



ENTRE EL BINOMIO FRASEOLÓGICO, LA COLOCACIÓN COMPLEJA Y LA COLOSTRUCCIÓN: UN EXAMEN «LARGO Y TENDIDO»

1 INTRODUCCIÓN

La existencia de una transición gradual entre la sintaxis y el léxico (Langacker 1987; Kay/Michaelis 2012; Michaelis 2019) no cuestiona la existencia de los prototipos más extremos, tanto de las combinaciones «libres» (*open choice principle*) como de la fijación fraseológica (*idiom principle*), solo pone en duda que dichas figuras se autoexcluyan de manera discreta. Es obvio que *comer patatas* o *beber agua* son sintagmas verbales, mientras que «marear la perdiz y pelar la pava son locuciones, porque, en ellas, la relación predicado-argumento no se actualiza cada vez que un hablante emplea dichas expresiones, sino que se encuentra petrificada» (Osorio Olave 2012: 113). En cambio, la ubicación exacta de otras secuencias multilexicales en dicha transición es más borrosa, no sólo para las colocaciones, sino para los *phrasal verbs*, las comparaciones estereotipadas y los esquemas fraseológicos (cf. Dobrovól'skij 2007; Pamies 2019; Mellado 2020; Pamies/Wang 2020). También provocan discrepancias las secuencias internamente fijas pero variables en la conexión con su contexto discursivo inmediato. Por ejemplo, el *Diccionario combinatorio del español contemporáneo* de Ignacio Bosque (2001) estipula que *a bombo y platillo* es una locución adverbial, pero añade que, en el discurso, sólo se combina con una cuarentena de verbos, que pertenecen a determinadas *clases léxicas* (2001: 26). En el caso de *a bombo y platillo* su colocabilidad exigiría:

- a) verbos que denotan difusión o transmisión de información (p.ej., *proclamar, anunciar, señalar, remachar*);
- b) verbos que designan la puesta en marcha de algo (*organizar, estrenar*);
- c) verbos que denotan elogio o conmemoración de algo (*celebrar, aclamar...*);
- d) verbos que designan transacciones económicas (*contratar, venderse*) ([REDES] 2004).

Cabe entonces preguntarse si *anunciar a bombo y platillo* es una unidad fraseológica o si, por el contrario, es una combinación libre entre un frasema adverbial y un verbo totalmente ajeno a la expresión (cf. Corpas 1996: 119). En cambio, el mismo

diccionario sólo cita tres verbos para *a tocateja*, que son *pagar*, *comprar* y *abonar*. Esta secuencia no presenta diferencias gramaticales «internas» con *a bombo* y *platillo* pero parece sufrir mayores restricciones con respecto a su entorno sintagmático. Si las palabras pueden sufrir fijación con sus vecinas formando frasemas, este fenómeno también puede afectar al frasema en su conjunto, y su inserción en su contorno puede a su vez sufrir una fijación secundaria. Es el caso de lo que Koike (2001, 2012) y García-Page (2011) llaman *colocación compleja*, y que nos devuelve a la eterna pregunta de ¿dónde empieza la fraseología? (cf. Colson 2015). Hablar de *colocación* implica, por definición, manejar datos cuantitativos (Firth 1957; Sinclair et al. 1970; Alonso Ramos 1994-1995), por ello, para «pescar» colocaciones en un corpus, se suelen emplear algoritmos estadísticos basados en el análisis frecuencial (Halliday 1961: 276). El presente trabajo intenta invertir el mecanismo: partimos de una expresión previamente etiquetada como frasema, la locución adverbial *largo y tendido*, y estudiamos su comportamiento frecuencial y sintáctico-semántico en un corpus textual de grandes dimensiones. De esta manera, los datos lingüísticos están en el punto de partida y los comportamientos frecuenciales son la incógnita que debería confirmarlos.

1.1 Precisiones terminológicas

Los fraseólogos llaman *binomios* a las combinaciones fijas que están formadas por dos palabras (y sus eventuales clíticos), sean yuxtapuestas o coordinadas. Su sentido literal suele ser redundante (*nunca jamás*; *al fin y al cabo*; *sin causa ni razón*, *dueño y señor*) o antitético (*a diestra y siniestra*; *día y noche*; *más o menos*, *de quita y pon*, *tarde o temprano*) (cf. García-Page 1998, 2008: 329-333). Dichas unidades son transversales a varias subclases fraseológicas (Čermák 2007: 414-429), y su fijación las hace internamente irreversibles; prueba de ello es que su traducción a otra lengua, aunque tenga los mismos componentes, puede a veces colocarlos en orden inverso (p.ej. el binomio ing. *black and white* y fr. *noir et blanc* se traduce por esp. *blanco y negro*).

Siguiendo a Corpas, el *binomio* se puede considerar como una subclase de la locución (1996: 96-112), ya sea nominal: *flor y nata*; *santo y seña*; *el oro y el moro* (p. 96), adjetiva: *mondo y lirondo*; *sano y salvo*; *de carne y hueso* (1996: 98), verbal: *ir y venir*; *coser y cantar*; *nadar y guardar la ropa* (1996: 102) o adverbial *cara a cara*; *poco a poco* (1996: 111), categoría en la cual precisamente incluye esta autora *largo y tendido* (1996: 102). Sin embargo, dada su elevada coocurrencia con *hablar*, se podría sospechar que *largo y tendido* fuese, a su vez, componente de una unidad fraseológica mayor. De hecho, la propia autora advierte que, cuando una locución coocurre a menudo con una misma palabra *es difícil decidir si ésta es un elemento de la locución* (1996: 119), citando ejemplos como *cumplir a rajatabla* y *dinero contante y sonante*¹.

Los pioneros de la lingüística computacional llamaron *collocations* a todas las coocurrencias léxicas cuya frecuencia estadística fuera superior a su cálculo normal de probabilidades (Firth 1957; Sinclair 1970, 1991). Sin embargo, tal definición parece

1 Véase también al respecto Írsula Peña (1994: 27-34), Corpas (2001: 91) y Bustos Plaza (2005).

demasiado vaga, porque no establece límites sistémicos con todas las combinaciones libres frecuentes ([REDES] 2001: XXII), ni tampoco con respecto a frasesemas de otras subclases. El término inglés *collocation*, para esta escuela, era prácticamente un hiperónimo equivalente a lo que hoy llamaríamos *frasesema*. Otros enfoques, especialmente en la Europa continental, defienden una visión más restrictiva de la colocación, entendida como una subclase fraseológica bien delimitada por su forma y su función:

[...] unidades fraseológicas formadas por dos unidades léxicas en relación sintáctica, que no constituyen, por sí mismas, actos de habla ni enunciados; y que, debido a su fijación en la norma, presentan restricciones de combinación establecidas por el uso, generalmente de base semántica: el colocado autónomo semánticamente (la base) no sólo determina la elección del colocativo, sino que, además, selecciona en éste una acepción especial, frecuentemente de carácter abstracto o figurativo. (Corpas Pastor 1996: 66);

[...] *word combinations in which some substitution is possible, but with some arbitrary limitations on substitution; in which at least one element has a non-literal meaning, and at least one element is used in its literal sense; and the whole combination is transparent* (Gouverneur 2008: 232).

Muchos especialistas prefieren hoy usar *multi-word expressions* (MWE) como hiperónimo fraseológico general (p. ej., Colson 2019), dejando a *collocation* este sentido más restringido (cf. Mel'čuk 1998, 2003, 2013; Hausmann 1979, 1998; Alonso Ramos 1994-95; Čermák 2007; Grossmann/Tutin 2003; Pamies 2017)², incluso criticando explícitamente el criterio frecuencial (p. ej. Bosque 2001), aunque quienes mantienen una postura más ortodoxamente sinclairiana siguen empleando *collocation* en su sentido amplio (cf. Alonso/Williams/DeCesaris 2019).

1.2 Largo y tendido

La expresión *largo y tendido* no figuraba en los primeros diccionarios generales «clásicos» ([COV] 1611, [AUT] 1726-1739, [TERR] 1786), aunque sí lo hacía en una de las obras pioneras de la fraseografía hispana, el *Vocabulario de frases y refranes* de Gonzalo Correas ([COR] 1627), que, aunque no la definía, citaba ejemplos de uso. La lexicografía general, que la define hoy como ‘extensamente’ y ‘sin prisa’, la vincula a un contexto verbal restringido: *hablar/escribir*~ (DLE 2021)³. La fraseografía moderna la define vinculándola siempre a unos contextos específicos. P. ej., ‘con profusión’ [*hablar* ~] ([DEC] 1997), ‘extensamente’ [*hablar* o *escribir* ~] ([DFDL] 2004), ‘durante mucho tiempo’ [*escribir, explicarse, playarse, hablar* ~] ([DFE] 2007). Dado que

2 La teoría de las *funciones léxicas* emplea criterios cualitativos, al afirmar que la colocación une elementos con funciones distintas: la *base* (generalmente literal), selecciona el *colocado* (generalmente metafórico), que le aporta, de manera sistemática, una modificación semántica específica (Mel'čuk 1998, 2003, 2013). Para el español, véase Alonso Ramos (2002).

3 [DLE] <https://dle.rae.es/largo?m=form> (consulta 06/02/2021).

estos diccionarios no siempre describen con precisión los contextos, es de suponer que conceden a estos casos particulares una representatividad elevada. Sin embargo, existe bastante variación de un diccionario a otro en la selección de dichos verbos, los que citaba Correas eran (curiosamente) *mentir, dormir, pasear, jugar y comer* ([COR] 1627).

Como se ha dicho, algunos fraseólogos consideran que una locución adverbial que se combina asiduamente con un mismo verbo (*llorar a moco tendido / llorar a lágrima viva*) constituye una subcategoría especial, que Koike (2001, 2012) y García-Page (2011) llaman *colocación compleja*. Este fenómeno también se podría considerar como un hecho (micro)sintáctico: la *imbricación* de un fraseologismo dentro de otro (Pamies 2013, 2017)⁴. Son muchas las locuciones adverbiales que se integran dentro de una colocación, dando lugar a una unidad de segundo nivel, donde hacen de colocativo y aportan diversas funciones léxicas a su verbo base: <reír> *a mandíbula batiente*, <dormir> *a pierna suelta*, <saber> *de buena tinta* (cf. Blanco 2001: 175). Es más, en principio, nada parece impedir que cualquier locución adverbial pueda imbricarse dentro de un frasema mayor, no sólo con un verbo sino incluso con otro frasema, solapándose o interrumpiéndolo, como se ha comprobado en muchos ejemplos de nuestro corpus (Fig.1).

- ...yo podría hablar largo y tendido de este tema
- ...la nueva controversia que traba el Senado da para largo y tendido
- ...los estatutos del mismo tendrán que ser aprobados por todos los municipios, lo que *va para largo y tendido*
- ...le dio largo y tendido a la cháchara...

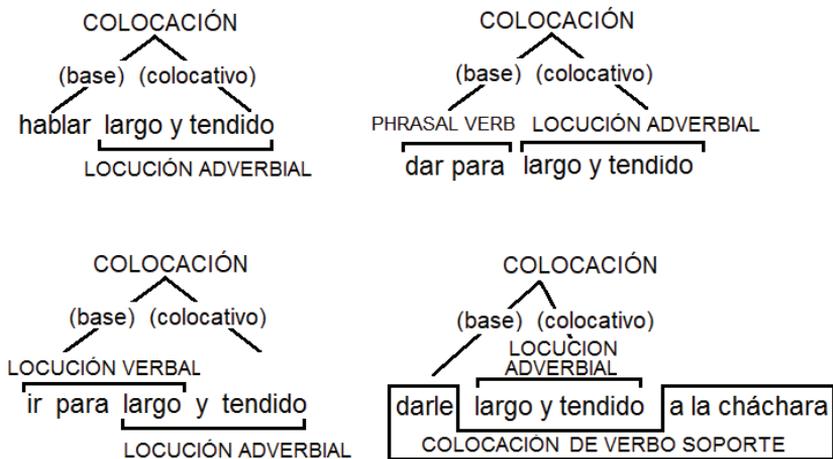


Figura 1. Imbricación entre frasemas

4 P. ej. una locución dentro de una paremia.

En estos casos, la expresión es doblemente idiomática: «internamente» y «externamente». Para distinguir una colocación compleja de la combinatoria libre que incluye una colocación, Xavier Blanco propone el criterio siguiente: la locución *a mares* formaría una colocación compleja con *llover*, *sudar* y *llorar*, porque sólo modifica estos verbos, a diferencia de *a mansalva*, que modifica muchos verbos (2001: 175). Para estudiar en detalle esta cuestión, hemos realizado una serie de búsquedas en el corpus *Spanish Web (EsTenTen11_Eu+Am) 2011* [ETT], que contiene casi 9500 millones de palabras (cf. Jakubiček *et al.* 2013, Kilgarriff/Renau 2013), accesible a través de la plataforma *Sketchengine* ([SKE] 2021)⁵.

2 EXPERIMENTO Y RESULTADOS

2.1 V + Loc._{Adv}

Mediante la opción *Concordancia > CQL*, obtenemos un listado inicial de 3542 ocurrencias de *largo* y *tendido*⁶, que sometemos a un examen «manual» para observar su contexto inmediato, sin ningún filtro previo que pudiera sesgar los resultados. Para estudiar la relación entre este colocativo y su base, descartamos de este *output* las secuencias cuya frecuencia de coocurrencia es inferior a 5, que podrían ser casuales o irrelevantes, en la medida en que, en un corpus tan inmenso, incumplirían por ello una característica estadística básica para cualquier tipo de frasema.

Al examinar los verbos, salta a la vista el predominio de la asociación con *hablar*, que, en posición antepuesta y continua, aparece en 1578 coocurrencias de la locución⁷ (*del tema de las reservas de los restaurantes hemos hablado* largo y tendido *en los últimos meses*).

Pero también aparecen otros contextos verbales que justifican un análisis cuantitativo más detallado. En primer lugar, abundan los sinónimos de *hablar* asociados a esta locución (con 556 coocurrencias por encima del umbral):

Tabla 1. Sinonimia verbal recurrente con la locución

hablar	1578
charlar	218
conversar	175
platicar	163
TOTAL	2134

5 Aunque el colocativo es seleccionado por la base, para la práctica lexicográfica resulta más eficaz partir del colocativo (Bosque 2019: 19-20), sobre todo cuando éste es a su vez un frasema en sí mismo.

6 Concordancia CQL [word="largo"] [word="y"] [word="tendido"].

7 Concordancia CQL [lemma="hablar"] [word="largo"] [word="y"] [word="tendido"].

El binomio *largo y tendido* también coocurre con hipónimos de *hablar*, que pertenecen igualmente a la clase semántica de los «verbos discursivos», o *verba dicenda*, que designan actos comunicativos (incluido *escribir*). De hecho, el diccionario combinatorio de Ignacio Bosque, que apenas trata esta expresión, observa aun así que, además de *hablar*, el binomio coocurre también con *otros verbos de lengua* [REDES 2004: 1253]. Estos hipónimos aportan otros 470 casos que superan el filtro frecuencial $f \geq 5$.

Tabla 2. Hiponimia verbal recurrente con la locución

discutir	126	chacharear	9	despachar	7
debatir	90	analizar	9	entrevistar	6
escribir	68	discurrir	5	responder	6
dialogar	51	tratar	7	disertar	5
comentar	25	extenderse	7	exponer	5
explicar	21	contar	7		
departir	9	opinar	7	TOTAL	470

A ello deben sumarse las formas discontinuas con el verbo separado o pospuesto con respecto al binomio. Lanzamos otras búsquedas *ad hoc* para estos verbos y encontramos 5 formas invertidas (todas con *hablar*) y 282 discontinuadas para el total de verbos que satisfacen el umbral $f \geq 5$ y con una ventana de 1 a 4 palabras⁸ (Tablas 3 a 5).

Tabla 3. Coocurrencias invertidas y discontinuas con *hablar*

LOC + V <i>hablar</i>	5
V <i>hablar</i> <X> LOC	188
TOTAL	193

Tabla 4. Coocurrencias discontinuas con sinónimos de *hablar*

conversar	22
charlar	14
platicar	12
TOTAL	48

8 [lemma="hablar"] []{1,4} [word="largo"] [word="y"] [word="tendido"], etc.

Tabla 5. Coocurrencias discontinuas con hipónimos de *hablar*

discutir	10
debatir	11
tratar	7
escribir	7
explicar	6
dialogar	5
TOTAL	46

...del que otro día más largo y tendido **hablaremos**
 ...**hablaremos** del aborto largo y tendido otro día
 ...**charlamos** sobre el tema, largo y tendido
 ...quienes los vieron **platicar** dicen que lo hicieron largo y tendido

Sumando todos los verbos discursivos, independientemente de su posición con respecto al binomio, se obtienen 2939 ítems (tabla 6).

Tabla 6. Sumas de variantes continuas, invertidas y discontinuas

	CONTIN	INVERT	DISCONTIN	TOTALES
con <i>hablar</i>	1578	5	188	1771
con sinónimos	604	0	48	652
con hipónimos	470	0	46	516
TOTALES	2652	5	282	2939

Aparte de este grupo, se observa otro formado por verbos «de pensamiento» (*pensar* o hipónimos: *reflexionar*, *meditar*, *especular*, *considerar*, *teorizar*, *filosofar*...), en 73 co-ocurrencias con *largo y tendido* (Tabla 7).

Tabla 7. Coocurrencias con verbos de pensamiento

V ^{PENSAM.} + LOC _{Adv} ^{largo y tendido}	f _{≥5}
<i>pensar</i>	40
<i>reflexionar</i>	16
<i>meditar</i>	12
<i>saber</i>	5
TOTAL	73

...me he permitido el lujo de **pensar** largo y tendido en ambos temas.
...recuerdo que lo **meditaste** largo y tendido y, de repente, te lanzaste

Merece comentario aparte el caso del verbo *saber*, pues cuando se asocia a *largo y tendido* ya no es para significar «durante mucho tiempo» sino simplemente⁹ «mucho»:

...como un zorro viejo que de trabajo táctico **sabe** largo y tendido¹⁰

Además de estas dos clases semánticas (verbos discursivos y verbos de pensamiento), aparecen otros verbos, semánticamente ecléticos y con frecuencias casi siempre muy bajas, formando en principio combinaciones libres, en las que el hablante selecciona el verbo por un lado y la locución adverbial por otro, y las junta de manera imprevisible: *trabajar, laburar, descansar, sentarse, esperar, bostezar, dormir, viajar, caminar, avanzar, seguir, proseguir, detenerse, entretenerse, escalar, caer, sonar, escuchar, atender, ocuparse, participar, probar, compartir, mostrar, ver, mirar, buscar, investigar, hurgar, leer, batallar, pelear, batir, ganar, cortar, gastar, rugir, putear, gobernar, ejercer (el poder), llorar, suspirar, reír, aplaudir, abrazarse, besar, disfrutar, jugar, bailar, silbar, tocar, chupar, festejar, almorzar, desayunar, cocinar, matear, dar para, ser, estar, haber, coexistir, moquear*¹¹:

...se cubrió el rostro y **lloró** largo y tendido
...el jugador balcánico **batalló** largo y tendido para eliminar al checo
...hay administraciones que en la práctica **han gobernado** largo y tendido; y ni así aprenden
...los Python de hecho se **rieron** largo y tendido de mendigos y viejos y enfermos y homosexuales.

Estos contextos suman 97 coocurrencias que alcanzan el umbral $f \geq 5$ (Tabla 8), muchos otros verbos son en cambio descartados porque sólo coocurren una vez con *largo y tendido*.

9 Acerca de la polisemia fraseológica en general, véase Mellado (2014).

10 <http://www.futbolprimera.es> [21/11/2008] (*apud.* [ETT]).

11 Recuérdese que Correas (1627) citaba también *dormir* y *mentir* (cf. *supra*). Éste último, aun designando un acto de habla, no aparece ni una sola vez en el corpus moderno, pese a las enormes dimensiones del mismo.

Tabla 8. Coocurrencias con otros verbos

CON OTROS VERBOS (f≥5)					
festejar	10	disfrutar	7	mirar	6
dar para	9	suspirar	7	seguir	6
leer	8	besar	6	batallar	5
dormir	8	llorar	6	esperar	5
trabajar	8	escuchar	6	TOTAL	97

Si sumamos todas las coocurrencias entre un verbo y la locución *largo y tendido* que igualen o superen el umbral, la cifra obtenida es de 3109 (Tabla 9).

Tabla 9. Coocurrencias entre un verbo y la locución por encima del umbral

V + LOC ^{largo y tendido}	f ≥ 5
con verbos discursivos	2939
con verbos de pensamiento	73
con otros verbos	97
TOTAL	3109

2.2 N + LOC_{Adj}

Al examinar los usos restantes, nos encontramos con un fenómeno curioso que no se menciona en los diccionarios (quizás por considerarse como preceptivamente *incorrecto*) pero cuya productividad se manifiesta de manera indiscutible en el corpus: *largo y tendido* puede aparecer con función adjetiva y con el significado de «muy largo» (en sentido temporal o espacial), con sustantivos diversos. Hicimos una búsqueda automática, obteniendo 161 resultados, aunque, al examinarlos uno por uno, se comprueba que muchos de ellos eran falsos positivos, porque el sustantivo era simplemente un complemento verbal intercalado entre el verbo y la locución adverbial, dentro una variante discontinua¹². Eliminamos manualmente estos casos (que ya habían sido contabilizados antes), reduciéndose la lista a 52 ejemplos, donde *largo y tendido* modifica un sustantivo en vez de un verbo:

...después de **un almuerzo** largo y tendido, había que bajar lo comido con un sabroso y oscuro café turco¹³...

...os recomiendo **un vistazo** largo y tendido a su página web¹⁴

...Haití, destruido por **un terremoto** largo y tendido¹⁵...

12 P. ej. *nos habló <en Tenerife> largo y tendido de todo ello* (www.coar.es) (apud. [ETT]).

13 <http://www.labuenavida.cl> (apud. [ETT]).

14 <http://fcom.us.es> (apud. [ETT]).

15 <http://www.conaleche.gob.do> (apud. [ETT]).

Dada la existencia de esta transcategorización en locución adjetiva, hemos supuesto que, por la misma razón, también debía haber formas flexionadas en género y/o número. Hicimos una búsqueda a tal efecto y aparecieron 18 variantes en plural, 83 en femenino singular y 17 en femenino plural, o sea, 118 formas adjetivas flexionadas, lo que eleva el total de epítetos a 170.

- ...realizar una larga y tendida **lista** de gestiones burocráticas¹⁶
- ...unos **trámites de divorcio** largos y tendidos¹⁷
- ...a pesar de las **veladas** largas y tendidas, acompañadas de vinito tinto¹⁸

Estos casos flexionados no se habían contabilizado antes, puesto que nuestra búsqueda inicial no lematizaba una forma que, al presuponerse adverbial, creíamos invariable por definición. Por si fuera poco, 29 de estos empleos adjetivales están en función de atributo (separados del sustantivo por un verbo copulativo)¹⁹, lo cual confirma la productividad de este uso adjetival. Las variantes adjetivales aportan conjuntamente 118 ocurrencias más de la locución, alterando el total inicial, que ya no sería de 3542 sino de 3660.

- ...el **debate** alrededor de estos conceptos **sería** largo y tendido
- ...el **apagón** también **fue** largo y tendido

Dichas variantes (tanto epítetos como atributos) aparecen con sustantivos muy diversos, a veces de manera un tanto extraña (como un coche con un *capó largo y tendido*). La mayoría forma combinaciones libres, apareciendo sólo una vez, pero si nos fijamos exclusivamente en las **70** ocurrencias que alcanzan o superan el umbral, vemos que éstas se concentran en un grupo muy reducido de sustantivos, que se refieren todos a actos de habla o comunicación, con un llamativo paralelismo con el grupo antes citado de los *verbos discursivos* (Tabla 10).

Tabla 10. Coocurrencias entre sustantivos y variantes adjetivas de la locución

N + LOC ^{Adj} (f≥5)	
conversación(es)	21
charla(s)	20
debate	12
tema	10
plática	7
TOTAL	70

16 <http://www.aulamagna.com.es> (*apud.* [ETT]).

17 <http://www.divorcios-express.com.mx> (*apud.* [ETT]).

18 *El Colombiano* <http://www.elcolombiano.com.co> (*apud.* [ETT]).

19 26 casos con *ser*, 3 con *estar*, ninguno con *parecer* o *semejar*. Con *estar* hemos descartado manualmente los casos en que es auxiliar de otro verbo en gerundio.

Sumando las secuencias adverbiales y adjetivales, y considerando como posibles colocaciones complejas todas las que igualen o superen el umbral $f \geq 5$, el total alcanzaría los 3179 candidatos (Tabla 11).

Tabla 11. Posibles colocaciones complejas

POSIBLES COLOCACIONES COMPLEJAS	($f \geq 5$)
V + LOC ^{Adv} <i>largo y tendido</i>	3109
N + LOC ^{Adv} <i>largo y tendido</i>	70
TOTAL	3179

Si a la cifra total de apariciones de *largo y tendido* (3660) le restamos estos 3179 ejemplos (86,9 %) quedarían 481 casos (13,1 %) que, por su frecuencia, no pueden ser candidatos, y, por lo tanto, no es descartable que constituyan combinaciones libres (Fig. 2).

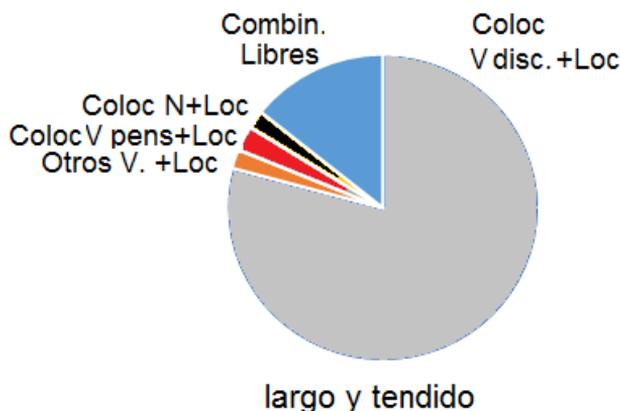


Figura 2. Proporción de coocurrencias de *largo y tendido* en el corpus

3 DISCUSIÓN

3.1 Aspectos cuantitativos

Según nuestros resultados en bruto, el binomio *largo y tendido* aparecería mayoritariamente como parte de unas *colocaciones complejas*, frente a una minoría de empleos en los que se combina con su núcleo de forma libre e imprevisible. Si las colocaciones se caracterizan por tener una mayor concordancia semántica o *fuerza de atracción* entre sus miembros (Apresjan 2009), ello debería reflejarse de alguna manera en las frecuencias absolutas de coocurrencia, pero un umbral mínimo quizá no sea una condición suficiente (aunque sí necesaria), sobre todo teniendo en cuenta que éste se fija de una manera relativamente arbitraria.

En principio, un cálculo preciso debería basarse en datos relativos y probabilísticos, puesto que la colocación se define como «dos palabras que coocurren más a menudo de lo que sus frecuencias respectivas y la longitud del texto permiten predecir»²⁰ (Sinclair/Jones/Daley 1970: 10)²¹. Obviamente, tales cálculos no serían necesarios para identificar una colocación con 1771 ocurrencias, como *hablar + largo y tendido*, pero las frecuencias más moderadas requieren mediciones precisas. Se utilizarán por ello varias medidas estadísticas que se describen más adelante. Se han propuesto diversas medidas estadísticas para indicar la atracción mutua que ejercen sobre sí los componentes de la colocación. Sin embargo, un tratamiento exclusivamente estadístico no basta para detectar colocaciones en corpus no filtrados: lo que extrae son *candidatos* que deberán ser comprobados individualmente a posteriori para acabar de deslindar las combinaciones libres de las colocaciones. Aquí, invertimos el sistema en cierto modo, puesto que aplicamos la estadística un listado de combinaciones previamente seleccionadas y cuyos datos de frecuencias se han obtenido después «manualmente». Calculamos 9 medidas estadísticas: Información Mutua (IM), IM2, IM3, z-score, t-score, log likelihood, Log Dice, Log Ratio y μ .

Información Mutua (IM) proporciona una indicación de la relevancia de la colocación. Obtiene valores más altos en casos de combinaciones exclusivas (A y B tienden a aparecer siempre juntos) independientemente de cuántas veces coocuran. Puede interpretarse como un indicador de la exclusividad de esa colocación (Brezina 2018: 70). Esa falta de relevancia de la frecuencia de coocurrencia en la IM la convierte en un buen indicador de la independencia de A y B, al destacar las combinaciones de palabras con bajas frecuencias frente a las de altas frecuencias de aparición. Es, en otra reformulación, la reducción en la incertidumbre sobre la ocurrencia de una palabra cuando aparece otra (Manning/Schütze 1999: 183). IM2 y IM3 son variantes heurísticas de la información mutua (IM) que intentan paliar el efecto amplificador de IM en el caso de colocaciones con componentes de reducida frecuencia individual. T-score es un indicador de lo fortuita que es una combinación AB: a mayores valores de t-score, menor es la probabilidad de que se trate de una combinación aleatoria. Las críticas que recibe el uso de esta medida es que asume una distribución normal de las probabilidades (Manning/Schütze 1999: 169). Z-score es el número de desviaciones estándar de la frecuencia media. Compara la frecuencia observada con la esperada. Log Likelihood es una medida que intenta solucionar los problemas asociados a IM y t-score. Compara la probabilidad bajo la hipótesis nula, de obtener los valores observados con la probabilidad en el caso de que exista dependencia entre los constituyentes de la combinación. El Coeficiente de Dice es conceptualmente similar a la información mutua, mejora las limitaciones de ésta y es considerada como una de las medidas de asociación más eficaces para detectar colocaciones (Rychlý 2008; Kolesnikova 2016: 340). Rychlý elabora una transformación del coeficiente de Dice que resulta más fácil

20 Traducción nuestra.

21 De la definición estadística de colocación se puede deducir su definición distribucional por inversión del razonamiento: «las colocaciones de determinada palabra son enunciados de los lugares habituales o convencionales que ocupa esa palabra».

de interpretar intuitivamente: Log Dice. En cuanto a Log Ratio, es una medida heurística también conceptualmente similar a la información mutua (IM) creada para facilitar su interpretación intuitiva. Cada punto adicional de Log Ratio equivale al doble de la diferencia entre la frecuencia del colocativo junto a la base y su frecuencia en otras posiciones (Hardie 2014). Finalmente, μ es la razón geométrica de la frecuencia real y la frecuencia esperada.

Las medidas estadísticas que hemos descrito pueden clasificarse en función de dos dimensiones: lingüística y estadística. La dimensión lingüística denota la *unicidad* (fuerza o estabilidad de la unidad); la dimensión estadística puede analizar el grado de asociación, la significancia (relevancia) de la asociación y la heurística (si esa medida se utiliza *ad hoc* por resultar intuitiva u operativa, sin estar necesariamente bien fundamentada desde el punto de vista teórico) (Pazienza *et al.* 2005). Mostramos esto en la Tabla 12, adaptada de Pazienza *et al.* (Ibid.: 260).

Tabla 12. Dimensiones de las medidas

Dim. Lingüística	Dimensión Estadística		
	Grado de Asociación	Significancia Asoc.	Heurística
Unicidad	IM Dice	z-score t-score Log Likelihood	IM2 IM3 Log Ratio Frecuencia Coocurrencia

Es posible, además, clasificar estas medidas desde otras dos dimensiones adicionales, situándolas en un sistema de referencia constituido por la exclusividad de la combinación (eje de ordenadas) y la frecuencia de la unidad (eje de abscisas), tal como se muestra en el siguiente gráfico adaptado de Brezina (2018: 74) (Fig. 3).

Estos marcadores estadísticos miden la relación entre las frecuencias absoluta de las secuencias, la frecuencia individual de sus componentes y el tamaño de corpus (medido en *tokens* y en *types*), para un corpus de 9.497.213.009 palabras (que, una vez lematizadas y contadas una sola vez cada una, suman 19.469.130), y para una frecuencia global de 3660 ocurrencias de *largo y tendido*.

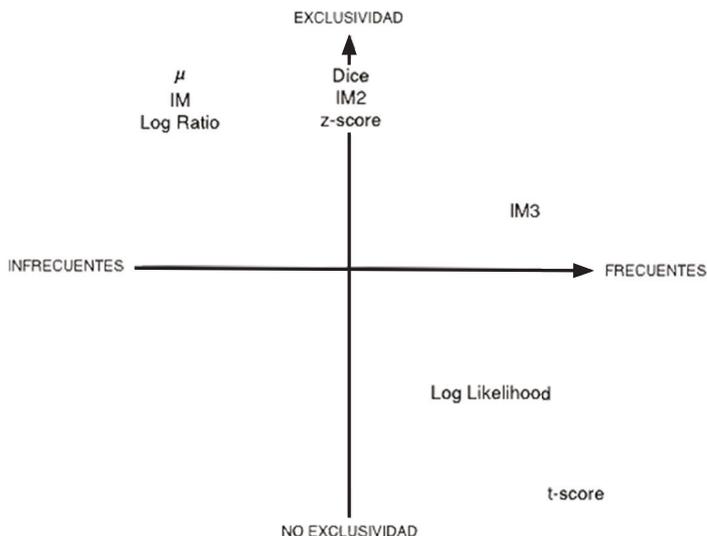


Figura 3. Clasificación de las medidas por las dimensiones de exclusividad y frecuencia

3.2 Aspectos lingüísticos

El hecho de que un colocativo coocurra también con los sinónimos e hipónimos de su base contradice en principio la definición Firthiana de colocación, puesto que contorno semántico-sintáctico del colocativo tiende a ser «heredado» por sus sinónimos e hipónimos (cf. Nazar/Renau 2019). Sin embargo, desde el punto de vista de las funciones léxicas, resultaría incluso lógico que el magnificador de una base también sirva para sus sinónimos e hipónimos. P.ej., la locución *en voz alta* puede formar colocación compleja con *hablar*, pero su colocativo es «heredado» por hipónimos de su base formando otras colocaciones: *decir en voz alta* / *leer en voz alta*, etc. (cf. Barrios 2009: 51-55).

Esta extensión puede incluso ser metonímica. Alain Polguère habla de «transplante colocacional» (*greffe collocationnelle*) para referirse a «colocaciones cuyo colocativo parece ‘prestado’ por otra colocación, y cuya base suele ser semánticamente cercana a la base a la que se injerta el colocativo prestado» (2007: 2)²². Así *soleil de plomb* (lit. *sol de plomo) puede reemplazarse por *soleil insoutenable* (lit. *sol inaguantable) por analogía con *chaleur insoutenable* (lit. *calor inaguantable) (2007: 11). En español, *asumir el trono* heredaría por analogía el colocativo *asumir la responsabilidad*, conceptualmente «cercano» a la base original (Uzcanga/Gómez 2009: 352). En las colocaciones complejas esta regla afectaría incluso a los derivados por nominalización (Blanco 2001: 177):

²² Traducción nuestra.

votar > votación a mano alzada;
asesinar > asesinato a sangre fría,
trabajar > trabajo por cuenta propia.

Sin embargo, como se ha dicho, dicha *herencia* también podría invocarse para negar que sean colocaciones, puesto que la sustitución sinonímica infringe la fijación y arbitrariedad teóricamente propia de las colocaciones (cf. Cserép 2017), la «concordancia semántica» de ciertos componentes los haría previsibles y hasta regulares (cf. Bosque 2019: 17). Pero la existencia de *series fraseológicas* ya era mencionada por el propio Bally (1905) desde los inicios de la teoría fraseológica. En lo que respecta al bloqueo sinonímico, es sin duda un rasgo de fijación, pero lo infringen también las locuciones con «variantes» (*apostar / jugar a caballo ganador; meter la pata / la pezuña / la gamba*) sin que por ello se cuestione su carácter fraseológico. Parece que el bloqueo sinonímico afecta sólo al colocativo: *dar un paseo* no permite *regalar un paseo (*echar la siesta* > *tirar la siesta), *tener miedo* > *poseer miedo, *romper a llorar* > *quebrar a llorar; *enemigo mortal* > *enemigo letal; *fumador empedernido* > *fumador endurecido, etc.) (cf. Alonso, Geoffrey/DeCesaris 2019:61). En cambio, la base suele infringir más fácilmente esta restricción (véase también Alonso Ramos 2019: 27, 34, 39). Se podría concluir que las colocaciones por imbricación son más proclives a la sustitución hiponímica de la base que las colocaciones «simples». P.ej., en *disparar a bocajarro* se puede reemplazar el verbo por *tirotear, ametrallar, acribillar, rematar*, etc. [REDES 2001: 14]. Incluso puede ocurrir una sustitución del colocativo (ej. *disparar a bocajarro* > *a quemarropa*), pero cuando la sustitución de un colocativo produce una expresión totalmente sinónima, el resultado es una *variante* y no una nueva colocación. En cambio, cuando es la base lo que se sustituye, da lugar a otra colocación, porque el significado global se ve afectado. P. ej., en *decir / plantear / expresar / soltar + a bocajarro*, el colocativo implica ('de improviso, inopinadamente'), y no 'desde muy cerca' como ocurre con *disparar*. Obsérvese que *de improviso* también es locución adverbial, pero no formaría colocación compleja porque puede combinarse con una lista abierta e impredecible de verbos (Penadés 2019: 88).

No por ello las colocaciones dejan de ser unidades restringidas, almacenadas y recuperadas en el lexicón mental de los hablantes (Corpas 2019: 180), pero las colocaciones por imbricación actúan en un nodo sintáctico superior, claramente descomponible, donde es toda la clase semántica del verbo base la que puede asociarse al mismo colocativo adverbial, por lo cual, la imbricación de un frasema en otro sería un hecho un poco más cercano al polo sintáctico del continuum construccional que al polo léxico.

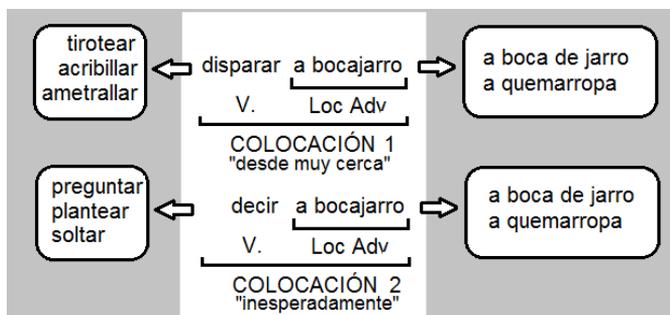


Figura 4. Conmutabilidad entre bases y entre colocativos

Estos usos adjetivales se pueden considerar como una nominalización de la base verbal, ya que no parece casual que se den principalmente con sustantivos como *conversación* (21 casos), *charla* (20), *debate* (12), *plática* (7), mientras que, entre los verbos que rigen su uso adverbial figuraban también *conversar* (175), *charlar* (218), *debatir* (90), *platicar* (163).

4 CONCLUSIONES

4.1 Filtrando con un umbral absoluto de al menos 5 coocurrencias, se obtienen 3138 candidatos a colocación frente a 522 combinaciones libres.

Tabla 13. Resumen del número y distribución de las ocurrencias de *largo* y *tendido* en el corpus

Posibles colocaciones complejas	LOC Adv + V. discursivo	2939 (80,3 %)
	LOC Adv + V. de pensamiento	73 (1,99 %)
	LOC Adv + V. de otra clase	97 (2,65 %)
	N + LOC. Adj	70 (1,91 %)
Posibles combinaciones libres		481 (13,14 %)
Total		3660 (100 %)

4.2 El análisis estadístico de los datos obtenidos del corpus confirma que todas las categorías satisfacen el requisito de coocurrencia mayor de la que podría esperarse por mera combinatoria libre (véase ANEXO). Sin embargo, las categorías [LOC Adv + V. de pensamiento], [N + LOC Adj] y [LOC Adv + V. de otra clase] obtienen siempre valores considerablemente menores a las de los verbos discursivos, «sólo» aparecen ≈ 40 veces más de lo que predice la combinatoria libre, ≈ 10 veces más en el caso de la categoría que incluye verbos de otra clase, y son las únicas que poseen valores negativos en

el marcador *Log Dice*. En lo que coinciden todas las medidas, aun así, es en confirmar el carácter colocacional de la coocurrencia de la locución *largo y tendido* con *hablar* y con los verbos discursivos en general. En efecto, en el análisis hemos observado que el verbo base *hablar* es reemplazable por otros elementos de su misma clase semántica. Por otra parte, el estudio de corpus hace aparecer abundantes usos adjetivales de *largo y tendido*, sobre todo complementando sustantivos cuyo referente es un acto comunicativo (*discusión, charla, conversación, diálogo...*), por tanto, probables nominalizaciones de dichos verbos. Por analogía, también hay verbos de pensamiento que conmutan entre sí como base del colocativo *largo y tendido*.

4.3 El que la colocación compleja permita sustituciones sinonímicas, hiponímicas, e incluso nominalizaciones, indica que su grado de fijación es mucho menor que el de las colocaciones «simples». Como la secuencia puede cambiar su sentido global, no se trata tampoco de *variantes* (cf. García-Page 2008: 244), sino de construcciones semi-composicionales. Un fenómeno similar ha sido señalado por Pejović y Valožić (2021) a propósito de la expresión *irse por los cerros de Úbeda*. Estas autoras concluyen que la locución propiamente dicha se limita a «los cerros de Úbeda», mientras que el verbo de movimiento formaría con ella una *construcción fraseológica*: [V^{MOV} +Prep +*los cerros de Úbeda*]. Hay dos razones que apoyan esta tesis: (a) el uso real demuestra que dicho verbo es variable (b) según qué verbo aparezca (*irse, salir, llegar, perderse, andar, estar, remontarse, desterrar*), las equivalencias contextuales de traducción al serbio de la secuencia global cambian necesariamente. En las 68 ocurrencias del corpus que utilizan (sumando CREA y CORPES XXI) se observa que los verbos no son sinónimos, pero tienen cierta afinidad conceptual, por lo que Mellado considera que se trata de *distintas formas léxicas (...) relacionadas entre sí de manera asociativa y creadas en virtud de un proceso de analogía* (2020a: 30).

4.4 Por este rasgo, estas secuencias se parecen más bien a los *esquemas fraseológicos* (Mellado 2020, Mura 2012), estructuras semifijas y semicomposicionales hoy llamadas *colostrucciones* (Stefanowitsh/Gries 2003), *frasemas construccionales*²³ (Booij 2013; Dobrovol'skij 2020) o *construcciones fraseológicas* (Mellado 2020b). La frontera entre *colocaciones* y *frasemas construccionales* quizá no sea aquí lo esencial, puesto que la oposición entre ambos es gradual (Colson 2018), y, como dice Wulff (2013: 285), «*effects of idiomatic variation are best observable in partially schematic complex constructions*». Lo más importante es que los datos del corpus confirman, con datos cuantitativos precisos, que la fijación no es un rasgo tan formal como tradicionalmente se pensaba, sino una restricción de naturaleza conceptual, y por ello resulta permeable a la conmutación entre componentes semánticamente afines. Esta especie de «fijación semántica», que Colson (2018) llama *entrenchment* y Carmen Mellado *fijación cognitiva*, (2020a: 19), es fruto de una tensión dialéctica entre dos fuerzas contrarias: la automatización léxica (fijación formal) y la analogía conceptual (procesos productivos a partir de un modelo), «cuanto mayor es la fijación cognitiva, mayor tendencia a la variación, a la productividad y a la esquematicidad» (Mellado 2020a: 19).

23 «*A constructional idiom is a (syntactic or morphological) schema in which at least one position is lexically fixed, and at least one position is variable*» (Booij 2013: 258).

Referencias bibliográficas

- ALONSO RAMOS, Margarita (1994-1995) «Hacia una definición del concepto de colocación: de J. R. Firth a I. A. Meľčuk». *Revista de Lexicografía* 1, 9-28.
- ALONSO RAMOS, Margarita (2002) «Explorando la frecuencia léxica para el Diccionario de colocaciones del español». En: T. E. Jiménez Juliá *et al.* (eds), *Cum corde et in nova grammatica: Estudios ofrecidos a Guillermo Rojo*. Santiago de Compostela: Universidad de Santiago, 19-40.
- ALONSO RAMOS, Margarita (2019) «Can collocations be deduced?». En: S. Torner/E. Bernal, E. (eds), *Collocations and Other Lexical Combinations in Spanish: Theoretical, Lexicographical and Applied Perspectives*. London: Routledge, 21-40.
- ALONSO, Araceli/WILLIAMS, Geoffrey / DeCESARIS, Janet (2019) «Studying lexical meaning in context: from collocation to collocational networks and resonance». En: S. Torner/E. Bernal, E. (eds), *Collocations and Other Lexical Combinations in Spanish: Theoretical, Lexicographical and Applied Perspectives*. London: Routledge, 41-74.
- APRESJAN, Juri (2009) «The theory of lexical functions. An update». En: D. Beck *et al.* (eds), *Proceedings [of the] Fourth International Conference on Meaning-Text Theory*. Montréal: Observatoire de linguistique Sens-Texte (OLST), 1-14.
- BALLY, Charles (†1905 [1951]) *Précis de stylistique. Esquisse d'une méthode fondée sur l'étude du français moderne*. Genève: Eggimann.
- BARRIOS, María Auxiliadora (2009) «Domain, domain features of lexical functions, and generation of values by analogy according to the MTT approach». En: D. Beck *et al.* (eds), *Proceedings [of the] Fourth International Conference on Meaning-Text Theory*. Montréal: Observatoire de linguistique Sens-Texte (OLST), 45-53.
- BLANCO, Xavier (2001) «Regroupements sémantiques dans un dictionnaire d'adverbes composés en espagnol». *Linguisticae Investigationes* 24(2), 167-182.
- BOOIJ, Geerd (2013) «Morphology in Construction Grammar». En: Th. Hoffmann/ G. Trousdale (eds), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press, 255-273.
- BOSQUE, Ignacio (2001) «Sobre el concepto de 'colocación' y sus límites». *LEA* 1, 9-40.
- BOSQUE, Ignacio (2019) «On the conceptual bases of collocations: Restricted adverbs and lexical selection». En: S. Torner/ E. Bernal, E. (eds), *Collocations and Other Lexical Combinations in Spanish: Theoretical, Lexicographical and Applied Perspectives*. London: Routledge, 9-20.
- BREZINA, Vaclav (2018) *Statistics in Corpus Linguistics: A Practical Guide* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781316410899>
- BUSTOS PLAZA, Alberto (2005) *Combinaciones verbonominales y lexicalización*. Frankfurt: Peter Lang.
- ČERMÁK, František (2007) *Frazeologia a idiomatika: česká a obecná*. Praha: Nakladatelství Karolinum.
- COLSON, Jean-Pierre (2015) «Editorial: Where does phraseology actually begin?». *Yearbook of Phraseology* 6, 1-2.

- COLSON, Jean-Pierre (2018) «Cognitive entrenchment, idiomaticity and corpora». En: A. Pamies/M. I. Balsas/A. Magdalena (eds), *Lenguaje figurado y competencia interlingüística (I): Aspectos teóricos*. Granada: Comares, 185-196.
- COLSON, Jean-Pierre (2019) «Multi-word Units in Machine Translation: why the Tip of the Iceberg Remains Problematic – and a Tentative Corpus-driven Solution», paper presented at *MUMTT2019, The 4th Workshop on Multi-word Units in Machine Translation and Translation Technology*. Málaga: Universidad de Málaga.
- CORPAS PASTOR, Gloria (1996) *Manual de fraseología española*. Madrid: Gredos.
- CORPAS PASTOR, Gloria (2001) «En torno al concepto de colocación». *Euskera* XLVI/1, 89-108.
- CORPAS PASTOR, Gloria (2019) «Collocations in e-bilingual dictionaries». En: S. Torner/E. Bernal, E. (eds), *Collocations and Other Lexical Combinations in Spanish: Theoretical, Lexicographical and Applied Perspectives*. London: Routledge, 173-199.
- COWIE, Anthony (ed), (1998) *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press
- CSERÉP, Attila (2017) «Idiom variation and decomposability Part I: Verbal variation». *Yearbook of Phraseology* 8, 95-122.
- DOBROVOL'SKIJ, Dmitrij (2007) «Cognitive approaches to idiom analysis». En: H. Burger *et al.* (eds), *Phraseologie: ein internationales Handbuch*. Berlin: De Gruyter, 789-818.
- DOBROVOL'SKIJ Dmitrij (2020) «German constructional phrasemes and their Russian counterparts: A corpus-based study». En: G. Corpas/J. P. Colson (eds), *Computational Phraseology*. Amsterdam: John Benjamins, 43-64.
- FIRTH, John R. (1957) *Papers in Linguistics 1934-1951*. Londres: Oxford University Press.
- GARCÍA-PAGE, Mario (1998) «Binomios fraseológicos antitéticos». En: G. Wotjak (ed), *Estudios de fraseología y fraseografía del español actual* Madrid: Vervuert/Iberoamericana, 195-202.
- GARCÍA-PAGE, Mario (2008) *Introducción a la fraseología española. Estudio de las locuciones*. Barcelona: Anthropos.
- GARCIA-PAGE, Mario (2011) «Collocations complexes (application à l'espagnol)». *Linguisticae Investigationes* 34/1, 67-111.
- GOUVERNEUR, Celine (2008) «The Phraseological Patterns of High-Frequency Verbs in Advanced English for General Purposes». En: F. Meunier/S. Granger (eds), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins, 223-246.
- GRIES, Stepan/STEFANOWITSCH, Anatol (2004) «Extending Collostructional Analysis: A Corpus-based Perspective on 'Alternations'». *International Journal of Corpus Linguistics* 9/1, 97-129.
- GROSSMANN, Francis/TUTIN, Agnès (eds), (2003) *Les collocations: analyse et traitement. Travaux et recherches en linguistique appliquée*. Amsterdam: de Werelt.

- HARDIE, Andrew (2014) «Log Ratio – an informal introduction». *ESRC Centre for Corpus Approaches to Social Science (CASS)*. <http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/>.
- HAUSMANN, Franz-Joseph (1979) «Un dictionnaire des collocations est-il possible?». *Travaux de Linguistique et de Litterature* 17/1, 187-195.
- HAUSMANN, Franz-Joseph (1998) «O diccionario de colocacións. Criterios de organización». En: X. F. Ruibal/M. J. Anido Silvosa/M. C. Lamela Villaravid/M. C. Viqueira Liñares (eds), *Actas do I Coloquio Galego de Fraseoloxía*. Santiago de Compostela: Centro Ramón Piñeiro.
- ÍRSULA PEÑA, Jesús (1994) *Substantive- Verb-Kollokationen. Kontrastive Untersuchungen Deutsch-Spanisch*. Frankfurt: Peter Lang.
- JAKUBIČEK Miloš/Adam KILGARRIFF/Vojtěch KOVÁŘ/ Pavel RYCHLÝ/Vít SUCHOMEL (2013) «The *TenTen* corpus family». En: A. Hardie/R. Love (eds), *7th International Corpus Linguistics Conference*. Lancaster: UCREL, 125-127.
- KAY, Paul/MICHAELIS, Laura A. (2012) «Constructional Meaning and Compositionality». En: C. Maienborn/K. von Heusinger/P. Portner (eds), *Semantics: An International Handbook of Natural Language Meaning*. Vol. 3. Berlin: de Gruyter, 2271-2296.
- KILGARRIFF, Adam/ Irene RENAU (2013) «*EsTenTen*, a Vast Web Corpus of Peninsular and American Spanish». En: *5th International Conference on Corpus Linguistics (CILC2013)*. <https://core.ac.uk/download/pdf/82198444.pdf>.
- KOIKE, Kazumi (2001) *Colocaciones léxicas en el español actual: estudio formal y léxico semántico*. Madrid: Universidad de Alcalá de Henares.
- KOIKE, Kazumi (2012) «Colocaciones complejas metafóricas». En: A. Pamies/J. M. Pazos/L. Luque Nadal (eds), *Phraseology and Discourse: Cross-Linguistic and Corpus-based Approaches*. Baltmannsweiler: Schneider, 73-80.
- KOLESNIKOVA, Olga (2016) «Survey of Word Co-occurrence Measures for Collocation Detection». *Computación y Sistemas* 20/3, 327-344.
- LANGACKER, Ronald (1987–1991) *Foundations of cognitive grammar*, Vol. I [1987]/ Vol. II [1991]. Stanford: Stanford University Press.
- LANGACKER, Ronald W. (2008) *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press.
- MANNING, Chris/Hinrich SCHÜTZE (1999) *Foundations of statistical natural language processing*. Cambridge: MIT Press.
- MEJRI, S. (2011) «Figement, collocation et combinatoire libre». En: J. C. Anscombe/S. Mejri (eds), *Le Figement linguistique : la parole entravée*. Paris: Champion, 63-77.
- MEJRI, Salah (2012) «Délimitation des unités phraséologiques». En: M. L. Ortiz (ed), *Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia*. Campinas, (São Paulo): Pontes, 139-156.
- MEL'ČUK, Igor A. (1998) «Collocations and lexical functions». En: A. J. Cowie (ed), *Phraseology: Theory, analysis and applications*. Oxford: University Press, 23-54.
- MEL'ČUK, Igor A. (2003) «Collocations : définition, rôle et utilité». En: F. Grossmann/A. Tutin (eds), *Les collocations: analyse et traitement. Travaux et recherches en linguistique appliquée*. Amsterdam: de Werelt, 23-31.

- MEL'CUK, Igor A. (2013) «Tout ce que nous voulions savoir sur les phrasèmes, mais...». *Cahiers de lexicologie* 102/1, 129-149.
- MELLADO BLANCO, Carmen (ed), (2008) *Colocaciones y fraseología en los diccionarios*. Frankfurt: Peter Lang.
- MELLADO BLANCO, Carmen (2014) «La polisemia en las unidades fraseológicas: génesis y tipología». En: V. Durante (ed), *Fraseología y paremiología: enfoques y aplicaciones*, Madrid: Instituto Cervantes, 177-195.
- 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) «Esquemas fraseológicos y construcciones fraseológicas en el continuum léxico-gramática». En: E. Tabares *et al.* (eds), *Clases y categorías en la fraseología de la lengua española*. Frankfurt: Peter Lang, 13-36.
- MENDÍVIL GIRÓ, José Luis (1991) *Las palabras disgregadas. Sintaxis de las expresiones idiomáticas y los predicados complejos*. Zaragoza: Pressas Universitarias de la Universidad.
- MICHAELIS, Laura A. (2019) «Constructions are Patterns and so are Fixed Expressions». En: B. Busse/ R. Moehlig (eds), *Patterns in Language and Linguistics*. Berlin: Mouton de Gruyter, 193-220.
- NAZAR, Rogelio/RENAU, Irene (2019) «A quantitative analysis of the semantics of verb-argument structures». En: Torner, S./Bernal, E. (eds), *Collocations and Other Lexical Combinations in Spanish: Theoretical, Lexicographical and Applied Perspectives*. London: Routledge, 114-136.
- OSORIO OLAVE, Gabriela (2012) «Colocaciones, compuestos sintagmáticos y locuciones nominales: hacia un intento de delimitación conceptual». *Lenguas Modernas* 39, 103-116.
- PAMIES, Antonio/WANG, Yuan (2020) «Frasemas verbales y metáfora gramatical en español y en chino». *LynX* 19 (Univ. Valencia), 89-145.
- PAMIES, Antonio (2007) «De la idiomatidad y sus paradojas». En: G. Conde Tarrío (ed), *Nouveaux apports à l'étude des expressions figées*. Cortil-Wodon: InterCommunications, 173-204.
- PAMIES, Antonio (2017) «Grammatical metaphor and functional idiomatity». *Yearbook of Phraseology* 8 (1), 69-104.
- PAMIES, Antonio (2019) «La fraseología a través de su terminología». En: J. J. Martín Ríos, (ed), *Estudios lingüísticos y culturales sobre China*. Granada: Comares, 105-134.
- PAZIENZIA, Maria Teresa/PENNACCHIOTTI, Marco/ZANZOTTO, Fabio M. (2005) «Terminology Extraction: An Analysis of Linguistic and Statistical Approaches». *Knowledge Mining* 185, 255-279.
- PEJOVIĆ, Andjelka/VALOŽIĆ Luiza M. (2021) «La construcción fraseológica [Vmov + prep + {los cerros de Úbeda}] y sus posibles equivalentes contextuales en serbio». *Komunikacija i kultura online* 12, 196-213.
- PENADÉS Martínez, Inmaculada (2012) *Gramática y semántica de las locuciones*. Alcalá de Henares: Universidad.

- PENADÉS Martínez, Inmaculada (2017) «El concepto de colocación a la luz de las colocaciones del tipo verbo más locución adverbial». *RILCE: Revista de filología hispánica* 33/3, 963-991.
- PENADÉS Martínez, Inmaculada (2019) «The fixedness of combinatory relationships». En: Torner, S./Bernal, E. (eds), *Collocations and Other Lexical Combinations in Spanish: Theoretical, Lexicographical and Applied Perspectives*. London: Routledge, 75-91.
- POLGUERE, Alain (2007) «Soleil insoutenable et chaleur de plomb: le statut linguistique des greffes collocationnelles». <http://olst.ling.umontreal.ca/pdf/Greffescollloc2007.pdf>.
- RYCHLÍ, Pavel (2008) «A Lexicographer-Friendly Association Score». En: P. Sojka/A. Horák (eds), *RASLAN 2008: Recent Advances in Slavonic Natural Language Processing, Karlova Studánka, Dec, 5-7, 2008*. Brno: Masaryk University, 6-9.
- SINCLAIR, John (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SINCLAIR, John; JONES, Susan/DALEY, Robert (1970) *The OSTI report*. Birmingham: University.
- STEFANOWITSCH, Anatol (2013) «Collostructional Analysis». En: Th. Hoffmann/G. Trousdale (eds), *The Oxford Handbook of Construction Grammar*. Oxford / New York: Oxford University Press, 290-306.
- UZCANGA VIVAR, Isabel/GOMEZ FERNANDEZ, Araceli (2009) «Les greffes collocationnelles en espagnol». En: D. Beck *et al.* (eds), *Proceedings [of the] Fourth International Conference on Meaning-Text Theory*. Montréal: Observatoire de linguistique Sens-Texte (OLST), 349-356.
- WULFF, Stefanie (2013) «Words and Idioms». En: Hoffmann, Th. / Trousdale, G. (eds), *The Oxford Handbook of Construction Grammar*. Oxford / New York: Oxford University Press, 274-289.
- ZULUAGA, Alberto (2002) «Los ‘enlaces frecuentes’ de María Moliner. Observaciones sobre las llamadas colocaciones». *LEA* XXIV, 97-114.

Diccionarios y corpus

- [AUT] (1726-1739) Real Academia Española *Diccionario de autoridades*. 6 vols. Madrid: Imprenta de Francisco del Hierro. (reed. facsímil). <http://web.frl.es/DA.html>.
- [COR] (1627. [1924]) Correas, Gonzalo *Vocabulario de refranes y frases proverbiales y otras fórmulas comunes de la lengua castellana...* [Reed. 1924 Tipografía de la Revista de Archivos, Bibliotecas y Museos]. <https://www.rae.es/obras-academicas/diccionarios/nuevo-tesoro-lexicografico-0>.
- [CORPES XXI] (2022) Real Academia Española: *Corpes XXI. Corpus del Español del Siglo XXI*. <http://www.rae.es>.
- [COV] (1611. [1995]) Covarrubias, Sebastián de *Tesoro de la lengua castellana, o española*. Madrid: Luis Sanchez.] [reed. de F.C.Maldonado, Madrid: Castalia 1995].
- [CREA] (2022) Real Academia Española: *CREA Corpus de Referencia del Español Actual* Banco de datos <http://www.rae.es>.

- [DFE] Cantera Ortiz De Urbina, Jesús y Gomis Blanco, Pedro (2007) *Diccionario de fraseología española: locuciones, idiotismos, modismos y frases hechas usuales en español*. Madrid: Adaba.
- [DEC] (1997) Martín Sánchez, Manuel *Diccionario del español coloquial: (dichos, modismos y locuciones populares)*. Madrid: Tellus.
- [DLE] (2019) Real Academia Española. *Diccionario de la lengua española*, Madrid, Espasa. <http://www.rae.es>.
- [DFDL] (2004) Seco, Manuel/Andrés, Olimpia/Ramos, Gabino *Diccionario fraseológico documentado del español actual: Locuciones y modismos españoles*. Madrid: Aguilar.
- [ETT] (2011) Corpus Spanish Web (*EsTenTen11_Eu+Am*) (accesible desde [SKE]).
- [REDES] (2004) Bosque, Ignacio (dir.) *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: SM.
- [SKE] (2021) *Sketchengine*. www.sketchengine.eu.
- [TERR] (1786) Terreros y Pando, Esteban de *Diccionario castellano: con las voces de ciencias y artes y sus correspondientes en las tres lenguas francesa, latina é italiana*. Madrid: Imprenta de la Viuda de Ibarra, Hijos y Compañía.

Resumen

ENTRE EL BINOMIO FRASEOLÓGICO, LA COLOCACIÓN COMPLEJA Y LA COLOSTRUCCIÓN: UN EXAMEN «LARGO Y TENDIDO»

Pese a una persistente diversidad terminológica, se puede decir que, desde el punto de vista conceptual, la taxonomía del ámbito fraseológico ha avanzado hacia un loable consenso en la última década. Quedan sin embargo algunas zonas polémicas o borrosas, como las colocaciones complejas y las «colostrucciones» (o ‘esquemas fraseológicos’), con las consiguientes dificultades técnicas para el reconocimiento de las mismas en un corpus electrónico. Ello repercute indirectamente sobre los límites del propio concepto teórico de *colocación*, que, dependiendo de cómo se defina, podría solaparse con estos tipos de frasema. Otra manera de enfocarlo es considerar que existe un mecanismo sintáctico productivo de *imbricación fraseológica*, por el cual un frasema puede encajarse en otro mayor sin constituir por ello una «nueva» categoría fraseológica. Por lo tanto, una locución adverbial podría formar parte de una colocación a mayor escala con un verbo (tanto por su semántica como por su frecuencia de coocurrencia), y seguir siendo una locución adverbial con otros verbos (Pamies 2017, 2019). Un caso muy representativo de esta indeterminación sería el de los *binomios fraseológicos*, que, en unos diccionarios fraseológicos aparecen clasificados como locuciones adverbiales (*a bombo y platillo*) pero, en otros, como locuciones verbales (*anunciar a bombo y platillo*). Para examinar de cerca este problema, hemos elegido la expresión *largo y tendido*, que, pese a sufrir una fijación total, se combina preferentemente con ciertos verbos, permitiéndose incluso sustituciones sinonímicas e hiponímicas entre los mismos. El estudio empírico (cualitativo y cuantitativo) de un gran corpus textual electrónico permite analizar todas las variaciones contextuales que afectan a su empleo,

poniendo a prueba, de paso, la metodología estadística de detección automatizada de las colocaciones. La heterogeneidad observada es mucho mayor de lo que cabría esperar, pero también pone de manifiesto ciertas tendencias regulares.

Palabras clave: colocación compleja, imbricación, esquema fraseológico, colostrucción, lingüística de corpus.

Abstract

BETWEEN THE PHRASEOLOGICAL BINOMIAL, COMPLEX PLACEMENT AND COLOSSTRUCTION: A “LONG AND EXTENDED” EXAMINATION

Despite a persistent terminological diversity, it can be said that, from a conceptual point of view, the taxonomy of phraseology has advanced towards a commendable consensus in the last decade. There remain, however, some controversial or blurred areas, such as complex collocations and ‘collostructions’ (or ‘constructional idioms’), with the consequent technical difficulties in recognizing them in an electronic corpus. This has indirect repercussions on the limits of the theoretical concept of collocation itself, which, depending on how it is defined, may overlap these types of phraseme. Another way to approach this question is to consider that there is a productive syntactic mechanism of phraseological imbrication, whereby a phraseme can be embedded in a larger phraseme without constituting a ‘new’ phraseological category. Thus, an adverbial idiom could be part of a larger-scale collocation with a verb (both in terms of semantics and frequency of co-occurrence), while remaining an adverbial idiom with other verbs (Pamies 2017, 2019). A highly representative case of this indeterminacy would be that of phraseological binomials, which, in some phraseological dictionaries, appear classified as adverbial idioms (*a bombo y platillo*) but, in others, as verbal idioms (*anunciar a bombo y platillo*). In order to examine closely this problem, we have chosen the expression *largo y tendido*, which, despite being totally fixed, is preferentially combined with certain verbs, even allowing synonymic and hyponymic substitutions between them. A (qualitative and quantitative) empirical study of a large electronic textual corpus makes it possible to analyze all the contextual variations that affect its use, testing, at the same time, the statistical methodology of automated detection of collocations. The heterogeneity observed is much greater than expected, but it also reveals certain regularities.

Keywords: Complex Collocation, Imbrication, Constructional Idiom, Collostruction, Corpus Linguistics

Povzetek
MED FRAZEOLOŠKIM DVOJČIČEM, KOMPLEKSNO KOLOKACIJO IN
KOLOSTRUKCIJO: ANALIZA STRUKTURE “*LARGO Y TENDIDO*”

Kljub trdoživi terminološki raznolikosti lahko rečemo, da je razvrščanje na frazeološkem področju s konceptualnega vidika v zadnjem desetletju napredovalo proti hvalevrednemu konsenzu. Vseeno pa ostaja nekaj polemičnih ali zabrisanih področij, kot so kompleksne kolokacije in »kolostrukcije« (ali »frazeeološke sheme«), s pripadajočimi tehničnimi težavami na področju njihovega prepoznavanja v elektronskem korpusu. To posredno vpliva na zamejitev samega teoretskega koncepta *kolokacije*, ki bi se, sicer odvisno od opredelitve, s temi vrstami frazemov lahko prekrival. Drugi možen pristop temelji na predpostavki o obstoju produktivnega skladijskega mehanizma *frazeeološkega sovpadanja*, v skladu s katerim se lahko določen frazem vdela v drugega, širšega, ne da bi se s tem oblikovala »nova« frazeološka vrsta. Tako bi prislovna stalna besedna zveza z določenim glagolom lahko tvorila širšo kolokacijo (tako s pomenskega vidika kot s stališča pogostosti sopojavljanja) in se z drugimi glagoli hkrati še vedno obnašala kot prislovna stalna besedna zveza (Pamies 2017, 2019). Zelo reprezentativen primer takšne nedoločnosti bi bili *frazeeološki dvojčiči*, ki se v nekaterih frazeoloških slovarjih uvrščajo med prislovne stalne besedne zveze (*a bombo y platillo* (≈ z rompom in pompom)), medtem ko jih drugi štejejo za glagolske stalne besedne zveze (*anunciar a bombo y platillo* (≈ naznaniti z rompom in pompom)). Da bi težavo podrobno preučili, smo izbrali izraz *largo y tendido* (≈ na dolgo in široko), ki ga kljub popolni ustaljenosti raje povezujemo z določenimi glagoli, med katerimi celo dopušča nadomeščanje s sopomenkami ali podpomenkami. Empirična (kvalitativna in kvantitativna) študija širokega elektronskega besedilnega korpusa omogoča analizo vseh kontekstualnih različic rabe, mimogrede pa tudi preizkus statistične metodologije avtomatiziranega zaznavanja kolokacij. Zaznana heterogenost je veliko večja, kot bi bilo pričakovati, vendar hkrati odraža nekatere ustaljene tendence.

Ključne besede: kompleksna kolokacija, sovpadanje, frazeološka shema, kolostrukcija, korpusno jezikoslovje

ANEXO

Gráficos de los valores alcanzados por cada categoría en las diferentes medidas estadísticas.

