

GLORIA CORPAS PASTOR (Universidad de Málaga)

Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas

Resumen

El presente capítulo aborda los desarrollos recientes de la *fraseología computacional*, la cual engloba tanto el procesamiento de expresiones multiverbales como la aplicación de técnicas de PLN para el análisis de unidades fraseológicas (UFs). Como telón de fondo, se parte del lugar que ocupa la fraseología en los distintos niveles competenciales del MCER, y se destacan las posibilidades que ofrecen las tecnologías lingüísticas para la resolución de cuestiones pendientes, como la detección, selección y análisis de UFs. A continuación se pasa revista a los tipos de unidades más frecuentemente estudiadas por la *fraseología* y el *procesamiento de expresiones multiverbales* a fin de tender puentes de unión entre ambas disciplinas. En este sentido, se presenta una metodología de corpus, de corte pragmático-discursivo, que permite el análisis exhaustivo de UFs, con especial referencia a sus límites formales y núcleos residuales, así como a la productividad y creatividad fraseológicas. Finalmente, la metodología propuesta se amplía para dar cabida a la comparación y al contraste de unidades fraseológicas, con vistas a su traducción. En este trabajo se emplea una amplia gama de aplicaciones de PLN que incluyen corpus en línea, sistemas de gestión de corpus web, bases de datos de conocimiento y sistemas CLIR. Finalmente, se aboga por un enfoque pragmático-tecnológico de la fraseología, basado en el análisis de datos reales.

Palabras clave: fraseología computacional, PLN, enfoques basados en corpus, unidad fraseológica, expresión multiverbal.

Abstract

This chapter deals with recent developments in *computational phraseology*, ie. the processing of MWE and NLP applications for the analysis of phrasological units. The starting point is a discussion of the status of phraseology within the CEFR's proficiency levels. It is established that language technologies can shed light on important—and yet—unsolved issues on the detection, selection and analysis of phrasological units. Next, frequently studied types of multiword units are revised in order to build bridges

between *phraseology* and *processing of multi-word expressions*. Based on extensive corpora and drawing on a discourse-pragmatic approach, a methodology for in-depth analysis is presented. The focus is on formal boundaries, fuzzy cores, productivity and creativity. Finally, the methodology is extended to compare and contrast phraseological units in the source and target languages, with a view to translation. The range of NLP applications used includes on-line corpora, web corpora and linguistic search engines, knowledge databases and CLIR systems. The concluding section argues for a data-driven/pragmatic and technology approach to phraseology.

Keywords: computational phraseology, NLP, corpus-based approaches, phraseological unit, multiword expression (MWE).

1 Introducción

Hoy día la sangre fraseológica no termina de llegar al río de la pragmática. Tampoco parece que se sustente sobre andamiajes tecnológicos lo suficientemente sólidos como para soportar la carga fraseológica ni la idiosincrasia de las unidades que pueblan los diversos universos fraseológicos. La cimentación es una parte muy importante de la estructura del edificio fraseológico, pues el terreno sobre el cual se asienta y que absorbe las cargas de la edificación es de tipo cambiante y propenso a la formación de asientos diferenciales. También es importante que el análisis se realice con las herramientas y técnicas de construcción más adecuadas para cada caso.

Los universos fraseológicos de las lenguas, por su propio dinamismo y riqueza de matices potenciales, solo pueden soportar –y, al mismo tiempo, dar soporte a– cimentaciones que se basen en el uso real de las unidades que los conforman (aproximación fraseopragmática). Actualmente, abundan las descripciones teóricas y las taxonomías *ad hoc*, en su mayoría trufadas de ejemplos “reales”, que se ajustan convenientemente a sus respectivos constructos teóricos. Esto es un buen comienzo, pero queda mucho por hacer aún. Para poder analizar de forma adecuada el comportamiento pragmático-discursivo de las unidades fraseológicas no se puede empezar la casa por el tejado, sino que hay que comenzar por establecer una cimentación profunda, robusta, que soporte los elementos verticales y los forjados de las distintas aproximaciones teóricas. Y todo esto debe ser previo a cualquier comparación y contraste de unidades fraseológicas.

En este capítulo nos vamos a ocupar de las herramientas disponibles para la localización y selección de los terrenos edificables, porque de la calidad del suelo dependerá la calidad final de la construcción. Nuestra perspectiva será eminentemente tecnológica. A modo de catálogo o repertorio, presentaremos una variedad de recursos, aplicaciones y herramientas que aportan nuevos materiales y nuevas vías de análisis para el estudio de la fraseología en contexto y su traducción en el par de lenguas inglés-español. Como telón de fondo analizaremos el emplazamiento de la fraseología en el contexto del Marco Común Europeo de Referencia para las Lenguas.

2 La fraseología en el marco competencial del MCER

La capacidad para reconocer, entender y utilizar las unidades fraseológicas correctamente en contexto forma parte innegable del conocimiento más avanzado de una lengua, ya sea materna, segunda o extranjera (cf. Howarth 1996). Es más, el dominio del componente fraseológico es uno de los indicadores más claros de uso nativo de un idioma, como veremos a continuación.

Según el Marco Común Europeo de Referencia para las Lenguas (MCER) –que, como su nombre indica, es el referente europeo para las políticas lingüísticas y de educación–, los hablantes para comunicarse hacen uso de sus capacidades generales así como de su competencia comunicativa más específicamente relacionada con la lengua. Esta, a su vez, comprende varios componentes: competencia lingüística, sociolingüística y pragmática. La competencia lingüística engloba la léxica, la fonológica, la gramatical, la ortográfica, la semántica y la ortoépica.

La fraseología se trata preferentemente dentro de la competencia lingüística y, dentro de esta, en la léxica¹. En el MCER se aborda la fraseología de pasada y desde una perspectiva eminentemente atórica, a tenor de las denominaciones utilizadas, que no son más que simples calcos del inglés, sin mayor precisión terminológica, y que, en muchos casos, inducen a confusión. No obstante, para este trabajo hemos preferido utilizar la propia terminología empleada en la versión española del Instituto Cervantes. Así pues, como parte de la competencia léxica se incluyen las expresiones hechas, de las cuales se dice que

.....
1 Por competencia léxica el MCER entiende “el conocimiento del vocabulario de una lengua y la capacidad para utilizarlo” (MECD 2002: 108). Comprende elementos léxicos y gramaticales.

“se componen de varias palabras que se utilizan y se aprenden como un todo” (MCER 2002: 108), y que engloban los siguientes tipos de unidades²:

- a) fórmulas fijas con funciones comunicativas (*how do you do?*, *buenos días*), arcaísmos residuales (*be off with you!*, *desfacer entuertos*), refranes y proverbios (*A stitch in time saves nine*; *No por mucho madrugar, amanece más temprano*);
- b) modismos (*a long shot*, *estirar la pata*) e intensificadores (*as white as snow*, *blanco como la pared*);
- c) estructuras fijas (*Please may I have...*, *¿Sería tan amable de + infinitivo...?*);
- d) otras frases hechas, como verbos con régimen preposicional o partículas (*to put up with*, *convencerse de*) y locuciones prepositivas (*in front of*, *delante de*); y
- e) régimen semántico³, esto es, palabras que habitualmente se utilizan juntas (*to make a speech/mistake*, *cometer un crimen/error*).

Dentro de la competencia sociolingüística (MCER 2002: 116 y ss.), que aborda aquellos usos de la lengua que no se tratan en otros apartados, se incluyen también unidades fraseológicas como parte de los marcadores lingüísticos de relaciones sociales: saludos (*Good morning!*, *Hasta luego*) y frases interjectivas (*Bloody hell!*, *¡Hay que ver!*); las expresiones de sabiduría popular, esto es, “fórmulas fijas” como los refranes y los modismos (tratados también como parte de la competencia léxica) y las frases estereotipadas (*It takes all sorts to make a world*, *De todo hay en la viña del Señor*). En cambio, no se menciona expresamente ningún tipo de unidad fraseológica ni se dan ejemplos de estas en las secciones dedicadas a la competencia pragmática –entendida como la capacidad del usuario para estructurar el mensaje desde el punto de vista discursivo, funcional y organizativo–, pero se sobreentiende el uso de fórmulas en diversas categorías de microfunciones relacionadas con las actitudes (por ejemplo, pedir disculpas), la persuasión (p. ej., dar ánimo), la vida social (saludos, despedidas, etc.), la estructuración del discurso, etc.

.....
2 Ilustraremos los tipos de unidades abordadas en inglés y en español; para ello vamos a utilizar las versiones del Marco Común de Referencia existentes para ambas lenguas (CEFR y MCER).

3 El régimen semántico también se menciona como parte de la competencia semántica (MCER 2002: 112). Sería asimilable a la noción de rango colocacional.

Por otra parte, dentro de los distintos niveles de competencia lingüística establecidos en el MCER⁴, se reconoce el papel de la fraseología en los niveles más altos. Así, en el cuadro de autoevaluación de niveles comunes de referencia⁵, para el C2 figura lo siguiente (MCER 2002: 31):

Tomo parte sin esfuerzo en cualquier conversación o debate, y conozco bien modismos, frases hechas y expresiones coloquiales. Me expreso con fluidez y transmito matices sutiles de sentido con precisión. Si tengo un problema, sorteo la dificultad con tanta discreción, que los demás apenas se dan cuenta.

Más específicamente se menciona la fraseología como indicador de riqueza léxica para C1 y C2, como se puede apreciar en la tabla resumen sobre adecuación sociolingüística del MCER (2002: 109), que reproducimos parcialmente a continuación:

	adecuación sociolingüística
C2	Tiene un buen dominio de un repertorio léxico muy amplio, que incluye expresiones idiomáticas y coloquiales; muestra que es capaz de apreciar los niveles connotativos del significado. Es plenamente consciente de las implicaciones de carácter sociolingüístico y sociocultural en el uso de la lengua por parte de los hablantes nativos, y sabe reaccionar en consecuencia. Media con eficacia entre hablantes de la lengua meta y de la de su comunidad de origen, teniendo en cuenta las diferencias socioculturales y sociolingüísticas.
C1	Tiene un buen dominio de un amplio repertorio léxico que le permite superar con soltura sus deficiencias mediante circunloquios; apenas se le nota que busca expresiones o que utiliza estrategias de evitación. Buen dominio de expresiones idiomáticas y coloquiales. Comprende las películas que emplean un grado considerable de argot y de uso idiomático. Utiliza la lengua con flexibilidad y eficacia para fines sociales, incluyendo los usos emocional, alusivo y humorístico.

La riqueza léxica requerida en el nivel B2 es lógicamente menor, en tanto se espera que el usuario tenga un amplio vocabulario pasivo o de lectura, que no incluye “modismos poco frecuentes” (MCER 2002: 81).

.....

- 4 Según la distinta habilidad lingüística, el MCER establece una escala de niveles que va del principiante (A1) al usuario que domina la lengua (C2).
- 5 El MCER establece seis niveles de competencia principales, agrupados en tres bloques: A1 (usuario básico: acceso); A2 (usuario básico: plataforma); B1 (usuario independiente: umbral); B2 (usuario independiente: avanzado); C1 (usuario competente: dominio operativo eficaz) y C2 (usuario competente: maestría).

Desde el nivel B2 se espera, además, que el hablante sea capaz de expresarse de forma apropiada al contexto situacional, lo cual incluye un mayor grado de corrección en el uso de los modismos. Al igual que ocurre con la competencia lingüística, en el caso de la competencia sociolingüística el nivel de exigencia se ve aumentado en los niveles superiores, que va desde ser capaz de reconocer un buen número de expresiones idiomáticas (C1) hasta tener un buen dominio de tales expresiones y apreciar los niveles connotativos (C2).

En cuanto a la interacción oral, no se precisa exactamente a partir de qué nivel el usuario debe ser capaz de reconocer y utilizar correctamente las unidades fraseológicas. Sí se indica claramente que no se espera que sea capaz de hacerlo en el nivel A1: “Responde en una entrevista a preguntas sencillas y directas sobre datos personales, si se habla muy despacio y con claridad sin modismos ni frases hechas” (MCER 2002: 81).

En resumen, las unidades fraseológicas empiezan a aparecer en el nivel avanzado (B2), hasta formar parte de la competencia integral –lingüística, sociolingüística y pragmática– del bloque de usuarios “competentes”, esto es, C1 (dominio operativo eficaz) y C2 (maestría). Sin embargo, en el MCER nada se dice acerca de las unidades fraseológicas que corresponderían a cada nivel de competencia, ni se proponen criterios claros, ni tampoco se especifican las capas semánticas ni los rasgos pragmático-discursivos que deberían conocerse de cada unidad por niveles.

3 La fraseología en el contexto de las tecnologías lingüísticas

Determinar qué secuencias de palabras constituyen unidades fraseológicas (UFs) y cuáles no, decidir qué unidades fraseológicas incluir para la enseñanza de idiomas o la descripción lexicográfica, a qué nivel de competencia corresponde una determinada unidad fraseológica, o con qué grado de detalle se debe tratar una unidad fraseológica concreta en cada uno de esos niveles, describir el comportamiento discursivo de las unidades en distintos registros y tipos de textos, analizar las valoraciones y las implicaturas asociadas a determinados tipos de UFs, etc. de forma automatizada son algunas de las posibilidades que se abren a partir del uso de las tecnologías lingüísticas (TL)⁶. A la

.....
6 Entendemos por *tecnología lingüística* “la derivada de aplicar los conocimientos sobre la lengua al desarrollo de sistemas informáticos” (<<http://oesi.cervantes.es/oesi/tls.jsp>>). Sobre este tema, consúltese la monografía

aplicación de TL para la detección, extracción, análisis y representación de unidades fraseológicas se le denomina “fraseología computacional” (*computational phraseology*)⁷ o, más comúnmente en el ámbito del PLN⁸, “procesamiento de expresiones multiverbales” (*processing of multi-word expressions [MWE]*).

Tradicionalmente, se distinguen dos grandes ámbitos dentro del estudio computerizado de la fraseología (Heid 2008: 336 y ss.). A un lado se sitúan los métodos y técnicas computacionales para procesar textos que contienen unidades fraseológicas. Aquí se inscribirían gran parte de los estudios realizados desde el campo de la traducción automática, la generación de lenguaje natural, las gramáticas y los lexicones computacionales. Estos suelen disponer de formatos para la descripción y representación de esas expresiones pluriverbales que dificultan los cuatro niveles de análisis, debido sobre todo a las dificultades de segmentación que entrañan, a su significado compositivo global y su defectividad gramático-sintáctica, así como a la ambigüedad adicional que plantean a la hora de procesar los textos en los cuales se insertan.

Al otro lado se hallan los métodos y técnicas computacionales, así como las herramientas informáticas utilizadas para describir, representar, clasificar o analizar unidades fraseológicas. Se trata de aplicaciones para el análisis asistido de la fraseología, que pueden ser de utilidad para usuarios de muy diverso tipo, por ejemplo, lexicógrafos, lingüistas, profesores de idiomas, traductores, correctores, posteditores... y, cómo no, fraseólogos. Aquí se distinguen dos tipos de herramientas y aplicaciones, según el tipo de tarea para la cual hayan sido desarrollados: a) para detectar y extraer UFs; y b) para localizar ejemplos ilustrativos de UFs y analizar sus características lingüísticas y de uso⁹.

de Lavid (2005).

7 Para esta breve introducción a la fraseología computacional, seguimos los trabajos de Rothkegel (2007) y Heid (2007, 2008).

8 El Procesamiento de Lenguaje Natural (PLN) es una subdisciplina de la Inteligencia Artificial y la rama aplicada de la Lingüística Computacional (LC), que se ocupa de investigar y formular mecanismos computacionalmente efectivos que permitan la comunicación fluida hombre-máquina. El PLN trata de simular la comunicación humana integral (comprensión y generación, pasando por la traducción) a partir de cuatro niveles básicos de análisis: léxico-morfológico, sintáctico, semántico y contextual o pragmático. Los trabajos recogidos en Dale, Moisl y Somers (2000) y en Indurkha y Damerou (2010) constituyen excelentes introducciones al PLN; para sus estribaciones computacionales, véase Mitkov (2003). Las tecnologías lingüísticas recubren los desarrollos computacionales del PLN, así como las técnicas de análisis, síntesis y generación del PLN.

9 Algunas de estas aplicaciones o herramientas pueden realizar ambos tipos de tareas.

3.1 Identificación de tipos de unidades

Términos como *fraseología*, *fraseología computacional*, *fraseologismos*, *unidad fraseológica* y similares no resultan demasiado frecuentes en los trabajos de PLN, donde se prefiere hablar más bien de *expresiones multiverbales* y de su procesamiento¹⁰. Sin embargo, salvo contadas excepciones, la noción de expresión multiverbal es equiparable a la de unidad fraseológica (Heid 2008: 338). La elección de uno u otro término depende, pues, de la orientación del investigador, ya que ambos parecen denominar fenómenos muy similares. Compruébese, si no, la reformulación de la definición de Sag *et al.* (2002) que ofrecen Baldwin y Kim (2010: 268):

Multiword expressions (MWEs) are lexical items that: (a) can be decomposed into multiple lexemes; and (b) display lexical, syntactic, semantic, pragmatic and/or statistical idiomaticity¹¹.

La identificación de unidades fraseológicas en los corpus constituye una de las dificultades por antonomasia de la fraseología computacional. Sin embargo, apenas ha habido contacto entre la Lingüística Computacional y la Fraseología: mientras los primeros han seguido un planteamiento estadístico inspirado en la lingüística de corpus, los segundos se han centrado en unidades menos frecuentes, fuertemente idiomáticas, las cuales identifican a través de la intuición (Tschichold 2008: 364).

Por este motivo, en el ámbito del PLN¹² se suelen tratar tipos muy heterogéneos de expresiones multiverbales estadísticamente significativas, al margen de cualquier taxonomía fraseológica al uso. La unidad de medida, como si dijéramos, es el *n*-grama, esto es, una secuencia de *n* palabras (como mínimo dos), entendiendo por palabra una secuencia de caracteres separada por espacios. Por ejemplo, dentro de la recuperación de información (RI) se aborda frecuentemente la identificación y clasificación de entidades nombra-

.....

10 En este trabajo utilizaremos indistintamente *expresión multiverbal*, *unidad fraseológica* o, simplemente, *unidad*.

11 Las expresiones multiverbales son elementos léxicos que: a) pueden descomponerse en varios lexemas; y b) muestran idiomática léxica, sintáctica, pragmática y/o estadística (la traducción es nuestra). Nota explicativa: el término *expresión multiverbal* no cuenta con una sigla institucionalizada en español, como sí ocurre en inglés (*MWE*).

12 Aunque la traducción automática (TA) se inscribe dentro del ámbito del PLN, preferimos abordarla en el apartado 3.2.

das multiverbales (en inglés, *MW named entities*)¹³. Se podrían considerar locuciones terminológicas o denominativas, ya que generalmente se refieren a personas (*Barack Obama, Julio Iglesias*), lugares (*New Orleans, San Sebastián*) u organizaciones (*United Nations Organization, Servicio Andaluz de Empleo*).

También se ha prestado especial interés a las expresiones multiverbales con significados categoriales en general, y, más específicamente, a los verbos con partícula, a los predicados complejos y a los compuestos nominales con o sin guión (*heart attack, mother-in-law, bolso de mano*). Muchos de estos se asimilarían a locuciones: prepositivas (*due to, en pos de*), conjuntivas (*as though, con tal de*), adverbiales (*in the nick of time, cuanto antes*), adjetivas (*long in the tooth, corto de medios*), verbales (*to take off, poner alg. todo de su parte*), y, dentro de los tipos señalados, comparaciones estereotipadas (*as safe as houses, más tonto que Abundio; to work like a dog, comer como una lima*) y binomios¹⁴ irreversibles (*now and then, high and mighty, to toss and turn, heads or tails; largo y tendido, coser y cantar, mondo y lirondo, vivo o muerto*).

La extracción de colocaciones constituye un tema de gran interés en PLN. Los trabajos se han centrado en los tipos de colocaciones más frecuentes, que suelen estar compuestas por dos palabras en función de verbo y complemento directo (*to draw attention, entablar amistad*) o complemento preposicional (*to jump to a conclusion, poner a prueba*) –incluidas las construcciones con verbos soporte o deslexicalizados¹⁵ (*to make an attempt, tener repercusión*)–, o bien forman sintagmas nominales compuestos de sustantivo más adjetivo (*reliable source, ignorancia supina*) o sustantivo en función adjetiva (*copycat criminal, viaje relámpago*).

Generalmente se recurre a medidas de asociación, que permiten calcular estadísticamente la probabilidad de que dos palabras aparezcan juntas, comparando la frecuencia real de coaparición del par de palabras analizado (bigrama) con su frecuencia estimada, esto es, según la frecuencia individual de aparición y distribución, por separado y en relación con su combinación con otras palabras (cf. Sinclair 1991). En las colocaciones, la frecuencia real debe ser superior al nivel de probabilidad convencional (significación estadística).

.....
13 Sobre este tema, consúltense los trabajos recogidos en Sekine y Ranchhod (2009).

14 En inglés se dan también trinomios irreversibles, esto es, secuencias compuestas por tres palabras léxicas unidas por una conjunción copulativa que no permiten el reordenamiento de sus componentes (*bell, book, and candle; calm, cool, and collected*).

15 Alonso Ramos (2004) los denomina “verbos de apoyo”.

Los tipos de unidades mencionadas hasta ahora han sido estudiadas desde el punto de vista de su posible identificación y extracción automatizada en los textos en los que aparecen. Son los candidatos por excelencia para herramientas de detección como Sketch Engine¹⁶ y Text-NSP¹⁷, en las cuales se usan medidas de asociación especialmente indicadas para el análisis de pares de palabras (o bigramas)¹⁸, a veces en combinación con el etiquetado morfo-sintáctico (cf. Seretan 2011). Precisamente, la mayor parte de los programas de gestión de corpus¹⁹ y sistemas integrados de análisis léxico, como WordSmith's Tools²⁰, AntConc²¹, o Compleat Lexical Tutor²², son igualmente de base estadística.

Aunque en menor medida, en el ámbito del PLN se han ocupado de otros tipos más complejos de expresiones multiverbales como *to twist so. round one's little finger, to put one's finger on sth. get hold of the wrong end of the stick, you know*, en inglés, y *llegarle uno a alg. a la altura de las babuchas, A buenas horas, mangas verdes, No se retire*, en español. La complejidad de tales unidades puede provenir de su longitud, que suele ser superior al bigrama, de su carácter discontinuo, de la presencia de casillas vacías actualizables actualizadamente, de su significado idiomático, de sus restricciones discursivas y condiciones de uso pragmáticas, etc. Se trataría de unidades perfectamente encuadrables entre las locuciones verbales, las clausales, las adverbiales con funciones discursivas (marcadores), los refranes, los dichos, las fórmulas rutinarias, etc. En estos casos, la fraseología computacional se centra más bien en el estudio descriptivo mediante corpus y la representación, en su caso, de las características lingüísticas (semánticas, sintácticas y pragmáticas) de las unidades analizadas.

Aquí la dificultad reside en detectar las unidades correspondientes, pues al estar compuestas de más de dos palabras, las medidas de asociación léxica al

.....

16 <<http://www.sketchengine.co.uk/>>

17 <<http://search.cpan.org/dist/Text-NSP/>>

18 Entre los métodos estadísticos más utilizados para determinar el grado de asociación léxica se encuentran el de información mutua (*MI-mutual information*) y la ratio de verosimilitud (G2), que en inglés se denomina *G-square*, *G score* y, también, *log-likelihood* (cf. Evert 2005; Baker 2006: 101-103).

19 Entendemos por *corpus* una colección de textos en soporte digital, que ha sido compilada conforme a unos criterios de diseño apropiados que le confieren representatividad (EAGLES 1996; Corpus Pastor 2008). Sobre programas de gestión de corpus, véase Corpus Pastor (2007b).

20 <<http://www.lexically.net/wordsmith/>>

21 <<http://www.antlab.sci.waseda.ac.jp/software.html/>>

22 <<http://www.lexutor.ca/>>

uso no son efectivas para calcular y contrastar la frecuencia de coaparición (real y estimada). Además, se precisan corpus muy extensos para localizar este tipo de unidades. Por ejemplo, en el corpus *BYU-BNC: British National Corpus*²³ (Davies 2004-) no se encuentra ningún ejemplo de *to out-Herod Herod*²⁴ ('ser peor que Herodes, comportarse incluso de forma más violenta y soez'), una unidad que suele aparecer recogida en los diccionarios y cuyo origen se remonta a Shakespeare (*Hamlet*, III.2).

Y es que las unidades fraseológicas con saliencia cognitiva y plenamente institucionalizadas no siempre son frecuentes estadísticamente hablando. De ahí la necesidad de contar con corpus muy extensos para el estudio de la fraseología. Para Tschichold (2008: 363), el tamaño adecuado de corpus estaría en torno a varios cientos (o incluso miles) de millones de palabras. Y añade la autora que, a pesar de que muchas lenguas disponen ya de corpus de tales dimensiones, todavía no se cuenta con las herramientas computacionales adecuadas para poder estudiar la variabilidad potencial de dichas unidades en esos corpus. Las posibilidades de variación y modificación complican aún más si cabe la identificación de las unidades, el establecimiento de sus formas canónicas o la capacidad de predecir sus posibles modificaciones (cf. los trabajos de Moon 1998 y Wry 2008). Volviendo al *BYU-BNC*, este solo contiene 8 ejemplos de *to twist so. round one's little finger*.

- (1) Mary knew she could *twist* him around her *little finger*.
- (2) She had always known that Liza could *twist* him round her *little finger*.
- (3) 'She can *twist* me round her *little finger*', is said with glowing pride.
- (4) I'm not saying that she *winds* Daddy round her *little finger*, she does not.
- (5) All the evidence suggests that she had her confessor *wrapped* around her *little finger*: while claiming the complete obedience appropriate to her spiritual supervisor she in fact got round every

23 El *BYU-BNC* es una versión web del *British National Corpus*, creada por Davies (2004-). El *BNC* es un proyecto liderado por Oxford University Press, en el cual participan las Universidades de Lancaster y Oxford, así como las editoriales Addison-Wesley Longman y Larousse Kingfisher Chambers. El *BNC* contiene unos 100 millones de palabras de inglés oral y escrito (1970-1993). Es quizá el corpus de referencia para la lengua inglesa más utilizado tanto en Lingüística Computacional como en Lingüística Aplicada.

24 "To go beyond even Herod in violence, brutality, or extravagant language. In the old miracle plays Herod was the type of tyranny and violence, both of speech and of action." (DPF).

attempt he ever made to make her moderate her life style; she persuaded him to confer his blessing and approval on her choices.

- (6) You were extremely fond of all the MacKay men—and they all let you *wrap* them round your *little finger*.
- (7) ‘Just because you could have *twisted* those three bimbos round your *little finger* this afternoon, don’t assume all women are cut from the same cloth.’
- (8) Spencer had been spoiled, treated for far too long as a baby and he had grown up knowing how to *twist* his mother around his *little finger*.

En 3 casos el componente verbal de la unidad es *twist*, pero en otros 2 se encuentran cuasi sinónimos del tipo *wind* (1) y *wrap* (2). Resulta curioso constatar que *twist* suele combinarse con un verbo modal (*can, could*), algo que no ocurre cuando el componente verbal es diferente del canónico. Se observa variabilidad también en relación a la preposición que forma parte de la unidad: *round* (5) o *around* (3). Aunque los ejemplos ilustrativos no son demasiado numerosos, salta a la vista la deixis típica de tercera persona que gobierna el uso discursivo de la unidad, así como la restricción pragmática de los participantes: el rol de agente se reserva típicamente a la mujer (*she, Mary, her*), mientras que son hombres los que suelen desempeñar el rol de paciente o experimentante (*him, confessor, spiritual supervisor*). La carga pragmática evidencia una diferencia de poder entre ambos participantes, siendo la mujer agente la que ocupa la posición dominante mediante una manipulación consciente del hombre, quien parece consentirlo a veces y aceptarlo con agrado, como sugieren las palabras y frases con las que coaparece la unidad en los ejemplos analizados: [she] *knows she can... knew/known she could..., got round..., persuaded him..., [he] to confer his blessing and approval..., let you..., with glowing pride*.

Cuando es un hombre el que desempeña el rol de agente, como en los dos últimos ejemplos, se ofrece una visión negativa de ambos: en el ejemplo 7 la manipulación del hombre es solo posible gracias a la estupidez de las mujeres con el rol de pacientes (*bimbos*)²⁵, pero se le advierte que no todas las mujeres

.....
25 En inglés, la palabra *bimbo* denomina a una mujer muy atractiva, pero poco inteligente y frívola. Pertenecce al nivel informal y tiene connotaciones peyorativas (cf. OSD).

están cortadas por ese mismo patrón (...*cut from the same cloth*); mientras que en el ejemplo 8 la imagen que se proyecta del hombre con el rol de agente es la de una persona inmadura y malcriada en un entorno sobreprotector (*spoiled, ...treated like a baby...*), que ha aprendido a manipular a su madre para conseguir lo que quiere.

Unos pocos ejemplos bastan para extraer datos relevantes acerca de una unidad fraseológica o expresión multiverbal “en acción”. Ahora bien, se necesita analizar muchas más líneas de concordancia para poder estudiar en profundidad el comportamiento pragmático-discursivo de una determinada unidad, según su distribución textual, su preferencia por determinados géneros o registros, etc.

Por otra parte, las unidades más complejas suelen presentar el problema añadido de la ambigüedad. Así, no resulta fácil distinguir cuándo una secuencia de *n* palabras constituye una unidad fraseológica o bien se trata simplemente de su homófono literal. Los ejemplos que siguen a continuación, extraídos del *Corpus de Referencia del Español Actual* (CREA)²⁶, ilustran el caso de la secuencia *meter un gol*, que puede ser una locución idiomática con el significado de ‘obtener un triunfo sobre él, a veces con engaño’ (DRAE) (ejemplos 9-12), o bien una colocación literal, que hace referencia a la entrada del balón en la portería (ejemplos 13-16)²⁷.

- (9) De haber sabido que iban a ser chavales, no habría dado la autorización... Sólo puedo reconocer que me han *metido un gol* y que habrá que pedir detalles en lo sucesivo.
- (10) En una cafetería atestada, Meg Ryan finge un orgasmo ante un paralizado Billy Crystal (Cuando Harry encontró a Sally, 1989). Quiere demostrarle que al más sagaz de los amantes se le puede *meter un gol*.
- (11) El Barcelona no ofrecía respuestas concluyentes, debido a que Zeljko Obradovic le había *metido un gol* a Aíto desde la pizarra cambiando las marcas esperadas al situar a “Skeeter” Henry –que

.....
26 El CREA consta de más de 150 millones de palabras de todas las variedades del español. Está disponible para su consulta en el sitio web de la Real Academia Española (<<http://www.rae.es/>>).

27 En el caso de la colocación se permite la sustitución del componente verbal *meter* (colocado generalista) por *marcar* (colocado especializado, propio de la prensa deportiva): “El dilema es el siguiente: O alguien marca un gol o habrá tanda de penaltis” (CREA). Sin embargo, no hemos encontrado ejemplos en el CREA de este tipo de sustitución paradigmática en el caso de la locución.

jugó cojo todo el partido– con Xavi Fernández y a Santos con Crowder.

- (12) Y es que con esta historia no sólo te has caído tú por los suelos. También me he caído yo. Ante ti, porque al engañarme me has *metido un gol*, has roto, aunque no seas consciente del todo, mi imbatibilidad.
- (13) El Real Madrid decidió no recurrir ante la UEFA la sanción de un partido de suspensión a Raúl por *meter un gol* con la mano ante el Leeds en la Liga de Campeones.
- (14) Francia acaba de caer eliminada del Mundial sin ni siquiera haber *metido un gol*.
- (15) El Atlético de Madrid cedió ayer el liderazgo al Espanyol al no ser capaz de *meter un gol* a un Sevilla al que arrolló durante casi todo el partido.
- (16) Y nada, y así, entonces es muy gracioso porque mientras Diana venía diciendo a casa: pues, mamá, me han elegido para hacer de virgen en el Belén de Navidad, no sé cuánto, y me tienes que comprar tal o cual, o sea, superfina, Alba venía diciendo: papá, he *metido un gol*, no sé qué, no sé cuántos.

Desde un punto de vista computacional, una manera de desambiguar la lectura semántica de la unidad sería aplicando técnicas de distribución léxica y textual, que prevean la lectura literal cuando se trate de crónicas deportiva o textos de la prensa especializada en deporte, que versen especialmente sobre fútbol, fútbol-sala, balonmano o waterpolo, y cuando aparezcan en contexto palabras (*partido*, *suspensión*) y colocaciones (*recurrir una sanción*, *ceder el liderazgo*, *caer eliminado*) típicas de tales dominios, así como entidades nombradas de instituciones, jugadores, ligas y clubes (*UEFA*, *Liga de Campeones*, *Mundial*, *Atlético de Madrid*, etc.). Sin embargo, habría ejemplos que no pasarían dichos filtros. Por ejemplo, el 11 actualiza la lectura idiomática al tratarse de otro deporte, en este caso, baloncesto; mientras que el 16 activa la lectura literal de la secuencia, aun cuando no se dan las condiciones descritas anteriormente.

Casos similares de ambigüedad potencial se producen cuando las expresiones multiverbales idiomáticas tienen como homófonos secuencias libres de palabras. Ello dificulta notablemente la identificación e interpretación de tales secuencias en PLN. Baldwin y Kim (2010: 289) mencionan el caso

típico de *to kick the bucket* ('morir'), de significado completamente idiomático y compositivo, cuya forma canónica podría utilizarse de forma literal en contexto (por ejemplo, "Kim kicked the bucket in frustration, and stormed out of the room").

Hemos buscado la unidad en el BYU-BNC, que arroja 11 resultados. En todos aparece la unidad fraseológica con su significado idiomático, ya sea en un uso lingüístico (17) o metalingüístico (18), pero también se da un caso de uso literal (19).

- (17) 'It's just that Uncle was a cautious old devil and—he looked away—he got the impression I was a bit of a spendthrift because—well, because I used to get through my allowance pretty rapidly when I was away at school, and... oh, hell, he wanted to make sure I was going to be dull and sensible about all that money when he finally *kicked the bucket*.
- (18) It has long been recognised that expressions such as to pull someone's leg, to have a bee in one's bonnet, *to kick the bucket*, to cook someone's goose, to be off one's rocker, round the bend, up the creek, etc. are semantically peculiar.
- (19) Jinny was so startled that she nearly *kicked the bucket* over.

Lo mismo ocurre en el caso de la unidad española *meter la pata*²⁸. Todos los ejemplos del CREA muestran el uso exclusivamente idiomático de la unidad, así como su preferencia por formas verbales concretas (*meter*, *metido*, *metiendo*, *metas*) y la nominalización (*metedura*). No obstante, frecuentemente hallamos casos (20-22) que requieren una doble interpretación: a) literal, ante la presencia implícita o explícita del sinónimo *pierna* (interpretación primaria); y b) idiomática (interpretación secundaria o residual), que evoca el significado unitario de la unidad en un segundo plano, creando de esta manera un efecto estilístico por modificación externa (cf. Corpas Pastor 1996).

- (20) El reto, dicho así en formato simple, es diseñar una nueva urbe tapando los agujerillos en los que llevamos *metiendo la pata* o *pierna* desde hace años y aprovechar el tirón para construir algo nuevo y reluciente: se acabó vivir en zonas degradadas, las

.....
28 "loc. verb. coloq. Hacer o decir algo inoportuno o equivocado" (DRAE, s. v. *pata*).

agresiones al patrimonio, las torpezas urbanísticas y la escasez de equipamientos.

- (21) Hay demasiada diferencia entre este Barça'95 y la elite europea. Y, además, el equipo demostró que está tocado física y psicológicamente. Lo demostró después de haber disimulado durante más de una hora, lo demostró cuando era el momento de no conformarse con la prórroga. Ahí salió el Barça primitivo, con los defensas intentando *meter la pata* para evitar lo que ya era irremediable.
- (22) Don Emilio Alonso ha hecho un acto de humildad, ejercicio tan infrecuente en un político como el del cumplimiento de las promesas electorales. Don Emilio ha dicho que su desahogo contra Convergencia y contra los periodistas cuando comentó el "caso Flick" fue una *metedura de pata*. "He *metido la pata*", ha dicho humildemente don Emilio... Bienvenida la rectificación o explicación, por la cual ya tenemos la evidencia de que don Emilio Alonso tiene *pata*, y que la *mete*, pero que también la saca, o sea, que está en el *metisaca*, o en el *saca y mete*, como la chica del 17 de la plazuela del Tribulete.

El ejemplo 22 contiene un caso curioso de interpretación múltiple. Las dos primeras instancias de la secuencia [meter la pata] activan su significado fraseológico idiomático habitual ('equivocarse, cometer un error') como interpretación primaria. Queda latente la interpretación literal, que se reactiva a partir de la tercera instancia de la secuencia: *la mete*. Aquí el pronombre *la* funciona como elipsis de objeto directo en referencia anafórica a la palabra *pata*, que, a su vez, constituye el objeto directo del verbo *tener*. Dicha combinación {tener + pata} desencadena, por así decirlo, la interpretación literal latente en las instancias anteriores de la secuencia y provoca una reinterpretación de la palabra *pata*, que, por analogía con la unidad fraseológica, se carga de un nuevo significado fraseológico ('error, equivocación'). Desde ese nuevo estatus autónomo, la palabra *pata* se vuelve productiva: ya está lista para la neología y la creatividad léxica –igual que se *mete la pata* ('se comete un error o equivocación'), se *saca la pata* ('se corrige el error o equivocación')–, y para el juego de palabras con *metisaca*²⁹, que hace

.....
29 "f. *Taurom*. Estocada imperfecta, en la cual el diestro clava el estoque en la res y lo saca rápidamente sin

referencia también a la rapidez de reacción ante un error en el estoque, y con *saca y mete*, que alude a un tipo de sexo rápido e inmediato.

A día de hoy, los sistemas actuales de identificación (detección y extracción) de expresiones multiverbales no disponen de las técnicas necesarias para poder procesar estos fenómenos fraseológicos, ni mucho menos realizar una interpretación adecuada en contexto.

3.2 Comparación y contraste de unidades

Mediante el análisis de corpus se puede estudiar el funcionamiento real de las unidades fraseológicas, sus preferencias semánticas, sus implicaturas y valoraciones pragmáticas, sus rasgos discursivos, sus tendencias a la modificación en contexto, etc. Cuando el análisis abarca las unidades de dos o más lenguas, nos hallamos en el terreno de la *fraseología contrastiva*. Es ahí donde se determinan las semejanzas y diferencias entre los respectivos sistemas fraseológicos, así como las relaciones de correspondencia e (in)equivalencia que se establecen entre sistemas y unidades. Cuando de los parámetros de comparación y sus resultados se pasa al estudio de las equivalencias textuales y los procedimientos y técnicas de traducción nos adentramos ya en el terreno de la *traductología*. Ambas disciplinas son necesarias a la hora de comparar unidades fraseológicas de lenguas diferentes mediante el nuevo paradigma de investigación que supone el PLN, dentro del cual tiene un lugar preferente la *lingüística de corpus*, tanto desde su vertiente gramático-léxica como desde la posición que ocupa dentro del inventario de recursos y técnicas del PLN.

Por limitaciones de espacio no podemos ofrecer un panorama completo del estado de la cuestión. Intentaremos, al menos, dar unas breves pinceladas para contextualizar. En PLN y LC se trabaja en el desarrollo de sistemas y aplicaciones basados en corpus para la detección y extracción automatizada de expresiones multiverbales equivalentes (cf. Corpas Pastor 2008: 75 y ss.). Al igual que en el caso monolingüe, las investigaciones se han centrado en las unidades constituidas por bigramas (frecuentemente colocaciones), trigramas y, ocasionalmente, en n-gramas de longitud superior. Especialmente relevantes son los trabajos realizados en el seno de los proyectos METIS³⁰ y ASSIST³¹, que

.....
soltarlo, por considerar imperfecto el golpe⁹ (DRAE, s. v. *metisaca*).

30 METIS: *Statistical Machine Translation Using Monolingual Corpora I y II* (UE, n.º de ref. IST-2001-32775). Para más información, consúltese la URL: <<http://www.ilsp.gr/metis/>> (fase I) y <<http://www.ilsp.gr/metis2/>> (fase 2).

utilizan corpus comparables, compuestos de textos originales equiparables en varias lenguas. Estos siguen la estela de las investigaciones llevadas a cabo a principios de los noventa, como, por ejemplo, el proyecto *Multilingual Corpora* (contrato nº 57/98, Consejo de Europa), bajo la dirección de John Sinclair; o el “Translators’ Workstation” (módulo de extracción), desarrollado en el seno del proyecto AQUILEX (cf. Marinai *et al.* 1991; Peters y Pichi 1995).

Pero donde tradicionalmente ha venido suscitando mayor interés la identificación de equivalentes multiverbales es en el ámbito de la traducción automática (TA). De hecho, fue ahí donde primero tuvieron que lidiar con las denominadas “unidades multiverbales” (*multiword units*) (cf. Arnold *et al.* 1994: 122 y ss.). Desde un principio, programas como SYSTRAN contaron con lexicones computacionales de expresiones multiverbales, así como con subrutinas para identificar y traducir dichas unidades. Posteriormente, la representación de las expresiones multiverbales en dichos lexicones se hizo más compleja para poder ser incorporada al análisis sintáctico y semántico profundo de tales sistemas, especialmente en los modelos basados en reglas o simbólicos (por ej., ATR, Vermobil, Eutrans, C-Star, etc.). Los modelos estadísticos y los basados en ejemplos (EBMT), como MOSES, se nutren de análisis de corpus y están especialmente diseñados para detectar las expresiones multiverbales.

En TA la extracción de unidades se suele realizar a partir de corpus paralelos, que contienen textos originales en la lengua L1 y sus traducciones a otra(s) lengua(s) (L2, L3, Ln). Aquí presentan evidentes similitudes con los sistemas de memorias de traducción (MT) y los sistemas de extracción de terminología bilingüe. Generalmente, primero se realiza la extracción monolingüe en la lengua 1 y luego se emplean distintas técnicas de alineación para extraer las unidades equivalentes en la lengua 2. Para la extracción de unidades se emplean modelos simbólicos basados en información categorial (Dagan y Church 1994; Okita *et al.* 2010), modelos estadísticos que emplean medidas de asociación para identificar candidatos (cf. Vintar y Fisier 2008) y modelos híbridos que combinan los dos anteriores (Wu y Chang 2004; Sharoff, Babych y Hartley 2006; Boulaknadel *et al.* 2008; Mundraya *et al.* 2008; etc.)³². Y para la alineación, se emplean las técnicas habituales basadas en las propuestas de

.....
31 ASSIST: *Automated Semantic Assistance for Translators* (EPSRC, nº de ref.: EP/C004574 y EP/C005902, 2005-2007). Para más información, consúltese la URL: <<http://ucrel.lancs.ac.uk/projects/assist/>>.

32 Últimamente se están propugnando modelos que amplían los anteriores con valores estadísticos de frecuencia de las unidades extraídos de su uso en Internet y se habla incluso de la lingüística de corpus basada en la red (cf. Colson 2010a, 2010b).

Gale y Church (1991) y Brown *et al.* (1991) (ápuđ Dagan y Church 1994). Actualmente se investiga en la mejora de la calidad en TA mediante el enriquecimiento fraseológico de tales sistemas (Anastasiou 2010; Bouamor, Semmar y Zweigenbaum 2012).

Los estudios sobre fraseología contrastiva también se nutren de corpus para el análisis. Aquí el interés no está en la identificación de unidades y extracción de equivalentes. El grueso de la investigación se centra más bien en torno al análisis de las distintas unidades en los corpus y no se suele recurrir a técnicas y herramientas más sofisticadas de PLN. Por lo general, se localizan fragmentos ilustrativos y se analizan líneas de concordancia para aquilatar con datos reales la descripción lingüístico-fraseológica de las unidades objeto de estudio. Investigaciones de este calibre son las que se llevan actualmente a cabo en el contexto nacional. Claros ejemplos son el grupo FRASEONET³³, liderado por M.^a Isabel González Rey, que trabaja en fraseodidáctica y fraseología contrastiva (español-francés) con metodología de corpus; FRASESPAL³⁴, dirigido por Carmen Mellado Blanco, que aplica corpus bilingües para el estudio contrastivo de la fraseología alemán-español y para la lexicografía bilingüe; y FRASYTRAM³⁵, con Pedro Mogorrón Huerta a la cabeza, que se ocupa del ámbito de la fraseología contrastiva y la fraseografía multilingüe.

Las investigaciones que se realizan en el seno de los equipos mencionados utilizan corpus de distintas características, compilados por ellos mismos, siguiendo criterios de diseño predeterminados o bien de forma *ad hoc*. Evidentemente, del diseño del material de análisis (o sea, el corpus) dependerán los resultados que se obtengan. La compilación suele ser de tipo manual, lo cual puede resultar una tarea muy tediosa, además de costosa, dado que para estudiar la fraseología se necesitan corpus de grandes dimensiones para todas las lenguas implicadas. Según Colson (2012: 1), esto complica el panorama hasta tal punto que el estudio de la fraseología a través de la lingüística de corpus podría llegar a parecer “misión imposible”:

Et pourtant, vouloir recourir aux corpus en phraséologie contrastive peut paraître une mission impossible. En effect, il est établi que des corpus gigantesques sont nécessaires afin de couvrir l'ensemble du spectre

.....
33 <<http://www.phraseonet.com/>>

34 <<http://www.usc.es/frasespal/>>

35 <<http://dti.ua.es/es/frasytram/grupo-de-investigacion-frasytram.html/>>

de la phraséologie, depuis les structures les moins idiomatiques (routines, collocations) jusqu'aux idioms ou phrasèmes ... Dans une perspective contrastive, cette exigence se retrouve multipliée par deux: le défi sera de constituer aux moins deux corpus comparables et de taille très importante³⁶.

Por ello, cada día más se recurre también a ejemplos sacados de Internet o bien se hace uso de herramientas que gestionan la red como un gigantesco corpus (cf. Corpas Pastor 2006, 2007a; Torrent-Lenzen 2009; Mellado Blanco 2012 y en este volumen; y gran parte de los trabajos recogidos en Ptashnyk, Hallsteindóttir y Bubenhofer 2010).

En esta sección vamos a mostrar un modelo de análisis fraseológico contrastivo con metodología de corpus. Para ello hemos seleccionado la unidad *Every cloud has a silver lining*. Tiene su origen en un poema de John Milton, titulado *Comus: A Mask Presented at Ludlow Castle* (1634)³⁷, pero la unidad fraseológica en sí se remonta a finales del siglo XIX, bajo la forma de dos variantes: *There's a silver lining to every cloud* y *Every cloud has a silver lining*³⁸. De acuerdo con el LDOEI, se trata de una paremia (*saying*)³⁹, cuya forma canónica es *Every cloud has a silver lining*⁴⁰, y la define de la siguiente forma "something good will come out from a difficult or unpleasant moment, situation, etc." Es decir, la interpretación estándar de la unidad incide en lo positivo que se esconde tras cada situación u acontecimiento, por muy negativos o tristes que estos pudieran parecer en un principio.

.....
36 "Y, sin embargo, querer utilizar corpus en fraseología contrastiva puede parecer una tarea imposible. En efecto, se ha constatado la necesidad de contar con corpus de enormes dimensiones para poder cubrir todo el espectro fraseológico, desde las unidades menos idiomáticas (fórmulas rutinarias, colocaciones) hasta las propiamente idiomáticas o frasesas... Desde una perspectiva contrastiva, esta exigencia se ve multiplicada por dos: el desafío consistirá entonces en compilar, por lo menos, un corpus comparable de gran tamaño para ambas lenguas" (la traducción es nuestra).

37 Como, *Máscara: Traducida Del Inglés de Milton en Verso Castellano* (1862). Kessinger Publishing, 2010. (Traducción de H. L. De Vedia).

38 Martin (s/f).

39 En el LDOEI (1979) el término *saying* se utiliza de forma genérica para referirse a todo tipo de paremias, desde las menos institucionalizadas (*You can't take it with you, There is always a next time*) hasta las más enraizadas y semánticamente opacas (*A bird in the hand is worth two in the bush, All work and no play makes Jack a dull boy*).

40 Otros diccionarios fraseológicos (EI, ODOEI y NTC) recogen también *Every cloud has a silver lining* como forma canónica de la unidad.

Para estudiar el comportamiento pragmático-discursivo de la unidad vamos a utilizar el corpus BYC-BNC y la red Internet, gestionada por WebCorp⁴¹. Lo primero que llama la atención es la variación formal de la unidad en el discurso. Si bien suele aparecer en su forma canónica frecuentemente, también se detectan numerosas modificaciones y variaciones creativas.

<p>Every cloud has a silver lining The clouds had a silver lining Every silver lining having a cloud a silver lining to a cloud a silver lining on the cloud find a silver lining in the blackest cloud have a silver lining a silver lining the silver lining silver lining</p> <p>(BYU-BNC)</p>	<p>Every cloud has a silver lining Not every cloud has a silver lining Behind every cloud there is a silver lining Behind every silver lining there is a dark cloud For every cloud, there is a silver lining There is a silver lining to every cloud There is a dark cloud behind every silver lining Is there a silver lining to the cloud? Every retailer has a silver lining the silver lining to the cloud the silver lining shining through the dark clouds dark cloud, silver lining dark cloud's silver lining silver lining in the cloud find a silver lining in the blackest cloud have a silver lining a silver lining the silver lining silver lining</p> <p>(WebCorp)</p>
--	--

Esta primera aproximación formal nos permite distinguir los nodos fraseológicos, que representamos aquí a modo de secuencias de búsqueda booleanas:

- a) [every .and. silver lining] y [cloud .and. silver lining], que darán como resultados instancias de la *paremia completa* (23), así como de sus modificaciones creativas por *recategorización*, *reducción*, *adición*, *sustitución* o *reordenamiento de componentes* (24); y

.....
⁴¹ Se trata de un programa gratuito de gestión de corpus *web*, desarrollado por el RDUES (Birmingham), que se encuentra disponible en la URL: <<http://www.webcorp.org.uk/>>. [Consulta: 10 diciembre 2012]. Para este estudio hemos utilizado la opción *.uk* de búsquedas avanzadas para restringir los resultados diatópicamente. A modo de ilustración, a lo largo de este apartado iremos incluyendo líneas de concordancia alineadas con WebCorp.

b) [silver lining .not. every], [silver lining .not. cloud], que solo arrojarán resultados para *silver lining*, lo cual permite analizar en detalle el núcleo fraseológico residual de la unidad (25).

- (23) *Every cloud has a silver lining*. When former Tincknells Fuels driver, Tony Norris, suffered a stroke in September 2010, rather than letting it get him down, he took the opportunity to change his career. [BYU-BNC]
- (24) Weather: *Could every cloud have a sinister lining?* So, farewell 2012 and, meteorologically speaking, good riddance, says David Randall for whom our weather is beyond a joke. [WebCorp]
- (25) *Silver linings* and bankers whining. Recent events have been as nectar for the political blogosphere ... Whilst I'm not sure about White's characterisation of the fundraising scandal as a "moving traffic offence", he makes *a pertinent every-cloud-has-a-silver-lining observation* about Hain's resignation. [BYU-BNC]
- (26) It's *a silver lining* the Chancellor desperately needs amidst the current spate of doom and gloom, most recently this weekend C B I report showing industry in recession. [WebCorp]

A continuación, comprobamos el estatus de las secuencias que hemos detectado en la base de datos StringNet 3.0⁴², a través de StringNet navigator, su interfaz de usuario. StringNet es una base de datos de conocimiento que alberga patrones combinatorios léxico-sintácticos representados por n-gramas híbridos, del tipo *be high time [pers pn] [vb past]*, constituidos por variantes morfológicas de una palabra (*time*, pero no *times*); lemas (*to be* y sus formas conjugadas: *is, are, was, been*, etc.); y/o categorías gramaticales (etiquetas POS como [pers pn] y [vb past]).

.....
42 <<http://nav.stringnet.org/>>. Para más información sobre StringNet, véase Wible y Tsao (2010).

The screenshot shows the StringNet search interface with the search term 'silver lining' entered. Below the search bar is a table with the following columns: No. Chunks, Frequency, Examples (with a magnifying glass icon), Parent (with an upward arrow), Child (with a downward arrow), Expand (with a double-headed arrow), and Contract (with a double-headed arrow). The table lists 8 chunks, with the most frequent being 'silver lining' (37 occurrences).

No. Chunks	Frequency	Examples <small>(from BNC)</small>	Parent <small>(more general versions)</small>	Child <small>(more specific versions)</small>	Expand <small>(longer versions)</small>	Contract <small>(shorter versions)</small>
1	8	every cloud have a silver lining	↑	↓	↔	↔
2	12	have a silver lining	↑	↓	↔	↔
3	27	[art] silver lining	↑	↓	↔	↔
4	37	silver lining	↑	↓	↔	↔
5	8	a silver lining	↑	↓	↔	↔
6	17	a silver lining	↑	↓	↔	↔
7	10	silver lining [prep]	↑	↓	↔	↔
8	9	the silver lining	↑	↓	↔	↔

Ilustración 1: Resultados para silver lining en StringNet.

En la ilustración 1 mostramos una captura de pantalla de la búsqueda realizada para el núcleo fraseológico residual [silver lining]. Los n-gramas (*Chunks*) incorporan datos de frecuencia en el BNC y permiten el acceso a los ejemplos correspondientes (*Examples*) en el BNC con solo clicar sobre el icono en forma de lupa. A su vez, cada n-grama está vinculado a otros que estén situados en el nivel superior (*Parent*) o inferior (*Child*) en la jerarquía. Por ejemplo, [art] silver lining es el nivel superior (*Parent*) de the silver lining; y viceversa, the silver lining es el nivel inferior (*Child*) de [art] silver lining. Finalmente, las etiquetas POS llevan directamente a las palabras que pueden ocupar dichas casillas, ordenadas alfabéticamente o por frecuencia: en el caso de silver lining [prep], el listado de preposiciones de ese patrón combinatorio son, por orden de frecuencia, to (4), on (2), seguidas de for, upon, in, with (1).

Los resultados obtenidos sugieren la coexistencia de dos unidades fraseológicas sobre la misma base: la pemia *Every cloud has a silver lining* (6-grama) y su forma acortada, *silver lining*, ya institucionalizada como locución nominal, que se presenta en forma de cuatrigrama (*have a silver lining*), trigramas ([art] silver lining, a silver lining, silver lining [pre], the silver lining) y bigrama (*silver lining*). Esos serán nuestros nodos de búsqueda en las líneas de concordancia (KWIC) que hemos analizado y cuyos resultados pasamos a resumir a continuación. No hacemos distinción entre las concordancias extraídas del BYU-BNC y de WebCorp porque los patrones combinatorios que

se observan son similares, con la salvedad de que mediante WebCorp se consiguen muchos más ejemplos para ambos nodos.

Para el análisis se han alineado las líneas de concordancia a la izquierda y a la derecha de ambos nodos. En las posiciones 1-3 a la izquierda, el nodo 1 [Every cloud has a silver lining] se suele colocar con presentadores, conectores y marcadores discursivos. Los presentadores son elementos deícticos contextuales que introducen la unidad y llaman la atención, a modo de balizas, sobre la presencia de esta en el discurso. En este caso, se trata de estructuras verbales moduladoras que hacen referencia a la supuesta verdad, consabida y sancionada, inherente a las paremias, que contienen verbos que expresan ‘ser’, ‘parecer’, ‘recordar’, ‘creer’, ‘conocer’, etc.: [So] *It seems [that], It may be that, Don't forget that, have you heard the phrase...?, everyone has heard the phrase...*; sustantivos y adjetivos que denotan verdad o creencias (*true, truth, belief, etc.*) o denominaciones metalingüísticas, del tipo *saying, proverb, phrase, cliché*.

focuses on avoiding problems. But	<u>every cloud has a silver lining</u>	, and risks often have potential
But, as the cliché goes:	<u>every cloud has a silver lining</u>	. Children loved having a day
analogue technology as obsolete. However,	<u>every cloud has a silver lining</u>	. CCTV is essentially a modular
crumbling market Motoring News. If	<u>every cloud has a silver lining</u>	then car companies such as
day in lost production. Now,	<u>every cloud has a silver lining</u>	and I guess for the
plans! But as they say;	<u>every cloud has a silver lining</u>	. So rather than panic we
from less resources.Of course	<u>every cloud has a silver lining</u>	. A significant amount of training
due to wet weather. Still,	<u>every cloud has a silver lining</u>	, as this gave us more
things? The truth is that	<u>every cloud has a silver lining</u>	. So first-time house buyers are
years of disenchantment, and yet	<u>every cloud has a silver lining</u>	, and we can take strength

Ilustración 2: *Nodo 1 (alineación 1-I)*.

Este sería un nivel 1 de “señalización” de la presencia de la unidad en el discurso. El segundo nivel de “señalización” corresponde a la función estructuradora y argumentativa de la unidad en el discurso, y que suele ser la de ratificar con carácter anafórico lo ya expresado, introducir catafóricamente un tema u opinión (especialmente en los títulos) o bien indicar contraste u oposición, para lo cual se coloca con adverbios y estructuras adverbiales de carácter negativo, adversativo o enfático (*not, but, however, still, now, so, of course*), o bien entra a formar parte de oraciones condicionales con *If... then*. La fuerza ilocutiva de la unidad sería argumentar, convencer, ratificar, o bien animar y consolar, y vendría condicionada por su patrón combinatorio⁴³. Las valoraciones pragmáticas serían negativas, pero no referidas a los interlocutores en sí, sino al contexto situacional de la comunicación o al tema sobre el cual se habla.

Alineado a la derecha, el nodo 1 sigue presentando patrones combinatorios parecidos, así como doble “señalización”. Así, en las posiciones D1-3, aparecen algunos presentadores (*[as] they say, as the [old] saying goes*, etc.), correlaciones condicionales con *If...then*, conectores y marcadores de contraste, adición, y consecuencia como *and* (especialmente frecuente en la posición D-1), *now, although, but, however, still, now, so*, etc., pero en mucho menor número que a la izquierda del nodo. Abundan, en cambio, los signos de puntuación (punto, coma, dos puntos, guión, signo de interrogación, signo de exclamación) que marcan claramente los límites de la unidad. Cuando no se dan tales límites, la unidad aparece modificada, generalmente por recategorización (con guión), lo que le permite asumir funciones sintácticas, frecuentemente en premodificación nominal (cf. el ejemplo 25 y la octava línea de concordancia en la ilustración 3).

.....
43 La noción de nodo y su patrón combinatorio es asimilable a la unidad semántica ampliada (*extended unit of meaning*) que propugna Sinclair (1996), esto es, la unidad multiverbal, flexible y variable, que comprende el nodo junto con sus colocaciones, coligaciones, preferencias semánticas y prosodia semántica.

from less resources.Of course	<u>every cloud has a silver lining</u>	. A significant amount of training
one's hands. As always, though,	<u>every cloud has a silver lining</u>	and I particularly enjoyed the
police arrived. Ho hum... Anyway,	<u>every cloud has a silver lining</u>	, and so the restaurant we
Dorset and Bedrock. However, ‘	<u>every cloud has a silver lining</u>	’, as the old saying goes,
Everyone has heard the saying	<u>every cloud has a silver lining</u>	but it's hard to believe
of travel restrictions. Mind you,	<u>every cloud has a silver lining</u>	, eh? Shares in major pharmaceutical
Well, on the	<u>every cloud has a silver lining</u>	side, at least it looks as if...
of the world. They say	<u>every cloud has a silver lining</u>	. So why waste your days
crumbling market Motoring News IF	<u>every cloud has a silver lining</u>	then car companies such as
It is often said that	<u>every cloud has a silver lining</u>	. Well this is exactly what

Ilustración 3: *Nodo 1 (alineación 1-D).*

El nodo 2 [silver lining] presenta una combinatoria más compleja. Por un lado, comparte la combinatoria del nodo 1 cuando forma parte de su inventario de elementos constitutivos. Pero, en el resto de casos, presenta patrones combinatorios propios y diferenciados. Así, cuando se alinean las líneas de concordancia a la izquierda (ilustración 4), en las posiciones L-1 y L-2 se obtienen determinantes (*a, the, one*), posesivos (*her, its, my, your*), adjetivos con valor negativo (*unpleasant, black*), verbos de estado, apariencia, entendimiento y visión (*be, find, see, be blind*) y el verbo *have*, con el cual forma la colocación estable: *to have a silver lining*. A partir de la L-3 nos encontramos con sustantivos, adjetivos y verbos que denotan deseo o esperanza (*desire, hope [for], hopeful, wish*).

could this cloud have a	<u>silver lining</u>	? Calvin posted this on Apr
since 2003. But there's a	<u>silver lining</u>	to all this. First-quarter earnings
Black Friday will have a	<u>silver lining</u>	. Preliminary Black Friday Survey Suggests
money for research is cloud's	<u>silver lining</u>	Salesforce.com: Cloud Computing is
appreciate the desire to find	<u>silver lining</u>	in bad situations, but the
prices may provide a small	<u>silver lining</u>	for the suffering retail industry,
other sources of income. The	<u>silver lining</u>	was the motivation it provided
I am hopeful that the	<u>silver lining</u>	of this economic trouble will
the dark clouds. And the	<u>silver lining</u>	is this - the current political
price being paid for the “	<u>silver lining</u>	” by the millions of Americans

Ilustración 4: *Nodo2 (alineación 1-I).*

De la misma manera, en la posición D-1, el nodo 2 no aparece anclado y delimitado en la oración mediante signos de puntuación, sino que se conecta con complementos formados por sintagmas prepositivos introducidos por *for*, *to*, *in*, *upon*, etc. Merece la pena señalar que tanto a la izquierda como a la derecha el nodo 2 presenta un comportamiento fraseológico-discursivo idiosincrásico, pues se coloca frecuentemente con adjetivos de valoración negativa (*difficult*, *black*, *suffering*, *bad*), y sustantivos (y algún adjetivo también) del ámbito de las finanzas, con especial referencia a la crisis económica de los últimos años: *economic (crisis)*, *recession*, *personal finance*, *source of income*, *prices*, *Black Friday*, *crisis*, *recession*, *retail industry*, *taxpayers*, *gold*, etc.

media seem blind to the	<u>silver-lining</u>	benefits from the current recession. ``
we're asking you: Where's the	<u>silver lining</u>	of this recession? There's no
a new strategic path, this	<u>silver lining</u>	could be gold... This entry
down relationship-all clouds have a	<u>silver lining</u>	, don't they say
A-Z Comment is free A	<u>silver lining</u>	for Japan The economic suffering
to try to find the “	<u>silver lining</u>	” in bad situations in general
14, 2009 - 16:16. A wonderful	<u>silver lining</u>	is the money going towards
that interests you, that has a story line with “	<u>silver lining</u>	” moments or endings.
I believe (and hope) a	<u>silver lining</u>	to the recession will be
with us and discover the	<u>silver lining</u>	to your training courses
Carmel sees	<u>silver lining</u>	upon season's end

Ilustración 5: *Nodo 2 (alineación 1-D).*

En estos contextos, el significado de la unidad *silver lining* se ha especializado para pasar a denotar indicios de mejora de la economía o posible salida de la crisis, y al mismo tiempo, se ha generalizado para denotar la parte positiva o lo bueno de algo (en contraposición a *cloud*). El sustantivo *cloud* (y también *clouds*, en plural) suele aparecer tanto a la derecha como a la izquierda del nodo 2, no como parte de la forma canónica de la unidad (nodo 1), sino como elemento evocador de la unidad completa que interviene en las variantes y modificaciones creativas de la unidad fraseológica en el discurso.

En este sentido, la carga semántico-fraseológica del sustantivo *cloud* es deudora no solo de la paremia de la que procede, sino también del propio origen de esta a partir del poema de John Milton: “Was I deceived, or did a sable cloud/ Turn forth her silver lining on the night?”. En la metáfora que encierran esos versos, el sintagma nominal *sable cloud* representa la adversidad, mientras que *silver lining* representa justo lo contrario. La creatividad fraseológica de ambos componentes viene de lejos, pues. Y, en el caso de *cloud*, está incluso recogida de alguna manera en las propias definiciones lexicográficas, como acepción figurativa de valor negativo, lo cual denota cierta institucionalización. Por ejemplo, la acepción segunda de *cloud* en el OED es figurativa y coincide también en el aspecto negativo: “used to refer to a state or cause of gloom, suspicion, trouble, or worry [...] a frowning or depressed look”⁴⁴. De hecho, en la lengua inglesa existen otras unidades fraseológicas con *cloud* que comparten ese mismo sentido negativo: *a cloud in the horizon* (‘causa de preocupación’), *to cast a cloud over/on sth.* (‘disminuir o hacer que disminuya la esperanza o la ilusión de que algo ocurra’) y *under a cloud* (‘bajo sospecha’, ‘desacreditado’, ‘en circunstancias poco claras’).

Por último, que la fuerza evocadora de la paremia sigue latente se evidencia también por el uso de comillas y algunos presentadores (*don't they say, phrase*), que siguen apareciendo a modo de balizas en el entorno léxico-distribucional de la unidad acortada (nodo 2) o bien introduciendo lo que, en realidad, son variantes del nodo 1. Digamos que la fuerza ilocutiva del nodo 1 se pierde en el nodo 2, aunque sigue presente en un segundo plano que se puede llegar a actualizar según los casos, activado por determinados patrones combinatorios. Sigue conservando, en cambio, la valoración negativa del tema o el contexto situacional.

.....
44 “Se usa para referirse a un estado o causa de desánimo, sospecha, problema o preocupación [...] un aspecto o mirada que refleja malestar o depresión” (la traducción es nuestra).

En español contamos con una paremia equivalente: *No hay mal que por bien no venga*. Eso es, al menos, lo que figura en diccionarios bilingües como el OSD y el CSD, sin más información al respecto. De hecho, la interpretación estándar es parecida, pues se incide en lo positivo que traerá una situación negativa, desde la paciencia, el ánimo y la resignación: “Da a entender que un suceso infeliz suele ser inopinadamente ocasión de otro venturoso, o que sobrellevado con resignación cristiana los males traen bienes seguros para el hombre.” (DR 1995).

A continuación, vamos a analizar la paremia española (nodo 3) también con metodología de corpus, lo cual nos permitirá determinar el nivel de equivalencia real de las unidades en ambas lenguas. Nos valdremos del CREA y de WebCorp, en los cuales realizamos búsquedas restringidas diatópicamente al español peninsular. Los resultados obtenidos a partir de ambos son similares, con la salvedad de que WebCorp recupera un número muy superior de ejemplos de uso. Además, es mucho más fácil extraer líneas de concordancias que incluyan la paremia española mediante WebCorp que a través del CREA. El principal problema con el CREA consiste en que no se pueden realizar búsquedas de la unidad completa, por error de sintaxis de la BD que interroga el corpus: “Error en la consulta especificada: No aparece el operador Y delante del operador NO en el campo consulta. Probablemente se escribió la palabra *no* sin entrecomillar”. Tampoco se pueden analizar los componentes aislados (*mal*, *bien*, *venga*, etc.) por ser palabras muy frecuentes, por lo que, de nuevo, el sistema de gestión se bloquea y se obtienen los siguientes mensajes de error: “No se pueden ver estadísticas. Demasiados documentos”; “No se puede filtrar. Demasiados aciertos y documentos” y “No se puede Recuperar. Demasiados documentos.” Los mismos mensajes de error se producen cuando se intenta recuperar la información para los n-gramas [no hay mal], [hay mal], [no venga]. Solo se pueden recuperar concordancias para [que por bien] y [por bien], posiblemente porque son menos frecuentes en la lengua y presentan un índice estadístico de información mutua superior a los otros.

La paremia española se muestra mucho más monolítica y fija que su correspondiente en inglés. Por eso su patrón combinatorio resulta bastante heterogéneo, con algunas referencias a la verdad consabida que sanciona la paremia y a su fuerza ilocutiva, en combinación con verbos de lengua (*decir*, *afirmar*, *sancionar*, *sentenciar*, etc.) y otras unidades fraseológicas (*No hay mal que cien años dure*, *tirar la toalla*, *verlo todo negro*, *tener razón de ser*, etc). A derecha e izquierda coaparece con sustantivos y adjetivos que denotan verdad, ánimo, adversidad (*cierto*, *verdad*, *embite*, *situación mala*, *ánimo*, *adversidad*, *resignación*), y con sustantivos relativos

al mundo empresarial y a la política (*empresario, crisis, Pyme, elecciones, proceso electoral*, etc.), entre muchos otros difíciles de clasificar en categorías.

Más llamativo es el hecho de que, generalmente, la paremia va delimitada a derecha e izquierda por signos de puntuación (punto, comas, dos puntos y signos de interrogación). A veces va entrecomillada, tras marcadores de reformulación (*en este sentido, es decir, ya ves, como el que dice*) o bien suele aparecer junto a presentadores de introducción (*el refrán que reza, sentencias, refranero, dicen que, el adagio popular, decían nuestros viejos sabios*, etc.), a la izquierda del nodo, o de delimitación, a la derecha (*dice el refranero, que acuñó alguien antes*), aunque estos suelen ser menos frecuentes. Discursivamente cumple funciones también similares persuasivas y estructuradoras del discurso, con preferencia por la catáfora (muy frecuente como título de artículos periodísticos o libros, o como oración que introduce un tema) frente a la anáfora.

política del famoso refrán que reza “	<u>No hay mal que por bien no venga</u>	” . El Gobierno, tras las elecciones
rido en su mensaje de fin de año: “	<u>No hay mal que por bien no venga</u>	” La autoría material de ETA
Logroño. (De nuestro corresponsal.)	<u>No hay mal que por bien no venga</u>	. El pueblecito riojano de Bergasa
no, mi reina, mira lo que te digo,	<u>no hay mal que por bien no venga</u>	. De esta forma sabré más o menos
“Van a misa todos los Dominguez”, “	<u>No hay GAL que por bien no venga</u>	” “Reinsértame mucho”, etcétera. En
. la certeza del dicho popular:	<u>no hay mal que por bien no venga</u>	. Pongamos como ejemplo el gigante
riesgos infinitos. Es decir, que	<u>no hay mal que por bien no venga</u>	. A pesar de que el disgusto del
veces vuestra augusta madre, de que	<u>no hay mal que por bien no venga</u>	. JUANA.- Ése es el destino de los
las adversidades viene la suerte.	<u>No hay mal que por bien no venga</u>	. No hay mal que cien años dure.
que todo lo ven negro. “	<u>No hay mal que por bien no venga</u>	”, decían nuestros viejos sabios.

Ilustración 6: *Nodo 3 (I-I)*.

La fuerza ilocutiva de la unidad sería de tipo persuasivo o argumentativo, a veces con la finalidad de insuflar ánimo al interlocutor desde la desde la

paciencia y la resignación. Las valoraciones pragmáticas serían negativas, pero no referidas a los interlocutores en sí, sino al contexto situacional de la comunicación o al tema sobre el cual se habla, más bien como una implicatura asociada a la interpretación estándar de la unidad.

A diferencia de la proverbia inglesa, la española no tiene tendencia a presentar modificaciones internas que afecten a la forma o al inventario de sus componentes, con la excepción de la sustitución de algún componente o la ampliación no invasiva para ajustar de forma explícita la interpretación estándar al contexto del discurso y crear efectos estilísticos o humorísticos (cf. la octava línea de concordancia de la ilustración 7). La proverbia española presenta una evidente autonomía textual, incluso cuando se inserta en el discurso, pues no va más allá de aparecer como cita literal o, como mucho, de funcionar como cláusula completiva de objeto directo o preposicional. No es de extrañar, pues, que las modificaciones sean entonces externas, sin cambios formales, y que solo afecten al plano semántico, por activación del significado literal de alguno de sus componentes ante la presencia de un determinado elemento del contexto (cf. la primera línea de concordancia de la ilustración 7).

Kahn escribió:	<u>No hay mal que por bien no venga</u>	, así pueden venir en patera
nombre de Marbella en boca de todos.	<u>No hay mal que por bien no venga</u>	, dice el refranero. Y es cierto
Diciembre 208 a las 8:18.	<u>No hay mal que por bien no venga</u>	Gisela aguantó el embate de
Cultura Gente y Estilo Motor	<u>no hay mal que por bien no venga</u>	, la Pyme podría beneficiarse de
Tiene una razón de ser.	<u>no hay mal que por bien no venga</u>	. Mucho ánimo, el 04 Feb
permite las herejías para espavilarnos. “	<u>No hay mal que por bien no venga</u>	”. “No te dejes vencer por
deberá venir por otra parte. “	<u>no hay mal que por bien no venga</u>	” sentencia. Subir ACTUALIDAD
Que están en tu mente.	<u>no hay mal que por bien no venga</u>	siempre al corazón, Bye, bye
a no tirar la toalla.	<u>No hay mal que por bien no venga</u>	. Todo se torció tras los
crónica económica. Joan F. Domene.	<u>No hay mal que por bien no venga</u>	.14.16.2005. La crisis de la

Ilustración 7: *Nodo 3 (1-D)*.

Los resultados del análisis realizado confirman que la unidad fraseológica inglesa constituye, en realidad, dos unidades emparentadas: una paremia (*Everycloud has a silver lining*) y una locución (*silver lining*). Ambas coexisten en el discurso, pero, aunque tienen un innegable aire de familia, se comportan de forma diferenciada. De hecho, la unidad *silver lining* es mucho más productiva y ágil, más propensa a formar parte de variantes de la paremia y a entrar en todo tipo de modificaciones creativas. *Every cloud has a silver lining*, en cambio, presenta una actitud más monolítica y textual, con una carga ilocutiva importante, que se mantiene latente también en la locución derivada de ella, con la cual comparte también las valoraciones negativas del entorno y la temática que se está tratando. La unidad acortada, a su vez, presenta especialización semántica de dominio en el ámbito financiero, y, al mismo tiempo, muestra una generalización semántica parecida al fenómeno de la banalización terminológica, que guarda relación con la interpretación estándar de la paremia (lo positivo en clave optimista). La unidad española es mucho más fija que la paremia británica, más inflexible textualmente hablando, con una menor capacidad para los usos creativos y con menor productividad. Sus patrones combinatorios tampoco están tan determinados como en el caso de las unidades inglesas, a pesar de que se detectan ciertas similitudes en cuanto a la “señalización” léxica y ortotipográfica de las unidades, sus funciones discursivas y aspectos pragmáticos.

Finalmente, la equivalencia formal y funcional entre la paremia española y la locución británica es mucho menor que en el caso de ambas paremias, que al menos presentan equivalencia formal. Las disfunciones se reflejan claramente en las traducciones de los fragmentos que contienen unidades de este tipo. En la ilustración 8 mostramos algunos ejemplos organizados a doble columna, que hemos extraído con Linguee⁴⁵, esto es, un sistema CLIR (*cross-language information retrieval*), que gestiona la red Internet como si fuese un enorme corpus paralelo multilingüe y que presenta los segmentos traducidos alineados a la manera de las memorias de traducción. La multiplicidad de equivalentes y de procedimientos de traducción empleados da idea de las diferencias que subyacen cuando se contrastan unidades fraseológicas, cuando se profundiza en su funcionamiento real en el discurso y se atiende a las peculiaridades semánticas y pragmáticas de este tipo de unidades.

.....
45 <<http://www.linguee.es/espanol-ingles/>>. [Consulta: 23 diciembre 2012].

On the other hand, we have seen recently that <i>there is a silver lining</i> .	Por otro lado, recientemente hemos visto que <i>existe un pequeño resquicio</i> .
At this juncture the information about this course seen at a poster advertisement was <i>really a silver lining in the cloud of my life</i> .	Sólo cuando supe a través de un afiche de las posibilidades de una beca en la Fundación Samuel <i>volvió mi esperanza</i> .
As the saying goes, <i>every cloud has a silver lining</i> .	Como suele decirse, ' <i>no hay mal que por bien no venga</i> '.
But as that great Englishman, Ken Dodd, once remarked, <i>every cloud has a silver lining</i> , plus VAT.	Pero, como afirmó una vez Ken Dodd, esa gran figura inglesa, <i>tras la tormenta llega la calma</i> , y el IVA.
In any case, though, <i>every cloud has a silver lining</i> .	No obstante, sucede que <i>toda parte mala tiene también su parte buena</i> .
There is a Polish saying along the lines of <i>every cloud has a silver lining</i> but this cloud is actually very black indeed.	Hay un dicho polaco similar al de ' <i>después de la tempestad, siempre llega la calma</i> ' aunque la verdad es que esta tempestad es muy fuerte.
The <i>silver lining</i> is that bedbugs don't transmit disease and don't pose any huge health problem, outside of stress and anxiety	El <i>consuelo</i> es que las chinches no transmiten enfermedades y no suponen un problema de salud de gran importancia, aparte de estrés y ansiedad
At the global level, I believe that, beyond its adverse impacts, the current crisis <i>has a silver lining</i> insofar as it stimulates creativity, initiative and a fighting spirit.	A nivel mundial, creemos que, más allá de sus consecuencias negativas, la actual crisis <i>tiene su lado positivo</i> por cuanto estimula la creatividad, la iniciativa y un espíritu de lucha.
But, rather than just waiting for its local suppliers to recover, it is helping them to do so and believes that, in the process, it can <i>pluck a silver lining</i> from the earthquake.	Sin embargo, en lugar de esperar simplemente que sus proveedores locales se recuperen, los está ayudando a lograrlo y estima que en el proceso, puede sacar <i>algo bueno</i> del terremoto.
The <i>silver lining of this dark cloud</i> is that Mrs Patrizia Toia's report is now on the agenda of the plenary session, since initiatives focusing on solidarity, as well as on social and regional cohesion, are especially important these days.	El <i>resquicio de esperanza</i> es que el informe de Patrizia Toia está ahora en la agenda de la sesión plenaria, ya que las iniciativas centradas tanto en la solidaridad como en la cohesión social y regional son especialmente importantes hoy en día.

Ilustración 8: Fragmentos traducidos que contienen los tres nodos analizados.

4 Corolario

En este trabajo hemos intentado tender un puente entre los distintos enfoques desde los cuales se puede abordar con éxito el estudio de las unidades fraseológicas en el discurso. La fraseología ocupa un lugar destacado en las políticas lingüísticas europeas, pero su tratamiento no pasa de ser una mera declaración de intenciones. No se proporciona información sobre qué tipos

de unidades se debe incluir en cada uno de los niveles que reconoce el MCER, ni con qué profundidad se debe tratar una determinada unidad en cada uno de los niveles, etc. Una manera de solventar el problema es recurrir al PLN, y más concretamente dentro de este, a la lingüística de corpus. No obstante, existe una brecha entre ambas disciplinas. La Fraseología y el PLN no han estado realmente en contacto, a pesar de ocuparse de fenómenos muy similares: unidades fraseológicas y expresiones multi-verbales.

El campo del PLN (incluidas la lingüística de corpus, las tecnologías lingüísticas y las tecnologías de la traducción) abre enormes posibilidades para la investigación de la fraseología en todas sus vertientes. Por el momento se vienen realizando estudios descriptivos monolingües y contrastivos con metodología de corpus, como hemos ilustrado en este capítulo. Incluso se emplea la red Internet como enorme pista de pruebas para la fraseología (monolingüe y multilingüe) y la traducción.

La reutilización de los recursos electrónicos y la aplicación de las tecnologías lingüísticas nos permitirá avanzar a pasos agigantados en la comprensión del fenómeno fraseológico desde todos sus ángulos. Y viceversa, el progreso en la investigación fraseológica ayudará a refinar las actuales técnicas y procedimientos del PLN y la lingüística de corpus, a mejorar el inventario de recursos disponibles (corpus, lexicones, bases de datos, etc.), lo que, a la postre, redundará en la mejora de los actuales sistemas de extracción de terminología, de memorias de traducción, de traducción automática, de generación de lenguaje natural, entre muchos otros. Quizá en un futuro no muy lejano podamos disponer de los corpus necesarios y las herramientas adecuadas para detectar, extraer, interpretar, representar y traducir todo tipo de unidades fraseológicas, así como sus posibles usos creativos y modificaciones (como los que hemos visto en este trabajo), sin que por el camino se pierdan valoraciones, interpretaciones en contexto, funciones discursivas y textuales, fuerza ilocutiva, saber compartido, rasgos pragmáticos, etc. Será entonces cuando contaremos con una teoría fraseológica bien construida, no desde el tejado, sino a partir de sus propios cimientos y con los medios técnicos adecuados.

Referencias bibliográficas

- ALONSO RAMOS, M. (2004): *Las construcciones con verbo de apoyo*, Madrid: Visor.
- ANASTASIOU, D. (2010): *Idiom Treatment Experiments in Machine Translation*, Newcastle upon Tyne: Cambridge Scholars Publishing.
- ARNOLD, D., BALKAN, L., HUMPHREYS, R. L., MEIJER, S. y SADLER, L. (1994): *Machine Translation. An Introductory Guide*, Londres: Blackwell.
- BAKER, P. (2006): *Using Corpora in Discourse Analysis*, Londres y Nueva York: Continuum.
- BALDWIN, T. y KIM, S. N. (2010): "Multiword Expressions", N. INDURKHYA y F. J. DAMERAU (eds.), *Handbook of Natural Language Processing*, 2.ª ed, Boca Raton, EE.UU.: CRC Press, 267-292.
- BOUAMOR, D., SEMMAR, N. y ZWEIGENBAUM, P. (2012): "Identifying bilingual Multi-Word Expressions for Statistical Machine Translation", *LREC 2012, Eighth International Conference on Language Resources and Evaluation*. Estambul, Turquía: ELRA, 674-679. Publicación electrónica en: <http://www.lrec-conf.org/proceedings/lrec2012/pdf/886_Paper.pdf>.
- BOULAKNADEL, S., DAILLE, B. y DRISS, A. (2008): "A Multiterm Extraction Program for Arabic Language", *LREC 2008, Proceedings from the Sixth International Conference. On Language Resources and Evaluation*, París: ELRA, 340-351.
- COLSON, J.-P. (2010a): "Automatic Extraction of Collocations: A new Web-based Method", S. BOLASCO, I. CHIARI y L. GIULLIANO (eds.), *Proceedings of IADT 2010. Statistical Analysis of Textual Data. Sapienza University, Rome, 9-11 June 2010*, Milán: LED Edizioni, 397-408.
- COLSON, J.-P. (2010b): "The Contribution of Web-Based Corpus Linguistics to a Global Theory of Phraseology", PTASHNYK, S., HALLSTEINDÓTTIR, E., BUBENHOFER, N. (eds.), *Corpora, Web and Databases. Computer-Based Methods in Modern Phraseology and Lexicography*, Hohengehren: Schneider Verlag, 23-35.
- COLSON, J.-P. (2012): "Phraséologie contrastive et linguistique de corpus: outils récents et perspectives", A. PAMIES, J.M. PAZOS BRETANA y L. LUQUE RAMOS (eds.), *Phraseology and Discourse: Cross Linguistic and Corpus-based Approaches* [Phraseologie und Parömiologie, 29], Baltmannsweiler: Schneider, 1-9.
- CEFR = CONSEJO DE EUROPA (2001): *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Estrasburgo: Council of Europe. Publicación electrónica en: <http://www.coe.int/t/dg4/linguistic/Source/Framework_en.pdf>.
- CORPAS PASTOR, G. (1996): *Manual de fraseología española*, Madrid: Gredos.
- CORPAS PASTOR, G. (2006): "La Xarxa com a corpus: un estudi pilot de dues unitats fraseològiques amb lügen en alemany", V. SALVADOR y L. CLIMENT (eds.), *El discurs prefabricat ii: fraseologia i comunicació social* (Col·lecció Estudis Filològics), Castellón de la Plana: Servicio de Publicaciones de la Universidad de Jaume I, 213-242.
- CORPAS PASTOR, G. (2007a): "Europäismen – von Natur aus phraseologische Äquivalente? Von blauem Blut und sangre azul", M. EMSEL y J. CUARTERO OTAL (eds.),

- Brücken: Übersetzen und interkulturelle Kommunikationen. Festschrift für Gerd Wotjak zum 65. Geburtstag*, Fráncfort: Peter Lang, 65-77.
- CORPAS PASTOR, G. (2007b): "Lost in specialised translation: the corpus as an inexpensive and under-exploited aid for language service providers", *Translating and the Computer 29. Proceedings of the Twenty-ninth International Conference on Translating and the Computer, 29-30 November 2007*, Londres. Aslib/IMI: Londres, 1-21.
- CORPAS PASTOR, G. (2008): *Investigar con corpus en traducción: los retos de un nuevo paradigma* [Studien zur romanische Sprachwissenschaft und interkulturellen Kommunikation, 49], Fráncfort: Peter Lang.
- CREA = REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. <<http://www.rae.es>> [04-01-2013].
- DAGAN, I. y CHURCH, K. (1994): "Termight: Identifying and Translating Technical Terminology", *Proceedings of the 4th Conference on ANLP*, Stuttgart, Germany, 34-40.
- DALE, R., MOISL, H. y SOMERS, H. (eds.) (2000): *Handbook of Natural Language Processing*, Nueva York y Basilea: Marcel Dekker.
- DAVIES, M. (2004-): *BYU-BNC*. [Basado en el British National Corpus de Oxford University Press]. Publicación electrónica en: <<http://corpus.byu.edu/bnc/>>.
- EAGLES (1996): *Preliminary Recommendations on Corpus Typology*. Documento técnico EAGLES EAG-TCWG-CTYP/P. <<http://www.ilc.cnr.it/EAGLES/pub/eagles/corpora/corpusyp.ps.gz>>.
- EVERT, S. (2005): *The Statistics of Word Cooccurrences – Word Pairs and Collocations*. Stuttgart: Universidad de Stuttgart. Publicación electrónica en: <<http://collocations.de/phd.html>>.
- HEID, U. (2007): "Computational Linguistic Aspects of Phraseology II." En H. BURGER, D. DOBROVOL'SKIJ, P. KÜHN y N. R. NORRICK (eds.) 2007. *Phraseologie – Phraseology. Ein internationales Handbuch der zeitgenössischen Forschung – An International Handbook of Contemporary Research*, vol. II, Berlín: Walter de Gruyter, 1036-1044.
- HEID, U. (2008): "Computational Phraseology: an Overview", S. GRANGER y F. MEUNIER (eds.) *Phraseology. An Interdisciplinary Perspective*, Ámsterdam y Filadelfia: John Benjamins, 337-360.
- HOWARTH, P. (1996): *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*, Tübinga: Niemeyer.
- INDURKHYA, N. y DAMERAU, F. (eds) (2010): *Handbook of Natural Language Processing*, 2.ª ed., Boca Ratón: CRC Press.
- LAVID, J. (2005): *Lenguaje y nuevas tecnologías. Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*, Madrid: Cátedra.
- MARINAI E., PETERS C. y PICCHI E. (1991): "Bilingual Reference Corpora: A System for Parallel Text Retrieval in Using Corpora", *Proceedings of 7th Annual Conference of the UW Centre for the New OED and Text Research*, Oxford: Oxford University Press, 63-70.
- MARTIN, G. (s/f): "Every cloud has a silver lining". *The Meanings and Origins of Sayings and Phrases*. Publicación electrónica en: <<http://www.phrases.org.uk/meanings/every-cloud-has-a-silver-lining.html>>. [Consulta: 20 diciembre 2012].

- MCER = MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE (2002): *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*, Madrid: Secretaría General Técnica del MECD-Subdirección General de Información y Publicaciones y Grupo ANAYA, S. A. [Traducción al español de Instituto Cervantes]. Publicación electrónica en: <<http://cvc.cervantes.es/obref/marco>>.
- MELLADO BLANCO, C. (2012): "Pragmatische Aspekte der Bedeutung von Sprichwörtern aus dem kognitiven Feld SCHWEIGEN am Beispiel des Sprachenpaares Deutsch-Spanisch", K. STEYER (ed.), *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie*. [Studien zur deutschen Sprache. Forschungen des Instituts für Deutsche Sprache 60], Tübinga: Gunter Narr Verlag, 165-204.
- MELLADO BLANCO, C. (en este volumen): "La gramaticalización de las restricciones y preferencias de uso de las unidades fraseológicas del español y alemán desde un enfoque cognitivo-pragmático".
- MITKOV, R. (ed.) (2003): *The Oxford Handbook of Computational Linguistics*, Oxford y Nueva York: Oxford University Press.
- MOON, R. (1998): *Fixed Expressions and Idioms in English*, Oxford: Clarendon Press.
- MUDRAYA, O., PIAO, S., RAYSON, P., SHAROFF, S. BABYCH, B. y LOFBERG, L. (2008): "Automatic Extraction of Translation Equivalents of Phrasal and Light Verbs in English and Russian", S. GRANGER y F. MEUNIER, F. (eds.), *Phraseology: an Interdisciplinary Perspective*, Ámsterdam: John Benjamins, 293-309.
- OKITA, T., GUERRA, M., ALFREDO GRAHAM, Y. y WAY, A. (2010): "Multi-word Expression Sensitive Word Alignment", *Proceedings of the 4th International Workshop on Cross – Lingual Information Access at COLING 2010*, Peckín, 26-34.
- PETERS, C. y PICCHI, E. (1995): "Capturing the Comparable: A System for Querying Comparable Text Corpora", S. BOLASCO, L. LEBART y A. SALEM (eds.), *JADT'95 – 3rd International Conference on Statistical Analysis of Textual Data, CISU*, Rome, 255-262.
- PTASHNYK, S., HALLSTEINDÓTTIR, E. y BUBENHOFER, N. (eds.) (2010): *Corpora, Web and Databases. Computer-Based Methods in Modern Phraseology and Lexicography*, Hohengehren: Schneider Verlag.
- ROTHKEGEL, A. (2007): "Computerlinguistische Aspekte der Phraseme I", H. BURGER, D. DOBROVOL'SKIJ, P. KÜHN y N. R. NORRICK (eds.), *Phraseologie – Phraseology. Ein internationales Handbuch der zeitgenössischen Forschung – An International Handbook of Contemporary Research*, vol. II, Berlín: Walter de Gruyter, 1027-1035.
- SAG, I. A., BALDWIN, T., BOND, F., COPESTAKE, A. y FLICKINGER, D. (2002): "Multiword Expressions: a Pain in the Neck for NLP", *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing, CICLing'02*, Londres: Springer, 1-15.
- SEKINE, S. y RANCHHOD, E. (ed.) (2009): *Named Entities: Recognition, Classification and Use*, Ámsterdam y Filadephia: John Benjamins.
- SERETAN, V. (2011): *Syntax-Based Collocation Extraction*. [Text, Speech and Language Technology series Volume 44], Dordrecht: Springer.

- SHAROFF, S., BABYCH, B. y HARTLEY, A. (2006): "Using Collocations from Comparable Corpora to Find Translation Equivalents", *Proceedings of LREC2006*, Génova, mayo 2006, 465-470.
- SINCLAIR, J. (1991): *Corpus, Concordance, Collocation*, Oxford: Oxford university Press.
- SINCLAIR, J. (1996): "The Search for Units of Meaning", *Textus: English Studies in Italy*, 9, 75-106.
- TORRENT-LENZEN, A. (2009): "Internet: Reto y revolución. Las definiciones de las locuciones en los diccionarios y su verdadero significado fraseológico", S. ROISS, C. FORTEA GIL, A. RECIO ARIZA, B. SANTANA LÓPEZ, P. ZIMMERMANN, I. HOLL (eds.), *En las vertientes de la traducción e interpretación del/al alemán*, Berlín: Frank und Timme.
- TSCHICHOLD, C. (2008): "A Computational Lexicography Approach to Phraseologisms", S. GRANGER y F. MEUNIER (eds.) *Phraseology. An Interdisciplinary Perspective*, Ámsterdam y Filadelfia: John Benjamins, 361-376.
- VINTAR, S. y FISIER, D. (2008): "Harvesting Multi-word Expressions from Parallel Corpora", *LREC 2008, Proceedings from the Sixth International Conference. On Language Resources and Evaluation*, París: ELRA, 387-392.
- WIBLE, D. y TSAO, N.-L. (2010): "StringNet as a Computational Resource for Discovering and Investigating Linguistic Constructions", *Proceedings of the NAACL HLT Workshop on Extracting and Using Constructions in Computational Linguistics*, 25-31. Publicación electrónica en: <<http://www.aclweb.org/anthology-new/W/W10/W10-0804.pdf>>.
- WRY, A. (2008): *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.
- WU, C. y CHANG, S J. (2004): "Bilingual Collocation Extraction based on Syntactic and Statistical Analyses", *Computational Linguistics* 679, 1-20.

Diccionarios

- CSD = AA. VV. (2011): *Collins Spanish Dictionary. 40th Anniversary Edition Complete and Unabridged*, Londres y Glasgow: Harper Collins.
- DPF = BREWER, E. C. (1898): *Dictionary of Phrase and Fable*. Publicación electrónica en: <<http://www.bartleby.com>>.
- DR = CAMPOS, J. C. y BARELLA, A. (2002): *Diccionario de refranes*. 3.^a ed. rev. Madrid: Espasa-Calpe.
- DRAE = REAL ACADEMIA ESPAÑOLA (2001): *Diccionario de la Lengua Española*, 22.^a ed., Madrid: Espasa-Calpe.
- LDOEI = SUMMERS, D. (ed.) (1979): *Longman Dictionary of English Idioms*, Harlow y Londres: Longman.
- NTC = BERTRAM, A. (comp.), SPEARS, R. A. (1993): *NTC's Dictionary of Proverbs and Clichés*, Lincolnwood, Illinois (EE.UU.): National Textbook Company.
- ODOEI = COWIE, A. P., MACKIN, R., MCCAIG, I. R. (1993): *Oxford Dictionary of English Idioms*, Oxford: Oxford University Press.

OSD = AA. VV. (2003): *Oxford Spanish Dictionary*, 3.^a ed., Oxford: Oxford University Press.

EI = SEIDEL, J. y MCMORDIE, W. (1988): *English Idioms*, 5.^a ed., Oxford: Oxford University Press.