguajes de especialidad: Similares a los de la lengua origen – Grado de creatividad metafórica: Similar al de la lengua origen – Fuentes: RAE; Diccionario Reverso en línea; Diccionario del Español Jurídico (DEJ) en línea; Gran Diccionario Oxford Inglés/Español, Español/Inglés (en papel) 	
Unidad fraseológica en lengua meta (FR) <i>Derrière des portes fermées</i>	
Fuente del término meta	Diccionario Oxford Inglés/Francés, Francés/Inglés (en papel)
Definición	‘Débats judiciaires hors de la présence du public. (Le huis clos peut être ordonné par le juge lorsque les débats présentent un danger pour l’ordre public ou les bonnes mœurs)’

Fuente de la definición	Diccionario Larousse
Ejemplo de uso de la UF en LM	Enfin, nous assistons ici, non à processus constitutionnel démocratique, mais à un jeu de pouvoirs qui se déroule derrière des portes fermées.
Fuente	Debate del Parlamento Europeo. Enlace
Observaciones	
<ul style="list-style-type: none"> – Tipo de correspondencia: Locución adverbial. La traducción se mantiene, tanto en la forma como en el sentido. Correspondencia formal y conceptual – Tema/Campo semántico: Legal, jurídico, político – Variantes en LO: À portes fermées – Sinónimos: À huis clos; toutes portes fermées; sans que le public soit admis; en petit comité; en secret – Antónimos: Publiquement; officiellement; de manière manifeste; partager des informations; communautairement; sans restriction – Registro: Formal e informal. En este contexto, formal – Pertenencia a lenguajes de especialidad: Similares a los de la lengua origen – Grado de creatividad metafórica: Similar al de la lengua origen – Fuentes: P. de Rieux, A. (2006): <i>Dictionnaire des idiomes anglais et américains</i>. Paris: Publibook, p. 50. Diccionario Larousse; Lexilogos; ABC Thesaurus; Diccionario Reverso en línea; Diccionario Oxford Inglés/Francés, Francés/Inglés (en papel) 	

Figura IV. Modelo de ficha fraseológica (EN, FR, ES) siguiendo el modelo propuesto en Serrano Lucas (*ut supra*)

3. Resultados del análisis

Tras llevar a cabo un extenso análisis, con una fase analítica similar a la que hemos esbozado en la muestra, pero con la totalidad de discursos que componen el corpus y ahondando más en las características lingüísticas que definen a las UFS, hemos obtenido los resultados siguientes:

– Uso de UFS por cada lengua de trabajo

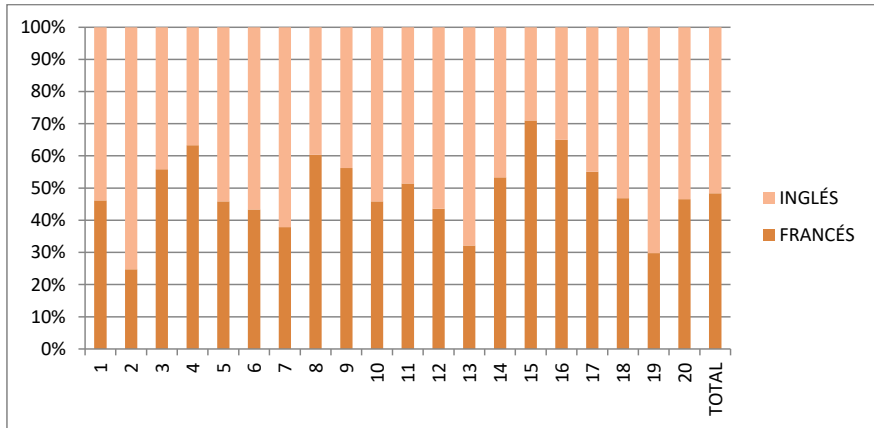


Figura V. Uso de las UFS por cada lengua de trabajo (EN, FR)

– Porcentaje de uso referido a las locuciones, colocaciones y pemiias

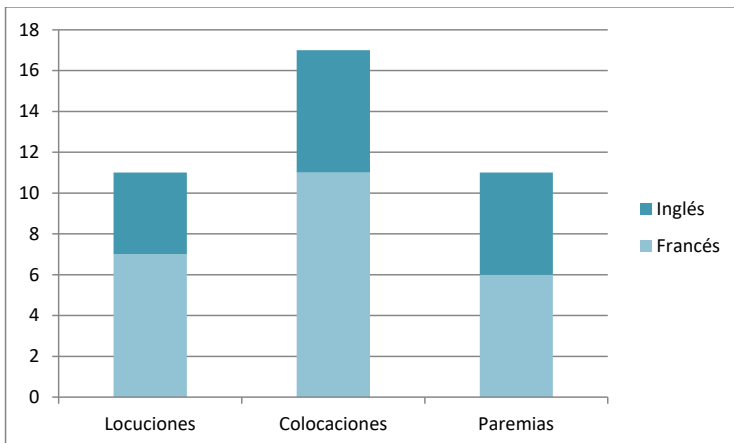


Figura VI. Uso de locuciones, colocaciones y pemiias (EN, FR)

– Porcentaje relativo al nivel semántico de las UFS: ontológicas o culturales

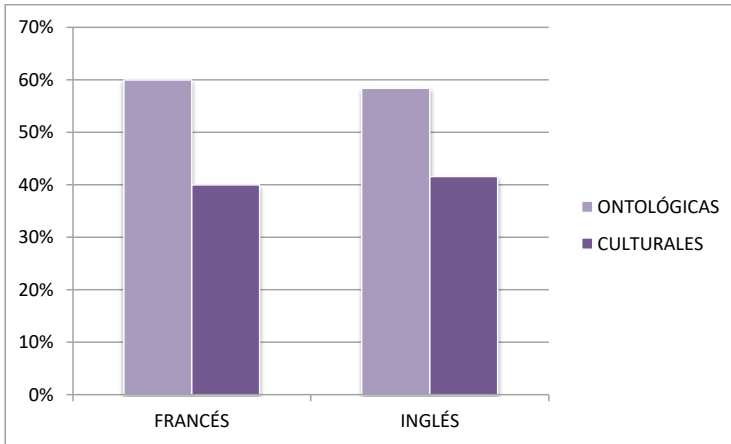


Figura VII. UFS relativas al nivel semántico (EN, FR)

4. Propuesta de glosario trilingüe de unidades fraseológicas halladas en discursos políticos (EN, FR, ES)

El análisis llevado a cabo en esta investigación nos ha permitido delimitar una búsqueda documental muy amplia, con el fin de valorar la cantidad de UFS resultantes del corpus en EN y en FR compilado, para este trabajo y proceder con la búsqueda de una solución traductológica apropiada para las lenguas metas correspondientes. La utilización de fichas fraseológicas ha facilitado el análisis de cada una de las UFS objeto de estudio, y ha simplificado la tarea de documentación para la disposición del glosario trilingüe final. Antes de entrar en materia, nos gustaría aclarar que existen, en muchos casos, otras opciones posibles de traducción aplicadas a otros contextos, pero la propuesta que hemos añadido en esta ocasión está adaptada, según nuestro mejor criterio, al marco comunicativo del texto origen de donde procede cada UF. En cuanto a la clasificación formal de estas, como podrá comprobarse, se han seguido de cerca los criterios expuestos con anterioridad de Corpas Pastor

(1997), Durán y Manjón (2002) y Montoro del Arco (2007). Ese ha sido el principio seguido para incluir las UFS que componen el corpus de trabajo. De este modo, podrán encontrarse principalmente todos aquellos paradigmas o modelos de UFS que se conciben como colocaciones (con modelos de colocación que siguen cualquiera de los patrones anteriormente mencionados, tales como: verbo + sintagma nominal (*porter des risques*, por ejemplo); verbo + sintagma preposicional, (*to die in vain*); locuciones (por ejemplo, *en lo referente a*, cuyo equivalente en francés [TO] sería *à l'égard de*, y cuya propuesta de traducción al inglés sería la de *with regard to*); unidades fraseológicas ontológicas y culturales (*the House of Commons*, por ejemplo). Este término supone un ejemplo de unidad terminológica indisociable, puesto que la realidad a la que alude carecería de sentido si se separaran cualquiera de los dos términos que la componen esta unidad. Además, forma parte de una realidad establecida culturalmente, ya que hace referencia a un término propio de la jurisdicción y política de Reino Unido, por lo que el propio término ya ayuda a contextualizar la cultura origen. Una vez terminada la investigación, se dispondrán todas las UFS (fichas fraseológicas y glosario) en un portal web que permitirá acceder a toda la información formal y semántica, así como a las diferentes propuestas de traducción (directa e inversa) de cada una de estas unidades, con la posibilidad de acceder a ellas mediante diferentes criterios de búsqueda. Como ya justificamos con anterioridad, por cuestiones de espacio no nos será posible detallar todas y cada una de las UFS localizadas en el corpus de trabajo, por lo que tan solo se han dispuesto aquellas que suscitan mayor interés y con una frecuencia de repetición más alta. Las UFS y las traducciones hacia los otros dos idiomas proceden de nuestro corpus de trabajo trilingüe.

Término en INGLÉS (EN)	Fuente (TO)	Término en FRANCÉS (FR)	Término en ESPAÑOL (ES)
A call to bear arms	EN-11	Un appel à porter les armes	Pedir que se levanten las armas
A failure of will	EN-18	Manque de volonté	Falta de voluntad
A new birth of freedom	EN-1	Une nouvelle liberté	Un nuevo concepto de libertad
All over again	EN-3	Encore une fois	De nuevo
Along with our allies	EN-6	Une collaboration avec nos alliés	Junto a nuestros aliados
At the forefront of	EN-2	À l'avant-garde	Al frente
At the highest level	EN-9	Au plus haut niveau	Al máximo
Behind closed doors	EN-2	Derrière des portes fermées	Detrás de puertas cerradas
Beyond a shadow of a doubt	EN-3	Sans aucun doute	Sin lugar a dudas
By our presence	EN-7	Avec notre présence	Con nuestra presencia
Cold War	EN-16	Guerre Froide	Guerra Fría
Common ground	EN-2	Un terrain d'entente	Punto de vista común
Drive someone's apart	EN-2	Mettre quelqu'un à l'Écarte	Apartar a alguien
Each day since	EN-5	Chaque jour depuis	Cada día desde entonces
Ever since	EN-1	Dès lors	Desde entonces
Follow the Constitution	EN-2	Respecter la Constitution	Respetar la Constitución
For over a	EN-2	Pour plus de	Durante más de un
For the last days/centuries	EN-6	Pour les derniers jours/siècles	Durante los últimos días/siglos
God bless America	EN-3	Dieu bénisse l'Amérique	Dios bendiga América ⁴

4. En español, se ha registrado una frecuencia mayor para traducir esta expresión protocolaria característica del epílogo del discurso por "Dios bendiga América". Sin embargo, debemos tener en cuenta que aunque se trata de la expresión probablemente más

God bless you	EN-3	Dieu vous bénisse	Dios te bendiga
House of Commons	EN-4	Chambre des communes	Cámara de los Comunes
Human beings	EN-2	Êtres humains	Seres humanos
In my belief	EN-2	Selon moi	En mi opinión
In the aftermath	EN-12	Par la suite	Después de esto
In the heights of victory	EN-5	Au sommet de la victoire	En la cumbre de la victoria
In their heart of hearts	EN-12	Au fond de leur cœur	En el fondo (de su corazón)
It is no coincidence that	EN-18	Ce n'est pas une coïncidence que	No es ninguna coïncidencia que
Let (me/us) begin (tonight/today) by	EN-18	Commencer (ce soir/aujourd'hui) par	Avanzar/Ponerse en marcha
Let (us) go forth to	EN-11	Aller vers/Avancer	Centrémonos en
March to the tune of pressure groups	EN-2	Marcher au rythme des groupes de pression	Seguir de cerca a los grupos de presión
Nuclear attack	EN-2	Attaque nucléaire	Ataque nuclear
Open mind	EN-15	Esprit ouvert	Mente abierta
Out of secondary	EN-7	Aller droit au but	Ir al grano
Outside the stockade	EN-4	En dehors de la palissade	Salir de la estacada
Tax burden	EN-18	Charge fiscale	Carga fiscal
The American dream	EN-5	Le rêve américain	El sueño americano
The best I can	EN-2	Le mieux que je puisse faire	Lo mejor que pueda
The direction we want to move in	EN-16	La direction dans laquelle nous voulons aller	Es a donde queremos llegar
The mounting arms race	EN-16	La course aux armements	La carrera armamentista

extendida, en realidad el término en inglés “America”, en este caso, considerando en todo momento el contexto de los discursos del corpus, se refiere a los “Estados Unidos” como país y no a “América” como continente.

The need will be upon us fast	EN-4	Le besoin sera vite là	La necesidad caerá sobre nosotros rápidamente
The pain falls drop by drop upon the heart	EN-17	La douleur tombe goutte à goutte sur le cœur	El dolor cae gota a gota sobre el corazón
The tactics of terror	EN-9	Les tactiques de la terreur	Las tácticas del terror
The Taliban regime	EN-9	Le régime Taliban	El régimen talibán
The trumpet summons us again	EN-11	La trompette nous convoque à nouveau	Ya suenan de nuevo los tambores
There is no doubt	EN-2	Il n'y a aucun doute	No cabe duda
To achieve political emancipation	EN-2	Parvenir à l'émancipation politique	Lograr la emancipación política
To be created equal	EN-1	Être créé égal	Ser iguales
To be honoured to be	EN-2	Avoir l'honneur d'être	Es un honor...
To be imposed upon (one nation)	EN-2	Être imposé à (une nation)	Imponerse a (una nación)
To be in a fight	EN-9	Être dans un combat	Estar en medio de una pugna
To be made in peacetime	EN-12	Être fait en temps de paix	Hacer algo en tiempos de paz
To be put on notice	EN-9	Être mis en garde	Ser informado de algo
To be singled out for unfair treatment	EN-9	Être distingué pour un traitement inéquitable	Ser señalado por un trato injusto
To become a Democrat	EN-9	Devenir démocrate	Hacerse demócrata
To breed fear	EN-10	Susciter la peur	Promover el miedo
To cause grave danger	EN-8	Causer un grave danger	Causar un grave peligro
To combat malaria	EN-15	Lutter contre le paludisme	Combatir la malaria
To comply with EU regulations	EN-4	Se conformer aux règlements de l'UE	Respetar la legislación de la UE
To curb the size and influence of	EN-3	Réduire la taille et l'influence des	Reducir el tamaño y la influencia de

To develop procedures	EN-1	Développer des procédures	Desarrollar procedimientos
To die in vain	EN-15	Décéder en vain	Morir en vano
To each other	EN-2	Les uns envers les autres	Entre nosotros
To embark on the consideration	EN-12	Entamer la réflexion	Llegar a considerar algo
To engage in violence against	EN-3	S'engager dans la violence contre	Promover/Ejercer la violencia contra
To expand health care coverage	EN-14	Étendre la couverture des soins de santé	Ampliar la cobertura en materia de salud
To experience the oppression of	EN-9	Faire l'expérience de l'oppression des	Experimentar la opresión de
To forge a grand alliance	EN-20	Forger une grande alliance	Forjar una gran alianza
To form a new administration	EN-16	Former une nouvelle administration	Consolidar una nueva administración
To go beyond these rather difficult times	EN-3	Aller au-delà de ces temps plutôt difficiles	Superar estos tiempos tan difíciles
To go into politics	EN-3	Faire de la politique	Entrar en política
To joke someone's only purpose	EN-10	Plaisanter sur le seul but de quelqu'un	Burlarse del propósito de alguien
To keep in mind	EN-11	Ne pas oublier que	Tener en cuenta
To look back on the victory	EN-12	Se souvenir de la victoire	Recordar la victoria
To look the urgency of the moment	EN-11	Valoir l'urgence du moment	Valorar la urgencia del momento
To lose the ability	EN-11	Perdre la capacité	Perder la capacidad
To lose the institutional memory	EN-13	Perdre la mémoire institutionnelle	Perder la memoria institucional
To meet in the aftermath of	EN-2	Se rencontrer à la suite de	Reunirse después de
To move forward	EN-2	Avancer	Mirar hacia adelante
To promote peace	EN-11	Promouvoir la paix	Fomentar la paz

To rejoice at our success	EN-9	Se réjouir de notre succès	Alegrarse de los éxitos
To renew the pledge of support	EN-7	Renouveler la promesse de soutien	Renovar la promesa de apoyo
To respect someone's faith	EN-12	Respecter la foi de quelqu'un	Respetar la fe de alguien
To stand as a beacon of hope	EN-2	Représenter la lueur d'espoir	Representar la esperanza
To stand for civil rights	EN-3	Défendre les droits civils	Defender los derechos civiles
To stand in line	EN-7	Faire la queue	Hacer cola
To support the victims	EN-9	Soutenir les victimes	Apoyar a las víctimas
To sustain the state	EN-20	Soutenir l'État	Mantener el estado
To take defensive measures	EN-9	Prendre des mesures défensives	Tomar medidas defensivas
To take some steps	EN-19	Prendre quelques mesures	Tomar medidas
To take sth by the heart(s)	EN-7	Prendre qq'ch au sérieux	Tomarse algo seriamente
To take tax exemptions away	EN-3	Supprimer les exonérations fiscales	Eliminar las exenciones de impuestos
To take an oath	EN-18	Prêter serment	Prestar juramento
To violate the rights	EN-15	Violer les droits	Violar los derechos
To wallow in the valley of despair	EN-14	Se vautrer dans la vallée du désespoir	Revolcarse en el valle de la desesperación (cit. lit.)
To welcome all of you	EN-19	Vous accueillir tous	Darles la bienvenida a todos
To work for an election	EN-13	Travailler pour une élection	Preparar las elecciones
With regard to	EN-16	En ce qui concerne/À l'égard de	Con respecto a/En lo referente a

Figura VIII. Glosario trilingüe (EN, FR, ES) a partir de las UFS en inglés (TO: EN)

Término en FRANCÉS (FR)	Fuente (TO)	Término en INGLÉS (EN)	Término en ESPAÑOL (ES)
À juste titre	FR-17	Quite rightly	Con justa razón
À très court terme	FR-3	In the very short term	A corto plazo
À un rythme soutenable	FR-3	At a sustainable pace	A un ritmo sostenible
Aborder une question fondamentale	FR-7	To address a fundamental issue	Abordar una cuestión importante
Afficher des taux de ...	FR-7	To display the rates of	Mostrar las tasas de
Apprendre au cours de quelque temps	FR-10	To learn over time	Ir aprendiendo con el tiempo
Au cours de cette conférence	FR-14	During this conference	Durante esta conferencia
Avec la plus grande franchise	FR-5	Frankly	Con total franqueza
Avoir la tentation de faire qq'ch	FR-8	To be tempted to do sth	Tener la tentación de hacer algo
Avoir tous les atouts pour faire qq'ch	FR-18	To have all the assets to get something	Tener ventajas para lograr algo
Baisser la pression fiscale	FR-5	To low the tax burden	Reducir la carga fiscal
Capacité à construire ensemble un monde meilleur	FR-2	To get the ability of building a better world together	Ser capaces de crear un mundo mejor entre todos
Ce n'est pas un hasard	FR-9	It is no accident that	No es ninguna casualidad que
Compter sur les compatriotes	FR-4	To rely on compatriots	Contar con los compatriotas
Constater à regret que	FR-6	To notice regretfully that	Observar con pesar que
Consulter le bilan	FR-8	To consult the balance sheet	Consultar el balance de los resultados

Contribuer au rayonnement de la France	FR-6	To contribute to France's influence	Promover la influencia de Francia
Depuis lors	FR-2	Since then	Desde entonces
Des mesures concrètes	FR-6	Concrete measures	Medidas concretas
Dissimuler l'ampleur de qq'ch	FR-2	To conceal the magnitude of sth	Ocultar la magnitud que tiene algo
En matière de	FR-3	In terms of	En materia de
En quelque sorte	FR-13	In other words	Por así decirlo
La hausse du chômage	FR-5	The rise of unemployment	El aumento de la tasa de desempleo
Éprouver la moindre complaisance	FR-2	To feel the slightest sense of complacency	Sentir la más mínima sensación de complacencia
Établir des relations	FR-6	To build a relationship	Establecer una relación
Être à l'appui des efforts menés	FR-2	To support the efforts of	Apoyar las medidas
Être au cœur de qq'ch	FR-2	To be at the heart of sth	Encontrarse en el centro de algo
Être conscient que	FR-2	To be aware of	Ser consciente de algo
Être d'accord avec	FR-3	To agree over something	Estar de acuerdo con
Être sur l'estrade	FR-3	To be on stage	Estar en el ruedo
Être un honneur	FR-4	To be an honour	Ser un honor
Exposer la vision de qq'ch	FR-3	To present the vision of sth	Exponer los argumentos sobre algo
Faires des progrès	FR-7	To make progress	Progresar
Faire reculer	FR-1	To back up	Retroceder
Faire valoir les lignes rouges	FR-15	To bring out the red lines	Tener en cuenta las líneas rojas
Favoriser la création de nouveaux emplois	FR-5	To promote the creation of new jobs	Promover la creación de nuevos empleos

Gérer le budget pour qq'ch	FR-3	To manage the budget	Gestionar presupuestos
Il reste tant à faire	FR-8	There's so much left to do	Quedan muchas cosas por hacer
Je vous remercie	FR-2	Thank you very much	Gracias
La mission que j'ai confié à	FR-6	To entrust a mission to	Confiar una misión a
La première urgence, c'est de...	FR-6	The first emergency is	La primera necesidad que tenemos es
Les règles de la concurrence	FR-5	The rules of competition	Las reglas de la competencia
Mettre à mal	FR-4	To harm	Dañar
Mettre beaucoup d'énergie à	FR-7	To put a lot of energy into	Gastar esfuerzos en
Mettre en doute	FR-2	To put sth in doubt	Poner en duda algo
Mettre fin au discours	FR-10	To get to the end of the speech	Acabar el discurso
Mettre des dollars sur la table	FR-6	To lay dollars down on the table	Poner dinero sobre la mesa
Montrer de l'intérêt pou	FR-5	To show an interest in	Mostrar interés por
Mutations économiques	FR-3	Economic changes	Cambios económicos
Nous comptons pour cela avec l'appui de...	FR-10	To count on the support	Contar con el apoyo de
Ouvrir la voie aux entreprises	FR-13	To pave the way for business	Abrir camino para los negocios
Partir du principe que...	FR-7	To assume that	Partir del principio de que
Porter dans le monde nos valeurs	FR-4	To bring our values to the world	Mostrar al mundo nuestros valores
Porter des risques	FR-6	To carry/involve risks/Be risky	Correr riesgos
Poursuivre l'objectif	FR-2	To pursue the goal	Perseguir el objetivo

Prendre des responsabilités	FR-4	To take responsibility	Asumir responsabilidades
Prendre l'initiative	FR-5	To take the initiative	Tomar la iniciativa
Relever les défis en matière de productivité	FR-18	To meet productivity challenges	Responder a los desafíos de la productividad
Rendre hommage à qq'ch	FR-10	To pay tribute to	Rendir homenaje a
S'inscrire dans la durée	FR-5	To be maintained over time	Durar mucho tiempo
Sans merci	FR-2	Mercilessly	Sin piedad
Savoir emprunter le chemin de la coopération	FR-12	To take the path for cooperation	Iniciar el camino de la cooperación
Se mettre en rapport avec qq'un	FR-1	To keep in touch with somebody	Ponerse en contacto con alguien
Se trouver en territoire étranger	FR-1	To be in foreign territory	Encontrarse en tierras extrañas
Souhaiter au fond de lui-même	FR-16	To wish deep down oneself	Desear algo en el fondo de uno mismo
Subsister un risque	FR-16	To remain a risk	Quedar/Haber todavía riesgos
Sur la scène internationale	FR-12	On the international scene	Internacionalmente
Susciter des inquiétudes	FR-3	To raise concerns	Aumentar las preocupaciones sobre algo
Témoigner de la confiance	FR-11	To demonstrate confidence	Demostrar confianza
Tenir la mer et continuer la lutte	FR-1	To weather the storm and the battle	Capear el temporal y la lucha
Travailler sur des sujets	FR-5	To work on a topic	Centrarse en un tema
Traverser un moment de doutes	FR-17	To go through a moment of doubt	Pasar por un momento de dudas

Figura IX. Glosario trilingüe (FR, EN, ES) a partir de las UFS en francés (TO: FR)

5. Conclusiones y nuevos retos de investigación

Como hemos podido comprobar, los rasgos generales que caracterizan a las UFS también se han localizado en las UFS propias del lenguaje político, concretamente, en discursos políticos. A tenor de los resultados obtenidos tras el análisis del corpus, observamos que la lengua que hace mayor alarde de UFS (según el corpus de trabajo) sería la lengua inglesa, con un total de hasta 2418 UFS halladas en el corpus en inglés, frente a las 2265 encontradas en el corpus en francés. No obstante, debemos considerar que estos resultados solo pueden aplicarse a la representatividad de este corpus de trabajo, por lo que estos datos son meramente orientativos y no pretenden forjar aquí un patrón definitivo. Evidentemente, estos resultados responden al corpus de trabajo seleccionado, pero deben reconocerse también algunas limitaciones: la longitud de cada discurso (que aunque se ha seleccionado de un modo similar en todos los discursos, no resulta fácil encontrar discursos con exactamente el mismo número de palabras); el contexto (formal o informal); etc. Por su parte, la lengua que ofrece un mayor porcentaje de UFS en lo referido a locuciones sería el francés (29 %), seguido nuevamente del francés para las colocaciones (46 %) y, finalmente, del inglés, con un 33 % de uso de paremias. En lo referente a UFS ontológicas y culturales, ahondando ahora en el terreno semántico, diremos que se ha encontrado, en general, una tendencia al empleo de las UFS ontológicas por encima de las culturales, con una representatividad del 58,4 % en inglés y un 60 % en francés para las primeras, y un 41,6 % en inglés y un 40 % en francés para las segundas. En general, se ha observado que el uso de las UFS culturales prolifera en el caso de discursos políticos que se han pronunciado en un contexto nacional, donde la función connotativa (Navarro, *ibid.*) es mucho más clara: la cercanía con el auditorio, el hecho de compartir una misma lengua y cultura y la formalidad que adquiere el evento en el que se pronuncia el discurso pasarían a ser, desde nuestro punto de vista, los factores fundamentales para la utilización de UFS culturales en el discurso político. En el caso de discursos pronunciados en contextos internacionales, aunque también suelen aparecer UFS culturales, el número disminuye drásticamente en comparación con discursos pronunciados en otros ámbitos mucho más familiares para el orador. No obstante, y a la luz de estos resultados, sería conveniente analizar si esta tendencia puede establecerse de

manera global según el par de lenguas o si, por el contrario, estriba en un uso característico que depende en exclusiva del orador político/a del que se trate.

Llegados a este punto, conviene esgrimir cuáles podrían ser los próximos retos de investigación:

- a. Incluir otro tipo de variables a raíz de esta investigación, tales como sexo, edad, país de procedencia, lugar de pronunciamiento del discurso, formalidad del acto en que se pronuncia, etc.
- b. Aumentar nuestro análisis de modo que se muestren las variables que se han tenido en cuenta para este estudio, más todas las que nos gustaría añadir: el número de UFS esgrimidas por cada discurso político que ha conformado el corpus, la estimación del porcentaje en relación con UFS ontológicas o culturales por cada discurso, las funciones llevadas a cabo en cada uno (connotativa, icónica, fraseológica), etc.
- c. Estudiar todas estas variables de uso de la palabra en el discurso político centrándonos en las UFS, pero concretando la utilización de las mismas en función de cada orador político/a, para conocer las tendencias que marcarían la media y la moda a la hora de establecer resultados definitivos de cara a cada político/a de nuestro corpus.
- d. Recopilar la totalidad de UFS, una vez aplicadas todas las variables, para terminar de completar el glosario en línea que se dispondrá en el portal web.

En definitiva, gracias al presente estudio, podemos afirmar que la traducción de UFS en el discurso político se concibe como una tarea de traslación complicada, ya que hay que considerar todos los parámetros culturales que envuelven a dicha UF y conocer cuál es la función que desempeña (ontológica y/o cultural) en el discurso político, para lo cual es indispensable estudiar la UF en su contexto y localizar una expresión equivalente en LM que garantice, además, que se respeta la función retórica y el sentido del mensaje del orador en LO. Para ello, el trabajo documental previo con fichas fraseológicas facilita notablemente esta labor, ya que nos permite acercarnos a la UF en cuestión desde un punto de vista tanto lingüístico como cognoscitivo, ahondando en las dificultades de traducción relacionadas con el trasvase cultural del discurso político del que se trate. El glosario ofrecido supone la consumación de todo el proceso descrito. Por su parte, el modelo de análisis fraseológico

propuesto resulta pertinente para asegurar la calidad en la traslación del mensaje como tarea previa al acto de traducción en sí, y su objetivo es garantizar que se respeta tanto el sentido y función de la UF del TO en el TM como la función discursiva perseguida por el orador en el discurso político en la LO.

Finalmente, confiamos en que la muestra de glosario político en las tres lenguas resultante de nuestra investigación sea un buen recurso documental y traductológico para especialistas, lingüistas y traductores/as.

Referencias bibliográficas

- ARCE CASTILLO, Ángela. (2006) *El lenguaje político. Recursos pragmático-discursivos en registros formales e informales*. Salamanca: Ratio Legis.
- ARELLANO, Frank, Natalie Vielma y Alexyca Carrero. (2013) "Fraseologías en el discurso político de columnistas de opinión venezolanos." *Lengua y habla*, 17, pp. 211-218.
- BRAZO MILLÁN, Ana Isabel. (2018) "La enseñanza-aprendizaje de expresiones idiomáticas francés-español a través de la traducción." En: García Peinado, Miguel Ángel e Ignacio Ahumada Lara (eds.) 2018. *Traducción literaria y discursos traductológicos especializados*. Berlín: Peter Lang.
- CORPAS PASTOR, Gloria. (1997) *Manual de Fraseología Española*. Madrid: Gredos.
- CORPAS PASTOR, Gloria. (2003) *Diez años de investigación en fraseología: estudios sintáctico-semánticos, contrastivos y traductológicos*. Madrid: Iberoamericana-Vervuert.
- COSERIU, Eugenio. (1987) "Lenguaje y Política." En: Alvar, Manuel (ed.) 1987. *El lenguaje político*. Madrid: Fundación Friedrich Ebert, Instituto de Cooperación Iberoamericana, pp. 9-31.
- FABBRI, Paolo y Marcarino, Aurelia. (2002) "El discurso político." *DeSignis*, 2, pp. 17-32.
- FERNÁNDEZ LAGUNILLA, Marina. (1999a) *La lengua en la comunicación política I: El discurso del poder*. Madrid: Arco Libros.
- FUENTES RODRÍGUEZ, Catalina y Esperanza Alcaide Lara. (2002) *Mecanismos lingüísticos de la persuasión*. Madrid: Arco Libros.
- GARCÍA ROMERO, Marina y Julia Sevilla Muñoz. (2005) "La enseñanza de paremias en contexto y su traducción a través de su presencia en las 'fables' de La Fontaine." *Anales de filología francesa*, 13, pp. 131-146.

- GARCÍA YELO, Marina. (2005) "Consideraciones sobre la didáctica de las expresiones idiomáticas francesas a invidentes." *Paremia*, 14, pp. 61-70.
- GARCÍA-PAGE SÁNCHEZ, Mario. (2008) *Introducción a la fraseología española*. Barcelona: Anthropol.
- GUTIÉRREZ DÍEZ, Francisco. (1995) "Idiomaticidad y Traducción." *Cuadernos de Filología Inglesa*, 4, pp. 27-42.
- HIGUERAS GARCÍA, Marta. (2009) "Aprender y enseñar léxico". En: Miquel, Lourdes y Neus Sans (coords.) *Didáctica del Español como Lengua Extranjera* 3. Madrid: Fundación Actilibre, pp. 111-126.
- HURTADO ALBIR, Amparo. (2001) *Traducción y Traductología. Introducción a la Traductología*. Madrid: Cátedra.
- LOZANO, Carlos Wenceslao. (1992) "Aproximación al problema de las expresiones idiomáticas y su traducción." *Sendebarr*, 3, pp. 141-156.
- LUQUE DURÁN, Juan de Dios y Francisco José Manjón Pozas. (2002) "Claves culturales del diseño de las lenguas. Fundamentos de tipología fraseológica." *Estudios de Lingüística del Español*, 16 [en línea]. Recuperado de: <http://elies.rediris.es/elies16/Claves.html>. [Última consulta: 14.09.2019].
- MARTÍNEZ MARÍN, Juan. (1996) *Estudios de fraseología española*. Málaga: Ágora.
- MESSINA FAJARDO, Luisa. (2016) *El lenguaje político*. Milano: Maggioli Editore.
- MOGORRÓN HUERTA, Pedro. (2014) "Las expresiones fijas diatópicas argentinas y mexicanas." En: González Rey, María Isabel. (ed.) *Didáctica y Traducción de las Unidades fraseológicas*. Santiago: Universidad de Santiago de Compostela, pp. 77-98.
- MONTORO DEL ARCO, Esteban Tomás. (2007) "El porqué de los dichos: la fraseología del vino." *Per Abbat: boletín filológico de actualización académica y didáctica*, n.º 2, pp. 131-136.
- NAVARRO, Carmen. (2002) "La fraseología en el discurso político y económico de los medios de comunicación". En: Cusato, D.A. y L. Frattale (coord.). *Atti del XX Convegno dell' AISPI*. Messina: Andrea Lippolis, pp. 209-211.
- NÚÑEZ CABEZAS, Emilio Alejandro y Susana Guerrero Salazar. (2002) *El lenguaje político español*. Madrid, Cátedra.
- RADULESCU, Romana Anca. (2006) "Estudio contrastivo de expresiones idiomáticas en español, inglés y rumano." *Universidad Complutense* [en línea], pp. 111-128. Recuperado de: https://www.uam.es/gruposinv/upstairs/upstairs2/curricula/trabajos/radulescu_2006_estudio.pdf. [Última consulta: 25.09.2019].

- SERRANO LUCAS, Lucía Clara. (2010) “Metodología para la enseñanza de la fraseología en traducción: la ficha fraseológica como tarea final.” *Paremia*, 19, pp. 197-206.
- SEVILLA MUÑOZ, Julia. (1997) “Fraseología y Traducción.” *Thelème*, 12, pp. 431-440.
- SEVILLA MUÑOZ, Julia y Carlos Antonio CRIDA ÁLVAREZ. (2013) “Las paremias y su clasificación”. *Paremia*, 22, pp. 105-114.
- TANNEN, Deborah. (1999) *Cultura de la polémica. Del enfrentamiento al diálogo*. Barcelona: Paidós.
- TIMOFEEVA, Larissa. (2012) “Sobre la traducción fraseológica.” *ELUA*, 26, pp. 405-432.
- VAN DIJK, Teun Adrianus. (1996) “Análisis del discurso ideológico.” *Versión* 6, pp. 15-43.
- ZULUAGA, Alberto. (1997) “Sobre las funciones de los fraseologismos en textos literarios.” *Paremia* 6, pp. 631-640.
- ZULUAGA, Alberto. (1980) *Introducción al estudio de las expresiones fijas*. Frankfurt: Peter Lang.

NOTA BIOGRÁFICA / BIONOTA

MARÍA DEL CARMEN LÓPEZ RUIZ es doctoranda en Traducción e Interpretación por la Universidad de Córdoba (España). Ha obtenido un contrato predoctoral de Formación para el Profesorado Universitario (FPU) otorgado por el Ministerio de Universidades y desarrolla sus labores investigadoras y docentes correspondientes en el seno del Dpto. de Traducción e Interpretación, Filología Francesa, Estudios Semíticos y Documentación de esta misma universidad. Traductora y correctora en la combinación de lenguas EN/FR/ES. Sus líneas de investigación se centran en la Traducción del Discurso Político, Análisis del Discurso y Didáctica de la Traducción. Entre sus publicaciones recientes destacan *La TAV como herramienta de aprendizaje en el estudio del Análisis Crítico del Discurso para la traducción institucional y la comunicación política*, en J.J. Gázquez Linares et al., Dykinson, 2020 y *Discurso político, pragmática y retórica: un estudio contrastivo desde una perspectiva traductológica de discursos EN/ES de Barack H. Obama y Donald J. Trump*, en A. Bueno García, Comares, 2020.

MARÍA DEL CARMEN LÓPEZ RUIZ is a PhD student in Translation and Interpreting at the University of Córdoba (Spain). She has been awarded a Four-Year PhD Scholarship by the Spanish Ministry of Universities, and develops the related teaching and researching tasks within the Department of Translation and Interpreting, Romance Languages, Semitic Studies and Documentation Science at this university. EN/FR/ES translator and proofreader. Her research interests in Translation Studies encompass different domains: Translation of Political Discourse, Discourse Analysis and Didactics of Translation. Amongst her recent publications are *La TAV como herramienta de aprendizaje en el estudio del Análisis Crítico del Discurso para la traducción institucional y la comunicación política*, in J.J. Gázquez Linares *et al.*, Dykinson, 2020 and *Discurso político, pragmática y retórica: un estudio contrastivo desde una perspectiva traductológica de discursos EN/ES de Barack H. Obama y Donald J. Trump*, in A. Bueno García, Comares, 2020.

ANEXO: CORPUS DE TRABAJO

A. Corpus de discursos en inglés (TO)

N.º discurso	Fecha	Orador	Fuente TO (EN)
1	19.11.1863	Abraham Lincoln	http://www.abrahamlincolnonline.org/lincoln/speeches/gettysburg.htm . [Consulta: 28.04.19].
2	04.06.09	Barack Obama	https://www.nytimes.com/2009/06/04/us/politics/04obama.text.html . [Consulta: 28.04.19].
3	27.07.16	Bill Clinton	https://edition.cnn.com/2016/07/27/politics/bill-clinton-speech-transcript/index.html . [Consulta: 28.04.19].
4	14.02.18	Boris Johnson	https://www.spectator.co.uk/article/full-text-boris-johnson-s-brexiteer-speech . [Consulta: 29.04.19].
5	30.01.18	Donald Trump	https://www.whitehouse.gov/briefings-statements/president-donald-j-trumps-state-union-address/ . [Consulta: 29.04.19].

6	08.12.53	Dwight Eisenhower	https://www.eisenhowerlibrary.gov/sites/default/files/research/online-documents/atoms-for-peace/atoms-for-peace-draft.pdf . [Consulta: 30.04.19].
7	04.03.33	Franklin D. Roosevelt	https://avalon.law.yale.edu/20th_century/froos1.asp . [Consulta: 30.04.19].
8	13.11.90	Geoffrey Howe	https://www.britpolitics.co.uk/speeches-sir-geoffrey-howe-resignation/ . [Consulta: 30.04.19].
9	21.09.01	George W. Bush	https://www.theguardian.com/world/2001/sep/21/september11.usa13 . [Consulta: 12.05.19].
10	04.05.07	Harold Macmillan	https://www.theguardian.com/theguardian/2007/apr/24/greatspeeches . [Consulta: 30.04.19].
11	20.01.1961	John F. Kennedy	https://avalon.law.yale.edu/20th_century/kennedy.asp . [Consulta: 12.05.19].
12	03.07.82	Margaret Thatcher	https://www.margaretthatcher.org/document/104989 . [Consulta: 13.05.19].
13	16.01.14	Mark Drakeford	http://www.wales.nhs.uk/sitesplus/documents/863/Health%20Ministers%20speech%20to%20NHS%20Confederation%20Annual%20Conference.pdf . [Consulta: 30.04.19].
14	14.10.64	Martin Luther King	https://www.nobelprize.org/prizes/peace/1964/king/26142-martin-luther-king-jr-acceptance-speech-1964/ [Consulta: 10.12.64].
15	10.05.94	Nelson Mandela	https://www.sanews.gov.za/south-africa/read-nelson-mandelas-inauguration-speech-president-sa . [Consulta: 13.05.19].
16	23.09.60	Nikita Khrushchev	https://digitalarchive.wilsoncenter.org/document/155185.pdf?v=9f7ac7df82c2cf1162b9f845c67ef067 . [Consulta: 13.05.19].

17	04.04.68	Robert F. Kennedy	https://www.jfklibrary.org/learn/about-jfk/the-kennedy-family/robert-f-kennedy/robert-f-kennedy-speeches/statement-on-assassination-of-martin-luther-king-jr-indianapolis-indiana-april-4-1968 . [Consulta: 14.05.19].
18	20.01.81	Ronald Reagan	https://www.reaganfoundation.org/ronald-reagan/quotes-speeches/inaugural-address-1/ . [Consulta: 14.05.19].
19	13.07.16	Theresa May	https://www.washingtonpost.com/gdpr-consent/?next_url=https%3a%2f%2fwww.washingtonpost.com%2fnews%2fworldviews%2fwp%2f2016%2f07%2f13%2ffull-transcript-may-promises-bold-new-positive-role-for-britain-after-brexite%2f . [Consulta: 15.05.19].
20	13.05.40	Winston Churchill	https://winstonchurchill.org/resources/speeches/1940-the-finest-hour/blood-toil-tears-and-sweat-2/ . [Consulta: 15.05.19].

B. Corpus de discursos en francés (TO)

N.º discurso	Fecha	Orador	Fuente TO (EN)
1	18.06.40	Charles de Gaulle	http://www.charles-de-gaulle.org/wp-content/uploads/2017/03/Appel-%C3%A0-la-r%C3%A9sistance-sur-les-ondes-de-la-BBC-Londres.pdf . [Consulta: 11.05.19].
2	18.11.15	Dominique de Villepin	https://www.lefigaro.fr/politique/le-scan/2014/04/08/25001-20140408ARTFIG00066-le-discours-de-villepin-sur-l-irak-a-l-onu.php . [Consulta: 10.05.19].
3	21.11.2017	Édouard Philippe	https://www.echodescommunes.fr/info_155_discours-de-m-edouard-philippe-premier-ministre-prononce-au-congres-des-maires-de-france.html . [Consulta: 10.05.19].

4	14.05.17	Emmanuel Macron	https://www.elysee.fr/emmanuel-macron/2017/05/15/discours-d-investiture-du-president-de-la-republique . [Consulta: 10.05.19].
5	17.11.09	François Fillon	https://www.vie-publique.fr/discours/176986-declaration-de-m-francois-fillon-premier-ministre-sur-la-politique-de . [Consulta: 10.05.19].
6	30.08.16	François Hollande	https://www.vie-publique.fr/discours/200262-declaration-de-m-francois-hollande-president-de-la-republique-sur-les . [Consulta: 10.05.19].
7	30.05.2000	François Legault	https://www.cmm.ca/fr/nouvelles/speech-guest-speaker-mr-francois-legault-ministre-detat-a-leducation-et-a-la-jeunesse/ . [Consulta: 11.05.19].
8	25.10.83	François Mitterrand	https://unesdoc.unesco.org/ark:/48223/pf0000057371_fre . [Consulta: 11.05.19].
9	28.02.70	Georges Pompidou	https://www.georges-pompidou.org/sites/default/files/documents/dossier_avenir_discours_long.pdf . [Consulta: 11.05.19].
10	10.10.18	Gautier Mignot	https://co.ambafrance.org/Fete-nationale-Discours-de-Monsieur-Gautier-Mignot-Ambassadeur-de-France-en . [Consulta: 12.12.19].
11	14.03.97	Jacques Chirac	https://www.vie-publique.fr/discours/133796-discours-de-m-jacques-chirac-president-de-la-republique-sur-les-relat . [Consulta: 19.05.19].
12	05.09.12	Jean Charest	https://www.lesoleil.com/archives/discours-integral-de-la-demission-de-jean-charest-9be7c81803568ee88cd56d4d7b7a8ee7 . [Consulta: 19.05.19].
13	28.08.17	Jean-Yves Le Dirian	https://co.ambafrance.org/Discours-de-Jean-Yves-Le-Drian-Ouverture-Semaine-des-ambassadeurs-28-aout-2017 . [Consulta: 19.05.19].
14	12.12.15	Laurent Fabius	https://co.ambafrance.org/Discours-de-Laurent-Fabius-lors-de-la-cloture-de-la-COP-21 . [Consulta: 19.05.19].

15	06.09.35	Léon Blum	http://www.lours.org/archives/defaultd53d.html?pid=508 . [Consulta: 19.05.19].
16	10.11.18	Marine Le Pen	https://rassemblementnational.fr/discours/intervention-de-marine-le-pen-au-monument-de-la-victoire/ . [Consulta: 19.05.19].
17	27.08.07	Nicolas Sarkozy	https://es.ambafrance.org/XVeme-conference-des-ambassadeurs . [Consulta: 19.05.19].
18	21.05.14	Philippe Couillard	https://webcache.googleusercontent.com/search?q=cache:aCVL-cuhwI8J:https://www.bibliotheque.assnat.qc.ca/DepotNumerique_v2/AffichageFichier.aspx%3Fidf%3D146878+&cd=8&hl=es&ct=clnk&gl=uk . [Consulta: 19.05.19].
19	04.01.19	Ueli Maurer	https://www.admin.ch/gov/fr/accueil/documentation/discours/allocutions-nouvel-an/2019.html . [Consulta: 20.05.19].
20	27.05.74	Valéry Giscard d'Estaing	https://www.elysee.fr/la-presidence/l-investiture-de-valery-giscard-d-estaing . [Consulta: 20.05.19].

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	hombre (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	clamar (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 light
	light a fire under someone
	light as a feather
	the light 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	Not included
Diasystematical marking (colloquial, jargon, etc.)	Not included
Not included	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	Not included

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).

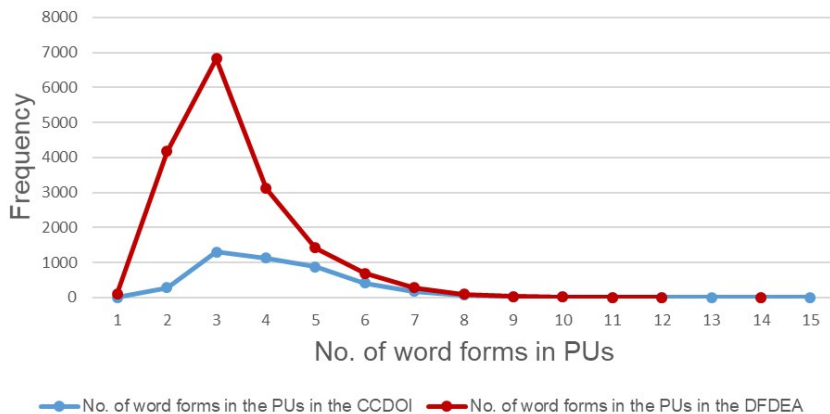


Figure 1. No. of word forms of the PUs in the DFDEA and the CCDOI.

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).

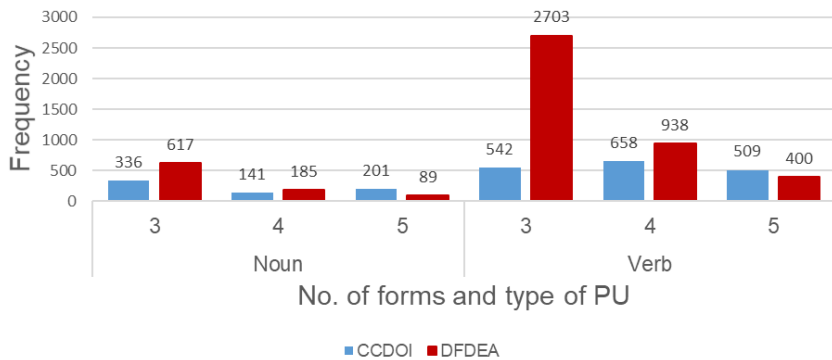


Figure 2. No. of nominal and verbal PUs in the analysis databases and their corresponding number of forms.

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).

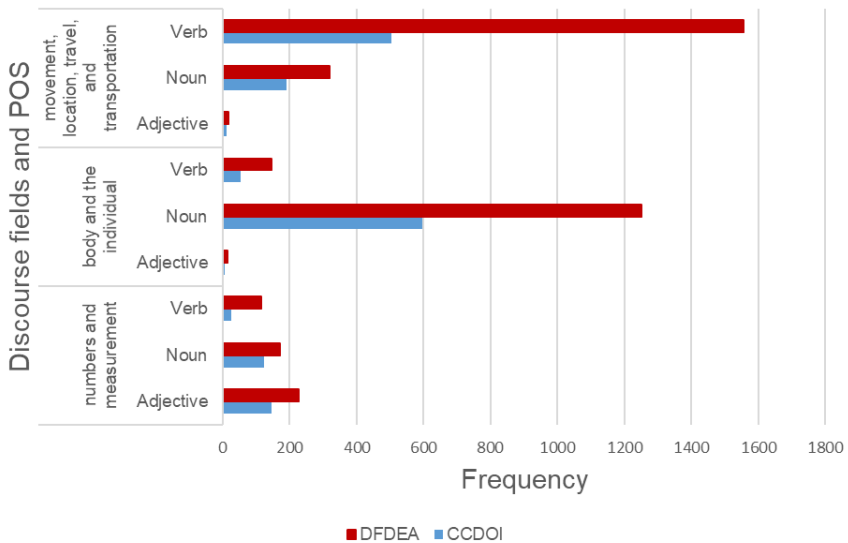


Figure 4. Distribution of main lexical categories (verbs, nouns, and adjectives) grouped by discourse fields.

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.

3.2. 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 %)
	Total DFDEA	420		1,365	
CCDOI	Noun	126	(31.8 %)	455	(35.3 %)
	Verb	270	(68.2 %)	835	(64.7 %)
	Total CCDOI	396		1,290	
<i>Total dictionaries</i>		816	(23.5%)	2,655	(76.5 %)
<i>Grand total</i>		3,471			

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).

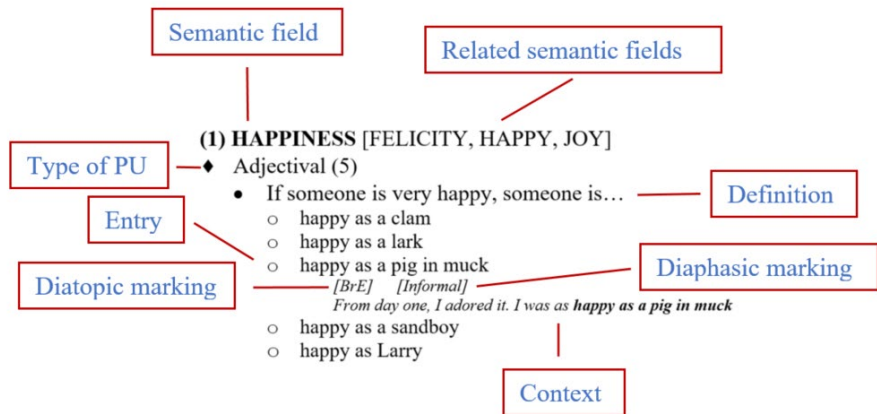


Figure 6. Example of an entry with an onomasiological/semasiological hybrid approach.

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 sentence "If you are happy as a pig in muck, you are very happy". The "Context:" text box contains the sentence "From day one I adored it. I was as happy as a pig in muck". At the bottom of the main area is an "OK" button. Below the main area is a footer 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>
- 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.

Recibido / Received: 14/11/2019

Aceptado / Accepted: 23/04/2020

Para enlazar con este artículo / To link to this article:

<http://dx.doi.org/10.6035/MonTI.2020.ne6.10>

Para citar este artículo / To cite this article:

Navarro-Brotons, Lucía. (2020) "Traducción de unidades fraseológicas gastronómicas francés-español peninsular/español de México." En: 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. 327-354.

TRADUCCIÓN DE UNIDADES FRASEOLÓGICAS GASTRONÓMICAS FRANCÉS-ESPAÑOL PENINSULAR/ ESPAÑOL DE MÉXICO

TRANSLATION OF GASTRONOMIC PHRASEOLOGICAL UNITS FRENCH-PENINSULAR SPANISH/MEXICAN SPANISH

LUCÍA NAVARRO-BROTONS

lucia.navarro@ua.es

Universidad de Alicante

Resumen

Cuando una misma lengua se habla en diversos países, las marcas dialectales, resultado de diferencias culturales, son más acusadas, pues cada variedad dialectal transmite la idiosincrasia de las distintas culturas. En este marco, el presente trabajo aborda la variación de las unidades fraseológicas especializadas del español gastronómico peninsular (EP) y el español de México (EM) utilizado en las recetas de libros y blogs de cocina usando como metodología la lingüística de corpus. Los objetivos perseguidos son: identificar las UFE en textos producidos en español de México (EM) y buscar las respectivas variantes en español peninsular (EP); analizar la variación de algunas de las UFE para reafirmar la importancia de la variación fraseológica en el ámbito de la traducción; esbozar un glosario de UFE contrastivo y bilingüe del EM, el EP y el francés destinado a especialistas, estudiantes y traductores gastronómicos.

Palabras clave: Variación fraseológica; Gastronomía; Variedad del español peninsular; Variedad del español de México; Fraseotraducción.



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

Abstract

When the same language is spoken in different countries, dialectal marks, the result of cultural differences, are more evident as each dialectal variety conveys the idiosyncrasies of different cultures. Within this framework, this paper deals with corpus linguistics methodology in order to study the variation of the gastronomic phraseological units of Peninsular Spanish and Mexican Spanish used in recipes from cooking books and blogs. The objectives pursued are: to identify the specialized phraseological units in texts produced in Mexican Spanish and to search their variants in Peninsular Spanish; to analyse the variation of some of the units in order to reaffirm the importance of phraseological variation in the field of translation; to outline a contrastive and bilingual glossary of Mexican Spanish, Peninsular Spanish and French for specialists, students and gastronomic translators.

Keywords: Phraseological variation; Gastronomy; Peninsular Spanish. Mexican Spanish; Phraseological translation.

1. Introducción

La variación es una realidad existente en cada lengua. Cuando esta, además, se habla en diversos países, las marcas dialectales, resultado de diferencias culturales, son más acusadas, pues cada variedad dialectal transmite la idiosincrasia de las distintas culturas. El español, presente en veintiún países, es una de las lenguas con mayor variedad lingüística. Cuenta de ello da el *Diccionario de americanismos* (2010), editado por la Asociación de Academias de la Lengua Española, que constituye un repertorio léxico que pretende recoger todas las palabras propias del español de América.

En este marco, el presente trabajo pretende abordar la variación de las unidades fraseológicas especializadas del español gastronómico peninsular (EP) y el español de México (EM) utilizado en las recetas de libros y blogs de cocina atendiendo a la necesidad de conocerlas para aplicarlas a la traducción francés-español.

Para entender mejor esta variación diatópica existente entre las dos comunidades lingüísticas estudiadas (México y España), el trabajo se apoya en tres elementos: 1) la corriente socioterminológica (Gaudin 2005) que comprende la terminología como una disciplina social en la que los diferentes factores socioculturales van a condicionar el uso de los términos; 2) la

Teoría Comunicativa de la Terminología (TCT) (Cabré 1999) que incluye la fraseología en su ámbito de estudio y reconoce la posibilidad de variación en las áreas especializadas y 3) la terminología cultural (Diki-Kidiri 2009) que sostiene que el componente cultural influye en la percepción de la realidad y, por ende, en la variación de los términos.

No se debe perder de vista que la gastronomía infiere marcas culturales de un pueblo, región o país. Como apunta González-Pastor (2012: 139) “los platos típicos de una región forman parte de sus costumbres más arraigadas y se transmiten de generación en generación”. Además, la gastronomía está presente en varias áreas de conocimiento como la literatura, la historia, la nutrición, etc. Nuestro estudio, cuya metodología es la lingüística de corpus, propone: 1) identificar las UFE en textos producidos en español de México y buscar las respectivas variantes en español peninsular; 2) analizar la variación de algunas de las UFE para reafirmar su importancia en el ámbito de la traducción; 3) esbozar un glosario de UFE contrastivo y bilingüe del EM, el EP y el francés destinado a especialistas, estudiantes y traductores gastronómicos. Todo ello mediante una perspectiva interdisciplinar que engloba la terminología y la fraseología desde el punto de vista de la variación.

2. Terminología, fraseología, variación y traducción

2.1. *Convergencia entre terminología y fraseología*

Cabré (2000) sitúa el término como objeto de estudio de la TCT. En la actualidad estos estudios se centran en el término y, más allá de este, en las unidades de conocimiento especializado (UCE). Las UCE se caracterizan por 1) ser unidades que representan y transmiten el conocimiento especializado; 2) estar formadas por un significado y un significante; 3) representar y fijar el conocimiento especializado y 4) transmitir el conocimiento especializado (cf. Cabré y Estopà 2005). Las UCE se pueden dividir en unidades del lenguaje artificial y unidades del lenguaje natural. A su vez, estas últimas se subdividen en 1) unidades morfológicas especializadas (UME); 2) unidades léxicas especializadas (ULE); 3) unidades fraseológicas especializadas (UFE) (que pueden ser nominales, verbales, adjetivales o adverbiales). Sin embargo, no se debe perder de vista que las ULE pueden ser monoléxicas o poliléxicas,

compuestas por dos o más vocablos, denominados también compuestos sintagmáticos (cf. Ruiz Gurillo 2002; Wojak 2006; Montoro 2008).

Cabe indicar que el estudio de las UFE en el ámbito de la terminología se ha estudiado durante décadas. Como señala Sevilla (2015: 240), Arntz y Picht (1989: 52-55) realizan una de las primeras consideraciones teóricas relacionadas con la fraseología especializada. Estos autores destacan que el lexicógrafo Schlomann incorporó de manera sistemática la fraseología en los diccionarios técnicos en la década de 1920. Desde ese momento, otros terminólogos como Pavel (1993); Lorente, Bevilacqua y Estopà (2002); Cabré y Estopà (2005) o Sanz (2011) han seguido trabajando en este campo. Sin embargo, “[...] desde la fraseología apenas se ha tratado el uso de las UF en las lenguas de especialidad [...]” (Sevilla 2015: 240).

Aguado de Cea (2007), por su parte, resalta la estrecha relación existente entre la fraseología y la terminología e indica que “Tanto los frasemas [...], como los términos [...] constituyen los nodos a partir de los que se estructura el conocimiento de un campo de especialidad” (2007: 54). La autora apunta también la falta de consenso en torno a los límites de la fraseología. Este ha sido un tema de discusión por excelencia del que surgen dos posturas que divergen al establecer los criterios de inclusión de las unidades fraseológicas (UF) en general y, por ende, las UFE en particular. Se trata, por una parte, de la concepción *ancha* de la fraseología y, por otra, de la concepción *estrecha* (cf. Ruiz Gurillo 1997; Corpas 2003 o García-Page 2004). Ambas concepciones se basan en el modelo centro-periferia (cf. Wojak 1988; Penadés Martínez 1999) establecido por la escuela de Praga. *Grosso modo*, la concepción *estrecha* solamente incluye en su objeto de estudio las UF dotadas de fijación e idiomatidad que no superan el sintagma. Su estudio se limita a lo que la escuela de Praga denomina *centro*. La concepción *ancha* de la fraseología, por su parte, abarca el centro y la periferia. De esta manera, considera también UF aquellas combinaciones de palabras sintagmáticas u oracionales que están provistas de fijación o de fijación e idiomatidad. Desde el punto de vista teórico puede parecer esclarecedor. Sin embargo, los límites entre el centro y la periferia, e incluso entre la periferia y el discurso libre, no son exactos, sino que, por el contrario, son maleables. Como apunta García-Page (2008: 38):

[...] uno de los más espinosos problemas de delimitación entre locuciones nominales y colocaciones (y compuestos) se plantea en la observación real de

ciertas combinaciones de difícil clasificación: lo que para unos es locución, para otros es compuesto o colocación.

Partimos de la base de que las UCE nominales son uno de los tipos predominantes en el ámbito gastronómico. En este sentido, uno de los principales escollos encontrados a la hora de abordar este trabajo gira en torno a la clasificación de estas UCE poliléxicas como locución, compuesto sintagmático o colocación. Los siguientes ejemplos *chile dulce*, *crema agria*, *frijol negro*, *jugo de limón*, *niño envuelto*, *cabellos de ángel* (EM); *comino molido*, *azúcar moreno*, *cebolla morada*, *brazo de gitano*, *cabello de ángel* (EP) muestran que, en su mayoría, carecen de idiomatidad. Los compuestos sintagmáticos y colocaciones, además, son dos tipos de unidades que se encuentran en la periferia entre la fraseología y la sintaxis.

La mayoría de UCE poliléxicas detectadas obedecen a las estructuras [N+adj.] y [N+prep.+N]. Estas estructuras responden a los esquemas prototípicos de determinadas locuciones nominales y son patrones regulares de los compuestos sintagmáticos y colocaciones (cf. García-Page 2008: 95). Siguiendo a Corpas Pastor (1996: 93) y Blasco Mateo (1999: 27, 54), optamos por tratarlos como UF desde una concepción *ancha* de la fraseología en la que la idiomatidad es un rasgo posible de la UF, pero no definitorio. Esta decisión trasladada al ámbito de la terminología, que nos ocupa en este trabajo, supone considerar las ULE nominales poliléxicas como UFE. Por su carácter monoléxico, desestimamos en este estudio las UME, unidades formadas a partir de la prefijación y la sufijación.

2.2. Variación de las unidades fraseológicas especializadas

Como explica García-Page (2008: 217) “[...] la variación, por su alianza ineluctable con la fijación, se convierte en un rasgo característico de la Fraseología [...]”. Centrándose en variantes geolectales o diatópicas, este autor, afirma que, en el caso del español, habida cuenta de los vastos dominios que ocupa, a veces tan distantes, el surgimiento de expresiones diversas para designar una misma realidad es una causa natural (2008: 253-254). Hace especialmente hincapié en que muchas de esas variantes serán fundamentalmente léxicas, en virtud de los distintos términos que se acuñan para nombrar los objetos oriundos que se convierten en índices socioculturales.

Por ejemplo: *maíz, calabaza* (EP) versus *elote, ayote* (EM) respectivamente. En este mismo sentido, Mogorrón apunta que el fenómeno de la variación diatópica pertenece a la geolingüística y que no es exclusivo del español (2018a: 111). Al igual que García-Page (2008), Mogorrón defiende que:

[...] los usuarios [...] que viven en determinadas zonas geográficas expresan los mismos conceptos en los actos de comunicación, pero con unas combinaciones plurilexicales que utilizan elementos léxicos o culturales propios a ese determinado país o incluso región, sumándole de esa forma un matiz local que no conocen o no utilizan las demás áreas geográficas. (2018a: 112)

Si atravesamos el puente y pasamos de la fraseología general a la especializada podríamos pensar, en primera instancia, que las teorías terminológicas clásicas no aceptarán la variación ni la sinonimia por comprender los términos como denominaciones especialmente constituidos para designar un concepto. Sin embargo, la práctica del trabajo terminológico, a partir de los años 90, ha tenido como finalidad adecuar la terminología a las realidades sociales, lingüísticas y culturales atendiendo a nuevas tendencias terminológicas. Trasladando todo lo anterior al ámbito especializado, partimos de la hipótesis de que existe una terminología de la variación. Esta variación, en el caso que nos ocupa, está motivada por la variación diatópica y es avalada por las corrientes socioterminológica (Gaudin 2005) y la terminología cultural Diki-Kidiri (2009). Presuponemos 1) la existencia de variación denominativa (VD) (Freixa 2002, 2005), es decir, formas alternativas para un mismo concepto, por ejemplo, *aceite de colza* (EP) y *aceite de canola* (EM) y 2) la existencia de variación conceptual (VC) (Ciapuscio 1993; Cabré 1998; Freixa 2002, 2005), es decir, conceptos alternativos para una misma forma dentro de un área especializada. Por ejemplo, la *salsa verde* (EP y EM) no hace referencia al mismo concepto en España y en México. En España la salsa verde es una “salsa hecha a base de perejil, usada especialmente para acompañar pescados” (RAE 2014); mientras que en México se trata de “Nombre que reciben distintas salsas de color verde, hechas por lo general con tomate, ajo, cebolla, chile verde y cilantro” (Muñoz Zurita 2013).

2.3. Importancia de la variación fraseológica especializada en la traducción gastronómica

En la traducción gastronómica tienen mucho peso las costumbres, la sociedad y la cultura de cada país o región y es ahí donde radica su dificultad. Entendemos por traducción gastronómica aquella que permite a una persona o empresa relacionada con la gastronomía acercar sus platos a personas de otros países. Como hemos visto anteriormente, la comida y la bebida son aspectos arraigados en la cultura de un país o región. Esta importante carga idiosincrásica que encontramos en la terminología gastronómica es la que hace que deba ser tratada con sumo cuidado con el fin de mantener, en la medida de lo posible, los matices que encierra. Hemos observado que la literatura sobre la traducción gastronómica no es excesivamente amplia si lo comparamos con otros ámbitos especializados (Bugnot 2006; Rodríguez Abella 2008; González-Pastor y Cuadrado-Rey 2014; Sánchez Rubio 2017; Díaz Flores 2017; entre otros) y que, en muchas ocasiones, se trabaja como un subgénero dentro de la traducción turística (Martínez-Sánchez e Inigo Ros 1998; González-Pastor 2018).

De acuerdo con Bugnot (2006: 10), los contextos que generan más encargos relacionados con la gastronomía son: 1) los folletos turísticos; 2) los menús y 3) las recetas, en las que incluimos libros, sitios web y blogs de cocina. Dos son los aspectos fundamentales que complican este tipo de traducción: las diferencias culturales entre países y la alta especialización de la terminología gastronómica. En este último sentido, es importante disponer de glosarios especializados adaptados a la lengua de destino, atendiendo a la variación, puesto que incluso entre países hispanohablantes hay enormes diferencias terminológicas para referirnos a los mismos productos. Catford (1965: 86), al hablar sobre las variedades de la lengua, afirma que poseen un núcleo común sobre el que cada variedad aporta sus rasgos específicos o marcadores. Además, este autor señala que en traducción es importante tener en cuenta que el número y la naturaleza de las variedades cambian de una lengua a otra (1965: 85) y especifica que una de las variedades relacionadas con las características permanentes del hablante es el dialecto geográfico. En esta línea, Larson (1984: 131) afirma que el traductor debe ser consciente de que la situación condiciona los significados de las palabras. En cuanto al

uso de terminología técnica y los usos regionales, pone de relieve que, si se traduce para los hablantes de un área local, se debe escoger la forma de hablar de esa área en concreto (1984: 136-7). Bugnot (2006: 19) cuando trata la traducción de las recetas apunta que uno de los factores que se debe considerar es la “adecuación al receptor” y toma como ejemplo la receta de los *crêpes* en México que dice “*Extendre la harina en una charola, [...]*” donde *charola* es un vocablo propio del EM cuyo equivalente en EP es *bandeja*. Hay que tener en cuenta, por tanto, a qué colectivo va dirigida la traducción. Por ello, la traslación de UFE gastronómicas del francés al español variará en función de los receptores y la *purée de patate douce* será *puré de boniato o batata* (EP) en la variante española y *puré de camote* (EM) en la mexicana; *faire revenir les oignons* se traducirá por *pochar la cebolla* (EP) y *acitronar la cebolla* (EM); *tripes à la provençale* será *callos a la provenzal* (EP) y *pancita a la provenzal* (EM).

3. Metodología: Lingüística de corpus

Centrándonos en el marco teórico anteriormente descrito, la metodología utilizada para analizar la variación fraseológica especializada en el campo de la gastronomía entre el EP y el EM y esbozar un glosario francés-español se basará en la lingüística de corpus.

La Lingüística de Corpus, definida como “the study of language base on examples of ‘real life’ language use” (McEnery y Wilson 1996: 1) ha ofrecido un gran abanico de posibilidades para el estudio del lenguaje. En el ámbito traductológico Baker (1996) inició la andadura del *Corpus-based Translation Studies*, metodología que se ha ido perfeccionando hasta llegar a ser imprescindible (Corpas Pastor 2008). A pesar de los avances en terminografía los traductores siguen teniendo dificultades a la hora de traducir textos especializados, aún más si la traducción debe adecuarse a la variante geográfica de la lengua del receptor. Dentro de la fraseología en general y, por ende, en la especializada, “no son muchos los estudios a gran escala sobre la fraseología diatópica y no existen (todavía) Atlas dialectales especializados en la fraseología” (Pamies 2017: 57). En concreto, la disciplina “adolecía hasta hace poco de escasa profesionalidad en materia de diatopía” (Pamies 2017: 58). En este sentido, algunos trabajos como los de Mogorrón (2014a, 2014b,

2014c, 2015a, 2015b); Cuadrado-Rey (2016, 2018); Navarro-Brotons (2011, 2018); Navarro-Brotons & Cuadrado-Rey (2016) pretenden completar este campo. Por su parte, los diccionarios dialectales generales son más coherentes pero las UF tienen en ellos un rol periférico (cf. Pamies 2017: 59). Ante este panorama de falta de recursos lexicográficos y terminológicos que respondan a las necesidades de los traductores, se hace imprescindible el uso de corpus con el objetivo de paliar el problema. Como apunta Mogorrón (2018b: 11):

En efecto, hoy en día, la lingüística de corpus ha adquirido tal importancia que ya ha influido notablemente [...] en la traducción, así como en el diseño de materiales [...] traductológicos permitiendo utilizar datos reales y exhaustivos que posibilitan por fin acercarse a las características de las lenguas realmente habladas y utilizadas por las comunidades de hablantes.

3.1. *Constitución del corpus*

Para la creación de nuestro corpus hemos recurrido a algunos de los pasos que describen Atkins *et al.* (1992): 1) planteamiento, 2) soporte técnico, 3) recopilación y preparación de los datos.

3.1.1. *Planteamiento*

Como ya dijimos, pretendemos usar la Lingüística de Corpus para identificar las UFE en textos producidos en español de México (EM) y buscar las respectivas variantes en español peninsular (EP); analizar la variación denominativa (VD) y conceptual (VC) de algunas de las UF para reafirmar la importancia de la variación fraseológica en el ámbito de la traducción y esbozar un glosario fraseoterminológico contrastivo y bilingüe del EM, el EP y el francés destinado a especialistas, estudiantes y traductores gastronómicos. Para ello, el primer paso ha sido determinar lo que sería un corpus mínimamente representativo de esas variantes lingüísticas del español. Tomamos como base el criterio de medida de Sardinha (2000: 346), quien propone la clasificación del corpus según el número de palabras como se resume en la siguiente tabla:

Cantidad de palabras/tokens	Clasificación/Tamaño
Menos de 80 000	Pequeño
Entre 80 000 y 250 000	Pequeño-mediano
Entre 250 000 y 1 000 000	Mediano
Entre 1 000 000 y 10 000 000	Mediano-Grande
10 000 000 o más	Grande

Tabla 1. Extensión del corpus según Sardinha

Con esta referencia hemos compilado un corpus de recetas mediano de 556 820 *tokens* compuesto por dos subcorpus, uno para cada variedad del español. Quedan repartidos así: 279 091 *tokens* para el corpus EM y 277 729 *tokens* para el corpus EP.

Para la selección de los textos gastronómicos hemos seguido los criterios siguientes: solo recetas (libros y blogs de recetas¹ publicados después del año 2007 en la web escritos por nativos de las variantes de la lengua de trabajo) por tratarse de un tipo textual técnico, altamente convencional en las dos variedades del español estudiadas y tener, a su vez, una estructura discursiva que parte de una convención generalizada. Como apunta Rosa Rabadán (1991: 164-176 cit. por Bugnot 2006: 18), las recetas se constituyen por: a) el título; b) la enumeración de los ingredientes; c) la preparación, paso a paso, del plato y la forma de servirlo; d) la presencia de una fotografía del plato generalmente ya terminado y, en ocasiones, e) el coste aproximado, el grado de dificultad y el tiempo medio que necesita la preparación del plato.

Las recetas escritas en español peninsular se extrajeron de los siguientes libros y blogs de la web:

1. Es cierto que pueden existir diferencias estilísticas entre los libros de recetas y las recetas de los blogs en la web. Sin embargo, como para nuestro trabajo solo consideraremos los ingredientes, utensilios y platos, la diferencia estilística no presenta un problema.

Libro web o blog	Número de <i>tokens</i>
100 recetas de cocina española de la web <i>Recetas.net</i>	14 164
Cocina regional española	9 799
https://www.javirecetas.com/	71 563
https://webosfritos.es/	182 203
Total	277 729

Tabla 2. Corpus de recetas del EP

Las recetas escritas en español de México fueron extraídas de los siguientes libros y blogs:

Libro web o blog	Número de <i>tokens</i>
El recetario: cocina mexicana	30 992
Catálogo de recursos gastronómicos de México	80 444
https://www.lacocinamexicanadepily.com/	42 206
https://www.gabrielaclavoycanela.com.mx/	125 449
Total	279 091

Tabla 3. Corpus de recetas del EM

3.1.2. Soporte técnico

Nuestro corpus, al tratarse de un proyecto piloto y ser técnico, no posee grandes dimensiones por lo que no ha sido necesario utilizar un *hardware* específico. En cuanto al *software*, el programa utilizado para explotar el corpus ha sido *Sketch Engine*.² Se trata de una herramienta de análisis y gestión de corpus que ofrece diferentes funcionalidades: extracción de datos de amplios corpus textuales como frecuencias de palabras, colocaciones, concordancias, búsquedas mediante expresiones regulares, etc. Por su parte, *Sketch Engine* además lematiza automáticamente los textos. De las funcionalidades que ofrece destacaremos para este estudio 1) herramienta terminológica, 2) concordancias y 3) análisis de comportamientos gramaticales y colocaciones.

2. <https://www.sketchengine.eu/>

3.1.3. Recopilación y preparación de los datos

El programa *Sketch Engine* ofrece la posibilidad de compilar corpus nutriéndose de textos de la web o bien, a partir de archivos propios. En aras de primar la calidad de los textos y de eliminar la mayor cantidad de ruido posible, hemos descargado los libros de recetas disponibles en formato *pdf*, los hemos transformado en *Word*, hemos procedido a limpiarlos y, posteriormente, los hemos transformado a *txt*. En el caso de los blogs, hemos utilizado la herramienta *Save Webpage As Word Document* para transformar en archivo *Word* las recetas de la web. Archivos que, posteriormente, hemos limpiado y transformado a formato *txt*.

4. Análisis de la variación

Cabré (1993: 246) afirma que “[...] hacer terminología [...] es [...] recoger las denominaciones que los usuarios de una lengua emplean para referirse a un concepto [...]”. Siguiendo esta premisa, el análisis que presentamos a continuación se centrará en los contrastes de las unidades fraseológicas especializadas del lenguaje gastronómico de España y de México. Se trata de un análisis que, aunque mostrará algunos puntos convergentes, persigue poner especial interés en subrayar las divergencias de estas unidades dentro del español. De esta manera el traductor podrá utilizar el término adecuado según la variedad de español a la que tenga que traducir.

Siguiendo a Cabré (1993: 176), desde una dimensión lingüística las UFE pueden tipificarse según la forma, la función, el significado y la procedencia. Nuestro análisis se centrará en la función, que nos ha ayudado en la selección de los términos a tratar, y el significado de las UF dentro del cual analizaremos la variación denominativa (VD) y la conceptual (VC).

4.1. Desde el punto de vista de la función

Desde el punto de vista de la función que las UFE desempeñan en el discurso, podemos clasificarlas en sustantivos, adjetivos, verbos y adverbios. Una vez compilado los corpus EM y EP, el primer paso para proceder al análisis se ha centrado en la elección de los términos a partir de los cuales hemos trabajado. El programa *Sketch Engine* nos ofrece la posibilidad de establecer un listado de términos ordenados según la frecuencia y filtrado por categoría gramatical. A

continuación, mostramos un gráfico comparativo (corpus EM y EP) donde se puede apreciar que la frecuencia de sustantivos, verbos, adjetivos y adverbios en cada corpus es simétrica.

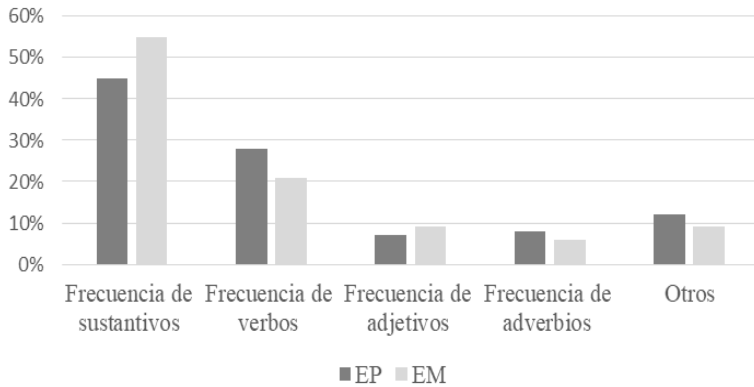


Gráfico 1. Frecuencia de categorías gramaticales en EP y EM

Debido a que la categoría gramatical más frecuente en ambos corpus son los sustantivos y, habida cuenta de que una gran cantidad de sustantivos del ámbito gastronómico se forman a partir de más de un lexema, el análisis se centrará en las UFE nominales del tipo [N+prep.+N]: *jugo de limón* (EM), *zumo de limón* (EP); *aceite de canola* (EM), *aceite de colza* (EP); *calabaza en pay* o *pay de calabaza* (EM), *tarta de calabaza* (EP) y del tipo [N+adj.]: *cebolla acitronada* (EM), *cebolla pochada* (EP); *azúcar impalpable* (EM), *azúcar glas* (EP); *carne molida* (EM), *carne picada* (EP).

4.2. Desde el punto de vista del significado

Desde el punto de vista semántico o del significado las UFE se pueden clasificar según el tipo de concepto que denominan (Cabré 1993: 180-181). Como apunta Usme López en toda receta encontramos, al menos, términos sobre los siguientes cuatro ámbitos (2010: 81) “[...] los ingredientes [...]; los utensilios y recipientes, [...]; las operaciones, verbos de acción y descripciones de los toques personales; los productos terminados y la denominación de los platillos obtenidos”. Basado en recetas, nuestro corpus cuenta con sustantivos relacionados con: ingredientes, utensilios, elaboraciones y platos, siendo

los ingredientes la categoría más numerosa como se puede comprobar en el gráfico que sigue.

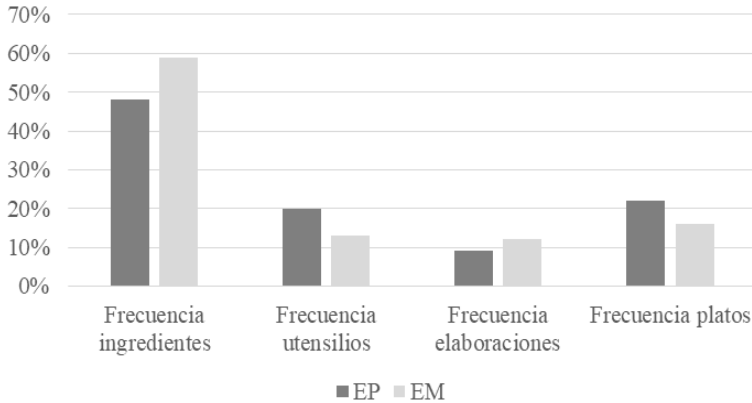


Gráfico 2. Frecuencia de los sustantivos relacionados con ingredientes, utensilios, elaboraciones y platos en EP y EM

Por ello, a partir del listado de los sustantivos que aparecen con mayor frecuencia, hemos creado una lista con los cien ingredientes más frecuentes en EM y en EP. Posteriormente hemos cruzado ambos con el fin de generar una lista de ingredientes comunes (47 en total): *aceite, canela, miel, oliva, pan, perejil, pimiento, pollo, sal, etc.* (EM y EP). Con los ingredientes no coincidentes hemos llevado a cabo una criba manual para seleccionar aquellos motivados por la variación léxica entre EM y EP: *chicharos* (EM) y *guisantes* (EP); *durazno* (EM) y *melocotón* (EP); *guajolote* (EM) y *pavo* (EP); *elote* (EM) y *maíz* (EP), entre otros.

4.2.1. Concurrencias en lemas comunes en EM y EP

Para explorar los patrones gramaticales, colocacionales y sintácticos de lemas comunes seleccionados usamos la función *Word sketch* del programa *Sketch Engine*. Por cuestiones de espacio mostraremos como ejemplo los lemas *azúcar* y *pollo*.

Así, de acuerdo con nuestro corpus, el lema azúcar (EM y EP) concurre con:

Concurrencia	Estructura	EM	EP
blanco ³	N+adj.	azúcar blanca	azúcar blanco
granulado	N+adj.	azúcar granulada	azúcar granulado
impalpable/ glas	N+adj.	azúcar impalpable	azúcar glas
invertido	N+adj.	azúcar invertida	azúcar invertido
moreno	N+adj.	azúcar morena	azúcar moreno
perlado	N+adj.	azúcar perlada	azúcar perlado
(a)vainillado	N+adj.	azúcar (a)vainillada	azúcar vainillado
capa	N+prep.+N	capa de azúcar	capa de azúcar
cucharada/ cucharadita	N+prep.+N	cucharada de azúcar/ cucharadita de azúcar	cucharada de azúcar/ cucharadita de azúcar
pizca	N+prep.+N	pizca de azúcar	pizca de azúcar

Tabla 4. Concurrencias del lema *azúcar* en EM y EP

La tabla 4 muestra dos tipos de variación denominativa entre el EM y el EP: variación léxica y variación morfológica.

La variación denominativa por cambio en el léxico se aprecia en ejemplos como *azúcar impalpable* (EM) frente a *azúcar glas* (EP).

Por su parte, la variación denominativa por cambio de morfemas se manifiesta con el uso de prefijos como *a-* para *azúcar avainillado* frente a *azúcar vainillado* y de morfemas de género: *azúcar blanca*, *invertida*, *morena* o *perlada* (EM) frente a *azúcar blanco*, *invertido*, *moreno* o *perlado* (EP).

En lo referente a las colocaciones del lema *azúcar* con términos relacionados con las medidas, vemos una completa similitud entre el EM y EP: *cucharada de azúcar*, *cucharadita de azúcar* o *pizca de azúcar*.

En todos los casos descritos anteriormente no existe variación conceptual (VC) pues todos los pares de UFE hacen referencia a los mismos conceptos.

El lema *pollo* (EM y EP) concurre con:

3. Marcamos con fondo de color las filas de la tabla donde se puede encontrar variación entre el EM y el EP.

Concurrencia	Estructura	EM	EP
adobo	N+adj./ N+prep.+N	pollo en adobo	pollo adobado
almendra	N+adj./ N+prep.+N	pollo almendrado	pollo con almendras
alita	N+prep.+N	alita de pollo	alita de pollo
brasa	N+prep.+art.+N	pollo a las brasas	pollo a la brasa
caldo	N+prep.+N	caldo de pollo	caldo de pollo
molleja	N+prep.+N	molleja de pollo	molleja de pollo
mano/pata	N+ prep.+N	manitas de pollo patas/patitas de pollo	patas de pollo patitas de pollo
Menudencia/ menudillo	N+ prep.+N	Menudencia(s) de pollo	menudillo(s) de pollo
pechuga	N+ prep.+N	pechuga de pollo	pechuga de pollo

Tabla 5. Concurrencias del lema *pollo* en EM y EP

La tabla 5 muestra tres tipos de variación denominativa entre el EM y el EP: variación sintáctica, léxica y morfológica.

Variación sintáctica: *pollo en adobo* [N+prep.+N] (EM) frente a *pollo adobado* [N+adj.] (EP) o *pollo almendrado* [N+adj.] (EM) frente a *pollo con almendras* [N+prep.+N] (EP).

Variación denominativa por cambio de morfemas de número: *pollo a las brasas* (EM) y *pollo a la brasa* (EP).

Variación léxica: *manitas de pollo* o *patas/patitas de pollo* (EM) frente a *patas/patitas de pollo* (EP); *menudencias de pollo* (EM) y *menudillos de pollo* (EP).

A pesar de las diferencias, también encontramos concurrencias idénticas con el lema *pollo* en ambas variedades del español: *alita de pollo*, *caldo de pollo*, *molleja de pollo* o *pechuga de pollo* (EM y EP).

Al igual que pasaba con la tabla 4, en todos los casos descritos anteriormente no existe variación conceptual (VC) pues todos los pares de UFE hacen referencia a los mismos conceptos.

Más allá de los lemas que hemos mostrado a título de ejemplo, el análisis arroja también otro tipo de variante denominativa como es la variante gráfica: *cacahuete tostado* (EM) y *cacahuete tostado* (EP) o *salsa de soya* (EM) y *salsa de soja* (EP). Igualmente, el análisis nos ha permitido detectar variaciones conceptuales (VC). Prueba de ello son las siguientes UFE *salsa verde*⁴; *salsa picante* o *agua fresca* (EM y EP): 1) la *salsa picante* no atiende al mismo concepto de un lado y otro del Atlántico si partimos de la base de que el concepto de *picante* no es el mismo en ambos países. En México el amplio abanico de chiles ofrece una mayor variedad de salsas picantes y una gradación en cuanto a la intensidad del picor, frente a la variedad y el concepto de salsa picante en España. 2) *Aguas frescas* en México hace referencia a una gran cantidad de bebidas refrescantes sin alcohol preparadas a base de agua con azúcar y pulpa de fruta que no se conoce en España.

Mención aparte merece el término *taco*, empleado generalmente en España como sinónimo de trozo y en México como término que denomina a uno de los platillos más conocidos de su gastronomía. De este modo, *tacos de pollo*, *tacos de pavo*, *taco de queso*, e incluso *taco de tortilla*⁵ serían interpretados de manera diferente en cada uno de estos países, pues conceptualmente no hacen referencia a la misma realidad.

4.2.2. Concurrencias en lemas diferentes en EM y EP

La variación denominativa en esencia pura se puede palpar en este grupo de ingredientes que para designar un mismo concepto utilizan lemas diferentes: *camote* (EM) y *batata o boniato* (EP), *crema* (EM) y *nata* (EP), *elote* y *maíz* (EP), *puerco* (EM) y *cerdo* (EP), *papa* (EM) y *patata* (EP), *tuna* (EM) o *higo chumbo* (EP), entre otros. Partiendo de que la base ya es distinta, las UFE que generen también lo serán.

Así, los lemas *papa* (EM) y *patata* (EP) coinciden en algunas concurrencias con términos de medida o elaboraciones como se puede ver en la tabla que sigue:

-
4. La diferencia conceptual de *salsa verde* en las dos variantes trabajadas ha sido expuesta en el punto 2.2 de este trabajo.
 5. A pesar de ser más usual en España la UFE *pincho de tortilla* también se utiliza *taco de tortilla*.

Concurrencia	Estructura	EM	EP
ensalada	N+prep.+N	ensalada de papas	ensalada de patatas
puré	N+prep.+N	puré de papas	puré de patatas
capa	N+prep.+N	capa de papas	capa de papas
kilo	N+prep.+N	kilo de papas	kilo de patatas

Tabla 6. Concurrencias de los lemas *papa* (EM) y *patata* (EP)

5. Esbozo de glosario EM, EP y francés

Con los datos obtenidos del análisis hemos podido compilar un listado de UFE divergentes en EM y EP que hemos completado con la ayuda del corpus francés *frTenTen* disponible en *Sketch Engine* y diccionarios/glosarios franceses generales (*Dictionnaire de l'Académie française* y *Trésor de Langue Française Informatisé*) y especializados en gastronomía (*Dictionnaire cuisine*; *Lexique de l'Académie culinaire*; *Dictionnaire gastronomique* y *Le grand Larousse gastronomique*). A continuación, mostramos una pequeña parte de lo que es el germen para la creación de un glosario especializado de UFE gastronómicas EM/EP<>Fr.

Español de México	Español Peninsular	Francés
aceite de canola	aceite de colza	huile de colza
azúcar impalpable	azúcar glas	sucre glace
azúcar morena	azúcar moreno	sucre roux
berenjena capeada	berenjena rebozada	aubergine panée
budín de camote	pudin de boniato/batata	pudding de patate douce
budín de elote	pudin de maíz	pudding de maïs
cabellos de ángel	cabello de ángel	confiture de courge
calabazas en pay/ pay de calabaza	tarta de calabaza	tarte à la citrouille
carne de puerco	carne de cerdo	viande de porc
carne de res	carne de vacuno/ ternera	viande de veau
carne molida	carne picada	viande hachée
cebolla acitronada	cebolla pochada	faire revenir les oignons

cebolla blanca	cebolla dulce	oignon doux
cebolla curtida	cebolla encurtida	oignon au vinaigre
cebollitas de Cambray	cebolla tierna	oignon tendre
cerveza oscura	cerveza negra	bière noire
chocolate de tablilla	chocolate de tableta	tablette de chocolat
crema de chícharos	crema de guisantes	purée de petit pois
crema fresca	nata líquida	crème fraîche
crema espesa	nata para montar	crème fraîche épaisse
cuadradito de chocolate	onza de chocolate	carré de chocolat
dulce de camote	dulce de boniato/batata	pâte de patate douce
elote	mazorca de maíz	épi de maïs
encurtido de camote	encurtido de batata	patate douce au vinaigre
ensalada Tampico	ensalada de surimi	salade de surimi
guajolote/pavo de campo	pavo de corral	dinde de ferme
guisado de res	guisado de ternera	ragoût de veau
hombro de cerdo/ trasero de cerdo	paleta de cerdo/ cuarto trasero de cerdo	épaule de porc
jarabe/néctar o miel de agave	sirope de agave	sirop d'agave
jitomate saladette	tomate de pera/ tomate pera	tomate olivette
jugo de durazno	jugo de melocotón	jus de pêche
jugo de limón	zumو de limón	jus de citron
jugo de naranja	zumو de naranja	jus d'orange
jugo de taronja	zumو de pomelo	jus de pamplemousse
leche sin grasa	leche desnatada	lait écrémé
lechuga orejona	lechuga romana	laitue romaine
limón dulce	limón	citron
niño envuelto o rollo	brazo de gitano	roulé
pancita a la provenzal	callos a la provenzal	tripes à la provençale
pasa de uva	uva pasa	raisin sec
pavo molido	pavo picado	viande de dinde hachée

polvo de hornear	levadura química	levure chimique
puré de camote	puré de batata/ boniato	purée de patate douce
puré de papas	puré de patatas	purée de pommes de terre
queso bajo en grasa	queso <i>light</i>	fromage <i>light</i>
queso de chiva/cabra	queso de cabra	fromage de chèvre
raja/rajita de canela	rama/ ramita de canela	bâton de cannelle
ramo de cilantro/ perejil	manejo de cilantro/ perejil	bouquet de coriandre/ persil
rosca de reyes	roscón de reyes	brioche des rois aux fruits confits
salsa de jitomate/tomate	salsa de tomate	sauce tomate
salsa de soya	salsa de soja	sauce de soya
tablilla de chocolate/ barra de chocolate	tableta/pastilla de chocolate	tablette de chocolat
tasajo/cecina de res	cecina de ternera	<i>cecina</i> / jambon cru de bœuf
torta de chocolate	pastel de chocolate	gâteau au chocolat

Tabla 7. Extracto del listado compilado

6. Conclusiones

Al inicio de nuestro trabajo nos habíamos marcado una serie de objetivos que estimamos haber cumplido. Hemos constatado que la lingüística de corpus puede ser una herramienta útil para la identificación de UFE en textos especializados. Gracias a ella hemos podido alcanzar nuestro primer y segundo objetivo: identificar UFE en recetas producidas en español de México y buscar variantes en recetas producidas en español peninsular.

El análisis de las UFE del español peninsular y mexicano evidencia la existencia de variaciones denominativas y conceptuales. Este hecho podría considerarse como un indicador de la importancia de crear glosarios especializados que contemplen ambos tipos de variación.

Desde un punto de vista general, el análisis de las unidades fraseológicas especializadas del español peninsular y del español mexicano del ámbito gastronómico refleja la complejidad y la riqueza de la lengua española en su

conjunto. En consecuencia, este hecho deja entrever la necesidad de elaborar glosarios especializados que reúnan UFE peninsulares con las variaciones diatópicas de los distintos países de habla hispana. Se reflejarían así las convergencias y las divergencias de las UFE de este idioma.

Desde un punto de vista más específico y propio de este estudio, el presente trabajo ha tratado de poner de relieve que el fenómeno de las variantes diatópicas requiere un examen detenido. Es obvio que un único trabajo no es suficiente para llegar a resultados concluyentes. Este estudio debería validarse con trabajos posteriores que abordaran las UFE en otros tipos de textos gastronómicos peninsulares y mexicanos, como por ejemplo cartas de restaurantes. Incluso hacerlo extensible a otras variedades del español. La suma de todo ello posibilitaría un análisis integral de las UFE gastronómicas del español.

Referencias bibliográficas

- AGUADO DE CEA, Guadalupe. (2007) “La fraseología en las lenguas especializadas”. En Alcaraz Varó, Enrique; José Mateo Martínez & Francisco Yus Ramos (eds.) 2007. *Las lenguas profesionales y académicas*. Barcelona: Ariel, pp. 53-65.
- ARNTZ, Reiner & Heribert Picht. (1989) *Einführung in die Terminologearbeit*. New York: Georg Olms Verlag. Citado en la traducción española de Amelia de Irazazábal; María José Jiménez; Erika Schwarz & Susana Yunquera (1995): *Introducción a la terminología*. Madrid: Fundación Germán Sánchez Ruipérez/ Ediciones Pirámide.
- ATKINS, Sue; Jeremy Clear & Nicholas Ostler. (1992) “Corpus Design Criteria”. *Literary and Linguistic Computing* 7:1, pp. 1-16.
- BAKER, Diana. (2015) “El Recetario: Cocina Mexicana.” <<http://www.recetasmexicanas.org/productos/El-Recetario-Diana-Baker.pdf>>
- BAKER, Mona. (1996) “Corpus-based translation studies: The challenges that lie ahead”. En Somers, Harold (ed.) 1996. *Terminology, LSP and Translation: studies in language engineering in Honour of Juan C. Sager*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, pp. 175-186.
- BLASCO MATEO, Esther. (1999) *Los límites de las perífrasis verbales y unidades fraseológicas verbales*. Barcelona: Universidad de Barcelona. Tesis doctoral inédita.

- BUGNOT, Marie-Ange. (2006) "La traducción de la gastronomía: textos y contextos (francés-español)." *TRANS* 10, pp. 9-22.
- CABRÉ CASTELLVÍ, María Teresa. (1993) *La terminología: teoría, metodología y aplicaciones*. Barcelona: Editorial Empúries.
- CABRÉ CASTELLVÍ, María Teresa. (1998) "Elementos para una teoría de la terminología: hacia un paradigma alternativo". *El lenguaraz. Revista académica del Colegio de Traductores Públicos de la ciudad de Buenos Aires* 1:1, pp. 59-78.
- CABRÉ CASTELLVÍ, María Teresa. (1999) *La terminología: representación y comunicación*. Barcelona: Universitat Pompeu Fabra.
- CABRÉ CASTELLVÍ, María Teresa. (2000) "Terminologie et linguistique: la théorie des portes". *Terminologie des portes* 21, pp. 10-15.
- CABRÉ CASTELLVÍ, María Teresa y Rosa Estopà Bagot. (2005) "Unidades de conocimiento especializado, caracterización y tipología". En Cabré Castellví María Teresa & Carme Bach Martorell (eds.) 2005. *Coneixement, llenguatge i discurs especialitzat*. Barcelona: IULA, Documenta Universitaria, pp. 69-94.
- CATFORD, John C. (1965) *A linguistic Theory of Translation*. Londres: Oxford University Press.
- CIAPUSCIO, Guiomar E. (1999) "Variación conceptual y grado de especialidad de los textos". *Revista Argentina de Terminología* 15, pp. 49-82.
- CORPAS PASTOR, Gloria. (1996) *Manual de fraseología española*. Madrid: Gredos.
- CORPAS PASTOR, Gloria. (2003) *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. Madrid: Iberoamericana.
- CORPAS PASTOR, Gloria. (2008) *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Frankfurt am Main: Peter Lang.
- CUADRADO-REY, Analía. (2018) "La variación fraseológica en el español hablado en Argentina: propuesta de clasificación". En Mogorrón Huerta, Pedro & Juan Antonio Albaladejo-Martínez (eds.) 2018. *Fraseología, Diatopía y Traducción. Phraseology, Diatopic Variation and Translation. [IVITRA Research in Linguistics and Literature, 17]*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 334-352.
- CUADRADO-REY, Analía. (2016) *Base de datos de UF del español de Argentina: propuesta de traducción al alemán*. Alicante: Universidad de Alicante. Tesis doctoral inédita.
- DÍAZ FLORES, Ismarie. (2017) *La terminología de la gastronomía puertorriqueña y su traducción al inglés*. Salamanca: Universidad de Salamanca. Tesis doctoral inédita.

- DIKI-DIKIRI, Marcel & Rodrigue Bigoundou. (2009) “Un enfoque cultural de la terminología”. *Debate Terminológico* 5. Versión electrónica: <<https://seer.ufrgs.br/riterm/article/view/23955/13885>>
- FREIXA, Judit. (2002) *La variació terminològica: anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient*. Barcelona: Universitat Pompeu Fabra. Tesis doctoral inédita.
- FREIXA, Judit. (2005) “Variación terminológica: ¿Por qué y para qué?”. *Meta* 50:4. Versión electrónica: <<https://www.erudit.org/en/journals/meta/2005-v50-n4-meta1024/019917ar.pdf>>
- GARCÍA-PAGE, Mario. (2004) “De los fines y confines de la fraseología”. En González Calvo, José Manuel; Jesús Terrón González & José Carlos Martín Camacho (eds.) 2004. *VII Jornadas de Metodología y Didáctica de la Lengua Española: las unidades fraseológicas*. Cáceres: Universidad de Extremadura, pp. 23-79.
- GARCÍA-PAGE, Mario. (2008) *Introducción a la fraseología española. Estudio de las locuciones*. Barcelona: Anthropos.
- GAUDIN, François. (2005) “La socioterminologie”. *Langages* 39:157, pp. 80-92. Versión electrónica: <http://www.persee.fr/web/revues/home/prescript/article/lgge_0458726x_2005_num_39_157_976>
- GONZÁLEZ-PASTOR, Diana. (2012) *Análisis descriptivo de la traducción de culturemas en el texto turístico*. Valencia: Universidad Politécnica de Valencia. Tesis doctoral inédita. Versión electrónica: <<https://riunet.upv.es/browse?value=Tesis+doctoral&type=type>>
- GONZÁLEZ-PASTOR, Diana & Analía Cuadrado-Rey. (2014) “El léxico en el ámbito de la gastronomía valenciana: un estudio de las técnicas de traducción de los culturemas”. En Vargas-Sierra, Chelo (coord.) 2014. *TIC, trabajo colaborativo e interacción en Terminología y Traducción*. Granada: Comares, pp. 345-357.
- GONZÁLEZ-PASTOR, Diana María. (2018) *¿Cómo se traducen los culturemas del ámbito turístico? Análisis de estrategias de traducción (español-inglés)*. Granada: Comares.
- GUZMÁN TENA, Pilar. (s.d.) “La cocina mexicana de Pily.” <<http://www.lacocina-mexicanadepily.com/publicidad/>>
- LARSON, Mildred L. (1984) *Meaning-based Translation. A Guide to Cross-language Equivalence*. Lanham: University Press of America.
- LORENTE CASAFONT, Mercè. (2001) “Terminología y fraseología especializada: del léxico a la sintaxis”. En Guerrero Ramos, Gloria & Manuel Fernando Pérez

- Lagos (eds.) 2001. *Panorama actual de la terminología*. Granada: Comares, pp. 159-180.
- LORENTE CASAFONT, Mercè; Cleci Regina Bevilacqua & Rosa Estopà Bagot. (2002) “El análisis de la fraseología especializada mediante elementos de la lingüística actual”. En Correira, Margarita (org.) 2002. *Terminologia, desenvolvimento e identidade nacional. Terminología, desarrollo e identidad nacional. Actas del VI Simposio Iberoamericano de Terminología (La Habana 1998)*. Lisboa: ILTE-Colibrí, pp. 647-666.
- MARTÍNEZ-SÁNCHEZ, Patricia C. & Marta Inigo Ros. (1998) “La traducción de términos culturales en los folletos turísticos: la gastronomía”. En Sánchez Macarro, Antonia (ed.) 1998. *Pragmática Intercultural*. Valencia: Universitat de València.
- MATEOS, Jimena. (2010) “Catálogos de recursos gastronómicos de México”. <https://patrimonioculturalyturismo.cultura.gob.mx/publi/Cuadernos_19_num/cuaderno17.pdf>
- MCENERY, Tony & Andrew Wilson (1996) *Corpus Linguistics*. Edimburgo: Edimburgh University Press.
- MOGORRÓN HUERTA, Pedro. (2015a) “Comprensión, competencia fraseológica y creatividad de las expresiones fijas con sus variantes diatópicas, ejemplos en francés y en español”. En Mogorrón Huerta, Pedro & Fernando Navarro Domínguez (eds.) 2015. *Fraseología Didáctica y Traducción*. Frankfurt am Main: Peter Lang, pp. 137-160.
- MOGORRÓN HUERTA, Pedro. (2015b) “La variation dans les Construction Verbales Figées de l’espagnol d’Amérique”. *Linguisticae Investigationes* 38:2, pp. 276-300.
- MOGORRÓN HUERTA, Pedro. (2014a) “Locuciones verbales, traducción y pérdida de equivalencia”. En Duarte, Vanda (ed.) 2014. *Fraseología y paremiología: enfoques y aplicaciones*. Madrid: Centro Virtual Cervantes, pp. 79-96.
- MOGORRÓN HUERTA, Pedro. (2014b) “Importancia (numérica) de las variantes diatópicas españolas y su tratamiento en los diccionario.”. *Yearbook of Phraseology* 5, pp. 123-144.
- MOGORRÓN HUERTA, Pedro. (2014c) “Las expresiones fijas diatópicas argentina y mexicanas.” En González Rey, María Isabel (ed.) 2014. *Didáctica y Traducción de las Unidades fraseológicas*. Santiago de Compostela: Universidade de Santiago de Compostela/ Servicio de Publicaciones, pp. 77-98.
- MOGORRÓN HUERTA, Pedro. (2018a) “La influencia de las lenguas indígenas precolombinas en las UF del español”. En Mogorrón Huerta, Pedro & Juan Antonio

- Albaladejo (eds.) 2018. *Fraseología, Diatopía y Traducción/ Phraseology, Diatopic Variation and Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 111-132.
- MOGORRÓN HUERTA, Pedro. (2018b) “Prólogo”. En Seghiri, Miriam (ed.) 2018. *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas*. Berlín: Peter Lang, pp. 9-12.
- MONTORO DEL ARCO, Esteban. (2008) “Relaciones entre morfología y fraseología: las formaciones nominales pluriverbales”. En Almela, Ramón & Esteban Montoro (eds.) 2008. *Neologismos y morfología*. Murcia: Ediciones de la Universidad de Murcia, 65-90.
- MUÑOZ ZURITA, Ricardo. (2013) “Diccionario enciclopédico de la Gastronomía Mexicana”. <<https://laroussecocina.mx/diccionarios/>>
- NAVARRO-BROTONS, Lucía. (2011) “Traduire la langue, traduire la culture. Les cas des proverbes multiculturels mexicains ayant des voix du nahuatl”. *Synergies Tunisie* 3, pp. 125-136.
- NAVARRO-BROTONS, Lucía & Analía Cuadrado-Rey. (2016) “Las locuciones verbales somáticas españolas y sus variantes diatópicas argentinas y mexicanas”. En Mogorrón Huerta, Pedro; Analía Cuadrado-Rey, Lucía Navarro-Brotóns & Iván Martínez Blasco (eds.) 2016. *Fraseología, variaciones y traducción*. Frankfurt am Main: Peter Lang, pp. 223-236.
- NAVARRO-BROTONS, Lucía. (2018) “Refranes españoles y sus equivalentes mexicanos”. En Mogorrón Huerta, Pedro & Juan Antonio Albaladejo-Martínez (eds.) 2018. *Fraseología, Diatopía y Traducción. Phraseology, Diatopic Variation and Translation. [IVITRA Research in Linguistics and Literature, 17]*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 380 – 400.
- PAMIES BERTRÁN, Antonio. (2017) “Fraseología y variación diatópica”. *Verba hispánica: anuario del Departamento de la Lengua y Literatura Españolas de la Facultad de Filosofía y Letras de la Universidad de Ljubljana* 25, pp. 58-81.
- PAVEL, Sylvia. (1993) “Vers une méthode de recherche phraséologique en langue de spécialité”. *Terminology Update* 26:2, pp. 9-13.
- PENADÉS MARTÍNEZ, Inmaculada. (1999) *La hiponimia en las unidades fraseológicas*. Cádiz: Servicio de Publicación de la Universidad de Cádiz.
- PÉREZ, Susana. (s.d) “Webos fritos”. < <https://webosfritos.es/>>
- RABADÁN, Rosa. (1991) *Equivalencia y traducción: problemática de la equivalencia transléctica inglés-español*. León: Universidad de León.

- REBUCHON, Joël (ed.) (2012) *Le grand Larousse gastronomique – nouvelle édition*. Paris: Larousse.
- RODRÍGUEZ ABELLA, Rosa María. (2008) “La traducción de los culturemas en el ámbito de la gastronomía (Análisis de los folletos de TURESPAÑA)”. *Revista internazionale di tecnica della traduzione* 10, pp. 47-69.
- RUIZ GURILLO, Leonor. (1997) *Aspectos de fraseología teórica española*. Valencia: Universitat de València. Cuadernos de filología anejo XXIV.
- RUIZ GURILLO, Leonor. (2002) “Compuestos, colocaciones, locuciones: intento de delimitación”. En Veiga, Alexandre; Miguel González Pereira & Montserrat Souto Gómez (eds.) 2002. *Léxico y gramática*. Lugo: Tris Tram, pp. 327-339.
- SANCHEZ RUBIO, Alicia. (2017) *Los problemas de equivalencia en la traducción gastronómica chino-español/español-chino*. Alicante: Universidad de Alicante. Tesis doctoral inédita.
- SANZ VICENTE, María Lara. (2011) *Análisis contrastivo de la terminología de la teledetección. La traducción de compuestos sintagmáticos nominales del inglés al español*. Salamanca: Universidad de Salamanca. Tesis doctoral inédita.
- SARDINHA, Tony Berber. (2000) “Lingüística de Corpus: Histórico e Problemática”. *D.E.L.T.A* 16 :2, pp. 323-367.
- SEVILLA MUÑOZ, Manuel. (2015) “Las unidades fraseológicas del discurso científico-técnico y su traducción (inglés-español)”. En Conde Tarrío, Conde; Pedro Mogorrón Huerta; Manuel Martí Sánchez & David Prieto García-Seco (eds.) 2015. *Enfoques actuales para la traducción fraseológica y paremiológica: ámbitos, recursos y modalidades*. Madrid: Centro Virtual Cervantes, pp. 239-256.
- USME LÓPEZ, Zuly Norbelia. (2010) *Cocina, texto y cultura: recetario para una semiótica culinaria*. Bogotá: Fundación Universitaria INPAHU.
- VARIOS AUTORES (Académie culinaire). (s.d.) “Lexique de l’Académie culinaire” <<https://www.dictionnaire-academie.fr/>>
- VARIOS AUTORES (Académie française). (s.d.) “Dictionnaire de l’Académie française”. <<https://www.dictionnaire-academie.fr/>>
- VARIOS AUTORES (Amiloquemegustaescocinar). “Cocina regional española”. <http://www.chefuri.net/usuarios/download/cocina_regional/Cocina-regional-espanola.pdf>
- VARIOS AUTORES (Asociación de Academias de la Lengua Española) “Diccionario de americanismos”. <<http://lema.rae.es/damer/?key=>>>
- VARIOS AUTORES (CNRS & Université de Lorraine). (1994) “Le trésor de la langue française informatisé”. <<http://atilf.atilf.fr/>>

- VARIOS AUTORES (Dictionnaire cuisine). (s.d.) “Dictionnaire cuisine”. < <http://www.dictionnaire-cuisine.com/>>
- VARIOS AUTORES (Dictionnaire gastronomique). (s.d.) “Dictionnaire gastronomique”. <http://www.dictionnaire-gastronomique.com/>>
- VARIOS AUTORES (Gabriela Clavo y Canela). (2009) “Gabriela Clavo y Canela.” < <https://www.gabrielaclavoycanela.com.mx/>>
- VARIOS AUTORES (Javi Recetas). (2018). “Blog. Javi Recetas.” <<https://www.javi-recetas.com/>>
- VARIOS AUTORES (Real Academia Española). (2014) “Diccionario de la lengua española. La 23ª edición”.<<https://www.rae.es/diccionario-de-la-lengua-espanola/la-23a-edicion-2014>>
- VARIOS AUTORES (Recetas.net). (s.d.) “100 recetas de cocina española.” < http://tutoriales.altervista.org/trabajos/Cocina_espanola.pdf>
- WOTJAK, Gerd. (1988) “Uso y abuso de las unidades fraseológicas”. En Peira, Pedro (ed.) 1988. *Homenaje a Alonso Zamora Vicente. Historia de la lengua. El español contemporáneo. Volumen 1*. Madrid: Castalia, pp. 535-548.
- WOTJAK, Gerd. (2006) *Las lenguas, ventanas que dan al mundo*. Salamanca: Ediciones de la Universidad de Salamanca.

BIONOTA / BIONOTE

LUCÍA NAVARRO-BROTONS posee una licenciatura en Traducción e Interpretación complementada con un máster en Traducción Literaria y el premio extraordinario de Doctorado en Traducción e Interpretación. Está acreditada igualmente como traductora jurada por el Ministerio de Asuntos Exteriores y Cooperación. Imparte clases de traducción en la Universidad de Alicante desde 2011. Desde 2008 es miembro del grupo de investigación Frasytram. Su investigación se centra principalmente en los temas relacionados con la traducción fraseológica en las áreas del francés y el español.

LUCÍA NAVARRO-BROTONS holds a degree in Translation and Interpretation, supplemented with a MA in Literary Translation and the PhD award in Translation and Interpretation. She has also the accreditation of sworn translator issued by the Ministry of Foreign Affairs and Cooperation. She is a lecturer of Translation at the University of Alicante since 2011. Since 2008,

she is a member of the research group Frasytram. Her research production focuses primarily on issues related to the translation of phraseological units, especially in the French and Spanish area.

Recibido / Received: 16/11/2019
Aceptado / Accepted: 22/04/2020

Para enlazar con este artículo / To link to this article:
<http://dx.doi.org/10.6035/MonTI.2020.ne6.11>

Para citar este artículo / To cite this article:

Cuadrado Rey, Analía. (2020) "Las lenguas originarias en la fraseología del español hablado en Argentina: un estudio basado en corpus." En: 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. 355-385.

LAS LENGUAS ORIGINARIAS EN LA FRASEOLOGÍA DEL ESPAÑOL HABLADO EN ARGENTINA: UN ESTUDIO BASADO EN CORPUS

INDIGENOUS LANGUAGES IN THE PHRASEOLOGY OF SPANISH SPOKEN IN ARGENTINA: A CORPUS-BASED STUDY

ANALÍA CUADRADO REY
analía.cuadradorey@ua.es
Universidad de Alicante

Resumen

El presente trabajo se centra en el estudio del componente indígena en la fraseología diatópica del español de Argentina. Metodológicamente se aplica la lingüística de corpus *EsTenTen11* y se emplea el software *Sketch Engine* para comprobar la frecuencia de una muestra de unidades fraseológicas (UF) del campo conceptual carácter-forma de ser-actitud-conducta no recogidas en el DRAE. Las fuentes empleadas proceden de diccionarios, que su vez se validan por informantes autóctonos. Se identifica los tipos de variaciones fraseológicas detectadas en el corpus periodístico y no registradas en los diccionarios. Los resultados revelan que el 5% de las UF tienen componentes de lenguas originarias y existe una relación directa entre la cantidad de UF aportadas por cada una de las lenguas y su frecuencia de uso. El quechua tiene el mayor número de UF muy frecuentes, seguido del guaraní y el mapuche. Se han detectado variantes interna léxicas, variantes internas morfo-léxicas, variantes estructurales, así como inflexiones internas morfológicas.

Palabras clave: Fraseología; variación diatópica; español hablado en Argentina; lenguas originarias.



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

Abstract

This paper focuses on the study of the indigenous component in the diatopic phraseology of Argentinean Spanish. Methodologically, EsTenTen11 corpus linguistics is applied and the Sketch Engine software is used to check the frequency of a sample of phraseological units of the conceptual field character-form-of-being-attitude-conduct not collected in the DRAE. The sources used are from dictionaries, which are in turn validated by indigenous informants. The types of phraseological variations detected in the journalistic corpus and not recorded in the dictionaries are identified. The results reveal that 5% of the phraseological units have components of native languages and that there is a link between the amount of them provided by each language and its frequency of use. Quechua has the highest number of very frequent phraseological units, followed by Guaraní and Mapuche. Internal lexical variants, internal morpho-lexical variants, structural variants, as well as internal morphological inflections have been detected.

Keywords: Phraseology; Diatopic variation; Spanish of Argentina; indigenous languages.

1. Introducción

La modalidad diatópica del español, es decir, «*la variedad en el espacio*» (cf. Coseriu 1990: 56) se expresa en las manifestaciones habladas y escritas en las que están también presentes Unidades Fraseológicas (UF) españolas que suelen presentar algunos cambios o modificaciones respecto de su forma peninsular o bien tener entre sus componentes vocablos autóctonos, propios que recuerdan las particularidades que adquieren las lenguas con una gran difusión territorial. En efecto, como señala el preámbulo de la vigésimo tercera edición del Diccionario de la lengua española de la Real Academia de la Lengua (DRAE), la difusión de la lengua española, para la que cada región hispanohablante ha contribuido con miles de vocablos, ha sido y es tan amplia que no puede esperarse de ella sino un léxico vastísimo con más de 88.000 palabras. En este sentido, la diversidad que aporta el español de América es producto, según Lipski (2004: 1118):

no sólo de su herencia europea y de la evolución interna, sino también de la variedad de contactos con lenguas autóctonas, de inmigración forzada y de inmigración voluntaria.

En cuanto al influjo de las lenguas originarias en el español de América, podemos afirmar que la influencia de las lenguas indígenas en el español afecta a todos los niveles de la lengua sin excepción, más allá de la incorporación de los préstamos léxicos (cf. Palacios 2008:18). En la misma línea, Rosenblat (2002:140), sostiene que:

desde 1492 estas lenguas han incorporado al español de América una serie de elementos como la entonación, los rasgos articulatorios, los sufijos, los nombres de flora y fauna y de vida material y espiritual, y en regiones bilingües hasta incluso moldes sintácticos.

Esta diversidad también se observa en el universo fraseológico del español en el que se reflejan las creaciones idiosincrásicas propias de cada una de las regiones lingüísticas hispanohablantes. Estas unidades de la lengua que contienen elementos culturales propios de cada región lingüística, si bien se expresan en la lengua común, no son, en todos los casos, comprensibles para los usuarios españoles o de las demás regiones hispanohablantes. Ante este panorama tan diverso, como afirma Moreno de Alba (1993:47):

resulta innegable la necesidad de explicar cómo y por qué el español en América, sin perder su unidad esencial con el español europeo, va adquiriendo en los niveles fonológico, fonético, gramatical y léxico, su propia fisonomía.

Así, la diversidad que aportan las lenguas originarias americanas, algunas extintas y otras en uso es una característica central y un factor clave de diferenciación en el español americano y no cabe duda de la influencia de estas en la conformación léxica, pero también fraseológica del español general. Sabemos, sin embargo, que este papel que han tenido y tienen las lenguas indígenas en el español de América varía en función de factores bien distintos como la intensidad del contacto, el bilingüismo o monolingüismo de los hablantes, el grado de adquisición del español, el aislamiento geográfico, entre otros.

En Argentina, desde el momento mismo de su conformación, la sociedad se configura como una comunidad de lenguas en contacto, en la cual los procesos migratorios generan identidades complejas producto de la diversidad étnica y cultural (cf. Siguan 2001:15). El carácter heterogéneo y constante de los movimientos migratorios procedentes de países europeos, en su mayoría

de Italia y España, llegados a partir del último tercio del siglo XIX a la región del Río de la Plata, en general, y a Argentina, en particular, junto al sustrato indígena ha generado un amplio espectro de situaciones de contacto interlingüístico e intralingüístico. Como consecuencia de este proceso se puede señalar que, si bien actualmente el español es la lengua oficial de Argentina, no es la única lengua que se habla en el país. El estudio de los términos foráneos y de la fraseología de una lengua que está presente en diversas regiones geográficas como es el caso de la lengua española nos permite, no sólo mejorar nuestros conocimientos de la lengua en cuestión, sino que además nos acerca a la cultura de los diversos pueblos de influencia, como afirma Sapir (1954: 259):

Un detenido estudio de tales préstamos constituye un interesante comentario sobre la historia de la cultura. Casi es posible llegar a precisar el papel que los diversos pueblos han desempeñado en el desarrollo y la difusión de ideas culturales por el grado en que su vocabulario se ha infiltrado en el de otros pueblos.

Estas huellas se observan indudablemente también en su universo fraseológico y muestran la capacidad de asimilación y de crecimiento del idioma. Sin embargo, según Koike (2001:110):

la lingüística hispánica apenas ha(bía) prestado atención a la variación fraseológica, debido fundamentalmente a la falta de datos fraseológicos que cubran toda el área hispana y con los que se puedan hacer investigaciones sobre el tema.

En efecto, el estudio de las variantes sociolingüísticas y, por ende, también de las diatópicas ha ocupado un lugar periférico dentro de los estudios de la fraseología, comparado con el tratamiento dado a las demás variantes fraseológicas (morfológicas, léxicas, sintácticas, etc.). Como señala Penadés Martínez (2012: 48):

si excluimos del estudio de las variantes fraseológicas las variedades sociolingüísticas [...] no se llega a obtener un auténtico análisis de la realidad de la lengua, que varía, cambia, se modifica, con el paso del tiempo, con el tránsito de una zona a otra, por los distintos hablantes (jóvenes, adultos, hombres, mujeres, con estudios, sin estudios, etc.) que la usan en diferentes situaciones comunicativas (formal, informal, etc.), y con distintas modalidades (oral, escrita, etc.).

En los últimos años, y en paralelo al auge que ha experimentado la fraseología, se han realizado grandes esfuerzos tendientes a recoger, sistematizar y describir las UF (Unidades Fraseológicas) del español gracias a la aportación de la lingüística de corpus a los estudios fraseológicos, ya que nos permite realizar comparaciones inter e intralingüísticas. En este sentido, hay que señalar la aparición de algunos trabajos recientes como los de Mogorrón (2014a, 2014b, 2015a, 2015b), Navarro Brotons (2011) Navarro Brotons & Cuadrado Rey (2016) que aportan los primeros resultados del análisis contrastivos y multilingües de las variantes mexicanas, españolas y argentinas a partir de la incorporación de aplicaciones informáticas y la utilización de corpus textuales en los estudios de las variantes diatópicas del español. A excepción de algunos trabajos (Corpas Pastor, 2018, Mogorrón, 2018, Bolaños-Fabres, 2015), es escasa la investigación llevada a cabo en el ámbito de la fraseología española que aborda la presencia de las lenguas originarias en las diferentes variantes americanas del español. Esta situación puede deberse a que los esfuerzos se concentraron tradicionalmente en los estudios léxicos frente al de las locuciones en la metodología tanto lexicográfica como dialectológica (cf. Haensch 1997: 286; Casado Velarde 2005: 689-690).

2. Propuesta de estudio

En el marco de nuestro trabajo abordamos la presencia de los componentes de las lenguas originarias en la fraseología del español hablado en Argentina. Consideramos, al igual que Mogorrón (2014a), que junto con la variedad mexicana esta variedad es una de las más complejas por extensión geográfica, número de hablantes, distribución de personas nativas que lo tienen como lengua materna, contacto con lenguas indígenas o como lengua de migrantes, entre otros. Nos proponemos como objetivos principales de esta investigación:

1. Establecer un estado de la cuestión en relación al desarrollo de la fraseología y la fraseografía del español hablado en Argentina.
2. Contextualizar la situación actual de las lenguas indígenas en Argentina.
3. Establecer la frecuencia de uso en textos periodísticos de un corpus de UF pertenecientes al campo conceptual carácter-forma de

ser-actitud-conducta y no recogidas en el *Diccionario de la lengua española* de la Real Academia de la Lengua (DRAE).

4. Identificar los tipos de variaciones de las UF del corpus detectadas en el corpus periodístico y no registradas en los diccionarios que reflejan la variedad argentina.

En cuanto a la organización, el presente trabajo se divide en siete apartados. En el primero se desarrolla una introducción, seguidamente, en el segundo apartado, se plantean los objetivos del trabajo. En el tercer apartado se desarrolla el primer objetivo y se establece un estado de la cuestión en relación al desarrollo de la fraseología y la fraseografía del español hablado en Argentina. En el cuarto apartado se desarrolla el segundo objetivo y se contextualiza la situación actual de las lenguas indígenas en Argentina. En el quinto apartado se describen la metodología, las fuentes, los criterios de selección y las herramientas a partir de las cuales se conforma y se contrasta el corpus de trabajo. En el sexto apartado se desarrollan el tercer (6.1) y cuarto objetivo (6.2) relacionados con la frecuencia de uso del corpus de UF en textos periodísticos y con los tipos de variantes fraseológicas identificadas. En el séptimo apartado se dedica a presentar las conclusiones del trabajo.

3. Fraseología y fraseografía del español hablado en Argentina

Entre las obras que se dedican al estudio de la fraseología de la variante argentina podemos citar el trabajo de Zamora (1993). El autor contrasta UF bonaerenses con sus equivalentes peninsulares y destaca puntos divergentes y comunes. La tesis doctoral de Sciutto (2005a) trata las UF somáticas del español rioplatense, posteriormente (Sciutto 2005b) estudia estas UF desde una vertiente contrastiva e interlingüística. En un trabajo reciente (Sciutto 2015) la autora amplía este corpus inicial de somatismos y los analiza desde una perspectiva morfosintáctica. Pauer (2008) trata la fraseología lingüística en el ámbito hispanorrioplatense desde una perspectiva contrastiva con el español peninsular y presenta similitudes y diferencia entre expresiones argentinas y uruguayas. Cuadrado Rey (2016) plantea un trabajo de recopilación de UF argentinas que no están registradas en los diccionarios y realiza una propuesta de traducción al alemán. En un trabajo posterior (Cuadrado Rey 2018) la autora parte del análisis de las UF relacionadas con el campo conceptual

carácter-forma-de-ser que contiene la base de datos del grupo de investigación Frasytram de la Universidad de Alicante; y establece un esquema variacional fraseológico aplicado al español de Argentina inexistente hasta el momento y ampliable a otras variedades del español. En relación a los estudios que tratan la fraseología de las variantes regionales argentinas podemos citar el trabajo de Soto de Matulovich (2008) que estudia la presencia de las UF en la prensa de la provincia de Salta. También tenemos noticia de que se está desarrollando un proyecto de investigación denominado Fraseología del español hablado en la provincia de Córdoba dirigido por la profesora Toniolo de la Universidad Nacional de Córdoba que tiene como objetivo realizar un inventario de combinaciones fijas en la modalidad culta cordobesa.

En cuanto a las obras lexicográficas más relevantes que reflejan la fraseología propia del español de Argentina, es decir, que registran únicamente aquellas voces y expresiones de uso exclusivo de Argentina, debemos destacar el *Diccionario fraseológico de uso* de Lisandro Segovia (1906). Se trata del primer diccionario exclusivamente fraseológico escrito en Argentina, ya que con anterioridad a esta obra se habían hecho apuntaciones de elementos fraseológicos incluidos en diccionarios de voces, pero no se trataba de obras dedicadas exclusivamente a la fraseología. Según Barcia (2010) esta obra sería el primer diccionario fraseológico hispanoamericano. Se trata de un diccionario fraseológico total, es decir, no contrastivo cuya edición se incorporó como una de las catorce secciones de las que se compone el *Diccionario de argentinismos* (1911) y que ha sido reeditado por la Academia Argentina de Letras en el año 2006. En este grupo de diccionarios, también cabe incluir el *Diccionario fraseológico del habla argentina, frases, dichos y locuciones* del año 2010 de Barcia y Pauer. Se trata de un diccionario fraseológico argentino no total sino contrastivo con España que se apoya en este aspecto, por un lado, en la edición de 2004 de la obra de Seco, Andrés y Ramos *Diccionario fraseológico documentado del español actual* y, por el otro, en el *Diccionario Espasa de dichos y frases hechas* (edición 2007) de Buitrago Jiménez. Este diccionario fraseológico cuenta con unos once mil artículos y registra UF gauchescas, las del lunfardo o modismos cotidianos que todavía se utilizan, y que a través del tiempo fueron cambiando de significado. También debemos reseñar un creciente interés por el acervo paremiológico argentino, como lo demuestra la publicación del *Refranero de uso argentino* de Barcia y Pauer (2013). En esta

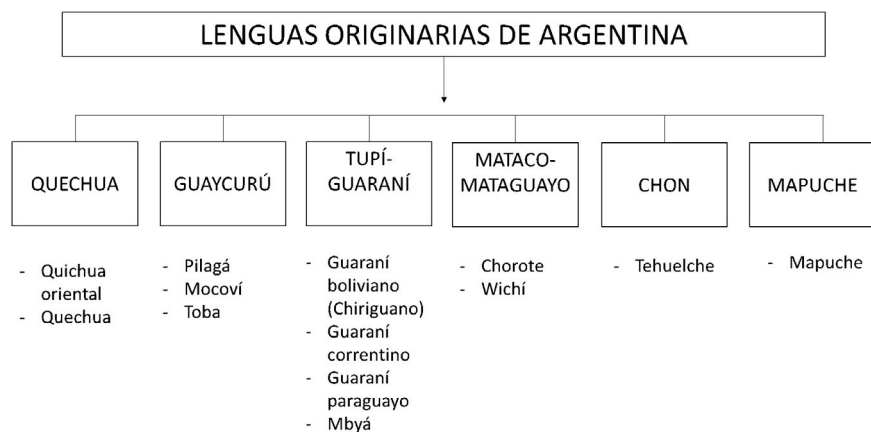
obra se realiza un completo estudio donde se analiza el origen de este tipo textual, sus antecedentes, características, vínculos con el español y bibliografía selecta de obras argentinas y españolas. Las obras lexicográficas citadas, así como las incipientes obras que tratan diversos aspectos de la fraseología argentina constituyen el inicio de un camino que llevará a reflejar y describir con mayor detalle la identidad de esta variedad.

4. La situación de las lenguas originarias en la actualidad en Argentina

Si bien la mayoría de los estudios dedicados a las lenguas indígenas en Argentina se refieren a las regiones de contacto con el guaraní (nordeste), el quechua (noroeste) y el mapuche (Patagonia) debemos tener en cuenta que, como señala Martínez (2008), en las últimas décadas, y como consecuencia de la gran influencia de migrantes procedentes de países limítrofes hablantes de guaraní o quechua, se ha activado el interés por los estudios de contacto entre el español y las lenguas autóctonas en la ciudad y en la provincia de Buenos Aires (Martínez, 2008: 255). Esta autora señala cuatro regiones diferenciadas en las que se produce el contacto de la lengua española con las lenguas indígenas en Argentina (Martínez, 2008: 256). Estas son: 1. Regiones con población indígena en las que además del bilingüismo español- lengua indígena, se manifiestan situaciones de monolingüismo en lengua indígena (Salta, Jujuy o Formosa). 2. Regiones de hablantes bilingües criollos como es el caso de algunas zonas de las provincias de Corrientes (bilingüismo español- guaraní) y de Santiago del Estero (bilingüismo español- quichua). 3. Regiones fronterizas con Paraguay y Bolivia, como, por ejemplo, la provincia de Misiones, en la que junto con el español se habla guaraní, la frontera de Formosa y Paraguay donde también el guaraní convive con el español o la región del noroeste, fronteriza con Bolivia, donde se manifiestan contactos con el quechua. 4. Conglomerados urbanos que han incrementado su heterogeneidad lingüística por la influencia de los inmigrantes bilingües quechua-español, procedentes de Bolivia y Perú; y guaraní-español, procedentes de Paraguay.

Vemos cómo el panorama multicultural y multilingüe de estos centros urbanos se ha enriquecido en las últimas décadas con el establecimiento de ciudadanos paraguayos, bolivianos y peruanos que se han sumado a los migrantes internos procedentes del litoral y del noroeste. Podemos afirmar,

sin embargo, que si bien Argentina en su devenir histórico ha invisibilizado la pluralidad etnolingüística que posee, en las últimas décadas esa tendencia se ha revertido, ya que innegablemente el contacto lingüístico es la regla y el monolingüismo la excepción (cf. Hecht 2011: 1). De este modo, en la actualidad la Argentina puede considerarse un país multilingüe poseedor de una enorme diversidad sociolingüística y sociocultural; donde, además del español y las lenguas de migración, existen doce lenguas indígenas pertenecientes a cinco familias lingüísticas diferentes, además del mapuche que no pertenece a ninguna familia (cuadro I).



Cuadro I. Distribución de las lenguas originarias en Argentina por familias lingüísticas

Si bien sabemos que cada una de las lenguas habladas en Argentina cuenta con un grado muy diverso de vitalidad y que se hallan en distintos grados en contacto con el español, no hay datos exactos sobre la población indígena en la Argentina, sobre la cantidad real de hablantes, en qué lengua aprenden a hablar los niños o si son mayoritariamente monolingües o bilingües. Los datos más actuales publicado que podemos tomar en consideración son los resultados del Censo del año 2010 en el que se continúa el trabajo desarrollado por primera vez en 2005 con la Encuesta Complementaria de Pueblos Indígenas (ECPI) del Instituto Nacional de Estadísticas y Censos (INDEC). El censo de 2010 es, según señala el *Instituto Nacional de Estadística y Censos*

de la República Argentina (INDEC), el segundo relevamiento sistemático que se realiza en el marco de un censo general de población sobre las características y condiciones de vida de la población que se reconoce perteneciente y/o descendiente de pueblos indígenas en el país. De acuerdo a los resultados el 2,38 % de la población total del país, es decir, 955.032 personas se reconocen pertenecientes y/o descendientes en primera generación de pueblos indígenas. A continuación, describimos la situación y la localización de cada una de las cinco familias en las que se agrupan las lenguas originarias habladas actualmente en Argentina.

4.1 *El quechua/ quichua santiaguëño*

El quechua fue la única modalidad indígena de América Latina que desempeñó el papel de una lengua de civilización antes de la llegada de los europeos; todavía hoy lo utilizan varios millones de personas distribuidas en distintos países (sur de Colombia, noroeste argentino, núcleos aislados del norte de Chile como resultado de emigraciones contemporáneas, Bolivia y, sobre todo, Ecuador y Perú). Es, por tanto, una de las familias lingüísticas más extensas tanto geográficamente como por la cantidad de hablantes que posee: alrededor de doce millones y medio en la región (Perú, Bolivia, Ecuador, Colombia, Chile y Argentina). Según Censabella (2005) existen en Argentina tres grupos diferenciados de hablantes del quechua: a) migrantes de nacionalidad boliviana y peruana, asentados en los últimos años en las grandes ciudades, principalmente en Capital y Gran Buenos Aires, b) indígenas denominados “collas” que habitan en el noroeste argentino, en las provincias de Salta y Jujuy y c) población criolla de Santiago del Estero que habla quichua santiaguëño.

4.2 *La familia guaycurú*

Dentro de la familia guaycurú, de procedencia chaqueña, los indígenas tobas constituyen la etnia más numerosa y se encuentran principalmente, en las provincias de Chaco, Formosa y, en menor cantidad, al este de Salta. En la última década se instalaron, como afirma Martínez (2008: 259), numerosas familias tobas en Santa Fe y Buenos Aires. En la localidad de Derqui, en la provincia de Buenos Aires, se están llevando a cabo investigaciones relacionadas con la vigencia de la lengua, el mantenimiento de la identidad toba y la educación.

4.3 *La familia tupí-guaraní*

El nordeste argentino, zona de contacto con Brasil y Paraguay, ha sido tradicionalmente denominado como Región Guaranítica (Vidal de Battini 1964). Se trata de una región fronteriza que se halla comprendida por las provincias argentinas de Corrientes, Chaco, Formosa y Misiones y el norte de Entre Ríos. En relación a la extensión de dominio de esta lengua, cabe recordar que cuando el español se instauró como la lengua oficial de la administración colonial, el guaraní era la lengua que muchas poblaciones ribereñas a lo largo del río Paraná utilizaban no sólo en sus relaciones interpersonales sino también para el intercambio comercial. Debemos señalar dos factores de relevancia en relación a la vitalidad del guaraní en la actualidad, por un lado, el estatus de lengua cooficial que ostenta en Paraguay y, por el otro, el alto grado de movilidad de sus hablantes propiciada por la libre circulación de personas que facilita el tratado MERCOSUR que integran Brasil, Argentina, Paraguay, Uruguay y Venezuela.

La lengua guaraní, pertenece, junto con las lenguas chiriguano y mbyá a la familia tupí-guaraní. Dentro de la familia, las lenguas específicamente denominadas guaraníes son ocho, y cuentan con más de cinco millones de hablantes en todo el continente. En Argentina se hablan cuatro lenguas de esta familia: 1. El chiriguano, hablado por chiriguano, chañes y tapietés en las provincias de Salta y Jujuy; 2. El guaraní correntino, hablado fundamentalmente en la provincia de Corrientes; también, en Chaco, Misiones, Formosa y, debido a migraciones, en Rosario y Buenos Aires; 3. El guaraní paraguayo, hablado por migrantes paraguayos; 4. El mbyá, hablado por los mbyás en Misiones. El guaraní correntino o Goyano es hablado en las provincias de Corrientes, Chaco, Misiones, Formosa y, producto de las migraciones, en otros centros urbanos como Santa Fe y Buenos Aires. A su vez, los migrantes paraguayos hablan el guaraní paraguayo. La lengua mbyá es hablada en Misiones por los indígenas mbyás mientras que el chiriguano es hablado por los indígenas chiriguano, tapietés y chanés de las provincias de Salta y Jujuy.

4.4 *La familia mataco-mataguaya*

Según Censabella (2005), otras lenguas chaqueñas habladas en Argentina, el wichí, el nivacle y el chorote, corresponden a la familia mataguaya. El wichí

es la lengua que cuenta con mayor número de hablantes alrededor de sesenta mil personas. Se localiza en las provincias de Chaco, Formosa y Salta. En tanto que los nivaclés y los chorotes, se encuentran en Salta. La lengua vilela también se conserva, aunque con muy pocos hablantes y es una lengua en serio riesgo de desaparición.

4.5. *El mapuche*

El pueblo mapuche habita en las provincias de Neuquén, Río Negro, La Pampa y Chubut. Hay también grupos en la provincia de Buenos Aires. Según Virkel (2004: 243), dada la vigencia actual y la difusión del mapuche, el mismo puede ser considerado como la lengua indígena más relevante de la América Austral. Además, señala que, en el medio rural de Chubut, se observa gran cantidad de préstamos de esta lengua relacionados con la flora, la fauna, la topografía, la alimentación, la vestimenta, los instrumentos musicales y la religión.

4.6 *La familia chon*

A esta familia lingüística de la Patagonia pertenecen las lenguas extintas de Tierra del Fuego, selk'nam y haush, así como el tehuelche, que se encuentra hoy en extinción y cuya gramática ha sistematizado Fernández Garay (2002).

5. Metodología, fuentes y corpus de trabajo

De entre todas las denominaciones empleadas en las investigaciones para hacer referencia a las unidades estudiadas por la fraseología convenimos utilizar la denominación de Unidad Fraseológica (UF). Esta denominación está estrechamente vinculada al término que designa la disciplina, disponen de una alta aceptabilidad internacional y engloban los diferentes tipos de construcciones fijas. Es decir, siguiendo a Corpas Pastor (1996) son unidades léxicas formadas por dos o más palabras gráficas en su límite inferior, cuyo límite superior se sitúa en el nivel de la oración compuesta. Debido a la gran variedad de tipos de UF presentes en las lenguas y ante la imposibilidad de poder tratar todas estas categorías fraseológicas, decidimos limitar nuestro estudio a las construcciones denominadas “construcciones verbales fijas” CVF (Mogorrón, 2008, 2010, 2012) y en las que se integran las siguientes

construcciones fijas: 1. locuciones verbales; 2. colocaciones verbales; 3. verbos soportes; 4. construcciones verbales comparativas.

El subcorpus que describimos a continuación y que utilizaremos como base de nuestro trabajo forma parte de la base de datos multilingüe del grupo de investigación *Frasyttram* de la Universidad de Alicante que cuenta con cerca de 45 000 UF registradas partiendo del español peninsular e incluyendo variantes diatópicas del español de América y sus equivalentes a numerosas lenguas.

5.1 Fuentes, criterio de delimitación, selección y clasificación

El corpus de trabajo centrado en UF del español de Argentina pertenecientes al campo conceptual carácter-forma de ser-actitud-conducta con componentes de las lenguas originarias proviene de dos fuentes. Por un lado, utilizamos obras lexicográficas y por el otro, recurrimos a datos aportados por informantes a partir de la realización de encuestas. Consideramos como hipótesis de partida que, al combinar dos tipos de fuentes, los datos con los que contamos son más completos y reflejarán UF de reciente creación o que no están reflejadas en las obras lexicográficas. A continuación, explicamos en detalle las fuentes utilizadas.

5.2 Las fuentes lexicográficas

En la práctica lexicográfica, la recolección de material léxico puede darse por dos vías principales: a través de fuentes secundarias, recogiendo el conjunto de materiales útiles de otros diccionarios y estudios lexicográficos, o por medio de fuentes primarias, aprovechando un corpus de materiales originales como textos, grabaciones, encuestas, entre otros (cf. Haensch 1997: 54).

Dado que es sabido que todo corpus debe cumplir con el criterio de autenticidad de los datos, hemos considerado relevante prestar especial atención a las fuentes utilizadas para recoger las UF del corpus. A tal fin optamos por utilizar dos tipos de fuentes diversas: Por un lado, hemos utilizado nueve fuentes lexicográficas diferentes que recogen UF argentinas y, por el otro, hemos trabajado con informantes argentinos. Así, hemos utilizado como fuentes lexicográficas los siguientes diccionarios: *Diccionario de coloquialismos y términos dialectales del español* (2011), *Diccionario fraseológico del habla*

argentina. Frases, dichos y locuciones (2010), *Diccionario de americanismos* (2010), *Gran diccionario de los argentinos* (2009), *Diccionario del habla de los argentinos* (2008), *Diccionario de hispanoamericanismos no recogido por la Real Academia* (2006), *Diccionario de uso del español de América y España. Diccionario Vox* (2003), *Diccionario del verbo español, hispanoamericano y dialectal* (2000), *Breve diccionario ejemplificado de americanismo* (1999).

Partiendo de estos rasgos de selección el corpus final incluye únicamente UF que no aparecen en el *Diccionario de la lengua española* (DRAE), editado por la Real Academia de la Lengua. Debemos destacar que la edición actual, que es la vigesimotercera edición, fue publicada en 2014 y, aunque nos ofrece un total de 93.111 entradas frente a las 88.431 de la edición anterior con un total de 195.439 acepciones y se han introducido cerca de 140.000 enmiendas que afectan a unos 49.000 artículos, no se registran las UF de nuestro corpus. Como resultado de la aplicación de estos criterios de delimitación ha quedado conformado un corpus lexicográfico de un total de 1989 UF. El gráfico 1 muestra el porcentaje de UF que aporta al corpus cada fuente lexicográfica.

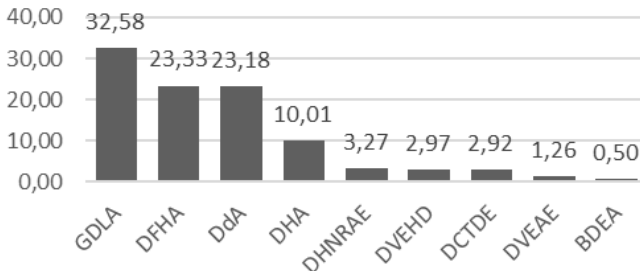


Gráfico. 1. Porcentaje de UF por fuente consultada (%).¹

1. GDLA. Gran diccionario de los argentinos. 32,58% (648)
DFHA. Diccionario fraseológico del habla argentina. Frases, dichos y locuciones. 23,33% (464)
DdA. Diccionario de Americanismos. 23,18% (461)
DHA. Diccionario del Habla de los argentinos. 10,01% (199)
DHNRAE. Diccionario de hispanoamericanismos no recogidos por la Real Academia Española. 3,27% (65)
DVEHD. Diccionario del verbo español, hispanoamericano y dialectal. 2,97% (25)
DCTDE. Diccionario de coloquialismos y términos dialectales del español. 2,92 % (58)

5.3 *Los informantes y las encuestas*

Dada la dinámica del lenguaje y debido a que las unidades de lengua que usa una comunidad lingüística no necesariamente se encuentran incluidas en los diccionarios, también hemos recurrido a la realización de encuestas a informantes residentes en Argentina y pertenecientes a diferentes grupos etarios y con distinto nivel académico. Estas encuestas fueron realizadas entre los meses de julio y agosto de 2015 en Buenos Aires. En total hemos recurrido a una treintena de informantes pertenecientes a tres grupos etarios.

El primer grupo está compuesto por ocho personas de entre veinte y treinta y cinco años de edad, el segundo grupo está integrado por doce personas de entre treinta y seis y cincuenta años de edad y el último grupo lo forman diez personas de más de cincuenta años.

En cuanto a la procedencia de los informantes seleccionados, la mitad, es decir, quince, son oriundos de Buenos Aires tanto de la Capital Federal (Ciudad Autónoma de Buenos Aires) como del conurbano bonaerense. Los quince restantes viven en localidades próximas situadas en la Provincia de Buenos Aires (La Plata, Luján, Mar del Plata, entre otras). Cuatro informantes nacieron en otras provincias como Córdoba, Santa Fe, Misiones o Tucumán, pero están afincados desde hace más de quince años en las localidades bonaerenses mencionadas.

En cuanto al nivel de estudios de los encuestados hemos constatado que el 60% (18) tiene estudios medios; aquellos con estudios superiores representan un porcentaje menor, del orden del 24% (7), seguido de los de estudios primarios con un 16% (5).

El cuestionario que utilizamos para nuestro trabajo tiene un doble planteamiento que contempla tanto el reconocimiento de las UF como el registro de posibles variaciones. En este contexto tomamos en consideración los criterios de juicio, de necesidad y de prestigio propuestos por Haensch (1993). Siguiendo el juicio de uso, los hablantes de una lengua saben perfectamente las unidades que se utilizan con frecuencia; juicio de necesidad, si esa unidad léxica se utiliza sólo en ciertos ámbitos, pero dentro de éstos es fundamental; juicio de prestigio, cuando una unidad se utiliza poco pero su uso es necesario

DUEAE. Diccionario de uso del español de América y España. 1,26% (59)

BDEA. Breve diccionario de americanismos. 0,50% (10)

en determinadas materias y acabará generalizándose (Haensch 1993). Este cuestionario de reconocimiento incluye las UF argentinas seleccionadas de las fuentes lexicográficas. Los informantes debían leer las UF que aparecían en el cuestionario y señalar las UF que les resultaran desconocidas o consideraran que están en desuso. Como resultado de la valoración de la totalidad de las respuestas proporcionadas por los encuestados podemos concluir que un 9% de las UF pertenecientes al corpus de trabajo pueden ser consideradas en desuso. Haciendo un análisis total de las UF sometidas a la encuesta y tomando como requisito mínimo que al menos cinco encuestados hubieran señalado la UF en desuso, podemos decir que este porcentaje se reduce al 5%.

5.4 Selección de las UF argentinas con componentes de las lenguas originarias

Teniendo en cuenta las dos fuentes y los criterios de delimitación y selección utilizados hemos seleccionado del corpus total recopilado las UF del español hablado en Argentina que contienen lexemas provenientes de lenguas originarias. Hemos obtenido un total de 101 UF lo que supone que un 5% del total inicial de UF tienen entre sus componentes términos de las lenguas originarias. Se utilizan un total de 44 términos provenientes de las siguientes lenguas originarias argentinas: tehuelche (*gualicho*), guaraní (*caracú, catinga, maraca, matete*), mapuche (*ají, guayo, laucha, pichi, pilcha, poncho, pucho*), quechua (*aca, antarca, añapa, api, cancha, carpa, caucho, chala, chancua, chancha, chocho, choclo, chongo, chúcaro, chucho, concha, cucuvica, gaucho, guacho, guampa, guanaco, guasca, mate, nana, ñaupa, ocote, ojota, papa, poroto, poroto, pupo, quincho, suri, yapa, yuyo, yuyo, zapallo*).

Observamos que únicamente están representadas en las UF del corpus 4 lenguas indígenas pertenecientes a 3 familias lingüísticas. No se han encontrado registros de UF que contuvieran términos pertenecientes a las lenguas que integran la familias guaycurú, así como tampoco de la familia mataco-mataguaya.

La distribución de términos por lenguas indígenas es, en orden descendente, la siguiente: quechua (67), guaraní (14), mapuche (13), tehuelche (7).

Vemos a continuación (Tabla 1) un ejemplo por cada término de las lenguas originarias que está integrado en una UF:

Término	UF	Lengua originaria
aca	<i>no valer (alguien) ni aca</i>	quechua
ají	<i>ser (alguien) más malo que un ají</i>	mapuche
antarca	<i>caerse (alguien) antarca</i>	quechua
añapa	<i>hacer (alguien) ñapa (algo)</i>	quechua
api	<i>hacer (alguien) api (algo)</i>	quechua
cancha	<i>abrir (alguien) cancha</i>	quechua
caracú	<i>meter (alguien) la pata hasta el caracú.</i>	guaraní
carpa	<i>hacer (alguien) (algo) con carpa</i>	quechua
catinga	<i>ser (alguien) un negro catinga</i>	guaraní
caucho	<i>ser (alguien) de caucho</i>	quechua
chala	<i>fumar (alguien) en chala (a alguien)</i>	quechua
chancua	<i>hacer (alguien) chancua (algo)</i>	quechua
chocho	<i>estar (alguien) chocho</i>	quechua
chúcaro	<i>ser (alguien) chúcaro/a</i>	quechua
chucho	<i>andar (alguien) con chucho</i>	quechua
concha	<i>hacer (alguien) concha (algo)</i>	quechua
cucuvica	<i>hacer (alguien) cucuvicas</i>	quechua
gaucho	<i>ser (alguien) gaucho</i>	quechua
gualicho	<i>dar (a alguien) gualicho</i>	tehuelche
guampa	<i>clavar (alguien) las guampas</i>	quechua
guanaco	<i>ser (alguien) un guanaco</i>	quechua
guasca	<i>dar (alguien) guasca (a alguien)</i>	quechua
guayo	<i>colgar (alguien) los guayos</i>	mapuche
laucha	<i>ser (alguien) una laucha</i>	mapuche
maraca	<i>ser (alguien) una maraca /un maraca.</i>	guaraní
mate	<i>estar (alguien) (bien/mal) del mate</i>	quechua
matete	<i>hacerse (alguien) un matete</i>	guaraní
nana	<i>hacerse (alguien) nana</i>	quechua
ñaupa	<i>ser (algo o alguien) del tiempo del ñaupá</i>	quechua
ocote	<i>estar (alguien) con el ocote afuera</i>	quechua

ojota	<i>parar (alguien) la ojota</i>	quechua
papa	<i>ser (algo) una papa</i>	quechua
pichi	<i>ser (alguien) un pichi</i>	mapuche
pilcha	<i>empilcharse (alguien)</i>	mapuche
poncho	<i>pisar (alguien) el poncho</i>	mapuche
poroto	<i>anotarse (alguien) un poroto</i>	quechua
pucho	<i>fiar(le) (alguien) un pucho (a alguien)</i>	mapuche
pupo	<i>rascarse (alguien) el pupo</i>	quechua
quincho	<i>tener (alguien) (un) quincho</i>	quechua
suri	<i>hacerse (alguien) el suri</i>	quechua
yacaré	<i>ser (alguien) yacaré</i>	guaraní
yapa	<i>estar (alguien) de yapa</i>	quechua
yuyo	<i>echarse (alguien) a los yuyos</i>	quechua
zapallo	<i>ser (alguien) un zapallo</i>	quechua

Tabla 1. UF argentinas con componente lenguas originarias

Mostramos a continuación fichas fraseológicas que hemos elaborado para cada una de las UF a partir de la información recogida de las fuentes nombradas:

1. UF	2. Definición de la expresión
<i>ser (alguien) yacaré</i>	Se utiliza para denominar a los oriundos de la provincia de Corrientes, por la presencia de este animal, especialmente, en la zona de los Esteros del Iberá.
	3. Fuente de la expresión
	Encuesta
4. Ejemplo de uso, fuente, fecha, enlace.	
Leonardo Mayer, es un yacaré bravo, que dejó a Argentina en un grupo mundial. http://diarioepoca.com/365259/mayer-un-yacare-bravo-que-dejo-a-argentina-eelgrupo-mundial/15.09.2016	
5. Observaciones	
Según el DRAE un <i>yacaré</i> es América Meridional un caimán, un reptil emidosaurio.	

Ficha fraseológica I. Lengua originaria guaraní

1.UF	2. Definición de la expresión
<i>abrir (alguien) cancha</i>	Coloquial. Apartar o eliminar los obstáculos para llegar a un lugar o para conseguir un objetivo
	3. Fuente de la expresión
	GDLA
4. Ejemplo de uso, fuente, fecha, enlace.	
Dejó afuera a Mauricio Pineda y se inclinó por jugar con líbero (Roberto Ayala) y stoppers (Nelson Vivas, quien entra por el marginado Gustavo Lombardi, y Roberto Sensini). Así le <i>abrió cancha</i> a Sergio Berti, quien se moverá como carrilero, pero con obligaciones defensivas, al igual que Diego Simeone. http://edant.clarin.com/diario/1998/04/22/r-06201d.htm , Clarín, 22.04.1998	
5. Observaciones	
El DRAE registra <i>cancha</i> con origen en el quechua <i>kancha</i> y con el significado de, recinto, patio. Además, el diccionario incluye cuatro locuciones que contienen el vocablo <i>cancha</i> , la primera es de uso general, <i>dar cancha a alguien</i> con el significado de 'reconocerle la capacidad de actuar conforme a su voluntad en un determinado asunto'. Las siguientes dos son de uso únicamente en Cuba, <i>estar alguien en cancha</i> 'estar bien preparado o entrenado para algo determinado' y <i>estar alguien fuera de cancha</i> . 'No estar preparado o entrenado para algo determinado'. La última locución citada es <i>estar en su cancha</i> 'estar en su elemento' es de uso en Chile, Nicaragua y Paraguay.	

Ficha fraseológica II. Lengua originaria quechua

5.5 Contraste de las UF con un corpus extenso

Una vez establecido el corpus de UF del español de Argentina del campo conceptual carácter-forma-de-ser-actitud-conducta que cuentan entre sus componentes con términos provenientes de las lenguas originarias nos proponemos establecer a partir de la utilización de grandes corpus, por un lado, la frecuencia de uso de cada UF que los integra en textos periodísticos y, por el otro, detectar posibles variantes no registradas aún.

Hemos utilizado la herramienta en línea *Sketch Engine* (SkE) que nos permite realizar un análisis textual de corpus haciendo especial énfasis en las concordancias y patrones distribucionales de las palabras (Kilgariff et al 2014). Esta herramienta dispone de numerosas aplicaciones que nos permite generar listados de palabras frecuentes, observar las concordancias, recuperar colocaciones y paquetes léxicos para estudiar la distribución de las palabras,

sus expresiones frecuentes o combinatoria léxica (Zanettin 2012). *Sketch Engine* (SkE) permite explotar diferentes tipos de corpus, por un lado, cuenta con Corpus en diversas lenguas ya compilados a los que da acceso; y, por otro lado, permite crear nuestro propio corpus subiendo archivos a la plataforma o crearlo automáticamente con *WebBootCat* a partir de palabras clave, entre otras funciones.

En nuestro caso, hemos decidido trabajar en base al corpus esAmTen-Ten11 (español de América) que cuenta con 8641717816 tokens, 7475579365 palabras y 17913256 documentos que ha sido alineado automáticamente mediante el programa *Freeling pipeline v5*. El corpus refleja 18 variedades nacionales del español de América, aunque hay que señalar que la variedad argentina es la más extensa y representa con 3338970830 tokens el 39% del total del corpus.

6. Resultados

6.1. Frecuencia de uso de las UF

La herramienta *Sketch Engine* permite buscar contextos en los que aparecen ciertas combinaciones de palabras o lemas incluyendo las formas flexionadas. Esta cuestión es de gran relevancia si tenemos en cuenta las particularidades de la variedad argentina relacionadas con, por ejemplo, el voseo y las consecuentes variaciones en el paradigma verbal. Además, la herramienta permite limitar nuestra búsqueda a páginas web de periódicos con las extensiones .edu, .gov, .org y .com del dominio .ar asignado a las páginas web argentinas. Hemos recuperado la información del corpus buscando las 101 UF argentinas que contienen componentes de las lenguas indígenas para determinar la frecuencia de cada una de ellas. Observado que las UF del corpus muestran tres rangos diferenciados de frecuencias; las hemos dividido, por lo tanto, en 3 categorías: muy frecuente (presencia en más de 400 páginas web/ frecuencia $\geq 0,12$); frecuente (presencia en más de páginas 146 web/ frecuencia $\geq 0,01$); poco frecuente (presencia en menos de 10 páginas web/ frecuencia de $\leq 0,01$).

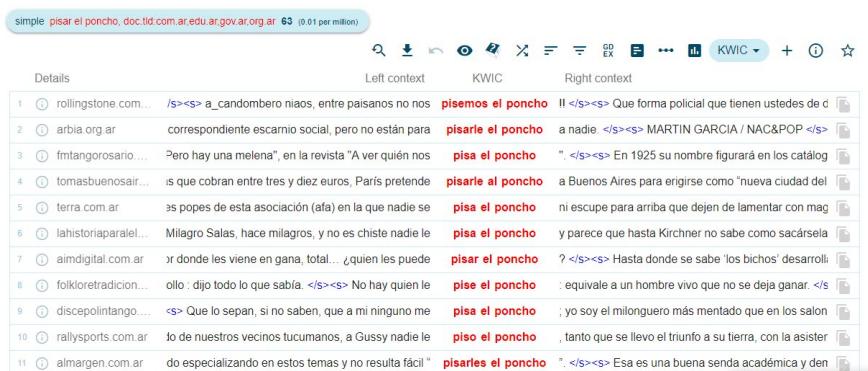
Los resultados muestran que el 22% de las UF del corpus, un total de 22, se utilizan muy frecuentemente en textos periodísticos, entre las que predominan las UF con términos provenientes del quechua.

Como muestra del rango de UF más frecuentes, vemos a continuación la UF *marcar la cancha* (*a alguien*) que se ha encontrado en 2000 páginas web y presenta una frecuencia de 0,6 por millón.



Captura de pantalla 1. UF *marcar* (*alguien*) *la cancha* (*a alguien*)-Sketch Engine

Hemos constatado que el 43% de las UF del corpus, es decir, 43 UF con componentes de las lenguas originarias, son de uso frecuente en los textos periodísticos argentinos, mientras que el 23% de las UF, un total de 23, reflejan poca frecuencia. Vemos a continuación un ejemplo de la UF *pisar(le)* (*alguien*) *el poncho* en la que podemos observar un registro poco frecuente ya que se ha encontrado en 63 páginas web argentinas y representa, por tanto, una frecuencia de 0,01 por millón.



Captura de pantalla 2. UF *pisar(le)* (*alguien*) *el poncho* (*a alguien*)-Sketch Engine

Un 13% de las UF no está registrado en el corpus consultado. De este modo, UF como *caerse (alguien) antarca* ‘caerse hacia atrás, sorprenderse mucho’, *hacerse (alguien) cucuvas* ‘dañarse, lastimarse’, *hacerse (alguien) el suri* ‘hacerse el tonto’, ‘desentenderse’, *estar (alguien) con el ocote afuera* ‘estar malhumorado’ o *colgar (alguien) los guayos* ‘morir’, están, según el corpus periodístico, en desuso.

Cabe destacar que se produce una coincidencia entre la información que aporta el corpus de textos periodísticos argentinos y los datos suministrados por los informantes ya que éstos, siguiendo el juicio de uso, habían señalada en la encuesta de reconocimiento las mismas 5 UF como en desuso.

En cuanto la frecuencia total de las UF con componentes de las lenguas indígenas argentinas, podemos decir que existe una relación directa entre la cantidad de UF aportadas por cada una de las lenguas y su frecuencia de uso en textos periodísticos. Es decir, el quechua, aporta 67 UF al corpus total y tiene el mayor número de UF muy frecuentes, seguido del guaraní y el mapuche que aporta un total de 14 UF y 13 UF respectivamente y tienen también un número menor de UF frecuentes. Entre las UF poco frecuentes predominan las que tienen entre sus componentes términos tehuelches. Sin embargo, entre las UF en desuso predominan las UF del quechua.

En la siguiente tabla vemos las UF con términos de las distintas lenguas en orden de importancia numérica decreciente, además de la contribución léxica relativa de cada una de ellas y la UF más frecuente.

Lengua originaria	Términos	UF más frecuente	UF totales
Quechua	37	<i>marcar (alguien) la cancha (a alguien) ‘ponerle límites a alguien’</i>	67
Guaraní	4	<i>hacerse (alguien) un metete ‘confundirse’</i>	14
Mapuche	7	<i>Pisarle (alguien) el poncho (a alguien) ‘abrumar o dominar a alguien, especialmente en una situación de rivalidad’</i>	13
Tehuelche	3	<i>hacerle un gualicho a alguien ‘embrujar a alguien’</i>	7

Tabla 2. Número de términos y UF aportados, UF más frecuentes por lengua

6.2 Variaciones detectadas

La importación de elementos o la copia de patrones (Heine y Kuteva 2005) de las lenguas originarias puede dar lugar a cambios inducidos por contacto (cambios indirectos) que modifican patrones de uso y estructuras ya existentes en la lengua, suprimen o amplían las restricciones lingüísticas de un fenómeno o amplían la frecuencia de uso de un fenómeno conocido en la lengua (Palacios 2010). La finalidad de estos cambios suele obedecer a estrategias comunicativas que permiten al hablante una comunicación más satisfactoria aprovechando recursos de ambas lenguas. Estos cambios pueden afectar también, como se ha observado, al plano fraseológico. Se trata de UF que no presentan diferencias en el significado, es decir, son parcialmente idénticas en su estructura y en sus componentes. En nuestro estudio nos hemos centrado en variantes que se producen mediante sustitución de una parte de la UF siendo siempre constantes los términos de las lenguas originarias y nos ceñimos únicamente a las variantes no registradas en las fuentes lexicográficas utilizadas para la conformación del corpus.

Con el fin de sistematizar los casos de variaciones detectadas a partir de la utilización del corpus EsTenTen11 y de la aplicación de la herramienta *Sketch Engine*, tomamos como base el esquema variacional de Montoro del Arco (2004) adaptado al español de Argentina.

Observamos, variantes internas léxicas, variantes internas morfo-léxicas, variantes estructurales, así como inflexiones internas morfológicas.

Dentro de las variantes internas léxicas encontramos tanto UF que sustituyen un lexema por otro (*marcarle/dibujarle (a alguien) la cancha, dar/hacer (a alguien) gualicho*) como las que sustituyen un lexema por un sintagma que completa el hueco funcional dentro de la sintaxis interna de la UF (*clavar (alguien) las guampas/el guampo/la guampada/de la guampa y los guampos*). En esta última variante *la guampa y los guampos* funciona como una única unidad léxica dentro de la UF.

En el caso de las variantes fraseológicas por alteración del núcleo verbal múltiple el hablante tiene varias opciones para sustituir el verbo que encabeza la UF a la hora de actualizar el discurso, por ejemplo, cinco en el caso de la UF *echarse/estrellarse/caerse/irse/mandarse (alguien) a los yuyos*.

Por el contrario, en el caso de las variantes fraseológicas por alteración del núcleo verbal simple, existe únicamente una alternativa posible como en la UF *marcarle* o *dibujarle* (*a alguien*) *la cancha*.

En cuanto a las variantes fraseológicas por alteraciones del núcleo nominal, hemos detectado variantes del tipo *ser* (*alguien*) *de caucho/ de goma* o *ser* (*alguien*) *del tiempo/de la época del ñaupa*.

Las variantes internas morfoléxicas, es decir, la modificación léxica de alguno de los componentes de la UF, que hemos identificado en el corpus periodístico coinciden con lo que afirma García-Page (1999) ya que la mayoría de estos cambios corresponden, más que a la morfología derivativa propiamente dicha, a la llamada afijación apreciativa (aumentativos, diminutivos, superlativos, etc.) como es el caso de *ser* (*alguien*) *gaucho/gauchito* o *hacerse* (*alguien*) *nana/nanita*.

Las variaciones estructurales se dan, por un lado, a partir de la modificación léxico-cuantitativa o por extensión son equiparables a las “abreviaturas fraseológicas” que define García-Page (1999) como expresiones que admiten bien incrementación-gramatical o léxica-, bien su reducción. Hemos detectado UF que presentan este tipo de variación como las siguientes: *ser* (*alguien*) (*un*) (*negro*) *catanga*, *ser* (*alguien*) (*un*) *pichi*, *ser* (*alguien*) (*un*)/ (*una*) *laucha*, *ser* (*alguien*) (*un*)/(*una*) *maraca*, *tener* (*alguien*)(*un*) *quincho*.

Por el otro, encontramos como variantes estructurales, las inflexiones internas morfológicas o sintácticas. Son ejemplos variantes estructurales con inflexiones internas morfológicas las que comportan un cambio de género o número como las siguientes: *ser* (*alguien*) *gaucho/gaucha*, *ser* (*alguien*) *gauchito/gauchita*, *pisarse* (*alguien*) *la(s) guasca(s)*.

Si bien el *vesre*, la inversión del orden de las sílabas en algunas palabras, es muy propio del habla argentina y es un caso claro de variante por cambios en el ordenamiento, sin embargo, no hemos identificado ninguna variante de este tipo en el corpus de UF con términos de las lenguas originarias. Esta ausencia puede tener relación con lo que afirman Gobello & Olivieri (2005) ya que, según estos autores, se trata de cierto modo de hablar peculiar porteño. Es decir, le adscriben un uso local restringido a la Capital Federal.

7. Conclusiones

A partir de las muestras seleccionadas y metodología empleada en esta investigación, se puede concluir que (1) la fraseología y la fraseografía del español hablado en Argentina son disciplinas cuyo desarrollo es incipiente y aún son necesarias nuevas aportaciones que permitan actualizar y describir con más detalle el acervo fraseológico. Se requieren, además, estudios diferenciales y contrastivos orientados a identificar puntos de convergencia entre las diferentes variantes del español y otras lenguas. (2) A pesar de la falta de datos exactos de la cantidad de hablantes de cada una de las lenguas, la contextualización de la situación actual nos permite identificar doce lenguas indígenas pertenecientes a cinco familias lingüísticas diferentes, e identificar las regiones de contacto especialmente en los centros urbanos. (3) En relación a la presencia de términos provenientes de las lenguas originarias en las UF de los corpus pertenecientes al campo conceptual carácter-forma de ser-actitud-conducta y no recogidas en el *Diccionario de la lengua española* de la Real Academia de la Lengua (DRAE), hemos identificado un total de 101 UF lo que supone un 5% del total inicial de UF. En cuanto a los términos utilizados, se han identificado 44 cuya distribución por lenguas indígenas es, en orden descendente, la siguiente: quechua (67), guaraní (14), mapuche (13), tehuelche (7). No se han encontrado registros de UF que contuvieran términos pertenecientes a las lenguas que integran la familia guaycurú, así como tampoco de la familia matak-mataguaya.

Al establecer la frecuencia de uso en textos periodísticos de un corpus de UF hemos constatado que, existe una relación directa entre la cantidad de UF aportadas por cada una de las lenguas y su frecuencia de uso en textos periodísticos. El quechua tiene el mayor número de UF más frecuentes, seguido del guaraní y el mapuche. Ambas lenguas tienen también un número menor de UF frecuentes. Entre las UF poco frecuentes predominan las que tienen términos tehuelches entre sus componentes. Sin embargo, entre las UF en desuso predominan las UF del quechua. Estas últimas han sido identificadas de modo coincidente por la información que aporta el corpus de textos periodísticos argentinos y los datos suministrados por los informantes.

El trabajo con el corpus *EsTenTen11* y la aplicación de la herramienta *Sketch Engine* nos ha permitido también (4) identificar los tipos de variaciones

de las UF del corpus registradas en el corpus periodístico, y que no están reflejadas en los diccionarios consultados. En cuanto a los tipos de variantes, se han detectado variantes interna léxicas, variantes internas morfo-léxicas, variantes estructurales, así como inflexiones internas morfológicas.

Por último, entre los resultados más destacados, podemos señalar, a nivel metodológico, que la combinación de fuentes lexicográficas que reflejan la variedad argentina, el trabajo con informantes apoyado en encuestas, la utilización del corpus extenso *EsTenTen11* y la herramienta *Sketch Engine*, nos ha permitido obtener datos relacionados con la frecuencia de uso de las UF con términos de las lenguas originarias argentinas y completar la información disponible de la variación fraseológica.

Si bien nuestro trabajo se restringió a la fraseología del español hablado en Argentina y a un único campo semántico, creemos que esta metodología puede replicarse también a otras variantes diatópicas con el fin de conocer el alcance de la influencia de las lenguas originarias en la fraseología actual del español.

Referencias bibliográficas

- BOLAÑOS-FABRES, Patricia. (2015) “Expresiones populares y el contacto de lenguas: Ecuador.” *Paremia* 24, pp. 137-144.
- CASADO VELARDE, Manuel. (2005) *Unidades fraseológicas y variación diatópica: para una fraseología dialectal extremeña*. Filología y lingüística. Madrid: CSIC, UNED & Universidad de Valladolid.
- CENSABELLA, Marisa (2005) *La revitalización de las lenguas y la educación bilingüe intercultural*. V Encuentro de Lenguas Aborígenes y Extranjeras. Salta: Departamento de Lenguas Modernas de la Universidad Nacional de Salta.
- Censo Nacional de Población, Hogares y Viviendas. (2010) Censo del Bicentenario. Resultados Definitivos, Serie B N°2. Tomo I. 2012. https://www.indec.gov.ar/ftp/cuadros/poblacion/censo2010_tomo1.pdf
- CORPAS PASTOR, Gloria. (1996) *Manual de fraseología Española*. Madrid: Gredos.
- CORPAS PASTOR, Gloria. (2018) “Laughing one’s head off in Spanish subtitles: a corpus-based study on diatopic variation and its consequences for translation”. En: Mogorrón Huerta, Pedro & Juan Antonio Albaladejo-Martínez (eds.) 2018. *Fraseología, Diatopía y Traducción. Phraseology, Diatopic Variation and*

- Translation*. [IVITRA Research in Linguistics and Literature, 17]. Amsterdam/Philadelphia: John Benjamins Publishing Company. pp.32-71.
- COSERIU, Eugenio. (1990) *Teoría del lenguaje y lingüística general*. Madrid: Gredos.
- CUADRADO-REY, Analía. (2016) *Base de datos de UF del español de Argentina: propuesta de traducción al alemán*. Alicante: Universidad de Alicante. Tesis doctoral.
- CUADRADO-REY, Analía. (2018) “La variación fraseológica en el español hablado en Argentina: propuesta de clasificación”. En: Mogorrón Huerta, Pedro & Juan Antonio Albaladejo-Martínez (eds.) 2018. *Fraseología, Diatopía y Traducción. Phraseology, Diatopic Variation and Translation*. [IVITRA Research in Linguistics and Literature, 17]. Amsterdam/Philadelphia: John Benjamins Publishing Company. pp. 334-352.
- FERNÁNDEZ GARAY, Ana. (2002) *Testimonios de los últimos ranqueles. Textos originales con traducción y notas lingüístico-etnográficas*. Buenos Aires: Facultad de Filosofía y Letras, Universidad de Buenos Aires.
- GARCÍA-PAGE, Mario. (1999) “Variantes morfológicas y unidades fraseológicas”. *Paremia*: 8, pp. 225-230.
- HEINE, Bernd & Tania KUTEVA. (2005) *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- INDEC. Encuesta Complementaria de Pueblos Indígenas (ECPI) 2004-2005 - Complementaria del Censo Nacional de Población, Hogares y Viviendas 2001. 2004-2005. Versión electrónica: <https://www.indec.gov.ar/ftp/indecinforma/nuevaWEB/cuadros/2/gacetilla_ecpi_05_07.pdf>
- KILGARRIFF, Adam; Vít BAISA; Jan BUŠTA; Miloš JAKUBÍČEK; Vojtěch KOVÁŘ; Jan MICHELFEIT; Pavel RYCHLÝ & Vít SUCHOMEL. (2014) “The Sketch Engine: ten years on”. *Lexicography* 1, pp.7-36.
- KOIKE, Kazumi. (2001) “Variación fraseológica del español. Varilex”. *Variación léxica del español en el mundo*, 9, pp. 77-92. Versión electrónica: <<http://lecture.ecc.u-tokyo.ac.jp/~cueda/varilex/public.htm>>
- LIPSKI, John. (2004) “El español en América: los contactos bilingües.” En: Cano, Rafael (ed.). *Historia de la lengua Española*. Barcelona: Ariel, pp. 1117-1138.
- MOGORRÓN HUERTA, Pedro. (2014a) “Locuciones verbales, traducción y pérdida de equivalencia”. En: Vanda Duarte (ed.) 2014. *Fraseología y paremiología: enfoques y aplicaciones*. Madrid: Centro Virtual Cervantes, pp. 79-96.

- MOGORRÓN HUERTA, Pedro. (2014b) “Importancia (numérica) de las variantes diatópicas españolas y su tratamiento en los diccionarios”. *Yearbook of Phraseology*: 5, pp. 123-144.
- MOGORRÓN HUERTA, Pedro. (2015a) “Comprensión, competencia fraseológica y creatividad de las expresiones fijas con sus variantes diatópicas, ejemplos en francés y en español”. Mogorrón Huerta, Pedro & Fernando Navarro Domínguez (eds.) 2015. *Fraseología Didáctica y Traducción*. Frankfurt am Main: Peter Lang, pp. 137-160.
- MOGORRÓN HUERTA, Pedro. (2015b) “La variation dans les Construction Verbales Figée de l’espagnol d’Amérique”. *Lingvisticae Investigationes*: 38:2, pp. 276-300.
- MONTORO DEL ARCO, Esteban. (2004) “La fraseología en la gramática española del siglo XX después de Casares”. *Anuario de Estudios Filológicos*, XXVII, pp.221-236.
- MORENO DE ALBA, José. (1993) *El español en América*. México: Fondo de Cultura Económica.
- NAVARRO-BROTONS, Lucía. (2011) “Traduire la langue, traduire la culture. Les cas des proverbes multiculturels mexicains ayant des voix du nahuatl.” *Synergies Tunisie*: 3, pp. 125-136.
- NAVARRO-BROTONS, Lucía & Analia CUADRADO-REY. (2016) “Las locuciones verbales somáticas españolas y sus variantes diatópicas argentinas y mexicanas”. Mogorrón Huerta, Pedro; Analia Cuadrado-Rey; Lucía Navarro-Brotóns & Iván Martínez Blasco (eds.) 2016. *Fraseología, variaciones y traducción*. Frankfurt am Main: Peter Lang, pp. 223-236.
- PALACIOS, Azucena. (2008) *El español en América: contactos lingüísticos en Hispanoamérica*. Barcelona: Ariel.
- PALACIOS, Azucena. (2010) “La lengua como instrumento de identidad y diferenciación: más allá de la influencia de las lenguas amerindias”. Castañer, Rosa María & Vicente Lagüéns (Coords.). *De moneda nunca usada: Estudios dedicados a José M^a Enguita Utrilla*, Zaragoza: Instituto Fernando El Católico, CSIC. pp. 503-514.
- PAUER, Gabriela. (2008) “Fraseología lingüística en el ámbito hispanorrioplatense: modismos que reflejan una identidad común y divergente.” *Estudios académicos*: 46, pp. 313-327.

- PENADES MARTÍNEZ, Inmaculada. (2012) *Gramática y semántica de las locuciones*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá de Henares.
- ROSENBLAT, Ángel. (2002) *El Español de América*. Caracas: Biblioteca Ayacucho.
- SAPIR, Edward. (1954) *El lenguaje: introducción al estudio del habla*. México: Fondo de Cultura Económica.
- SCIUTTO, Virginia. (2005a) “Unidades fraseológicas: un análisis contrastivo de los somatismos del español de Argentina y del italiano”. Versión electrónica: <http://cvc.cervantes.es/literatura/aispi/linguistica_contrastiva.htm>
- SCIUTTO, Virginia. (2015b) “Enunciados fraseológicos: Perspectiva morfosintáctica de los somatismos verbales del español de Argentina”. *E-Aesla*. Versión electrónica: <<http://cvc.cervantes.es/lengua/eaesla/pdf/01/57.pdf>>
- SIGUAN SOLER, Miguel. (2001) *Bilingüismo y lenguas en contacto*. Madrid: Alianza Editorial.
- SOTO DE MATULOVICH, Edda. (2008) “Unidades fraseológicas en la prensa escrita de Salta”. Actas del 10º Congreso REDCOM Conectados, hipersegmentados y desinformados en la Era de la Globalización. Salta: Universidad Católica de Salta. Facultad de Artes y Ciencias.
- VIDAL DE BATTINI, Berta. (1964) *El español de la Argentina*. Buenos Aires: Consejo Nacional de Educación.
- VIRKEL, Ana. (2004) *Español en la Patagonia. Aportes para la definición de un perfil sociolingüístico*. Buenos Aires: Academia Argentina de Letras.
- ZAMORA, Francisco. (1993) “Expresiones fraseológicas en una variedad del español estándar”. *Anuario de lingüística hispánica* 9, 347-357.
- ZANETTIN, Federico. (2012) *Translation Practices Explained: Translation-driven Corpora*. Manchester & Kinderhook (NY): St Jerome Publishing.

Diccionarios

- ACADEMIA ARGENTINA DE LETRAS. (2008) *Diccionario del habla de los argentinos*. Buenos Aires: Emecé Editores.
- ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA. (2010) *Diccionario de americanismos*. Madrid: Santillana.
- BARCIA, Pedro & Gabriela PAUER. (2010) *Diccionario fraseológico del habla argentina: frases, dichos y locuciones*. Buenos Aires: Emecé Editores.
- BARCIA, Pedro; Gabriela PAUER. (2013) *Refranero de uso argentino*. Buenos Aires: Emecé.

- BATTANER, Paz. (2003) *Diccionario de uso del español de América y España*. Barcelona: Spes Editorial.
- FITCH, Roxana. (2011) *Diccionario de coloquialismos y términos dialectales del español*. Madrid: Arco-Libros.
- GOBELLO, José & Marcelo OLIVERI. (2005) *Novísimo diccionario lunfardo*. Buenos Aires: Corregidor.
- REAL ACADEMIA ESPAÑOLA. (2014) *Diccionario de la lengua española*. Madrid: Planeta
- RENAUD, Richard. (2006) *Diccionario de hispanoamericanismos no recogidos por la Real Academia Española*. Madrid: Cátedra.
- SECO, Manuel; Olimpia ANDRÉS & Gabino RAMOS. (2004) *Diccionario fraseológico documentado del español actual. Locuciones y modismos españoles*. Madrid: Aguilar.
- SEGOVIA, Lisandro. (1911) *Diccionario de argentinismos, neologismos y barbarismos*. Buenos Aires: Coni.
- SUANCES-TORRES, Jaime. (2000) *Diccionario del verbo español, hispanoamericano y dialectal*. Barcelona: Herder.
- STEEL, Brian. (1999) *Breve diccionario ejemplificado de americanismos*. Madrid: Arco-Libros.
- TORNADÚ, Beatriz. (2009) *El gran diccionario de los argentinos: el uso del español actual de la Argentina*. Buenos Aires: Clarín. Arte Gráfico/ Ed. Argentino.

Corpus

esTenTen11, The Sketch Engine. Versión electrónica: < <http://www.sketchengine.co.uk>>

NOTA BIOGRÁFICA / BIONOTE

ANALÍA CUADRADO REY es doctora en Traducción e Interpretación por la Universidad de Alicante, Licenciada en Filología Alemana, Especialista universitaria en didáctica del alemán como lengua extranjera y postgraduada en Traducción Institucional. Es profesora ayudante doctora en el Departamento de Traducción e Interpretación de la Universidad de Alicante e integrante del grupo de investigación FRASYTRAM de la misma Universidad. Sus publicaciones más recientes se centran en la fraseología diatópica del español, la

didáctica de la fraseología y su traducción en el contexto de textos turísticos y literarios.

ANALÍA CUADRADO REY has a PhD in Translation and Interpreting from the University of Alicante, a Bachelor's Degree in German Philology, a University Specialist in the teaching of German as a foreign language and a Master's Degree in Institutional Translation. She is an assistant lecturer at the Department of Translation and Interpreting at the University of Alicante and is a member of the FRASYTRAM research group. Her most recent publications focus on the diatopic phraseology of Spanish, the didactics of phraseology and its translation in the context of tourist and literary texts.

Recibido / Received: 12/11/2019

Aceptado / Accepted: 03/02/2020

Para enlazar con este artículo / To link to this article:

<http://dx.doi.org/10.6035/MonTI.2020.ne6.12>

Para citar este artículo / To cite this article:

Recio Ariza, M^a Ángeles. (2020) "Equivalences of phraseological units in corpora." 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. 386-415.

EQUIVALENCES OF PHRASEOLOGICAL UNITS IN CORPORA

M.^a ÁNGELES RECIO ARIZA

recio@usal.es

University of Salamanca

Abstract

Language teaching in the education of translators-to-be must necessarily be based on Comparative Linguistics since the contrastive study of language is *conditio sine qua non* to explain and deepen the aspects that unite different languages and those that tell them apart. Translation and Phraseology studies have shown the role and importance of equivalences as well as their intercultural value. The search for equivalence, in this sense (Corpas Pastor, 1997; Sevilla, 2004, 2005; Mogorrón, 2008, etc.), becomes indispensable. This work attempts to analyse the treatment of Phraseological Units and their translation in German language teaching for translation purposes, in particular those PUs that appear in various corpora, which can shed light not only on their equivalences but also on their treatment in said corpora. Additionally, we will attempt to establish an inventory of *normal phraseological conventions* through the evidence found in the selected corpora, conducting a qualitative study.

Keywords: phraseological units; Comparative Linguistics; German language teaching for translation; communicative equivalence.

Resumen

La enseñanza de lenguas en la disciplina de Traducción e Interpretación debe basarse necesariamente en la Lingüística Comparada, ya que el estudio contrastivo de la lengua es *conditio sine qua non* para explicar y profundizar tanto en los aspectos que



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

unen a las diferentes lenguas como en los que las diferencian. Los estudios de traducción y de fraseología han demostrado la importancia de las equivalencias, así como su valor intercultural. En este sentido, la búsqueda de equivalencia (Corpas Pastor 1997; Sevilla 2004, 2005; Mogorrón 2008, etc.) se hace indispensable. Este trabajo pretende analizar el tratamiento de las unidades fraseológicas y su traducción en la enseñanza de la lengua alemana para su traducción, en particular, aquellas UF que aparecen en diversos corpus que pueden arrojar luz sobre las equivalencias y su tratamiento en estos. Además, intentaremos establecer un inventario de las convenciones fraseológicas normales, mediante la evidencia en los corpus seleccionados, realizando un estudio de carácter más cualitativo que cuantitativo.

Palabras clave: Unidades fraseológicas; lingüística comparada; enseñanza de la lengua alemana para la traducción; equivalencia comunicativa.

1. Introduction

This paper aims to analyse, through different corpora, the treatment of equivalences among the following phraseological units (hereafter, PU): *dein Wort in Gottes Ohr* [lit. your word in the ear of God], *mir fehlen die Worte* [lit. words fail me], *in anderen Worten* [lit. in other words], *ein gutes Wort einlegen* [lit. to put a good word] and *Wort für Wort* [lit. word by word], as well as the importance and usefulness of corpora in their translation.

In the translation process, central questions are posed regarding the approach to adopt and the role of the languages involved, both the mother tongue and the foreign language. In the specific case of phraseology, the objective is threefold: to transmit the knowledge and recognition of phraseological units (PUs), their syntax, and their translation.

The seemingly inextricable relationship between German and Spanish has drawn, for centuries, the attention of dozens of linguists, grammarians, lexicographers, translators or second language teachers who have considered it an exciting challenge to delve deeper into a study that is as essential as it is complex.

The selected PUs seem paradigmatic to us and they can help to establish an inventory of *normal phraseological conventions*¹. The evidence from various corpora may help us in this process by conducting a qualitative study.

In order to carry out the study below, the most relevant literature in the field is presented first, followed by a detailed analysis with specific examples, together with their treatment in the different monolingual and parallel corpora.

Finally, we will reflect on the advantages that the use of corpora as a tool in translation settings can offer, as well as their benefits in helping translators choose the most adequate equivalence.

2. The term *equivalence* in Translation and Phraseology Studies

Both Translation Studies and Phraseology Studies have shown, on the one hand, the role and importance of equivalence and, on the other, its intercultural value. In this sense, the search for equivalence (Corpas Pastor, 1997; Dobrovol'skij, 2000; Sevilla, 2004; Mogorrón, 2008, etc.) becomes essential, as well as the need to verify the relations of similarity and dissimilarity between languages. In recent years, equivalences have continued to be widely studied (Corpas Pastor (1996, 2008, 2017 and 2018); Coste Didier (2018); Sabban (2010); Mira Álvarez (2011); Trivic (2013), Leal (2008); Muñoz Medrano (2015); Szyndler (2014); Sozinho Quiraque (2016).

This work relies primarily on the research mentioned above, as well as on corpus linguistics, that is, the integration (*Einbettung*) of equivalence through its contributions, conclusions, and possible applications for translation.

Texts can change or lose their significant potential when they are translated inasmuch as certain aspects of the original language might get lost when transferred to the target language. In fact, not everything is translatable, and what may sound logical in a language might be completely meaningless in another. This is fundamentally due to the idiosyncrasy of each culture (Recio Ariza 2002). In this regard, we believe that phraseology is paradigmatic.

Phraseology is a branch of linguistics that encompasses several interdisciplinary paths (Corpas Pastor, Gloria & M^a Luisa Ortiz Álvarez 2017):

1. Hanks (2015) describes *normal phraseological conventions* as the patterns followed by the elements involved in the phraseological process.

contrastive and translation studies, constructional grammar, cognitive and psycholinguistic studies, and finally, applications of corpus linguistics and Natural Language Processing. For *Corpas Pastor* (2017: 273), there are a significant number of corpus-based studies in phraseology, both monolingual and contrastive, whose aim is either to exemplify theories or extract new premises and results. In our opinion, one of the greatest advantages of corpora is their ability to retrieve representative data, which is why they have become essential in translation and phraseology.

Corpas Pastor (2018: 16) indicates that the CREA and CORPES corpora, for example, cover peninsular and American varieties of Spanish. In her opinion, these reference corpora prove the difficulty of studying world Spanish. She also highlights that these corpora do not (or barely) take into account the diatopic restrictions of PUs and stresses that what makes the online corpora different is that they offer more possibilities for comparative analyses of frequency and use across the different varieties of Spanish. In her research, she carries out a study of the PU *laughing one's head off* in Spanish subtitles by means of a corpus analysis using the corpus manager tool Sketch Engine. By confirming that diatopic variation is avoided in the choice of translation equivalents, she concludes that there is a clear tendency to normalise and standardise the PU in subtitles. In our opinion, this can be extrapolated, to a certain extent, to translation in general.

2.1. *Equivalence of translated PUs*

As far as the research that has dealt with the study of equivalence in recent years is concerned, it is worth highlighting *Coste* (2018), who advocates introducing the concept of *reconversion*. In his view, a more normative than descriptive treatment of the term “equivalence” has been made, which implies paradoxes and potential contradictions:

This is how a text could be said to be equivalent or not to another and its functional value might be measured in comparison. The notion of equivalence in Translation Studies, insofar as it is always more normative than plainly descriptive, involves a number of presuppositions, implications and ambiguities that make it teem with paradoxes and potential contradictions.

According to this author, the good translator is the one who achieves the same impact on their readers as the original author.

In this sense, and quoting Sabban, translation always implies a loss to a certain extent, which takes the author to speak of *Verlusttopos* (2010: 193). In her opinion, phraseological units are contextualised in the translation of texts, so it is necessary to take into account the context and the variability of phraseological units that can be verified and studied much more easily by means of a corpus, especially given the fact that corpora provide clues on the semantic variability of a PU in its source text (ST). By the same token, there are no words without a context and, therefore, PUs should only be considered in combination with the speech and situational act. In this same line, Mira Álvarez (2011: 107) argues that what makes the task of establishing equivalents at the lexical and syntactic level especially complex is the need to consider the relevant connotative intentions, style, effect on the recipient, etc. Because something is always lost in the translation process and it is very difficult for full equivalence to exist, “translation becomes a search for equivalences”. Likewise, the author states that contrastive analyses help characterise the type of equivalence achieved in the translation of phraseological units.

Regarding PUs, Mira Álvarez agrees with García-Page (in Mira Álvarez 2011: 108), who considers that the classification of locutions formulated by Corpas Pastor (1997) and adopted by many scholars is valid as a starting point. To do so, he reviews the three spheres of PU proposed by this researcher, namely collocations, locutions, and phraseological statements. Traditionally, a locution is defined as a stable group of two or more words that functions as a lexical unit with its own meaning, which is not derived literally from the words that form it. Following the model proposed by Corpas Pastor (1996):

Locutions overlap in some aspects with free combinations of words and other complex units. Fundamentally, this type of PU differs from free word combinations by its institutionalisation, syntactic stability, and denotative function² (Corpas Pastor, 1996:89).

2. All references in Spanish have been translated into English by the author.

In her opinion, it is complicated to establish a distinction between locutions and compound words. Due to the lack of adequate criteria to establish this distinction clearly, Corpas Pastor (1997: 93) advocates “considering all lexical units formed by the graphic union of two or more bases as compounds”, while locutions would be “those units that (...) do not show a graphic union”.

One of the underlying difficulties of translating a PU is that its meaning does not coincide with the sum of the independent meanings of each of its individual components. Mira Álvarez claims (2011: 110) that “the analysis of the pragmatic value of the PU in the speakers’ discourse is carried out on the basis of Grice’s theory of conversational cooperation”. In his view, the maxims of quality, quantity, relation, and manner are *violated* when using phraseological units, while expressive, appreciative, and humorous connotations, etc., predominate in them. Mira Álvarez claims, therefore, that there is an indirect assessment of social interaction and behaviour according to the conventional and codified implications presented by Grice. As Mira Álvarez states (*ibid*), equivalence “is associated with the greater or lesser preservation of factors that are related to the semantic and pragmatic levels of these units”.

Some scholars, such as Hanks (2015), have even claimed that words themselves have no meaning, but have a potential meaning. Mira Álvarez argues that context is phraseology, hence we should try to find a way to observe and measure phraseology so as to know the patterns that are followed. In this regard, corpus analysis is paramount. According to the author, the conventional implications that underlie conversational cooperation are not grammatical or syntactic, nor are words. This may be one of the reasons why dictionaries do not contain enough phraseology. Similarly, the author states that words are ambiguous in nature. However, if they are contextualised, this ambiguity can be reduced or even eliminated.

Trivic (2013: 69) emphasises that absolute symmetry occurs when there is full equivalence in all three planes: formal, semantic and functional. We agree with this author in that the first phraseological studies dwelt mainly on delimitation and theoretical definition, as well as on the search for phraseological universals³. In this sense, the author cites Corpas Pastor (2003: 216):

3. Like many other phraseology scholars such as Corpas Pastor (1996, 2003) Trivic (2013), Dobrovel’skij (2000).

The search for interlinguistic correspondence represents the third and final phase in the process of translating a PU, which comes immediately after the phases of recognition or identification of a PU and its correct interpretation in context.

Like Trivic, we believe that lexical or phraseological gaps —the asymmetry between working languages— tend to reflect the idiosyncratic aspect of culture, as mentioned above. The author goes on to argue that this may be one of the reasons why lexical or phraseological gaps are neither conceptualised nor lexicalised in the form of PU in the target language. If we translate two of the examples the author gives into German: *decíamos ayer* [lit. as we said yesterday]⁴ and *contestar a la gallega* [lit. to give a Galician answer]⁵, presented also in Leal's study (2008), we can observe that they belong to the same semantic field: *sprechen, reden, Worte*. The latter is the example that will allow us in this study to illustrate and develop our hypothesis with respect to its occurrence in the various corpora consulted.

In this regard, Leal (2008: 104) claims that “what is interesting these days is the stylistic, discursive, textual and pragmatic aspects of phraseology” as well as “the psycholinguistic mechanisms of PUs”: their processing, their metaphorical and cultural base. In short, the historical, literary, ethnographic, and social realities of cultures underlie the difficulty in identifying the existence of equivalents. As we have already mentioned, the idiosyncrasy of each language and its peculiarities come into play here and this is why an equivalent is not always found. Furthermore, this explains the existence of full, partial or zero equivalence, as some PUs are unique to the culture of the language and its heritage.

In Muñoz Medrano's (2015) view, there has been an effort to determine and define the concept of equivalence in Translation Studies from several theoretical approaches. With regard to this, she quotes Vinay & Darbelnet (1976), who conducted an exhaustive study on the semantic equality between two French-English translated texts. She also includes the term *dynamic equivalence*, defended by Nida & Taber (1986), from an ethnolinguistic approach.

4. Famous quote by Fray Luis de León in the 16th century during his first class in five years at the University of Salamanca after being released from prison where he was held by the Spanish Inquisition.

5. To reply to a question with another question.

Finally, the author quotes Rabadán (1991), who refers to *translemic equivalence*, in the sense of adequacy of the TT to the ST and its acceptability in the target culture.

According to Muñoz Medrano (2015: 168):

A theory that pursues communicative equivalence and grants maximum fidelity to the conceptual programme of the author of the ST has been defended. However, due to the intersubjective and intercultural character of the bilingual communication process, formal equivalence does not ensure communicative equivalence.

Therefore, we can state that equivalence can only be achieved through pragmatic and semantic identity, or just through pragmatic identity. According to this author, equivalence depends primarily on extra-linguistic factors rather than on the correspondence of linguistic systems, norms or language uses. Thus, she suggests two ways of approaching the attempt to define the concept of equivalence of PUs in translation: the equivalence in the system and at the textual level. Like most scholars, she considers that full equivalence is usually achieved if there is the same denotative and connotative meaning, that is, if the same basic metaphor is found. As we have found in previous studies (see Recio Ariza 2010), the equivalence will be partial if the image or the idiomatic meaning is equivalent but still differs in the syntactic or lexical structure. It will be zero equivalence, however, if the idiomatic condition is lost. Therefore, we must consider communicative equivalence, given the fact that formal equivalence does not guarantee communicative equivalence and we translate texts instead of isolated words. In this regard, and as mentioned before (see Recio Ariza 2006), the categorisation of a word or even its modality can also play a decisive role.

As it has become clear by now, most of the scholars agree with the fact that it is necessary to limit the term and its field of action, as well as giving due importance to pragmatics and the idiosyncrasy implicit in the treatment and translation of equivalence.

All this shows that the task of finding equivalence and, therefore, of translating is complex and must be approached from multiple perspectives.

2.2. *Phraseology in Translation*

Regarding phraseology, it should be noted that it is a scientific discipline that has been consolidated over the past years thanks to a variety of studies, scientific advances and other contributions in this field of knowledge. However, this does not prevent the field from remaining controversial and with different approaches. In our opinion, these approaches complement rather than contradict each other, so the holistic approach that underlies other disciplines, such as cognitivism, can be very useful to shed light on phraseology studies, as well as on the translation and equivalence of PUs.

For some scholars, such as Szyndler (2014: 256), there are essentially two types of PUs: those that share an image, but differ in meaning, and those that have almost the same meaning, but differ in images, which make it possible to speak of *pseudo-phraseological equivalence*. She distinguishes the following:

- i. Inequivalence by contrasts of meaning;
- ii. Inequivalence by contrasts in mental images

Szyndler (2014: 265) believes that “the image component that is part of the plane of content of a given phraseologism and, at the same time, is the pivotal factor for its use is considered one of the *responsible* elements for semantic pseudo-equivalence”. All this highlights the importance of taking into consideration not only semantic and syntactic properties, but also functional properties, i.e. the consequences of phraseological meaning in discourse in the study of PU equivalences. For other authors such as Sozinho Quiraque and de Paula (2016), however, there is no morphological equivalence and only semantic equivalence is obtained from a perspective of interrelation between language and culture.

In Katan's view (2009: 287), the intercultural competence is metacognitive. Since students need to know the reason behind their translation decisions, cognitive skills provide the first decisive step for translating. While biculturalism cannot be taught, levels of understanding and awareness, metacognitive *awareness*, and reflection, can in fact be induced in the classroom (inductive method). All these factors need to be considered in the teaching of PUs and their correct acquisition and assimilation for future translators and interpreters.

Over the last years, attention has been drawn to the need of considering the cultural competence when translating in general and when translating phraseology in particular. Thus, as we have observed in previous studies (see Recio Ariza 2002), phraseology is largely a reflection of a culture and its conventions (Holmes 1972, Baker 1990, Toury 1995, Rica Peromingo 2007, Corpas Pastor 2008, 2018; Fantinuoli 2018, Krüger 2012, Holzinger 2018). In this regard, we believe that only a combination of cultural and linguistic proficiency can probably guarantee a successful translation.

3. Corpus linguistics and translation

We share the opinion of several scholars who emphasise the importance and usefulness of corpora in phraseology and its translation (Holmes 1972, Baker 1990, Toury 1995, Rica Peromingo 2007; Corpas Pastor 2008, 2018; Fantinuoli 2018, Krüger 2012, Holzinger 2018). The systematic use of corpora in translation emerges mainly with the studies carried out by Mona Baker, Holmes, and Toury, for whom corpus-based translation is included in the descriptive branch of the discipline.

In this sense, Rica Peromingo (2007: 10) states in his doctoral dissertation that “one of the possibilities offered by corpus linguistics is precisely its ability to store the students’ oral and written production, analyse this production and present it in the classroom”.

For Fantinuoli (2018: 133), the analysis tools offered by corpora are very useful, since they enhance the quality of the translator’s preparation. In the field of Translation Studies, it is well known that the translator and the interpreter cannot be knowledgeable about all areas, thus they must look for strategies to *supplement* or *compensate for* this knowledge in depth. In recent years, both the usefulness and the need for corpus linguistics and computational linguistics have been demonstrated, especially in the terminological field (see Pym 2011, Tripepi Winteringham 2010, Corpas Pastor 2016). Furthermore, Fantinuoli is of the opinion that terminological correspondence is quite difficult to establish. The high variability and constant evolution of specialised communication make it practically impossible to find sources or glossaries ready to be used in all areas. Another factor that hinders this is the lack of contextual information, since it is usually based on use frequency

rather than contexts, genres, etc. In Fantinuoli's opinion, the advantages of comparative corpora lie in the fact that they are thorough and comprehensive, they also include a wider variety of linguistic material and offer possible translation solutions as they contain real examples of language use. Yet, Fantinuoli points out that one of the disadvantages could be that they do not usually offer coincidences that could help verify the working hypotheses, especially in interpreting settings.

However, other authors, such as Krüger (2012), believe that adopting corpus linguistics in Translation Studies allows researchers to have a more significant contextualisation and control of texts, and provides greater representation and possibilities for generalisation, as well as greater replicability. In other words, corpus linguistics allows us to find features of translated language. In Krüger's opinion, there are two ways of approaching a corpus: on the one hand, from the perspective of its use, in order to learn how to translate; on the other, using a corpus as a translation tool. In this sense, we share Hanks' reflection: "Use introspection to interpret data, but not to create data", that is, data should not be created to confirm, but should be used to provide information and reflect on it.

As Corpas Pastor (2008: 47) puts it, "the automatic management of large volumes of text required the development of procedures to implement applications for management, analysis, and exploitation". Parallel corpora render data on the decisions made during the translation process, thus reflecting the expert knowledge of the professional translator. In this way, corpora offer an ideal framework for the search for equivalence in translation.

In this regard, Iglesias Iglesias (2018: 473) explains that:

Corpus-based analysis can contribute substantially to knowledge about the functioning of languages, the use that speakers make of them and the world in which speakers interact, not only, but mainly at the linguistic level.

Like the abovementioned authors, we believe that corpora are an indispensable tool for translators and interpreters nowadays, given their proven usefulness and the support they provide in making decisions when performing a translation.

4. Example of PUs with *Wort / Worte* in corpora

The following is an example of *Wort / Worte* ['word'] in the different corpora investigated. In order to do this, we used the Sketch Engine corpus manager⁶. Since trying to count all the possible uses of a word like *Wort / Worte* is virtually impossible and our intention was to carry out a qualitative rather than a quantitative study, we opted to approach the study delving into the usual and frequent phraseology of a word. We decided to choose PUs in which *Wort / Worte* appear because the different plural forms of this lexical unit have a different meaning. *Wort-Wörter* is equivalent to 'word', for instance, in the dictionary. According to the *Diccionario de la Lengua Española (RAE)*, "a word is a linguistic unit, generally endowed with meaning, which is separated from the other words by potential pauses in pronunciation and blanks in writing". Yet, *Wort-Worte* is also equivalent to "the ability to speak, oratorical aptitude", the third and fourth entries in this same dictionary, respectively.

In DWDS, the former is defined as "einsilbige oder mehrsilbige selbstständige sprachliche Einheit mit einem bestimmten Bedeutungsgehalt" [autonomous lexical units with one or more syllables and with specific meaning], while the latter is equivalent to "pl. Worte: mündlich oder schriftlich formulierte, sinnvolle Äußerung, Bemerkung" [expressions conveyed orally or in writing]. That is to say, the former, *Wort-Wörter*, is equivalent to 'linguistic unit', while the latter is equivalent to 'speech' in the Saussurian sense of the *parole-langue* dichotomy (sense-meaning). The translation of both lexical units is always 'word' in Spanish, which sometimes makes it difficult to translate it into German.

- i. The PUs chosen are the following:
- ii. *dein Wort in Gottes Ohr*;
- iii. *mir fehlen die Worte*;
- iv. *in anderen Worten*;
- v. *ein gutes Wort einlegen*;
- vi. *Wort für Wort*;

6. <http://www.sketchengine.co.uk> (subscription is required).

4.1. Corpus selection and frequency of use

For this study, DWDS⁷, the *DeTenTen13*, *Europarl3*, *Opus2*, *OpenSubtitles2011* and *EsTenTen18* corpora were selected and consulted through the Sketch Engine corpus manager. We selected this corpus manager because, among other functions, it allows us to study the chosen PUs and their translations directly in a parallel German-Spanish corpus.

4.2. Results obtained

The search in DWDS allows you to obtain data about frequency and the corpora used. In the case of *Worte* we found, for example, a total of 17,602 occurrences in subtitles, 55,850 in the *DWDS-Kernkorpus* (1900-1999), 4,792 in the *DWDS-Kernkorpus* (2000-2010), and 167,683 in *Die Zeit* (1946-2018). The frequency in journalistic language was much higher.

In the search of the PU *Wort für Wort* (131) the result is higher than, for instance, the PU *mir fehlen die Worte* (7).

Regarding the results obtained in the parallel corpus OPUS2 through Sketch Engine, the PU *in anderen Worten* generated the following data:

Table 1. Results obtained for *In anderen Worten*⁸

	Subcorpus	Frequency	Rel [%] ⁹
1	Europarl3	25	216.8
2	OpenSubtitles11	10	76.9
3	KDEdoc	5	3 260
4	KDE4	2	342.4
5	ECB	2	16.8

7. DWDS is a corpus that provides systematised lexical information of the German language, both current and historical. *Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart [Historical-contemporary Dictionary of German Language]*.

8. See Appendix, Table 3 + Table 4.

9. Relative frequency (or normalised frequency) refers to the frequency of a word per million words. It is used to compare frequencies between corpora of different sizes.

On the other hand, the PU *ein gutes Wort* appears much more frequently in the subtitles subcorpus than in the European Parliament subcorpus, unlike the previous PU.

Table 2. Results obtained for *Ein gutes Wort*¹⁰

	Subcorpus	Frequency	Rel [%]
1	OpenSubtitles2011	17	319.6
2	Europarl3	1	21.2

Regarding the PU *Wort für Wort*, the results are the following:

Table 3. Results obtained for *Wort für Wort*¹¹

	Subcorpus	Frequency	Rel [%]
1	OpenSubtitles2011	17	159.8
2	Europarl3	16	169.6
3	KDE4	2	418.5
4	OpenOffice3	1	864.9

The results for *Wort / Worte / Worten* are the following:

Table 4. Results obtained for *Wort / Worte / Worten*¹²

	Subcorpus	Frequency	Rel [%]
1	Europarl3	13 980	260.6
2	OpenSubtitles2011	4 203	69.5
3	ECB	746	13.5
4	KDE4	638	234.8
5	OpenOffice	321	483.2
6	OpenOffice3	292	444.2

10. See Appendix, Table 5 + Table 6.

11. See Appendix, Table 7.

12. See Appendix, Table 8.

7	MultiUN	141	15.2
8	KDEdoc	97	136.0
9	EUconst	38	186,0
10	EMEA	13	0,7

As we can see, the highest frequency, more than three times higher, occurs in the *Europarl3* subcorpus.

With regard to the Spanish concordances, *EsTenTen18* provides the frequency from a variety of websites, which may help to measure the frequency of use on the web.

In order to do this, we searched for the Spanish equivalents: *sin palabras* (*ohne Worte*) [speechless], *en otras palabras* (*in anderen Worten*) [in other words], and *palabra por palabra* (*Wort für Wort*) [word for word]. All of them included frequency data, as well as context (see Appendix). This allowed us to find the translation equivalent, and to verify the relevant translation suggestion.

Table 5. Results obtained for *Sin palabras* (6 754)

	Website	Frequency	Frequency per million
1	foroactivo.com	295	0.01
2	fanfiction.net	140	< 0.01
3	forosactivos.net	122	< 0.01
4	amor-yaoi.com	108	< 0.01
5	potterfics.com	79	< 0.01
6	foroactivo.mx	78	< 0.01
7	wikia.com	77	< 0.01
8	saintseyayaoi.net	72	< 0.01
9	deviantart.com	69	< 0.01
10	fanfic.es	67	< 0.01

Table 6. Results obtained for *En otras palabras* (184 873)

	Website	Frequency	Frequency per million
1	wikipedia.org	3,583	0.18
2	vlex.com.co	3,319	0.16
3	conicyt.cl	2,933	0.14
4	vlex.com.mx	2,568	0.13
5	scielo.org.mx	2 (14)	0.12
6	scielo.org.com	2,227	0.11
7	absta.info	2,166	0.11
8	vlex.es	1,746	0.09
9	scielo.org.ar	1-549	0.08
10	unam.mx	1,463	0.07

Table 7. Results obtained for *Palabra por palabra* (5 868)

	Website	Frequency	Frequency per million
1	wikipedia.org	246	0.01211
2	xn--elajacpihecdall.xn--plai	62	0.00305
3	noticiasalminuto.es	41	0.00202
4	wikihow.com	36	0.00177
5	foroactivo.com	36	0.00177
6	absta.info	32	0.00158
7	elpais.com	31	0.00153
8	infocatolica.com	30	0.00148
9	vlex.com.mx	28	0.00138
10	foroactivo.com	28	0.00138

In the present study we only display the examples collected from the corpora that seem more relevant and appropriate. In order to describe trends regarding equivalences of PUs, however, we will discuss some of the PUs below as paradigmatic examples.

Since we started from the hypothesis of an equivalence of PUs in corpora, in the case of the PU *Wort für Wort* we first analysed its use in the corpus *DWDS-Kernkorpus* in order to compare it, later on, with the parallel corpora available through the Sketch Engine manager. This analysis casts light on its use and also on its translation, as we can deduce through the parallel corpora. According to what can be seen in *DWDS-Kernkorpus*, this PU can appear in very different contexts, including texts on economy or sociology, or literary texts (*palabra por palabra*) [word by word]¹³. However, it is used both with a specific and a more abstract meaning (see Appendix *DWDS-Kernkorpus*). With regard to its specific meaning, it is worth highlighting the fact that it is generally used to refer to literalness as a *fact*, while the abstract meaning refers to the underlying *idea* of literalness, as we can see in the examples *aber es wurde Wort für Wort gedruckt* (*fue impreso palabra por palabra*) [it was printed word by word] (specific meaning) and *deshalb wurde sie Wort für Wort geglaubt* (*la creyeron palabra por palabra*) [they believed her word by word] (abstract meaning). In the first example, the translation would be *palabra por palabra* [lit. “word by word”], whereas in the second one the only possible PU is *a pie juntillas* [lit. with both feet together; fig. without questioning], since the non-literal meaning is being used.

Table 7 illustrates all the occurrences of the idiom in the DE-ES parallel subcorpus of *OpenSubtitles2011*, in which we can see that in most cases the translation strategy was equivalence, and more specifically full equivalence.

Its frequency of use is also much higher than the results obtained with the search of the PU *mir fehlen die Worte*, for example. In this case, the *DWDS-Kernkorpus* shows the number of occurrences for this search. Although the sample may not be very large, the results show that the highest frequency is found in literary texts. Regarding this PU, it is worth noting that the

13. <https://dle.rae.es/?id=RUI938s> It means “in a literal manner” according to *Diccionario de la lengua española. Edición del Tricenterario*. It is an adverbial phrase that means “literally”, last accessed 07/10/2019.

translations that generally appear in parallel corpora tend to be paraphrased so that the meaning of the PU is translated into Spanish.¹⁴ The most common translation strategy observed was omission, as we can see in the following example: *Endlich habe ich Sie vor mir und mir fehlen die Worte*, which has been translated as *por fin puedo hablar con Ud. y no se me ocurre nada* [I can finally talk to you and I cannot think of anything to say]. In most of the translations that we can see in Table 2 of the Appendix, the most common translation strategy was omission, or even paraphrasing with the expression *no sé qué decir* [I do not know what to say].

In the case of the PU *in anderen Worten* in the corpora of EuroParl, OpenSubtitles11, KDEdoc, KDE4 and ECB¹⁵, the relative frequency of use (compared with the total number of words in the corpus or the subcorpus) was much higher in the first one (216.8%), and relatively low in the last one (16.8%). Its use can be mainly found in contexts in which something that has been said is summarised and expressed in similar terms, which accounts for the equivalence strategy that is generally applied. In most cases, in fact, it is translated as *dicho de otro modo* [to put it another way] or *en otras palabras* [in other words].

As far as the PU *Wort in Gottes Ohr* is concerned, the results are summarised in the following table:

14. See Appendix, Table 2.

15. The multilingual parallel macrocorpus OPUS in Sketch Engine contains the following corpora: European Central Bank corpus (ECB), European Medicines Agency documents (EMA), The European constitution (EUconst), European Parliament Proceedings (v3) (Europarl3), PHP manual corpus (PHP), A parallel corpus of the Balkan languages (SETIMES2), Stockholm Parallel Corpora (SPC), Regeringsförklaringen – Declarations of Government Policy by the Swedish Government 8RF), Belgisch Staatsblad corpus (MBS), OfisPublik, TedTalks, hrenWaC, The Tehran English-Persian subtitle corpus (TEP), KDE4 localisation files (KDE4), KDE manual corpus (KDEdoc), OpenOffice, OpenOffice3, OpenSubtitles2011 – Open Subtitles corpus (2011 version), Tatoeba, Translated UN documents (UN) and Translated UN documents (MultiUN). See Corpus Pastor, 2018: 10.

Table 8. Results obtained for *Wort in Gottes Ohr*¹⁶

Europarl3	Ihr Wort in Gottes Ohr , oder wo auch immer das dann gemacht wird, was Sie vorgetragen haben.	Palabras y buenos deseos es lo que usted nos ha presentado.
Europarl3	Gestatten Sie, daß ich Ihnen ausgehend von meiner weltlichen Überzeugung sage: "Ihr Wort in Gottes Ohr "	Permítame que le diga desde mi convicción laica: » Dios le oiga ».
OpenSubtitles2011	Dein Wort in Gottes Ohr .	No sabes cómo te lo agradezco.
OpenSubtitles2011	- Lhr Wort in Gottes Ohr . - Bei dem Polytheismus hier ...	¡Que Dios te oiga! Desgraciadamente, aquí reina el politeísmo pero no te preocupes, todo irá bien.
OpenSubtitles2011	- Dein Wort in Gottes Ohr .	Desde tus labios hasta los oídos del cuarto de control.
OpenSubtitles2011	Dein Wort in Gottes Ohr .	Lo dices tú.

With respect to this PU, we can observe that both the omission strategy and paraphrasing or the equivalence *Dios te oiga* [may God hear you] are used. The same applies to *Fehlen die Worte* (see Appendix Table 2) or to *Ein gutes Wort einlegen* (see Appendix table 4). Moreover, *in anderen Worten* (see Appendix Table 3) and *Wort für Wort* (see Appendix table 5) are usually translated as *en otras palabras* [in other words], i.e. the translation strategy has almost always been literal translation which, in this case, corresponds to full equivalence as well.

We are aware that this reflection must be carried out by means of a deep and detailed analysis, so our intention is not to give a definitive solution, but to point out possible future avenues of research in this field.

16. See Appendix, Table 1.

5. Conclusions

The present study shows that translators must necessarily work with all the tools at their disposal, which not only includes dictionaries, glossaries, automatic translators such as *Linguee* or *DeepL*, for instance, but also available corpora. In this sense, corpus managers, like Sketch Engine, can be very useful since they display actual instances of language use and translational behaviour. Available corpora usually represent the lexicon and the contexts used by each community of speakers, so they inevitably reflect language at all times. In addition, let us not forget that the strategies that we must offer to the future translator and interpreter must be those that give them the ability and willingness to carry out their translational work in a correct way.

By using parallel corpora, translators have at their disposal not only a significant number of data that supply information about equivalence in translation, but also translation solutions. In this regard, corpora constitute an exceptional tool for professional translators and interpreters, since they allow them to systematise the search for equivalence and reflect the decisions made in the translation process, besides providing the expert knowledge of a professional translator.

An attempt has been made to ensure that the results obtained in this study can be not only reproduced and extrapolated, but also provide new avenues of development for research in this field.

References

- BAKER, Mona. (1990) *In other words*. London: Routledge.
- CORPAS PASTOR, Gloria. (1997) *Manual de Fraseología Española*. Barcelona: Gredos.
- CORPAS PASTOR, Gloria. (2008) *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Frankfurt am Main: Peter Lang.
- CORPAS PASTOR, Gloria & Fern Lily May. (2016) A Survey of Interpreters' Needs and their Practices Related to Language Technology. Technical report, Universidad de Málaga.
- CORPAS PASTOR, Gloria & Ortiz Álvarez, María Luisa. (2017) Fraseología y Paremiología: una entrevista con Gloria Corpas. *ReVEL* 15 (29). 271-280.

- CORPAS PASTOR, Glora. (2018) *Laughing one's head off* in Spanish Subtitles: a corpus-based study on diatopic variation and its consequences for translation. In Pedro Mogorrón Huerta & Antonio Martínez Albaladejo (eds.), *Fraseología, Diatopía y Traducción, Phraseology, Diatopic Variation and Translation*, 32-71. Amsterdam: John Benjamins.
- COSTE, Didier. (2018) Conversion is the limit or the receding horizon of equivalence. Aarhus "Equivalence(s)" Symposium 20.-21. September 2018. *Diccionario de la lengua española. Edición del Tricenterario* <https://dle.rae.es/?id=RUI938s> Last accessed 07/10/2019.
- DOBROVOL'SKIJ, Dmitriy. (2000) Idioms in contrast: a functional view. In Gloria Corpas Pastor (ed.), *Las lenguas de Europa: Estudios de Fraseología, Fraseografía y Traducción*, 367-388. Granada: Comares.
- FANTINUOLI, Claudio. (2018) The use of comparable corpora in interpreting practice and training. *The Interpreters' Newsletter*, 23. 133-149.
- HANKS, Patrick. (2015) Meaning, Phraseology and Lexicography: A Corpus-Driven Approach. Paper presented at the International Conference *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. University of Málaga, 29 June-2 July. <https://slideplayer.com/slide/10059021/>
- HOLMES, James. (1972) The Name and Nature of Translation Studies. In: James Holmes. 1988. *The Name and Nature of Translation Studies. Translated!: Papers on Literary Translation and Translation Studies*. 66-80. Amsterdam: Rodopi.
- HOLZINGER, Herbert. (2018) Unikale Elemente oder phraseologisch gebundene Wörter? Antworten aus korpuslinguistischer Sicht. *Revista de Filología Alemana*, 26. 199-213.
- IGLESIAS IGLESIAS, Nely (2018) ¿Qué puede aportar la lingüística de corpus al ámbito de la enseñanza y/o aprendizaje de lenguas extranjeras? (Ejemplificado a partir del alemán, concretamente de las unidades léxicas *jedenfalls, auf jeden Fall, in jedem Fall*). *Actas do Congreso Internacional de Lingüística Xeral*. 467-474.
- KATAN, David. (2009) Translator Training and Intercultural Competence. In Stefania Cavagnoli, Elena Di Giovanni & Raffaella Merlini (eds.), *La ricerca nella comunicazione interlinguistica. Modelli teorici e metodologici*, 282-301. Milano: Franco Angeli.
- KRÜGER, Ralph. (2012) Working with corpora in the translation classroom. *Studies in Second Language Learning and Teaching* 2 (24). 505-525.

- LEAL RIOL, María Jesús. (2008) Contraste fraseológico: similitudes y diferencias existentes entre las unidades fraseológicas del español y del inglés. *ES: Revista de filología inglesa* 29. 103-116.
- MIRA ÁLVAREZ, Germán Darío. (2011) La equivalencia en la traducción de las unidades fraseológicas. Un estudio empírico. *Íkala* 16 (27). 105-131.
- MOGORRÓN HUERTA, Pedro. (2008) Compréhension et traduction des locutions verbales. *Meta*, 53 (2). 378-406.
- MUÑOZ MEDRANO, Cándida (2015) Acerca de la equivalencia de los fraseologismos taurinos en los repertorios lexicográficos. *Moenia* 21. 167-185.
- PEJOVIC, Andjelka. & Trivic, Aneta. (2018) Algunos elementos idiosincrásicos en la fraseología serbia en comparación con la Española. In Antonio Pamies, Isabel M^a Balsas & Alexandra Magdalena (eds.) *Lenguaje figurado y competencia interlingüística (I)*. 101-114. Granada: Editorial Comares.
- PYM, Anthony. (2011) What technology does to translating. *Translation & Interpreting* 3 (1), 1-9.
- RECIO ARIZA, María Ángeles. (2002) De la teoría a la práctica: Fraseología y coloquialismos. Propuestas didácticas para su aplicación en el aula. In *Hacia la unidad en la diversidad: Difusión de las lenguas europeas* (Año europeo de las lenguas 2001), 404 - 415. Salamanca: Diputación provincial de Salamanca.
- RECIO ARIZA, María Ángeles. (2006) El uso de las unidades léxicas ‘schließlich’ y ‘endlich’ en alemán y su correspondencia en español. *Estudios Filológicos Alemanes*, 11. 485-496.
- RECIO ARIZA, María Ángeles. (2010) O emprego da fraseoloxía como recurso específico na Publicidade. *Cadernos de Fraseoloxía Galega*, 12. 233-245.
- RICA PEROMINGO, Juan Pedro. (2007) *Estudio fraseológico del uso de colocaciones gramaticales y grupos léxicos en textos argumentativos nativos y no nativos: análisis de corpus de estudiantes*. Madrid: Universidad Complutense dissertation.
- SABBAN, Annette. (2010) Zur Übersetzung von Idiomen im Wörterbuch und im Text: die Rolle von Kontextsensivität. *Trans-kom* 3 (2). 192-208.
- SEVILLA MUÑOZ, Julia. (2004a) O concepto “correspondencia” na traducción paremiolóxica. *Cadernos de fraseoloxía galega* 6. 221-230.
- SEVILLA MUÑOZ, Julia. (2004b) La enseñanza de paremias en contexto y su traducción a través de su presencia en las “fables” de La Fontaine. *Anales de Filología Francesa* 13, 131-146.

- SOZINHO QUIRAQUE, Zacarirs & de Paula, María Helena. (2016) Mecanismos de (Não) equivalência em fraseologias da Língua Tewe e do Português. *Revista do SELL* 5 (1), 1-16.
- SZYNDLER, Agnieszka. (2014) Sobre la pseudo-equivalencia fraseológica desde una perspectiva cognitivista. *Anuario de Estudios Filológicos* 37, 251-267.
- TOURY, Gideon. (1995) *Descriptive Translation Studies*. Amsterdam: John Benjamins.
- TRIPEPI WINTERINGHAM, Sara. (2010) The usefulness of ICTs in interpreting practice. *The Interpreters' Newsletter* 15. 87-99.
- TRIVIC, Aneta. (2013) Correspondencia recíproca en la fraseología. Análisis contrastivo español-serbio. *Colindancias* 4. 69-80.

APPENDIX

Parallel concordance

Table 1
Wort in Gottes Ohr

Europarl3	Ihr Wort in Gottes Ohr , oder wo auch immer das dann gemacht wird, was Sie uns vorgetragen haben.	Palabras y buenos deseos es lo que usted nos ha presentado.
Europarl3	Gestatten Sie, daß ich Ihnen ausgehend von meiner weltlichen Überzeugung sage: „ Ihr Wort in Gottes Ohr “.	Permítame que le diga desde mi convicción laica: «Dios le oiga».
OpenSubtitles2011	Dein Wort in Gottes Ohr .	No sabes cómo te lo agradezco.
OpenSubtitles2011	- Lhr Wort in Gottes Ohr . - Bei dem Polytheismus hier ...	¡Qué Dios te oiga! Desgraciadamente, aquí reina el politeísmo pero no te preocupes, todo irá bien.
OpenSubtitles2011	- Dein Wort in Gottes Ohr .	Desde tus labios hasta los oídos de el cuarto de control.
OpenSubtitles2011	Dein Wort in Gottes Ohr .	Lo dices tu.

Table 2
Fehlen die Worte

Europarl3	Herr Präsident! Mir fehlen die Worte !	Señor_Presidente, ¡válgame Dios!
Europarl3	Mir fehlen die Worte dafür, aber ich möchte die Präsidentschaft und Kommissar Fischler dringend auffordern, neue Methoden zur Bekämpfung zu finden, bei der Impfungen, zunächst vereint mit vorbeugender Beseitigung, ausprobiert werden.	Me parece una locura, pero quiero hacer un llamamiento urgente a la Presidencia y a el Comisario_ Fischler para que busquen nuevos métodos para combatir la enfermedad, poniendo a prueba vacunas, de momento en combinación con el sacrificio preventivo de animales.
OpenSubtitles2011	Ihm fehlten die Worte .	“Disculpe, dice nuestro amigo, pero...” “- Y quedó sin palabras .”
OpenSubtitles2011	Mir fehlen die Worte .	Es todo lo que puedo decir.
OpenSubtitles2011	Endlich habe ich Sie vor mir und mir fehlen die Worte .	Por fin puedo hablar con usted y no se me ocurre nada .
OpenSubtitles2011	Dem Chef fehlten die Worte .	El director no contest.
OpenSubtitles2011	Mir fehlen die Worte .	No puedo encontrar las palabras
OpenSubtitles2011	- Mir fehlen die Worte! ... grün!	Me sentí absolutamente verde.
OpenSubtitles2011	Mir fehlen die Worte!	¡No hay palabras!
OpenSubtitles2011	Mir fehlen die Worte!	Válgame Dios.
OpenSubtitles2011	Mir fehlen die Worte .	Cielo santo, no sé qué decir .
OpenSubtitles2011	Mir fehlen die Worte .	¡ No sé qué decir!

Table 3
In anderen Worten

	Subcorpus	Frecuencia	Rel [%]
1	Europarl3	25	216.8
2	OpenSubtitles11	10	76.9
3	KDEdoc	5	3 260
4	KDE4	2	342.4
5	ECB	2	16.8

Table 4

Europarl3	Dennoch scheinen weder die Kommission noch der Rat entschlossen, den Geist dieser Verpflichtung einzuhalten, in anderen Worten , unsere neuen Prioritäten zu finanzieren, ohne unsere traditionellen Prioritäten in Mitleidenschaft zu ziehen.	Ahora_bien, ni la Comisión ni el Consejo parecen decididos a respetar el espíritu de ese compromiso o, dicho de otro modo , afinanciar nuestras nuevas prioridades sin comprometer nuestras prioridades tradicionales.
Europarl3	In anderen Worten , jeder Änderungsantrag, der diese Ausgewogenheit beeinträchtigt, ist in unseren Augen gefährlich, und insbesondere wäre es meiner Auffassung nach falsch, die Auseinandersetzung über die völlige und zwingende Unentgeltlichkeit der Blutspenden wiederholen zu wollen, denn wenn als einziges Charakteristikum die Freiwilligkeit der Blutspende durchgesetzt	En otras palabras , toda enmienda que contemple la posibilidad de cuestionar este equilibrio nos parece peligrosa y, en particular, diría que reanudar la batalla sobre la gratuidad total y obligatoria de la donación nos parece un error, porque imponer como única referencia la donación gratuita es como priorizar un Sistema nacional de recogida sobre otro, y ello en contradicción total con los tratados.
Europarl3	Nach allem, was ich gehört habe, auch gestern von den Mitgliedstaaten und den neuen Mitgliedstaaten, ist man sich im Klaren darüber, dass die Entscheidungen so getroffen werden müssen, um die Probleme zu vermeiden, von denen Herr Maaten gesprochen hat, d. h. in anderen Worten Probleme mangelnder Beschlussfähigkeit.	Por todo lo que he escuchado hoy y ayer, de boca de los Estados miembros actuales y futuros, me parece que somos conscientes de la necesidad de tomar decisiones para evitar los cuellos de botella a los que se ha referido el Sr. Maaten o, en otras palabras , los problemas que conlleva la incapacidad de tomar decisiones.
Europarl3	In anderen Worten : eine wirkliche Fischereipolitik beschränkt sich nicht auf die Verabschiedung von Vorschriften hinsichtlich der Fischbestände, denn die Fischerei ist eine Wirtschaftstätigkeit mit starken sozialen Auswirkungen.	Dicho de otro modo , una verdadera política pesquera no se limita a la aprobación de normas sobre recursos pesqueros: tiene que ver también con una actividad económica de gran repercusión social, y esto es algo que no debemos olvidar.
Europarl3	In anderen Worten , wir nehmen für uns in Anspruch, zwischen Regimen zu wählen, anstatt Staaten anzuerkennen.	De hecho, estamos generalizando, si se me permite, la forma de justicia que estamos aplicando a Iraq. En otras palabras , nuestro propósito es escoger regimens en_lugar_de reconocer Estados.

Europarl3	Eine restriktive Drogenpolitik, in anderen Worten Nulltoleranz, ist die einzig mögliche Haltung, wenn es um Drogen geht.	Una política restrictiva en_materia_de drogas, es decir , la tolerancia cero, es el único posicionamiento posible en_materia_de drogas.
Europarl3	Erstens sollten diese Änderungen den dem gemeinsamen Markt zugrunde liegenden Gedanken widerspiegeln oder, in anderen Worten , Diskriminierung verhindern.	Primero, dichas enmiendas deberían seguir la lógica de el concepto de Mercado interior, es decir , evitar la discriminación.
Europarl3	Hauptgrund hierfür ist der Massenexodus von qualifiziertem Personal, oder in anderen Worten , der hoch qualifizierten Ärzte und Krankenschwestern, in reiche Länder.	La principal razón de ello es el éxodo masivo de trabajadores cualificados o, en otras palabras , de médicos y profesionales de la enfermería con una excelente formación, a los países ricos.
Europarl3	In anderen Worten , der Binnenmarkt baut auch auf gemeinsamen sozialen Normen auf und entspricht somit dem Geist der Verträge.	En otras palabras , el Mercado único se construye también sobre unas normas sociales comunes, inscribiéndose así en el espíritu mismo de los Tratados.
Europarl3	Statt die bestehenden Probleme – mangelndes Wachstum, Arbeitslosigkeit, Spannungen in den ländlichen Gebieten und den Städten – in Angriff zu nehmen, favorisieren die Kommission und das Parlament Dialoge und Foren, in anderen Worten „leeres Gerede“.	En_vez_de atacar los problemas: la falta de crecimiento, el desempleo y las tensiones en el campo y en la ciudad, la Comisión y el Parlamento preconizan el diálogo, los foros, en breve «lacháchara».
OpenSubtitles2011	In anderen Worten , dein Vater ist praktisch hilflos.	En otras palabras , tu padre está casi imposibilitado. - ¿Se va a morir?
OpenSubtitles2011	- In anderen Worten ...	lujuriosa. - En otras palabras ...
OpenSubtitles2011	- In anderen Worten ...	- En otras palabras ...
OpenSubtitles2011	Sie haben einen typischen Körper, der eine höhere Transpiration aufweist, in anderen Worten , Sie schwitzen stark.	- Gracias. Posee el típico cuerpo con alta transpiración, en otras palabras , suda mucho.
OpenSubtitles2011	In anderen Worten , alle.	En otras palabras , todos.
OpenSubtitles2011	In anderen Worten , Schweigen ist Gold.	En otras palabras , el silencio vale oro.
OpenSubtitles2011	In anderen Worten :	En otras palabras , los nueve_mil_millones pueden ser creados de la nada.
OpenSubtitles2011	In anderen Worten :	En otras palabras :
OpenSubtitles2011	In anderen Worten :	En otras palabras , casi cada dólar que existe eventualmente debe ser devuelto a el banco también con un interés pagado.
OpenSubtitles2011	In anderen Worten :	En otras palabras , el dinero no vino de sus activos existentes.

Table 5
Ein gutes Wort (18)

	Subcorpus	Frecuencia	Rel [%]
1	OpenSubtitles2011	17	319.6
2	Europarl3	1	21.2

Table 6

OpenSubtitles2011	- Majestät, wenn Majestät ein gutes Wort einlegen?	- Majestad. Si Su_Majestad tuviera a bien interceder . - ¿Por usted?
OpenSubtitles2011	Ich werd ein gutes Wort für Sie einlegen.	Le hablaré bien de usted. Estoy segura.
OpenSubtitles2011	Ich könnte im Inn ein gutes Wort einlegen.	Veré qué puedo hacer por ti en la posada.
OpenSubtitles2011	Leg ein gutes Wort bei Sheila für mich ein.	Si surge la ocasión, puedes hablarle bien a Sheila de mí.
OpenSubtitles2011	Können Sie ein gutes Wort einlegen?	Quería decírselo , pero Vd.
OpenSubtitles2011	Legen Sie ein gutes Wort für mich ein?	¿ Le he ofendido?
OpenSubtitles2011	Bitte legt ein gutes Wort für mich ein beim Sheriff von Nuttingham.	Quiero tenerlo a el Sheriff_de_Rottingham.
OpenSubtitles2011	Göth hat mir versprechen müssen, dass er für Sie ein gutes Wort einlegt.	Le hice prometer a Goeth que le recomendaría .
OpenSubtitles2011	Ich muß jetzt zur Party der Geldgeber, ich werde ein gutes Wort für dich einlegen.	Estoy atrasado para la Fiesta de recaudación de fondos y ... Yo te digo unas palabras .
OpenSubtitles2011	Wenn du ein gutes Wort für mich einlegst, drückt er vielleicht ein Auge zu.	Si dices una palabrita en mi favor para que olvide toda esta tontería ...
OpenSubtitles2011	Wenn ihr eure Sprache aufpoliert, leg ich ein gutes Wort für euch ein.	Y si mejoran su lenguaje ... le diré algo bueno de ustedes.
OpenSubtitles2011	Legst du ein gutes Wort für mich ein?	Necesito unos días la próxima semana. ¿Por qué? Le prometí un viaje a mi mujer. Échame una mano. Hablaré con el jefe de personal.
OpenSubtitles2011	Leg ein gutes Wort für mich ein.	Háblale de mí .
OpenSubtitles2011	Leg ein gutes Wort für mich ein.	Háblale bien de mí . - Deja de decir tacos.
OpenSubtitles2011	Leg ein gutes Wort für mich ein.	Defiéndeme .
OpenSubtitles2011	Du musst ein gutes Wort für mich bei Emma einlegen.	Esa Mercedes está equivocada respecto_a ti. - ¿Qué dijo ella?
OpenSubtitles2011	Vielleicht kannst Du ein gutes Wort für mich einlegen?	- Ah. Cariño, eres mi hermana. Te apoyare en todo lo que quieras hacer

Europarl3	in diesem Parlament am Freitag, dem 17. Dezember 1999, also im letzten Jahrhundert, die Stimme bis zu den Fischen gelangt ist, die mir - da sie bekanntlich nicht in der Lage sind, zu sprechen - schriftlich gesagt haben: „ Bravo, Du hast ein gutes Wort für uns eingelegt!	... he de decir que, tras mi intervención de el viernes_17_de_diciembre_de_1999, por_lo_tanto, de el siglo pasado, en que formulé mi explicación de voto ante el Pleno, la noticia ha llegado hasta los peces que, por escrito, pues como sabemos, no pueden hablar, me han dicho: “! Muy bien, has hablado a nuestro favor!
-----------	--	---

Tabla 7
Wort für Wort

	Subcorpus	Frecuencia	Rel [%]
1	OpenSubtitles2011	17	159.8
2	Europarl3	16	169.6
3	KDE4	2	418.5
4	OpenOffice3	1	864.9

OpenSubtitles2011	Ich kann nicht glauben, dass du alles kopiert hast, Wort für Wort .	No puedo creer que copiaras todo esto palabra por palabra.
OpenSubtitles2011	Ich soll Ihnen Wort für Wort sagen, was er mir sagt.	Pide que te diga lo que dice palabra por palabra.
OpenSubtitles2011	Wir haben sie Wort für Wort zusammengesetzt.	Lo hemos reproducido palabra por palabra.
OpenSubtitles2011	Sag Wort für Wort, was ich dir gesagt hab.	Di exactamente lo que te dije, palabra por palabra.
OpenSubtitles2011	Wort für Wort!	Literalmente. ¿ Entendido?
OpenSubtitles2011	Ich hab mich Wort für Wort , an den Text gehalten.	Seguí el guión palabra por palabra.
OpenSubtitles2011	Wort für Wort!	Literalmente. ¿ Entendido?
OpenSubtitles2011	Und zwar Wort für Wort, ja?	Dile a Molly lo que te digo palabra por palabra.
OpenSubtitles2011	Er hat dann Wort für Wort aus dem Wörterbuch übersetzt ... es war grausam.	Se puso a traducir el libro palabra a palabra , diccionario en mano. ¡ Qué cosa más patética!
OpenSubtitles2011	Lesen Sie es Wort für Wort ab.	Solo léalo palabra por palabra.
OpenSubtitles2011	Wort für Wort.	Eso es exactamente lo que les he dicho.
OpenSubtitles2011	Wort für Wort.	Palabra por palabra.
OpenSubtitles2011	Lesen Sie es Wort für Wort ab.	Léelo completo.
OpenSubtitles2011	Überbring die Nachricht, Laufbursche, ... und zwar Wort für Wort.	Entrega el mensaje, cadete... .. palabra por palabra.
OpenSubtitles2011	Wir haben sie Wort für Wort zusammengesetzt.	Lo hemos reproducido palabra por palabra.
OpenSubtitles2011	Wir haben sie Wort für Wort zusammengesetzt.	Lo hemos reproducido palabra por palabra.

OpenSubtitles2011	Wort für Wort .	Palabra por palabra.
Europarl3	Es macht doch keinen Sinn, die zahlreichen guten Empfehlungen des Ausschusses Unabhängiger Sachverständiger Wort für Wort in Ihren Bericht zu übernehmen.	A_el_fin_y_a_el_cabo, no tiene sentido que se recojan las numerosas buenas recomendaciones de el Comité_de_Sabios en su informe, palabra por palabra .
Europarl3	Wir wussten alle, und heute wissen wir es Wort für Wort , wie weit Herr Barak in seinem aufrichtigen Willen, ein gerechtes Friedensabkommen mit den Palästinensern zu schließen, gegangen ist, und wir wissen, dass einzig und allein der Wille Arafats und der ihn umgebenden Demokraten diesen Friedensprozess sabotiert und zerstört hat.	Todos supimos, y hoy lo sabemos literalmente , hasta dónde llegó el Sr.Baraken su voluntad profunda de realizar un acuerdo de paz justo con los palestinos, y sabemos que sólo la voluntad de el Sr.Arafat y de los demócratas que le rodean saboté y destruyó dicho proceso de paz.
Europarl3	Auf dieser Grundlage erarbeitete die Europäische Kommission den Vorschlag, der uns heute vorliegt und der beinahe Wort für Wort den Text der Vereinten Nationen aufgreift.	En base a ella, la Comisión_Europea ha presentado la propuesta que hoy tenemos entre manos, la cual recoge casi palabra por palabra el texto de las Naciones_Unidas.
Europarl3	Wir wissen, dass die Regierungskonferenz den Entwurf nicht Wort für Wort übernehmen wird.	Sabemos que la CIG no aprobará el borrador exactamente tal_como está.
Europarl3	Ein umständliches, langwieriges Genehmigungsverfahren für jeden einzelnen Werbespruch aus 25 Mitgliedstaaten – einzureichen natürlich Wort für Wort in alle zwanzig Amtssprachen übersetzt!	Están creando un monstruo burocrático: ¡un procedimiento de autorización laborioso y largo para cada lema publicitario de los 25 Estados miembros, que obviamente tendrá que traducirse a las 20 lenguas oficiales!

Tabla 8
Wort/ Worte/ Worten

	Subcorpus	Frecuencia	Rel [%]
1	Europarl3	13 980	260.6
2	OpenSubtitles2011	4 203	69.5
3	ECB	746	13.5
4	KDE4	638	234.8
5	OpenOffice	321	483.2
6	OpenOffice3	292	444.2
7	MultiUN	141	15.2
8	KDEdoc	97	136.0
9	EUconst	38	186.0
10	EMEA	13	0.7

BIONOTE/NOTA BIOGRÁFICA

M^a ÁNGELES RECIO ARIZA es doctora en Filología Alemana y Profesora Titular en el Departamento de Traducción e Interpretación de la Universidad de Salamanca. Es funcionaria del cuerpo de Profesores de EE OO II. Sus áreas principales de investigación son la fraseología y cognitivism, así como enseñanza de la lengua extranjera desde un punto de vista traductológico. Es autora de numerosas publicaciones nacionales e internacionales. Ha traducido artículos y literatura, además de trabajar como intérprete para empresarios salmantinos. Ha colaborado en varios proyectos nacionales e internacionales. Cofundadora y codirectora de la revista *CLINA*. Directora del Servicio de Orientación de la Universidad de Salamanca de 2011 a 2014. Coordinadora del Grupo E-Sphaera y del Portal Colaborativo FOCO. Ha impartido cursos y conferencias en las Universidades de Heidelberg, Ceske Budejovice (República Checa) y Novi Sad (Serbia). Coordinadora de la doble titulación de máster internacional MATEM entre las Universidades de Salamanca y Heidelberg.

M^a ÁNGELES RECIO ARIZA holds a degree and a PhD in German Philology. She is an Associate Professor at the Translation and Interpreting Department of the University of Salamanca. She is a civil servant at the Spanish National School of Languages. Her research interests and her publications (both national and international) focus on Phraseology and Cognitivism and the teaching of foreign language from both an educational and a translational point of view. She has translated articles and literature and has worked as a professional interpreter. She has collaborated in several national and international research projects. She is the co-founder and editor of the journal *CLINA*. She was Head of the Student Guidance Service of the University of Salamanca (2011-2014). She is the Coordinator of the E-Sphaera Group and FOCO. She has lectured at the University of Heidelberg, Ceske Budejovice (Czech Republic) and Novi Sad (Serbia). She is the Coordinator of the International Double Master's Degree MATEM, University of Salamanca and University of Heidelberg.

AIMS / OBJETIVOS / OBJECTIUS

MonTI (*Monographs in Translation and Interpreting*) is an academic, peer-reviewed and international journal fostered by the three public universities with a Translation Degree in the Spanish region of Valencia (Universitat d'Alacant, Universitat Jaume I de Castelló and Universitat de València).

Each issue will be thematic, providing an in-depth analysis of translation- and interpreting-related matters that meets high standards of scientific rigour, fosters debate and promotes plurality. Therefore, this journal is addressed to researchers, lecturers and specialists in Translation Studies.

MonTI will publish one issue each year, first as a hard copy journal and later as an online journal.

In order to ensure both linguistic democracy and dissemination of the journal to the broadest readership possible, the hard-copy version will publish articles in German, Spanish, French, Catalan, Italian and English. The online version is able to accommodate multilingual versions of articles, and it will include translations into any other language the authors may propose and an attempt will be made to provide an English-language translation of all articles not submitted in this language.

Further information at:

<http://dti.ua.es/es/monti-english/monti-contact.html>

MonTI es una revista académica con vocación internacional promovida por las universidades públicas valencianas con docencia en traducción e interpretación (Universidad de Alicante, Universidad Jaume I de Castellón y Universidad de Valencia).

Nuestra revista pretende ante todo centrarse en el análisis en profundidad de los asuntos relacionados con nuestra interdisciplina a través de monográficos caracterizados por el rigor científico, el debate y la pluralidad. Por consiguiente, la revista está dirigida a investigadores, docentes y especialistas en estudios de traducción.

MonTI publicará un número monográfico anual, primero en papel y a continuación en edición electrónica. Igualmente y con el fin de alcanzar un equilibrio entre la máxima pluralidad lingüística y su óptima difusión, la versión en papel admitirá artículos en alemán, castellano, catalán, francés, italiano o inglés, mientras que la edición en Internet aceptará traducciones a cualquier otro idioma adicional y tratará de ofrecer una versión en inglés de todos los artículos.

Más información en:

<http://dti.ua.es/es/monti/monti.html>

MonTI és una revista acadèmica amb vocació internacional promoguda per les universitats públiques valencianes amb docència en traducció i interpretació (Universitat d'Alacant, Universitat Jaume I de Castelló i Universitat de València).

La nostra revista pretén sobretot centrar-se en l'anàlisi en profunditat dels assumptes relacionats amb la nostra interdisciplina a través de monogràfics caracteritzats pel rigor científic, el debat i la pluralitat. Per tant, la revista va dirigida a investigadors, docents i especialistes en estudis de traducció.

MonTI publicarà un número monogràfic anual, primer en paper i a continuació en edició electrònica. Igualment, i a fi d'aconseguir un equilibri entre la màxima pluralitat lingüística i la seua difusió òptima, la versió en paper admetrà articles en alemany, castellà, català, francès, italià o anglès, mentre que l'edició en Internet acceptarà traduccions a qualsevol altre idioma addicional i tractarà d'oferir una versió en anglès de tots els articles.

Més informació a:

<http://dti.ua.es/es/monti-catalan/monti-contacte.html>

MAIN INSTRUCTIONS FOR AUTHORS

MonTI publishes one yearly issue. The contributions, which should be original and unpublished, will be strictly subjected to the following norms:

1. Maximum length: 10,000 words, including works cited.
2. Font and spacing: Font: Times New Roman; Size: 11 pt.; Line spacing: single.
3. Language options: Catalan, English, French, German, Italian or Spanish.
4. The title page should include the following information in this same order:

Title of the article, followed by a blank line. An English translation of the title should be included if this is not in English. Author(s). <e-mail>. Affiliation, followed by a blank line. Abstracts in English and in any of the other four languages. (Maximum length: 150 words each). Keywords: up to five subject headings in each of the same two languages.

5. Parenthetical citations: *MonTI* follows one of the main variants of the Chicago-style citation: Surname(s) (year: pages) or (Surname(s) year: pages).

6. Works cited / references: This section will only include works really cited in the text and will begin after the article has come to an end. The list will be arranged in alphabetical order by author and year of the first edition, and according to the following pattern: Monographs: Author (Surname(s), complete first name). Year (in brackets) *Title* (in italics). City: Publisher. Journal article: Author (Surname(s), complete first name). Year (in brackets). "Title of the article" (with quotation marks). *Name of the journal* (in italics). Volume: Issue, first page-last page (preceded by the abbreviation pp.)

At *MonTI*'s website (<http://dti.ua.es/es/monti-english/monti-authors.html>) numerous examples of each of these variants are available.

7. Deadline: The deadline will be May 31. The contribution and a short CV (a maximum of 150 words) for each of the authors in a separate file will be sent as an attachment (in Word or any other word processor compatible with Word) addressed to the Managing Editor of *MonTI*: <email: monti.secretaria@ua.es>

8. After requesting the editors' approval and receiving the reports from the referees, the journal will provide the authors with a reasoned statement regarding the acceptance of their contributions.

PRINCIPALES NORMAS DE REDACCIÓN

MonTI edita un número anual. Los trabajos originales e inéditos que se propongan para su publicación en la revista se someterán estrictamente a las siguientes normas:

1. Extensión máxima: 10.000 palabras, incluida la bibliografía.
2. Tipo de letra: Times New Roman. Tamaño de letra: 11 pt. Interlineado: sencillo.
3. Lenguas vehiculares: alemán, castellano, catalán, francés, inglés o italiano.
4. La primera página incluirá, por este orden y en líneas sucesivas, lo siguiente:

Título del trabajo, seguido de una línea en blanco de separación. Deberá aportarse, además, la traducción del título al inglés, si el artículo no está escrito en esta lengua. Autor(es). <Correo electrónico>. Centro de procedencia, seguido de una línea en blanco de separación. Resúmenes en inglés y en otra de las lenguas vehiculares (extensión máxima de 150 palabras cada uno). Palabras clave: se aportarán cinco términos en los dos idiomas de los resúmenes.

5. Remisión a la Bibliografía. Se seguirá una de las principales variantes del estilo Chicago de citas: Apellido(s) del autor (año: páginas) o (Apellido(s) del autor año: páginas).

6. Bibliografía: este epígrafe sólo recogerá los trabajos citados en el artículo, y aparecerá después del final del texto. Se ordenará alfabéticamente por autor y año del siguiente modo. Monografías: Autor (apellido(s), nombre completo). Año (entre paréntesis) *Título* (en cursiva). Ciudad: Editorial. Artículo de revista: Autor (apellido(s), nombre completo). Año (entre paréntesis). Título del artículo (entre comillas). *Nombre de la revista* (en cursiva). Volumen: fascículo, páginas de comienzo y fin del artículo (antecedidas por la abreviatura pp.).

En la página web de *MonTI* (<http://dti.ua.es/es/monti/normas-de-redaccion.html>) se puede acceder a numerosos ejemplos de cada una de las variantes de referencia bibliográfica.

7. Envío de originales: el plazo de recepción finalizará el 30 de junio. Los textos –y un breve currículum (150 palabras máximo) de los autores en otro documento– se remitirán en soporte informático (Word o cualquier programa de tratamiento de textos compatible con Word) dirigidos al Secretario de la revista: <e-mail: monti.secretaria@ua.es>

8. La Dirección de la revista, vistos los informes de los asesores y el parecer de los editores, comunicará a los autores la decisión razonada sobre la aceptación o no de los trabajos.

MON TI

Multidisciplinary analysis of phraseological variation
in translation and interpreting
(English & Spanish versions) (pp. 7-64)

Approccio cognitivo alla variazione fraseologica:
alcune concettualizzazioni della pazzia in italiano e implicazioni
per la loro traduzione in spagnolo e tedesco (pp. 65-93)

Modelización computacional de las locuciones verbales
del español chileno a partir de la propuesta
de la léxico-gramática (pp. 94-120)

Estudio de caso de la fraseología empleada en Colombia:
análisis crítico de un corpus *ad hoc* de locuciones (pp. 121-153)

Google Translate vs. DeepL: analysing neural machine
translation performance under the challenge
of phraseological variation (pp. 154-177)

Standardisation in translated language:
A study of collocations in Peninsular and Colombian Spanish
translations of *The Picture of Dorian Gray* (pp. 178-209)

Term and translation variation of multi-word terms (pp. 210-247)

La recepción y traducción de unidades fraseológicas
en el discurso político. Análisis
de un corpus *ad hoc* (EN, FR > ES) (pp. 248-286)

From head to toe: A lexical, semantic,
and morphosyntactic study of idioms in phraseological
dictionaries in English and Spanish (pp. 287-326)

Traducción de unidades fraseológicas gastronómicas
francés-español peninsular/español de México (pp. 327-354)

Las lenguas originarias en la fraseología del español hablado
en Argentina: un estudio basado en corpus (pp. 355-385)

Equivalences of phraseological units in corpora (pp.386-415)

Mogorrón Huerta, Pedro

Cataldo, Silvia

Koza, Walter
& Natalia Rivas Folch

Zuccalà, Daniele
& María Bernarda Espejo Olaya

Hidalgo-Ternero, Carlos Manuel

Valencia Giraldo, M.^a Victoria

León Araúz, Pilar
& Melania Cabezas García

López Ruiz, María del Carmen

Rojas Díaz, José Luis

Navarro-Brotons, Lucía

Cuadrado Rey, Analía

Recio Ariza, M.^a Ángeles