

En el *Manual de lingüística del hablar* se presenta el estado actual de la investigación sobre el discurso y la comunicación en el espacio de las lenguas románicas. El lector podrá acceder de una forma sencilla y clara a los principales objetos de estudio de una amplia gama de disciplinas y ramas de la Lingüística, como la Pragmática, el Análisis del Discurso, la Teoría de la comunicación, la Lingüística del texto o la Filología. El *Manual* supone un esfuerzo colectivo para explicar problemas teóricos complejos, concernientes a la comunicación, al discurso y a sus productos, los textos. Las descripciones contrastivas en lenguas románicas sirven para mostrar el alcance de las posiciones teóricas y para enriquecer las explicaciones con ejemplos para un lector con intereses distintos y atraído por lenguas diferentes. Asimismo, contiene análisis sobre formas de aplicación y transferencia de la investigación acerca del hablar en distintas áreas específicas y profesionales.

THE SERIES: MANUALS OF ROMANCE LINGUISTICS

The new series *Manuals of Romance Linguistics* (MRL) aims to present a comprehensive, state-of-the-art overview of Romance linguistics. It will comprise approximately 60 volumes that can either be consulted individually or used as a series of books providing a detailed overall picture of the current state of research in Romance linguistics. A special focus will be placed on the presentation and analysis of the smaller languages, the *linguae minores*.

MRL
28

Óscar Loureda, Angela Schrott (Eds.)
MANUAL DE LINGÜÍSTICA DEL HABLAR

DE GRUYTER

REFERENCE

Óscar Loureda, Angela Schrott (Eds.)

MANUAL DE LINGÜÍSTICA DEL HABLAR

MANUALS OF
ROMANCE LINGUISTICS



9 783110 334883

www.degruyter.com
ISBN 978-3-11-033488-3

DE
G

DE
G

ISBN 978-3-11-033488-3
e-ISBN (PDF) 978-3-11-033522-4
e-ISBN (EPUB) 978-3-11-039366-8

Library of Congress Control Number: 2020940757

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2021 Walter de Gruyter GmbH, Berlin/Boston
Cover image: © Marco2811/fotolia
Typsetting: jürgen ullrich typesatz, Nördlingen
Printing and binding: CPI books GmbH, Leck

www.degruyter.com



Carmen Mellado, Gloria Corpas y Katrin Berty

25 El hablar y el discurso repetido: la fraseología

Resumen: Este capítulo muestra la interrelación entre fijación y variabilidad en las unidades fraseológicas desde distintos puntos de vista. En primer lugar, realizamos un análisis detallado del concepto de «discurso repetido» de Coseriu, que ya considera en su origen la idea de cambio creativo, para después ofrecer una panorámica de la evolución de la fraseología en relación a la lingüística textual. En segundo lugar, se presenta una clasificación de la tipología de la variación fraseológica, ilustrada con ejemplos de corpus lingüísticos y centrada en los niveles del sistema y habla, así como en la intencionalidad del hablante. En tercer lugar, tratamos el tema de la variabilidad fraseológica y el giro que ha tomado la noción de «fijación» desde que se dispone de datos masivos de corpus. En este contexto, las magnitudes de frecuencia absoluta, normalizada y de significación estadística desempeñan un papel fundamental para el grado de fijación.

Palabras clave: unidad fraseológica, desautomatización, corpus, variabilidad fraseológica, procesamiento de lenguaje natural

1 Fraseología y discurso repetido en la lingüística del hablar

1.1 Lingüística de texto integral, técnica libre del discurso y discurso repetido

La «lingüística del sentido» de Coseriu se concibe como una lingüística de texto holística integrada en una «lingüística del hablar», que abarca las distintas disciplinas lingüísticas (cf. Loureda Lamas 2007; 2009) y que ya fue esbozada por Coseriu en los años 1950 (*Determinación y entorno*, 1955; cf. Coseriu 2007, 88). La lingüística del sentido es «tres veces lingüística del texto» (71 El hablar y las lenguas). Los textos son hechos individuales, pero no absolutamente particulares, pues presentan, además, una dimensión universal, que incluye aquellos rasgos de la textualidad, y una dimensión histórica que hace que compartan tradiciones y moldes expresivos» (Loureda Lamas 2009, 73, con referencia a Coseriu/Loureda Lamas 2006). Justamente en esta última dimensión tradicional del nivel del texto se localiza el fenómeno del «discurso repetido», junto a los tipos de texto y las tradiciones discursivas (cf. Loureda Lamas 2009, 83ss.). En el marco de la lingüística de texto de Coseriu, el concepto del discurso repetido se asocia con las relaciones de «sentido» surgidas entre un texto o un seg-

mento textual y otros textos ya existentes (cf. Coseriu 2007, 200–204). Coseriu lo describe por primera vez en su ponencia «Structure lexicale et enseignement du vocabulaire», independientemente de su lingüística de texto (1^{er} Colloque de Linguistique Appliquée, Nancy 1966, publicación: Coseriu 1967, 1973). El concepto del discurso repetido y su par «técnica del discurso» constituyen una de las siete distinciones que son fundamentales para la lingüística de Coseriu.¹ La técnica del discurso (o del hablar) se refiere a «las unidades léxicas y gramaticales (lexemas, categoremas, morfemas) y a las reglas para su modificación y combinación en la oración» que los hablantes tienen a su disposición para construir nuevos enunciados (Coseriu 1981b, 113, cf. también Coseriu 1976, 27). El discurso repetido comprende, en cambio, unidades lingüísticas que la tradición del hablar ha fijado como «expresión», «giro», «modismo», «frase» o «locución», y, por extensión, todas aquellas unidades cuya forma ya no se corresponde necesariamente con las posibilidades del sistema actual de una lengua (Coseriu 1981b, 113, cf. también Coseriu 1973, 27 y Coseriu 1976, 25 y 66); en otras publicaciones, Coseriu incluye en esta serie de fijaciones los refranes y las «fórmulas tradicionales de parangón», flexibilizando a la vez la idea de lo fijado o fosilizado: el discurso repetido comprende, así, «todo aquello que, en el hablar de una comunidad, se repite en forma más o menos idéntica, como discurso ya hecho o como combinación más o menos fija, como fragmento, por largo o breve que sea, de «ya dicho»» (Coseriu 1981a, 298).

La clasificación de las unidades del discurso repetido la desarrolla Coseriu en el marco de su concepción estructural y funcional del léxico, describiendo cada tipo con vistas a su conmutabilidad con unidades de distintos niveles de la técnica libre del hablar. En este sentido, Coseriu distingue entre cuatro tipos de unidades del discurso repetido (Coseriu 1981b, 115–118; 1981a, 300–302; cf. también Coseriu 1973, 29ss.; ejemplos *ibidem*):

1. las unidades funcionalmente equivalentes a oraciones («frasemas» o «textemas») como «frases metafóricas, proverbios, dichos, sentencias, «wellerismos», refranes» y también «citas de textos más o menos conocidos o, al menos, históricamente identificables» (esp. *cada palo aguante su vela*, it. *tanto va la gatta al lardo* (*che ci lascia lo zampino*);
2. las unidades que se corresponden con sintagmas, los «sintagmas estereotipados» (esp. *atar cabos*, fr. *se moquer du tiers comme du quart*);
3. las unidades que equivalen a palabras simples llamadas «perífrasis léxicas» (esp. *echar en cara* → *reprochar*, fr. *au fur et à mesure* → *successivement*);

1 Estas distinciones son a) «cosas» y lenguaje, b) lenguaje primario y metalenguaje, c) sincronía y diacronía, d) técnica del discurso y discurso repetido, e) arquitectura y estructura de la lengua, f) sistema y norma, g) relaciones de significación y relaciones de designación (Coseriu 1967, sobre todo Coseriu 1981b, 94s.; cf. también Coseriu 1973, 8).

4. y las fórmulas de parangón, el único tipo que funciona en el nivel de los morfemas (esp. *más vivo que un rayo*, it. *buono como il pane*).

En el contexto del análisis estructural-funcional, esta clasificación tiene el propósito de seleccionar ciertas unidades lingüísticas: solo las unidades del tercer tipo, las perífrasis léxicas, funcionan como lexemas y pueden ser analizadas en el marco de la lexicología estructural (Coseriu 1973, 31; 1981b, 116s.); las unidades de carácter sintagmático y oracional (los tipos 2 y 1) serían objeto de estudio de la sintagmática y de «las ciencias literarias» (Coseriu 1981b, 116) o la lingüística de texto y la filología, respectivamente (Coseriu 1981a, 301).

1.2 Discurso repetido y fraseología

Coseriu reconoce que el análisis de las unidades del tipo 1 se realiza «en parte» por la paremiología y propone analizar los tipos 2, 3 y 4 como «una amplia sección autónoma del «saber idiomático»» (Coseriu 1981a, 301s.). Aunque Coseriu mismo no identifica el discurso repetido con el inventario fraseológico, este área autónoma de las lenguas corresponde al objeto de estudio de aquella disciplina de la lingüística que, más allá del límite univocal, se ocupa de las unidades lingüísticas con significado integral que se reproducen como conjuntos en el discurso: la fraseología. En los trabajos de Coseriu sobre el discurso repetido, sin embargo, el análisis de la estructura léxico-morfológica de los productos del discurso repetido, primordial en los estudios fraseológicos hasta entonces, no se considera prioritario. El binomio conceptual distintivo de la técnica libre del hablar y del discurso repetido sirve más bien para localizar un fenómeno lingüístico universal en la descripción del lenguaje. No obstante, el concepto y la clasificación de las unidades del discurso repetido fueron retomados en el ámbito de los estudios fraseológicos de las lenguas romances desde mediados de los años 1970, como p.ej. en los trabajos de Hausmann (1974) en el marco de la lingüística del juego de palabras del francés, y Ettinger (1977) sobre el lenguaje bíblico, también del francés. En un trabajo orientado al francés, italiano, rumano y español sobre los «productos» del discurso repetido, resumidos bajo el término de «fixierte Wortgefüge» ('conjuntos léxicos fijados'), el discípulo de Coseriu Thun (1978), explica que el discurso repetido no es un problema unitario, sino que abarca toda un área del lenguaje (Thun 1978, 26). A diferencia de Coseriu, Thun equipara explícitamente este fenómeno lingüístico global con la fraseología; sin embargo, aclara que la perspectiva de Coseriu sobre las unidades de este fenómeno es la del análisis estructural-funcional, con una atención especial al análisis semántico y a la clasificación según criterios semánticos (cf. Thun 1978, XV y 170–172).

También en la lingüística hispánica, donde hasta los años 1980 apenas existen enfoques propios para el estudio de la fraseología del español, se retoma el concepto de discurso repetido. Las ideas de Coseriu influyen en los estudios fraseológicos, promoviendo la atención de la lingüística hacia las unidades fijas del vocabulario español (cf.

Martínez Marín 1996, 75; Corpas Pastor 1996, 37; Ruiz Gurillo 1997, 51). Así, en uno de los primeros estudios sistemáticos de la fraseología del español, Zuluaga (1980) aborda el concepto del discurso repetido de Coseriu antes de llevar a cabo su análisis de la gramática y semántica de las unidades lingüísticas que él denomina «expresiones fijas». En otro trabajo fundamental para el estudio de la fraseología del español, Corpas Pastor (1996) discute el concepto del discurso repetido de Coseriu en el marco de una exposición detallada de las clasificaciones más relevantes de las unidades fraseológicas. Tanto Zuluaga (1980, 75s.) como Corpas Pastor no consideran que la categorización de las unidades del discurso repetido sea apta para el estudio de la fraseología del español, ya que los criterios de clasificación no serían suficientes y en la categoría del tipo 1 se tomarían en cuenta unidades «que quedan fuera de los límites de la fraseología»; además, la distinción entre los tipos 2 y 3 no resultaría del todo clara, como Coseriu mismo reconocería más adelante (Corpas Pastor 1996, 37). Del mismo modo, Ruiz Gurillo (1997, 56), en un trabajo sobre la teoría de la fraseología del español, considera que la distinción entre discurso repetido y técnica libre del hablar es demasiado amplia para una aplicación adecuada al estudio lingüístico de la fraseología. Sin embargo, a la crítica de la clasificación de Coseriu de las unidades del discurso repetido se puede objetar que, aunque esta categorización quizás no sea precisa para el análisis sintáctico y semántico de las unidades fraseológicas y, en general, para los fines de la investigación fraseológica (de entonces), sí fue suficiente para señalar las posibilidades combinatorias a nivel léxico y gramatical y las combinaciones recurrentes de elementos lingüísticos consideradas como unidades discursivas que se repiten por la tradición lingüística, y también para proponer la oposición entre técnica libre del hablar y discurso repetido como una de las distinciones fundamentales en la descripción lingüística.

El hecho de que Coseriu, en el marco de su organización de las unidades del discurso repetido, se sirva en parte de las categorías con las que trabaja la fraseología como disciplina lingüística, también puede ser una de las razones por las que el concepto original del discurso repetido primero no se haya desarrollado en la dirección que Coseriu le dio al incorporarlo al contexto global de su lingüística de texto como lingüística del sentido. En este marco conceptual, Coseriu caracteriza el discurso repetido como un tipo de relaciones de signos lingüísticos de un texto con signos de otros textos que se producen mientras los hablantes usan «secuencias de signos (= de combinaciones «hechas» de signos) que se transmiten íntegramente», tal y como son –o incluso modificados de forma creativa– para hacer referencia a otros textos, segmentos de textos o secuencias de signos ya producidos y conocidos por la comunidad lingüística. De ese modo, los hablantes crean sentido en nuevos textos y logran los correspondientes efectos de interpretación (Coseriu 2007, 200–204).

Coseriu atribuye también esta función a las « citas célebres », textos conocidos por la comunidad lingüística, en parte palabra por palabra (Coseriu 2007, 202 y 203). Aunque las citas célebres no siempre pueden diferenciarse con claridad del discurso repetido, como en el caso de los proverbios (Coseriu 2007, 203), los dos conceptos presentan ideas parecidas: desde la perspectiva de la lingüística de texto de Coseriu, ambos

fenómenos de lo lingüístico no se entienden en primer lugar como secuencias de signos, sino como relaciones intertextuales.

A partir de los años 1990, en la investigación fraseológica, considerada como un ámbito de investigación independiente tanto a nivel teórico como metodológico, son cada vez más las propuestas de estudiar las unidades fraseológicas desde el punto de vista de su funcionalidad en el discurso o en el texto, más allá de sus características léxico-morfológicas. Muestra de este cambio de tendencia son, en el ámbito de la lingüística española, los trabajos de Ruiz Gurillo (1998), García-Medall (2006), Olza/Losada (2011), Zamora Muñoz (2014) y las obras colectivas de Mellado Blanco (2008) sobre la fraseología textual y de Olza/Manero Richard (2013) sobre la «fraseopragmática». Evoluciones más recientes demuestran, además, que el concepto del discurso repetido ya no se iguala *a priori* ni con la fraseología ni con sus objetos de estudio. Teniendo en cuenta la orientación del concepto a la lingüística de texto que le dio Coseriu, el discurso repetido se identifica más bien con la totalidad de las relaciones de signos en textos con signos en otros textos (cf. p.ej. Mellado Blanco/Berty/Olza 2017), y se aplica a los enfoques de una fraseología centrada en el discurso y, sobre todo, en el texto, los cuales tienen como objetivo el análisis de las funciones discursivas y textuales de las unidades fraseológicas. En este contexto se analiza p.ej. la función argumentativa del discurso repetido en el lenguaje de las tertulias políticas de televisión (González Sanz 2017) o la función contextual de las unidades fraseológicas en textos literarios (Manero Richard 2017), todo ello a partir de análisis de corpus escritos y orales. Mediante el concepto de discurso repetido quedan patentes las dimensiones de lo fraseológico en el hablar, pues el impacto de lo tradicional se ve complementado por los efectos de lo innovador.

1.3 La historicidad del hablar y la modificación de las unidades fraseológicas

Con respecto a la historicidad del hablar, el fenómeno de la modificación, re-metaforización, remotivación o desautomatización es de particular interés para las perspectivas discursivas y textuales en la investigación fraseológica (724 Las tradiciones discursivas). Sin trazar paralelos directos con las perspectivas fraseológicas, la posibilidad de convertir una unidad fijada en «una combinación libre que se refiere de diversas maneras a la fraseológica» (Ruiz Gurillo 1997, 75) ya fue considerada por Coseriu en relación con las funciones generales del discurso repetido. Para Coseriu, las unidades modificadas se conciben como textos conocidos y modificados (cf. Coseriu 1976, 25s.). Esto significa que, por un lado, una expresión fija modificada en un discurso concreto debe ser conocida por los hablantes como una unidad que puede ser sujeto y referencia de la modificación creativa; por otro lado, el hablante que modifica la expresión para constituir un sentido concreto se orienta primero al nivel de los textos de la comunidad lingüística a los que pertenece la unidad fijada, y no a los niveles del sistema y la norma de la lengua.

Vemos así que la noción de discurso repetido, ligada intrínsecamente al rasgo de la «reproducibilidad» en el discurso (cf. Zuluaga 1975, 226), conlleva en sí misma la variación. Ello se debe en primer término a la estructura plurimembre de las UF (unidades fraseológicas), que favorece en conjunto: (1) su variabilidad sintáctico-estructural (cf. § 2.1), (2) la divisibilidad semántica y la autonomización de componentes (rasgos fundamentales en la desautomatización, cf. § 2.2), (3) el carácter difuso de su significado, (4) la formación de sinonimia estructural y variantes léxicas (cf. § 2.1 y § 2.2), así como (5) la fuerte expresividad a través del carácter figurado y las connotaciones que porta la imagen (cf. Fleischer 1997, 214).

La variación, como magnitud complementaria a la fijación, tiene muchas y variadas facetas que trataremos de sistematizar en el apartado 2 de acuerdo con el plano en el que desarrollan.

2 Tipología de la variación

En este apartado distinguimos entre la variación que afecta al plano de la lengua, al plano del discurso y la que atañe a ambos niveles.

2.1 Variación en el plano de la lengua como sistema (*langue*)

El fenómeno de la variación en el plano de la lengua como sistema (*langue*) se expresa en variantes lexicalizadas consensuadas por los miembros de la comunidad lingüística:

1. Se dan dentro de la misma lengua funcional y no conllevan diferencias de sentido (cf. Corpas Pastor/Mena Martínez 2003; Montoro del Arco 2005, 129): variantes fónicas (esp. *en un pis pas/en un plis plas*, it. *mano a mano/man mano*), gráficas (esp. *a mata caballo/a matacaballo*, pt. *apesar de/a pesar de*), morfológicas (esp. *dueño y señor/dueña y señora*, fr. *jouer du coude/jouer des coudes*), gramaticales (esp. *en/por aquel entonces*, it. *a cavallo tra/di*), léxicas (*coger/tomar las de Villadiego*, pt. *meter o pé na argola/na poça*), sintácticas (esp. *¡Qué sé yo!/¡Yo qué sé!*, fr. *C'est dommage!/Quel dommage!*), variaciones por extensión o supresión de un componente (esp. *estar hasta los [mismísimos] cojones/estar hasta los mismísimos [cojones]*, it. *[bella/brutta] gatta da pelare*), «citas interrumpidas» (esp. *¡Dichosos los ojos [que te ven]!*, cf. Zuluaga 1980, 100, pt. *Ladrão que rouba ladrão [tem 100 cem anos de perdão]*) y listas fraseológicas (esp. *estar de buenas/malas*, fr. *parler bas/parler haut*).
2. Cuando aparecen en distintas lenguas funcionales estamos ante la variación marcada de acuerdo con Penadés Martínez (2012, 56): variantes diatópicas (esp. *ser uña y carne/ser uña y mugre*, pt. *Em boca fechada não entra mosca (PB)/Em boca fechada não entram moscas (PPT)*), diastráticas (esp. *estar hasta la coronilla/hasta los cojones*, it. *rompere le scatole/le palle*) y diacrónicas (esp. *importar un pito/*

un maravedí, fr. *donner la peau/chair de poule*) (cf. Pamies Bertrán 2017; Berty/Mellado Blanco/Olza 2018).

3. Variación surgida por transformación gramatical: nominalización (esp. *meter la pata* → *metedura de pata*, pt. *meter o pé na argola* → *uma argolada*).

2.2 Variación en el plano del discurso

En la variación en el plano del discurso distinguimos:

1. Variación resultante de la adaptación al contexto, la cual no persigue efectos lúdicos ni expresivos:² pronominalización, anaforización, transformación en pasiva, topicalización, inversión de los constituyentes, etc. Un ejemplo de topicalización y pronominalización se ve en (1a y b):

(1a) Al final, el gato al agua se lo llevaron los italianos, que pudieron ocupar Libia y las islas del Dodecaneso (CORPES XXI. Díaz Villanueva, Fernando: «1912: se hundió el Titanic...». *Libertad Digital. Historia*, consultado el 23.09.2009, Madrid, historia.libertaddigital.com).

(1b) Se non mancasse un anno e mezzo alla fine della legislatura, passerebbe come «foto», un sorriso amaro, e via. Ma siamo nel bel mezzo del «bordello» e di ridere non viene voglia a nessuno e certo non aiuta Mario Oliverio che l'altra sera le staffe le ha perse per davvero (*Il Fatto di Calabria*, 22.05.2018, <https://www.ilfattodicablabria.it/apertura/2018/05/bordello-campanella-e-tallini-abbracciamoci-tutti-ormai/>, consultado el 30.12.2019 [explicación: *perdere le staffe*: 'enfadarse', 'perder la paciencia']).

2. Variación como recurso estilístico por parte del hablante. Estamos ante las modificaciones intencionadas, que provocan desautomatización.³ Mena Martínez (2003, 6) distingue entre «desautomatización» propiamente dicha, cuando la intencionalidad de la variación es percibida como tal por parte del receptor, y «modificación», cuando esto no sucede. Además, según Mena Martínez (2003, 6), la desautomatización (1) ha de ser un cambio ocasional, voluntario e intencionado por parte del hablante (cf. Mena Martínez 2004, 501; Timofeeva 2009, 252–253;

² Este tipo de variación es interpretada por Penadés Martínez como desautomatización formal que no conlleva cambio en el significado (2014, 288) y es una consecuencia del interesante fenómeno que la autora describe como «fijación estructural». Al no presentar intencionalidad no la consideramos como desautomatización.

³ El concepto de «desautomatización», equivalente a otros términos como «deslexicalización» (García-Page Sánchez 2008, 306) o «manipulación creativa» (Corpas Pastor 1996), es original de los formistas rusos. Zuluaga (1975) lo introduce en el español posiblemente desde Coseriu (para una historia del término cf. Martí Sánchez 2015, 118).

Martí Sánchez 2015), (2) la desviación con respecto a la forma originaria ha de ser suficiente para que el cambio pueda ser percibido, (3) la forma originaria debe ser recuperable y reconocible por el oyente. En el sentido de la Teoría de la Relevancia, el fenómeno es también ostensivo, pues el hablante, cuando desautomatiza, manifiesta esa intención. La desautomatización, muy vinculada a la reliterarización, es una prueba de que ni la imagen subyacente a las UF ni los significados externos de los constituyentes fraseológicos desaparecen en la mente de los hablantes tras el proceso de fraseologización. Asimismo, dicho «potencial textual», al ser inherente a las UF como unidades plurimembres y signos de segunda nominación, está situado en el plano de la *langue* (cf. Dobrovol'skij 1987), aunque luego se desarrolle en el discurso. En la desautomatización distinguimos:

a. Desautomatización sin modificación formal de la UF. Este fenómeno conlleva remotivación del significado literal (reliterarización) de algunos o todos los constituyentes de la UF (ejemplos 2a y 2b). Un subtipo interesante de estas modificaciones consiste en un cambio en la selección semántica de los argumentos de la UF, p.ej. [-hum], en lugar de [+hum] para el sujeto, o viceversa.

(2a) En cualquier caso replicó Taboullier, nunca es aconsejable *tirar la casa por la ventana*, porque te quedas con una ventana pero sin casa [CORPES XXI. Del Val, Luis: *Estamos dentro*, Madrid, Espasa Libros, 2010].

(2b) *Uma mão lava a outra* e as duas lavam o Relvas (Comendador Marques Correia: «Os novos Provérbios Portugueses adaptados e revistos pela Troika», in: *As melhores 30 Crónicas em 30 anos*, Lisboa, Edições do Coleccionador do Expresso, 2020 [explicación: 'unos encubren a otros y todos encubren a Relvas, un político que originó un escándalo en su día']).

b. Desautomatización con modificación formal de la UF. Aquí entran la sustitución, método más frecuente según Martí Solano (2011) (cf. ejemplos 3a y 3b), la supresión de constituyentes, la permutación, la contaminación de varias UF (ejemplo 4a: entre la UF *pedir peras al olmo* y los versos del poema *A un olmo seco* de Antonio Machado,⁴ ejemplo 4b: entre a UF *La balle est dans [votre] camp* y *se renvoyer la balle*), la acumulación de UF (ejemplos 5a y 5b), o varias de estas técnicas a un tiempo.

(3a) Cría cuervos y *te comerán los cojones*, agrega con falso aire de intelectualidad popular (CORPES XXI. La Torre Lagares, Elidio: *Correr tras el viento*, Viejo San Juan, Terranova Editores, 2011).

⁴ Los versos del poema *A un olmo seco* (1912), en los que se inspira el ejemplo (4) de contaminación, rezan: «Al olmo viejo, hendido por el rayo y en su mitad podrido, con las lluvias de abril y el sol de mayo algunas hojas verdes le han salido.»

(3b) Dalle stelle alle *sardine*, e il Pd? (*La7*, Otto e Mezzo, 01.02.2020 [explicación: *dalle stelle alle stalle*: de una situación muy favorable a otra netamente peor]).

(4a) Pero no pretendas, querido Ignacio, que un relativista entienda estos conceptos. Es *como pedirle peras al olmo seco, hendido por el rayo y en su mitad podrido* (CORPES XXI. De Prada, Juan Manuel: «Zapatero, el relativista», *ABC.es*, 16.01.2006, Madrid, abc.es).

(4b) Plutôt uni sur le constat, le PS se fait même un malin plaisir de *renvoyer la balle de la discorde dans le camp de la majorité*. François Hollande a ainsi proposé à Nicolas Sarkozy de disjoindre les deux volets du texte, et de faire voter rapidement la partie concernant le Parlement (EUROPRESSE, *La Croix*, 31.10.2007 [explicación: *La balle est dans votre camp* 'la pelota está en su tejado', *se renvoyer la balle* 'devolver la pelota a alguien, pagar con la misma moneda', cf. Fortes/Merceur 2003]).

(5a) Sí, señor, no sé si tiene usted coRazón, pero seguro que tiene razón y está en su derecho de *hacerse de su capa un sayo y de su picha un lío* (CORPES XXI. Amezttoy, Begoña: *Escuela de mujeres*, Madrid, Oberon/Grupo Anaya, 2001).

(5b) So que nos tempos modernos, nesta sociedade extremadamente competitiva, e, mais do que isso, materializada *dos pés à cabeça*, há que *ter o «pão nosso de cada dia»* para alimentar ese fervor espiritual (<http://alf.clul.clul.ul.pt/CQPweb/portugal/context.php?batch=51&qname=gbp96tmt0A&uT=y>, consultado el 30.12.2019).

c. Lapsus o error (como p.ej. esp. *pelos de gallina*, en lugar de *carne/piel de gallina* (cf. ejemplo 6), o fr. **Je donnai un coup d'œil* en lugar de *Je jetai un coup d'œil* (ejemplo tomado de González-Rey 2016, 175).

(6) En la película de mi vida, tú en este momento me detendrías, [...] y dejaríamos que la música incidental nos pusiera los *pelos* de gallina (CORPES XXI. Valenzuela, José Ignacio: *El filo de tu piel*, México D. F., Santillana Ediciones Generales, 2009).

A diferencia de la desautomatización con o sin modificación formal de la UF, el lapsus o error no es intencionado. Para algunos estudiosos, como Martí Sánchez (2015, 120), el lapsus es una «desautomatización inintencionada» y en calidad de desautomatización se sitúa en el plano dialéctico del análisis, de la creatividad y de la remotivación frente a la síntesis y al lenguaje repetido y no creativo. Ciertamente es que al observar el error desde fuera no se puede concluir si se trata de una alteración accidental o de una desautomatización deseada. Para Dobrovol'skij (1999, 28), el papel del oyente es fundamental para poder llevar a cabo esta diferenciación. Así, para que exista juego creativo han de darse tres condiciones (1) que el oyente conozca la existencia de la UF, (2)

que esté dispuesto a aceptar las reglas colaborativas del juego y (3) que reconozca la UF a pesar de su desviación de la forma canónica. Según este autor, solo la condición (2) da la pauta para distinguir la modificación creativa del error: si el oyente no contribuye al juego, no hay juego, por lo que la desviación es interpretada como error. No obstante, Elspass (2002) pone de manifiesto en su estudio con informantes para el alemán que el umbral de tolerancia del oyente es muy alto en cuanto a la percepción de una UF como errónea o mal utilizada. Que una alteración sea percibida como error depende en primer término del tipo de transgresión, siendo las sustituciones y las contaminaciones los tipos de lapsus que antes se identifican y tienden a ser subsanados por el oyente. Otro factor importante es el grado de complejidad cognitiva de la UF: a mayor complejidad, mayor tolerancia y disponibilidad a interpretar la desviación como si de un uso canónico se tratara. Asimismo, los informantes no lingüistas se mostraron en este estudio más cooperativos y menos estrictos a la hora de valorar errores.

Tanto la noción de error, como la de desautomatización intencionada, ya sea con finalidad lúdica o seria –por ejemplo, con el fin de realizar una crítica irónica– presupone que existe una forma canónica previa con respecto a la cual se dan las alteraciones. El problema surge cuando queremos predeterminar la forma canónica real de una UF, su forma regular de uso, sus constituyentes, su flexión, etc., como hacemos con los monolexemas o compuestos. El grado de complejidad del referente denotado por la UF, desde un punto de vista cognitivo, tiene gran relevancia en este fenómeno, además del carácter pluriverbal de las UF. Así, las locuciones idiomáticas verbales largas con un referente complejo en la realidad extralingüística son especialmente escurridizas y la disparidad de lemas y acepciones en los distintos diccionarios refleja fielmente esta dificultad de encontrar una forma canónica única. A modo de ejemplo, ¿cuál es la forma canónica del binomio cognitivo *paja en el ojo ajeno* y *viga en el propio*? Una vista rápida en el CORPES XXI nos mostrará el problema de determinar un lema válido para los ejemplos del corpus, pues la variación formal impera en la mayoría de ellos (cf. § 3.2). En cuanto a la UF *no pedirle peras al olmo* (cf. supra ejemplo 4a), en los corpus aparece *no pedirle/exigirle peras al olmo*; *no esperar peras del olmo*. Cuando estos verbos están vinculados cognitivamente entre sí no puede quizá hablarse de sustitución léxica, sino de *entrenchment* o fijación cognitiva en torno a una imagen abstracta que se verbaliza de maneras diversas (cf. Mellado Blanco 2019; 2020a; 2020b). Aquí no se dan los requisitos básicos de la desautomatización en cuanto a la intencionalidad.

En este orden de cosas, cabe asimismo preguntarse si algunos de los tipos tradicionales de desautomatización serían realmente tales o simplemente mecanismos de adaptación al contexto, sin repercusión en el significado fraseológico (cf. Häcki Buhofer/Burger 1992, 15). Así sucede en:

- Expansión a través de un modificador, ya sea un adverbio o adjetivo (7a y 7b) o un sintagma preposicional (8) (cfr. Dobrovolskij 2009).

(7a) Quizá *hago montañas de un minúsculo grano*, pero también pecaría de inexacto si no lo hiciera [...] (CORPES XXI. Gomes, Miguel: «El vuelo de Se-

bastián Da Silva», in: *Un fantasma portugués*, Caracas, Otero Ediciones, 2004).

- (7b) [...] nonostante il tentativo di fermare l'avanzata della Lega *hanno fatto un enorme buco nell'acqua* (Ravenna Notizie, 28.10.2019, <https://www.ravenna-notizie.it/politica/2019/10/28/la-lega-commenta-il-successo-in-umbria-ci-at-tendono-grandi-vittorie-anche-in-emilia-romagna-e-a-faenza/>, consultado el 30.12.2019).
- (8) Por dar un ejemplo muy cercano, tenemos el efecto de *elefante en cacharrería de los gobiernos españoles* al intervenir en el mercado de la electricidad (CORPES XXI. Schwartz, Pedro: *La economía explicada a Zapatero y a sus sucesores*. En *dos tardes*, Madrid, Espasa, 2011).

- Expansión por medio de determinantes, ya sean artículos, posesivos, demostrativos, numerales (ejemplo 9), indefinidos, cuantificadores (ejemplos 10a y 10b), interrogativos o exclamativos, que permiten la integración de la UF en el contexto:

- (9) Pero dónde están los superratones que les *pongan los cascabeles a esos tres gatos* (CORPES XXI. Cuéllar, Fidel H.: «DÉFICIT, GRADUALISMO Y CHOCQUE», *El Tiempo*, Bogotá, eltiempo.com, 07.11.2003).
- (10a) «Una de sus mayores equivocaciones ha sido la de *criar numerosos cuervos*, que hoy rivalizan entre sí para *sacarles los ojos*», concluye (CORPES XXI. «Disgustados formarían nuevo partido», *ABC Digital*, Asunción, abc.com.py, 12.10.2005).
- (10b) À la dernière convocation, après un mois et demi, je répondis par un simple mot: «J'estime avoir à peu près répondu, et crois que les instances théoriques du Parti, qui ont *beaucoup de pain sur la planche*, feraient bien de cesser ce procès et de s'occuper d'affaires plus urgentes.» Et je ne vins pas (FRANTEXT. Louis Althusser: *L'avenir dure longtemps*, 1985 [explicación: *avoir du pain sur la planche* 'tener trabajo para rato/tela que cortar', cf. Fortes/Merceur 2003]).

Investigaciones llevadas a cabo en otras lenguas, como el inglés, parecen corroborar este punto (Keizer 2018, 193–194): No estamos en estos casos ante desviaciones de la forma canónica con la intención de conseguir un efecto lúdico o creativo, sino ante procesos de integración textual:

[U]nlike what has so far been assumed in the literature, anaphoric reference by or to part of a non-decomposable idiom does not necessarily result in wordplay, and that, even if there is a certain degree of wordplay involved, such reference may still be systematic and productive. This means that these occurrences cannot be dismissed as deviations and need to be accounted for in the grammar of a language.

Asimismo, tampoco la reducción, cuando no conlleva cambios de significado (p.ej. en esp. *A buen entendedor...*, en lugar de *A buen entendedor, con pocas palabras bastan*, it. *Occhio non vede... [cuore non duole]*), puede considerarse como un tipo de desautomatización. Especialmente en los refranes muy conocidos, la segunda parte resulta redundante precisamente por su alta probabilidad de ocurrencia y no necesita ser explicada (cf. Zuluaga 1975; Mellado Blanco 2016).

2.3 Variación que afecta al sistema y al plano textual⁵

Aquí se encuentran los esquemas fraseológicos o construcciones fraseológicas (cf. Mellado Blanco 2020a), caracterizadas por tener una parte fija y otra variable, y un significado global. Aunque la estructura como tal forma parte de la *langue*, la actualización léxica de los slots tiene lugar en el texto, siguiendo un paradigma que puede ser más o menos abierto, sometándose a restricciones categoriales, morfológicas o semánticas. A modo de ejemplo, el esquema fraseológico [DET ADJ de POS SUST], con un potencial ilocutivo de CRÍTICA por parte del hablante, se actualiza en el discurso mediante instancias como esp. *el listo de mi marido* (cf. ejemplo 11), *la guapa de tu vecina*, *la tonta de mi suegra*, pt. *o espertalhão do teu irmão*, *a burra da minha tia*:

- (11) Y ahora tú quieres devolverme ahí, a ese montón de gilipollas, para que se pasen todo el día restregándose que tenían razón y que *el listo de mi marido*, el moro, me ha abandonado (CORPES XXI. Conejero López, Alberto: «Oriente», in: *La Ratona*, El Entrelgo, Oris Teatro, 2009).

La delimitación entre variantes léxicas múltiples y esquemas fraseológicos no siempre es fácil, como sucede en *[no] importar un(a) bleo/comino/pimiento/rábano/pito/...*, que puede interpretarse como una sucesión de variantes léxicas sinónimas, pero también como un esquema fraseológico *[(algo/alguien) (no) importa + SN[DET_{NUMERAL} X] a alg.]*: ‘no importar absolutamente nada’, en el que el slot libre se satura léxicamente –según nos dice el corpus consultado *esTenTen18*– mediante sustantivos que indican frutos, objetos y monedas de escaso valor, o bien que están tabuizados por su contenido sexual (p.ej. *No me importa una mierda/un cojón/un carajo/...*) (cf. Mellado Blanco 2020b), y esto tanto en el español peninsular como en las variedades hispanoamericanas.

Desde la lingüística de corpus se están desarrollando métodos estadísticos para la diferenciación entre variantes léxicas y esquemas fraseológicos, dependiendo de la

⁵ No consideramos aquí las llamadas UF con casillas vacías, del tipo *estar a mi/tu/su bola*, porque entonces deberíamos considerar todas las UF verbales por su variabilidad flexiva: *No me tomes el pelo*; *Me tomas el pelo*, etc.

productividad de los slots afectados. Así, para Stumpf (2016), un porcentaje mayor a un 50 % de variación léxica de un fraseologismo es índice de esquema productivo y no de variantes léxicas. El problema de esta evaluación es que para poder determinar la desviación debe existir una forma fija a partir de la cual se pueda diverger, y eso es precisamente lo que no está claro en muchas UF.

3 Variabilidad fraseológica y lingüística de corpus

Tanto el concepto coseriano de discurso repetido y como la noción, más o menos afin, de fraseología, descansan sobre el aspecto de estabilidad, fijación e institucionalización que presentan determinadas secuencias de palabras. Dichas secuencias se repiten en el discurso con valores semánticos, discursivos y/o pragmáticos específicos. Clave para entender y analizar dichas secuencias es el grado de fijación o repetición de sus elementos constitutivos, el grado de centralidad y obligatoriedad de estos, así como su tipicidad, frecuencia y saliencia cognitiva.

Digamos que estas secuencias más o menos estables de palabras se definen por una relación dialéctica entre lo fijo y lo variable, entre la abstracción inherente a los sistemas lingüísticos y los datos reales que modelan tales sistemas. Hasta muy recientemente, los análisis de este tipo de secuencias se realizaban de manera introspectiva y utilizando, en el mejor de los casos, datos limitados. No obstante, con el surgimiento de la lingüística de corpus, ha sido posible realizar estudios empíricos basados en volúmenes ingentes de datos. Ello ha propiciado el surgimiento de nuevos enfoques centrados en el procesamiento de dichos datos y su análisis mediante técnicas estadísticas y de PLN (Procesamiento de Lenguaje Natural). Hoy día, las corrientes basadas en corpus y de corte computacional se abren camino como la tendencia predominante dentro de una disciplina completamente consolidada. Véanse, por ejemplo, los volúmenes editados por Corpas Pastor (2016), Mitkov et al. (2018), Corpas Pastor/Mitkov (2019), Corpas Pastor/Colson (2020), entre otros.

Los estudios de fraseología actuales, por su interdisciplinariedad, permiten demostrar cuantitativamente ideas que se asumían de forma cualitativa. Los resultados obtenidos hasta la fecha confirman, sin lugar a dudas, la existencia del principio de idiomatidad inherente a la comunicación humana (cf. Pawley/Syder 1983; Sinclair 1991). En este nuevo escenario, el concepto de texto se disuelve en conglomerados de patrones, construcciones, etc., que surgen del análisis de datos masivos. Y estos conglomerados se evidencian en secuencias de palabras «repetidas», esto es, instancias dentro de un discurso que se revela altamente prefabricado, pre-establecido y eminentemente fraseológico.

Una consecuencia importante de este giro empírico es el replanteamiento del lugar que ocupan las UF y otras secuencias repetidas en la lengua. De hecho, se pasa de considerarlas curiosidades o excepciones a darles un lugar preponderante. Esto es, estarían al mismo nivel que otras estructuras lingüísticas, como parte del continuum

idiomático que vertebró la esencia eminentemente fraseológica de las lenguas (cf. Kay/Fillmore 1999).

3.1 Detección automática

La detección de UF es uno de los aspectos más controvertidos dentro de estos nuevos enfoques empíricos, de corte computacional y de corpus (741 El hablar y la lingüística empírica). Se utilizan para ello diversas pruebas o criterios como, por ejemplo, la imposibilidad de sustitución paradigmática (**dar gato por conejo*) o la existencia de equivalentes de traducción (*dar gato por liebre* ≈ EN *pull the wool of someone's eyes*) (cf. Ramisch 2015; Monti et al. 2018), aunque el criterio más empleado es el de frecuencia (de aparición y/o de coaparición).

Se ha comprobado que si bien hay tipos de UF muy frecuentes en el discurso, como, por ejemplo, las colocaciones (esp. *prestar atención*, it. *prestare attenzione*; esp. *cometer un error*, fr. *commettre une erreur*, etc.), otras UF no parecen serlo tanto, como es el caso de las locuciones idiomáticas (esp. *dar gato por liebre*, pt. *vender gato por lebre*; esp. *tomar las de Villadiego*, it. *scappare a gambe levate*, etc.), aunque se reconocen fácilmente como UF (esto es, presentan un alto grado de saliencia cognitiva). Por ejemplo, la tabla 1 muestra la presencia de estas UF del español en el *esTenTen18*, un megacorpus de 17.500 millones de palabras de español compilado de forma automática y disponible a través de SketchEngine (cf. Kilgariff et al. 2014). Recogemos tanto la frecuencia absoluta (FA: número total de apariciones en el corpus) como la frecuencia normalizada (FN: número de apariciones por millón de palabras del corpus) de las UF y de sus elementos integrantes.

Tabla 1: Análisis de frecuencias de UF y elementos integrantes (esTenTen18)

UF	FA	FN
tomar las de Villadiego	504	0,02
<i>tomar</i>	9.139.726	450,09
<i>Villadiego</i>	4.771	0,23
dar gato por liebre	3.159	0,16
<i>dar</i>	30.734.964	1.513,50
<i>gato</i>	727.992	35,80
<i>liebre</i>	43.722	2,20
prestar atención	339.085	16,70
<i>prestar</i>	2.705.238	133,20
<i>atención</i>	4.390.299	216,20

cometer un error	428.926	21,12
<i>cometer</i>	1.326.993	65,35
<i>error</i>	2.461.587	121,20

Hay que distinguir, pues, entre la frecuencia de aparición (lo que se mide en la tabla 1) y la frecuencia de coaparición de los elementos integrantes de una UF, esto es, la probabilidad estadística de que dos o más palabras aparezcan juntas, comparando la frecuencia real de coaparición de la secuencia de palabras analizadas (n-grama) 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. En el caso de las UF, la frecuencia real debe ser superior al nivel de probabilidad convencional (significación estadística). Para calcular la significación estadística se emplean medidas de asociación léxica, como, por ejemplo, el índice de información mutua (*MI-mutual information*) o la ratio de verosimilitud (G2, en inglés *G-square*, *G score* y, también, *log-likelihood*) (cf. Corpas Pastor 2013; 2017; Kolesnikova 2016). Los resultados de estas medidas de asociación léxica mejoran cuando se combinan con análisis de patrones sintácticos y etiquetado gramatical (POS tagging), como ocurre en los sistemas híbridos de extracción (Seretan 2011).

Las medidas de asociación estadística facilitan la detección de UF superiores al bigrama, de carácter discontinuo o con sus elementos constitutivos alejados en el discurso. Sin embargo, estas no se muestran tan eficaces para identificar UF desautomatizadas o modificadas (ejemplo 12), ni para detectar secuencias homófonas –literal y fraseológica– de forma automática (ejemplos 13 y 14), como se puede observar en las siguientes concordancias extraídas de *esTenTen18*:

- (12) ¿Qué quieres pesado? – Somos pareja y bien combinados. Ya sabes que en política estoy hace rato, y aunque se me critica *cojo liebre por gato*.
- (13) Bueno, al menos no han hecho como Zapatero que *se bajó los pantalones* en la Eurocámara y renunció al derecho de veto que había conquistado el gobierno anterior.
- (14) La víctima, dolorida, se quejó rápidamente con el árbitro y *se bajó los pantalones* para mostrarle la lesión.

Por otra parte, los enfoques computacionales aspiran a computar también los aspectos semánticos e interpretativos de las UF. De hecho, la idiomatidad (*idiomaticity*) se entiende como un aspecto idiosincrásico y transversal propio de estas unidades. Por ejemplo, Sag et al. (2002) definen las unidades fraseológicas (*multiword expressions* o MWE, en su terminología) como unidades compuestas por varios lexemas que presentan idiomatidad léxica, sintáctica, pragmática y/o estadística. Los lingüistas computacionales son conscientes de las peculiaridades semánticas de dichas unidades, y de la dificultad que ello añade a la detección automática. Pero también

investigan la manera de poder determinar dichas peculiaridades semánticas a partir de reglas probabilísticas. Los modelos computacionales de semántica distribucional construyen representaciones semánticas de manera dinámica en forma de espacios vectoriales multidimensionales a través del análisis estadístico de los contextos en que cada palabra aparece (Mikolov et al. 2013). Dichos modelos, que requieren corpus de grandes dimensiones, crean vectores que puntúan el grado de composicionalidad de una secuencia dada (*compositional scores*). Se toman como criterios el grado de tipicidad y probabilidad de las relaciones semánticas entre los componentes de una UF mediante análisis semántico latente (*morder el polvo* vs. *morder la manzana*), o la asimetría en dichas relaciones, representada por un digrafo, que determina la direccionalidad de la interpretación semántica (*salir* etc. *con el rabo entre las piernas* → *salir* etc. ‘vencido, avergonzado y humillado’).

3.2 Productividad fraseológica

Directamente relacionado con la detección automática de la fraseología está el grado de fijación y productividad de las UF. Desde muy pronto se ha observado la tendencia de estas unidades a aparecer modificadas en el discurso (Corpas Pastor 1996). Sustituciones, adiciones, acortamientos, activación de doble lectura (literal y figurada), etc. son algunas de las maneras en las que se transforman las UF para obtener diversos efectos expresivos en el texto (cf. § 2.2). Este fenómeno se ha explicado tradicionalmente por la propia fijación de tales unidades, que las hacen especialmente susceptibles para este tipo de rupturas o disrupciones. La manipulación creativa de unidades en el discurso supone un acto consciente por parte del emisor, que consigue evocar tanto la unidad en sí como la propia interpretación creativa en el discurso. Así, en el ejemplo 12 se activan simultáneamente la UF canónica (*dar gato por liebre* ‘engañar a alg. en la calidad o valor de una cosa’) y la UF manipulada (*coger liebre por gato*, que conllevaría una interpretación aproximada en contexto de ‘quedarse con algo valioso haciendo creer que no lo es tanto’, o sea, algo así como ‘apropiación indebida’ con matices de engaño, dolo o estafa).

Ahora bien, la supuesta estabilidad y fijación de las UF, responsable en parte de esos usos creativos, presupone, por un lado, la baja o nula productividad de las UF respecto a los patrones y estructuras gramático-sintácticas de una lengua; y, por el otro lado, implica la existencia de una forma estable o «canónica». En este sentido, la lingüística de corpus ha demostrado que el análisis de grandes cantidades de texto cuestiona ambos aspectos. Si se analiza la UF anterior en el *esTenTen18*, se observa una gran diversidad de elementos léxicos que pueden actualizar la casilla correspondiente al verbo (V: *vender*, *dar*, *colar*, *meter*, *pasar*, *ser*, *ofrecer*, *comprar*). Véanse algunos ejemplos en la tabla 2.

Tabla 2: Concordancias para el nodo *gato por liebre* (*esTenTen18*)

son una basura total, Ford está vendiendo	gato por liebre	y eso lo podemos constatar en una pequeña
que proponía su adversario. En realidad fue	gato por liebre	, ya que ni bien llegó a la presidencia
grandes palabras para dar de la mejor manera	gato por liebre	y engatusar al personal, y otra lo que se aplica
quien sabe que durante mucho tiempo le pasaron	gato por liebre	.° °En este diálogo breve, Zoé Valdés explica
donde tienes serias sospechas de que te van a dar	gato por liebre	.° °Yo recomiendo que sobre todo tengan un sano
es que los poderes de dominación nos metieron	gato por liebre	también en eso, nos inventaron caudillos y nos
... o sea, me molesta que sea tan fácil pasar "	gato por liebre	" a nuestros pueblos...° °Bueno señoras y señores
a hablar de economía, pero no venga a ofrecer	gato por liebre	. Personalmente creemos que esto fue agrede, a
al CETA.° °"Votar por el CETA es comprar	gato por liebre	" recuerda la eurodiputada socialista belga
digo por si alguien se hace líos o intenta colar	gato por liebre	.° °Bromas aparte, está claro que el hecho de que

No todas las UF presentan el mismo grado de productividad ni de posibilidades «canónicas» (variantes fraseológicas), sino que más bien se sitúan en diversos puntos de una escala gradual de productividad y variabilidad. Por ejemplo, Corpas Pastor (2018) ofrece un análisis de la construcción $[V_{\text{reflex}} \text{ de risa}]$ en el corpus *esTenTen18* y los distintos subcorpus correspondientes a la variedad peninsular (.es) y a las siguientes variedades americanas: Argentina (.ar), Bolivia (.bo), Chile (.cl), Colombia (.co), Costa Rica (.cr), Cuba (.cu), República Dominicana (.do), Ecuador (.ec), El Salvador (.es), Guatemala (.gt), Honduras (.hn), México (.mx), Nicaragua (.ni), Panamá (.pa), Paraguay (.py), Perú (.pe), Uruguay (.uy) y Venezuela (.ve). Los resultados obtenidos apuntan a más de 70 formas verbales (*slot fillers*) posibles: $[V_{\text{reflex}} = \text{descojonarse, mearse, tirarse, troncharse, partirse, morirse, caerse, etc. + DE RISA}]$. De todas estas variantes, solo *morirse* se encuentra en los 19 subcorpus. El resto presentan una distribución desigual: desde *partirse* y *desternillarse* (presentes en 15 y 11 variedades respectivamente) a *pillarse* y *derrumbarse*, que indican Argentina y Perú, respectivamente. Además, distintas combinaciones de variantes (conglomerados) marcan variedades: por ejemplo, *cagarse de risa* es propio de 10 variedades, pero en combinación con *caerse de risa* solo se encuentra en 3 (Colombia, México y España).

La media de variantes está en 4,5 formas verbales distintas, aunque con diferencias diatópicas muy importantes, que no parecen correlacionarse con el tamaño de los distintos subcorpus. El porcentaje más alto de variantes se encuentran en los corpus más extensos: peninsular (50,74%), argentino (34,32%) y mexicano (26,85%). No obstante, el subcorpus peruano, que es al menos seis veces menor que los anteriores, presenta el 32,83% de variantes, por debajo solo del subcorpus argentino, mientras que la variedad chilena, con un subcorpus cuatro veces mayor que el peruano, solo presenta el 11,94% de variantes verbales totales. Estos datos indican que la productividad fraseológica es sensible a la diatopía, entre otros factores.

Por último, la variabilidad fraseológica que se descubre a través del análisis de corpus de grandes dimensiones no solo evidencia la existencia de productividad, sino también procesos de gramaticalización asociados. Véase, por ejemplo, el estudio de Corpas Pastor (2017) sobre [V_de miedo]. El análisis de corpus indica un proceso de deslexicalización del nodo, que pasa a adquirir valores de intensificación adverbial. El grado de variabilidad verbal es directamente proporcional al grado de gramaticalización y coerción semántica de esta construcción. Ello abre las posibilidades combinatorias con verbos con propiedades distributivas diferentes y, por ende, al cambio semántico y a la polisemia del nodo.

4 Conclusiones

A lo largo de este trabajo hemos incidido en la importancia de la perspectiva textual para el estudio de la fraseología. Ligada conceptualmente a la noción coseriana de «discurso repetido», esta disciplina lingüística tuvo que esperar hasta bien adentrada la década de los 90 para empezar a ser analizada en las distintas facetas y géneros discursivos. En la actualidad, el análisis estadístico de corpus puede descubrir las interesantes posibilidades combinatorias y de fijación de los lexemas, su atracción mutua, pero también la tendencia intrínseca de lo fijo a ser modificado, a coaparecer con colocados inesperados, aumentando así su productividad, y a ser objeto de manipulaciones creativas por parte del hablante. Todo ello no hace más que corroborar la interrelación y tensión dialéctica entre el principio idiomático y el del «open choice» en el momento del habla (Sinclair 1991). En este sentido, la perspectiva textual ligada al concepto del discurso repetido en el ámbito de la fraseología hace patente las dimensiones de lo fraseológico en el hablar y las variaciones que el discurso «repetido» puede ofrecer. Justamente esta integración de perspectivas sería la contribución de la fraseología al desarrollo de una lingüística de texto integral y por lo tanto de una lingüística del hablar holística. Y es en este reto donde la lingüística de corpus emerge como recurso inagotable para descubrir nuevas características y vías de estudio.

5 Bibliografía

- Berty, Katrin/Mellado Blanco, Carmen/Olza, Inés (edd.) (2018), *Fraseología y variedades diatópicas*, Pamplona, Eunsa, Ediciones Universidad De Navarra.
- Corpas Pastor, Gloria (1996), *Manual de fraseología española*, Madrid, Gredos.
- Corpas Pastor, Gloria (2013), *Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas*, in: Inés Olza/Elvira Manero Richard (edd.) *Fraseopragmática*, Berlin, Frank & Timme, 335–374.
- Corpas Pastor, Gloria (ed.) (2016), *Computerised and Corpus-based Approaches to Phraseology. Monolingual and Multilingual Perspectives (Full papers) – Fraseología computacional y basada en corpus. Perspectivas monolingües y multilingües (Trabajos completos)*, Genève, Tradulex.
- Corpas Pastor, Gloria (2017), *Collocational Constructions in Translated Spanish. What Corpora Reveal*, in: Ruslan Mitkov (ed.), *Computational and Corpus-Based Phraseology. Second International Conference, Europhras 2017, London, UK, November 13–14, 2017. Proceedings*, Berlin, Springer, 29–40.
- Corpas Pastor, Gloria (2018), *Laughing one's head off in Spanish subtitles: a corpus-based study on diatopic variation and its consequences for translation*, in: Pedro Mogorrón Huerta/Antonio Albaladejo-Martínez (edd.), *Fraseología, Diatopía y Traducción*, Amsterdam/Philadelphia, Benjamins, 54–106.
- Corpas Pastor, Gloria/Colson, Jean-Pierre (edd.) (2020), *Computational Phraseology*, Amsterdam/Philadelphia, Benjamins.
- Corpas Pastor, Gloria/Mena Martínez, Flor (2003), *Aproximación a la variación fraseológica de las lenguas alemana, inglesa y española*, ELUA. Estudios de Lingüística de la Universidad de Alicante 17, 181–201.
- Corpas Pastor, Gloria/Mitkov, Ruslan (edd.) (2019), *Computational and Corpus-Based Phraseology. Third International Conference, Europhras 2019, Malaga, Spain, September 25–27, 2019. Proceedings*, Berlin, Springer.
- Coseriu, Eugenio (1955), *Determinación y entorno. Dos problemas de una lingüística del hablar*, *Romanistisches Jahrbuch* 7, 29–54. Re-publicado en: Coseriu, Eugenio (1962), *Teoría del lenguaje y lingüística general. Cinco estudios*, Madrid, Gredos.
- Coseriu, Eugenio (1967), *Structure lexicale et enseignement du vocabulaire*, in: *Les théories linguistiques et leurs applications*, Strasbourg, Conseil de l'Europe, 9–51.
- Coseriu, Eugenio (1973, ¹1970), *Einführung in die strukturelle Betrachtung des Wortschatzes*, aus dem Franz. übers. von Erich Braun, in: Eugenio Coseriu, *Einführung in die strukturelle Betrachtung des Wortschatzes*, Tübingen, Narr, 1–104.
- Coseriu, Eugenio (1976), *Das romanische Verbalsystem*, hrsg. und bearb. von Hansbert Bertsch, Tübingen, Narr.
- Coseriu, Eugenio (1981a, ¹1977), *Lecciones de lingüística general*, traducción del italiano por J. M^a. Azáceta y García de Albéniz, revisada y ampliada por el autor, con un prólogo del autor, Madrid, Gredos.
- Coseriu, Eugenio (1981b, ¹1977), *Principios de semántica estructural*, traducción del alemán y del francés por M. Martínez Hernández, con un prólogo del autor, Madrid, Gredos.
- Coseriu, Eugenio (2007), *Lingüística del texto. Introducción a la hermenéutica del sentido*, ed., anotación y estudio previo de Óscar Loureda Lamas, Madrid, Arco/Libros.
- Coseriu, Eugenio/Loureda Lamas, Óscar (2006), *Lenguaje y discurso*, pról. de Johannes Kabatek, Pamplona, Eunsa.
- Dobrovolskij, Dmitrij (1987), *Textbildende Potenzen von Phraseologismen*, *Linguistische Studien A* 164, 69–85.

- Dobrovol'skij, Dmitrij (1999), *Haben transformationelle Defekte der Idiomstruktur semantische Ursachen?*, in: Nicole Fernández Bravo et al. (edd.), *Phraseme und typisierte Rede*, Tübingen, Stauffenburg, 25–37.
- Dobrovol'skij, Dmitrij (2009), *La variación léxico-sintáctica en la fraseología: la introducción del atributo en la estructura de la locución idiomática*, *Language Design* 1, 29–65.
- Elspass, Stefan (2002), «*Alles unter den Tisch gekehrt*» – *Phraseologische Gebrauchsauffälligkeiten im Urteil von Sprachbenutzern*, in: Dietrich Hartmann/Jan Wierer (edd.), *Wer A sagt, muss auch B sagen*, Hohengehren, Schneider, 127–160.
- Ettinger, Stefan (1977), «*Wiederholte Rede*» und *Bibelsprache. Bemerkungen zur deutsch-französischen Übersetzung biblischer Zitate*, *Linguistica Biblica* 40, 1–20.
- Fleischer, Wolfgang (1997), *Phraseologie der deutschen Gegenwartssprache*, Tübingen, Niemeyer.
- Fortes, Marie/Merceur, Marie Christine (2003), *Expressions idiomatiques français-espagnol*, Barcelona, Espasa Libros.
- García-Medall, Joaquín (ed.) (2006), *Fraseología e ironía. Descripción y contraste*, Lugo, Axac.
- García-Page Sánchez, Mario (2008), *Introducción a la fraseología española: estudio de las locuciones*, Madrid, Anthropos.
- González-Rey, María Isabel (2016), *Une approche analogique à la compétence phraséologique: une double compétence, intégrale et intégrée*, in: Antonio Pamies/Philippe Monneret/Salah Mejri (edd.), *Analogie, figement et polysémie*, *Language Design, Special Issue*, 165–188.
- González Sanz, Marina (2017), *El discurso repetido en la tertulia periodística de tema político*, in: Carmen Mellado Blanco/Katrin Berty/Inés Olza (edd.), *Discurso repetido y fraseología textual (español y español-alemán)*, Madrid, Iberoamericana, 167–187.
- Häcki Buhofer, Annelies/Burger, Harald (1992), *Gehören Redewendungen zum heutigen Deutsch?*, *Fremdsprachen Lehren und Lernen* 21, 11–32.
- Hausmann, Franz Josef (1974), *Studien zu einer Linguistik des Wortspiels. Das Wortspiel im «Canard enchaîné»*, Tübingen, Niemeyer.
- Kay, Paul/Fillmore, Charles J. (1999), *Grammatical constructions and linguistic generalizations: the «What's X doing Y?» construction*, *Language* 75, 1–33.
- Keizer, Evelien (2018), *Idiomatic expressions and anaphoric reference in functional discourse grammar*, *Quaderns de filologia, Estudis lingüístics* 23, 177–195.
- Kilgarriff, Adam, et al. (2014), *The Sketch Engine: ten years on*, *Lexicography: Journal of ASIALEX* 1/1, 7–36.
- Kolesnikova, Olga (2016), *Survey of Word Co-occurrence Measures for Collocation Detection*, *Computación y Sistemas* 20/3. <http://dx.doi.org/10.13053/cys-20-3-2456>.
- Loureda Lamas, Óscar (2007), *Presentación del editor. La Textlinguistik de Eugenio Coseriu*, in: Eugenio Coseriu (2007), *Lingüística del texto. Introducción a la hermenéutica del sentido*, ed., anotación y estudio previo de Óscar Loureda Lamas, Madrid, Arco/Libros, 19–74.
- Loureda Lamas, Óscar (2009), *Qué 30 años no es nada... tradición y modernidad de la lingüística del texto de Eugenio Coseriu*, in: María Azucena Penas Ibáñez/Rosario González (edd.), *Estudios sobre el texto, Nuevos enfoques y propuestas*, Frankfurt a. M. et al., Lang, 65–88.
- Manero Richard, Elvira (2017), *Relaciones co(n)textuales de las unidades fraseológicas en el discurso literario. El capitán Alatriste*, in: Carmen Mellado Blanco/Katrin Berty/Inés Olza (edd.), *Discurso repetido y fraseología textual (español y español-alemán)*, Madrid, Iberoamericana, 189–208.
- Martí Sánchez, Manuel (2015), *La búsqueda de sentido en la desautomatización fraseológica*, in: Pedro Mogorrrón/Fernando A. Navarro (edd.), *Fraseología, Didáctica y Traducción*, Frankfurt a. M. et al., Lang, 117–135.
- Martí Solano, Ramón (2011), *Phraseological Units in English: variation through lexical insertion*, *Rassegna Italiana di Linguistica Applicata* 43/1–2, 389–404.
- Martínez Marín, Juan (1996), *Estudios de fraseología española*, Málaga, Librería Ágora.
- Mellado Blanco, Carmen (ed.) (2008), *Beiträge zur Phraseologie aus textueller Sicht*, Hamburg, Kovač.
- Mellado Blanco, Carmen (2016), «*En boca cerrada no entran moscas*»: *significado social, rasgos pragmáticos y desautomatización de las paremias de CALLAR en alemán y español*, *Language Design* 18, 107–150.
- Mellado Blanco, Carmen (2019), *Phrasem-Konstruktionen kontrastiv Deutsch-Spanisch: ein korpus-basiertes Beschreibungsmodell anhand ironischer Vergleiche*, *YOP. Yearbook of Phraseology* 10, 65–88.
- Mellado Blanco, Carmen (2020a), *La desautomatización desde el prisma de la Gramática de Construcciones: un nuevo paradigma de la variabilidad fraseológica*, *Nasledje* 45, 17–34.
- Mellado Blanco, Carmen (2020b), «*(No) me importa un comino*» y sus variantes diatópicas. *Estudio de corpus de una construcción fraseológica desde la Gramática de Construcciones* in: Pedro Mogorrrón Huerta/Ana Cuadrado Rey (edd.), *ELUA: Fraseología y variaciones (socio)lingüísticas y diatópicas*, Anexo VII, 87–109. <https://doi.org/10.14198/ELUA2020.ANEXO7.06>.
- Mellado Blanco, Carmen (en prensa), *Esquemas fraseológicos y construcciones fraseológicas en el continuo léxico-gramática*, in: Encarnación Tabares et al. (edd.), *Clases y categorías en la fraseología de la lengua española*, Frankfurt a. M., Lang.
- Mellado Blanco, Carmen/Berty, Katrin/Olza, Inés (edd.) (2017), *Discurso repetido y fraseología textual (español y español-alemán)*, Madrid, Iberoamericana.
- Mena Martínez, Flor (2003), *En torno al concepto de desautomatización fraseológica: aspectos básicos*, *Tonos digital* 5, 1–12. <https://digitum.um.es/xmlui/bitstream/10201/50794/1/En%20torno%20al%20concepto.pdf> (consultado el 30/12/2019).
- Mena Martínez, Flor (2004), *Los efectos semánticos producidos por la desautomatización de las unidades fraseológicas*, in: Ramón Almela Pérez et al. (edd.), *Homenaje al profesor Estanislao Ramón Trives*, Murcia, Universidad de Murcia, 501–518.
- Mikolov, Tomas, et al. (2013), *Efficient estimation of word representations in vector space*, *Computing Research Repository*, abs/1301.3781.
- Mitkov, Ruslan, et al. (edd.) (2018), *Multiword Units in Machine Translation and Translation Technology*, Amsterdam/Philadelphia, Benjamins.
- Monti, Joanna, et al. (2018), *Multiword Units in Machine Translation and Translation Technology*, in: Ruslan Mitkov et al. (edd.), *Multiword Units in Machine Translation and Translation Technology*, Amsterdam/Philadelphia, Benjamins, 2–37.
- Montoro del Arco, Esteban T. (2005), *Hacia una sistematización de la variabilidad fraseológica*, in: Miguel Á. Pastor Millán (ed.), *Estudios lingüísticos en recuerdo de Juan Martínez Marín*, Granada, Universidad de Granada, 129–152.
- Olza, Inés/Losada, María Carmen (2011), *Apoyos co(n)textuales en el empleo discursivo de las unidades fraseológicas*, *Signos* 44/76, 132–144.
- Olza, Inés/Manero Richard, Elvira (edd.) (2013), *Fraseopragmática*, Berlin, Frank&Timme.
- Pamies Bertrán, Antonio (2017), *Fraseología y variación diatópica en español*, *Verba Hispanica* 25, 55–81.
- Pawley, Andrew/Syder, Frances (1983), *Two puzzles for linguistic theory. Nativelike selection and nativelike fluency*, in: Jack C. Richards/Richard W. Schmidt (edd.), *Language and Communication*, London, Longman, 191–226.
- Penadés Martínez, Inmaculada (2012), *Gramática y semántica de las locuciones*, Alcalá de Henares, Servicio de Publicaciones de la Universidad de Alcalá.
- Penadés Martínez, Inmaculada (2014), *Fixación estructural e desautomatización das locuções*, *CADERNOS DE FRASEOLOGIA* 16, 273–301.
- Ramisch, Carlos (2015), *Multiword Expressions Acquisition. A Generic and Open Framework*, Berlin, Springer.

- Ruiz Gurillo, Leonor (1997), *Aspectos de fraseología teórica española*, València, Univ. de València.
- Ruiz Gurillo, Leonor (1998), *La fraseología en el español coloquial*, Barcelona, Ariel.
- Sag, Ivan A., et al. (2002), *Multiword expressions. A pain in the neck for NLP*, in: Alexander Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing. CICLing 2002. Lecture Notes in Computer Science*, vol. 2276, Berlin, Springer, 1–15.
- Seretan, Violeta (2011), *Syntax-Based Collocation Extraction*, Dordrecht, Springer Netherlands.
- Sinclair, John (1991), *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.
- Stumpf, Sören (2016), *Modifikation oder Modellbildung? Das ist hier die Frage – Abgrenzungsschwierigkeiten zwischen modifizierten und modellartigen Phrasemen am Beispiel formelhafter (Ir-)Regularitäten*, *Linguistische Berichte* 247, 317–342.
- Thun, Harald (1978), *Probleme der Phraseologie. Untersuchungen zur wiederholten Rede mit Beispielen aus dem Französischen, Italienischen, Spanischen und Rumänischen*, Tübingen, Niemeyer.
- Timofeeva, Larissa (2009), *La desautomatización fraseológica: un recurso para crear y divertir*, *ELUA. Estudios de lingüística Universidad de Alicante* 3, 249–270.
- Zamora Muñoz, Pablo (2014), *Los límites del discurso repetido: la fraseología periférica y las unidades fraseológicas pragmáticas*, *Verba* 41, 213–236.
- Zuluaga, Alberto (1975), *La fijación fraseológica*, *Boletín del Instituto Caro y Cuervo* 30, 225–248.
- Zuluaga, Alberto (1980), *Introducción al estudio de las expresiones fijas*, Frankfurt a.M. et al., Lang.