

El fraseólogo internauta: cómo pasarlo pipa en la red

Gloria CORPAS PASTOR
Universidad de Málaga
gcorpas@uma.es

Resumen: En este trabajo se aborda la posibilidad de utilizar corpus y SRI en el ámbito de la investigación fraseológica. Los diccionarios de español al uso no consiguen describir las unidades fraseológicas (UF) de forma adecuada, pues suelen clasificarlas de forma errónea, no incluyen datos sobre frecuencia de uso o variantes, tienden a ofrecer ejemplos irrelevantes o rebuscados, etc. Los corpus de español disponibles son escasos y demasiado pequeños como para poder proporcionar un número suficiente de concordancias que faciliten la descripción lingüística de una determinada unidad. En este contexto, los sistemas de recuperación y gestión de la información electrónica podrían ser una alternativa complementaria válida. Los motores de búsqueda proporcionan al usuario gran riqueza de información, lo que reduce el silencio documental al tiempo que aumenta el ruido documental. Si bien el enfoque de la red como corpus podría paliar los efectos de la infoxicación resultante, los sistemas de gestión web presentan aún bastantes problemas. Una mejor solución de compromiso podría ser combinar varios tipos de recursos electrónicos a la hora de recuperar información fraseológica.

Palabras clave: Fraseología. Lexicografía. Diccionario.

Title: «*The phraseologist-Internet User: cómo pasarlo pipa en la red*».

Abstract: This paper addresses the possibility of using corpora and IR systems in the field of phraseology research. Spanish dictionaries fail to describe the use of phraseological units properly, because they are often misclassified, do not include data on frequency of use or variants, tend to offer irrelevant or obscure examples, etc. Available electronic corpora for Spanish are limited and too small so as to provide enough concordance lines to facilitate the linguistic description of a given phraseological unit. In this context, information retrieval and management systems could be a valid, complementary alternative. Search engines provide the user with a wealth of information, what reduces documentary silence but promotes documentary noise at the same time. While the web-as-a-corpus approach could mitigate the effects of the resulting infoxication, web management systems still present many problems. A better compromise might be to combine various types of electronic resources to retrieve phraseological information.

Keywords: Phraseology. Lexicography. Dictionary.

Titre : « *Le phraséologue internaute : cómo pasarlo pipa* ».

Résumé : Dans ce travail nous avons étudié les possibilités d'utilisation des corpus et des systèmes de recherche d'information (SRI) dans le cadre de la recherche phraséologique. Les dictionnaires espagnols ne décrivent pas correctement l'utilisation des unités phraséologiques, car ils les classent incorrectement, n'incluent pas les informations sur la fréquence d'utilisation ou les variantes, ont tendance à offrir des exemples non pertinents ou lambiqués, etc. Les corpus disponibles pour l'espagnol sont peu nombreux et bien trop courts pour procurer un nombre suffisant de concordances qui faciliteraient la description linguistique d'une unité phraséologique donnée. Dans ce contexte, les systèmes de recherche et de gestion de l'information pourraient être une alternative

valide et complémentaire. En effet, les moteurs de recherche mettent à la disposition de l'utilisateur une quantité remarquable d'informations, ce qui permet de réduire le silence documentaire, mais aussi d'augmenter le bruit documentaire. Tant que les corpus Web ne pourront atténuer les effets de l'infocication résultante, les systèmes de gestion Web présenteront encore de nombreux problèmes. Combiner différents types de ressources électroniques pour rechercher des informations phraséologiques pourrait s'avérer être un meilleur compromis.

Mots-clés : Phraséologie. Lexicographie. Dictionnaire.

INTRODUCCIÓN

En apenas 40 años, la Lingüística de corpus ha pasado de cenicienta a princesa del cuento. No hay disciplina lingüística que se resista a su encanto, ya sea como simple metodología de análisis y fuente de ejemplos que ilustren tal o cual teoría (investigación basada en corpus) o bien como inspiración y campo de estudio para elaborar teorías nuevas, cercanas a la etnolingüística y centradas en el uso real de la lengua en contexto. La lexicografía y la gramática, la enseñanza de lenguas asistida por ordenador (CALL), el procesamiento de lenguaje natural, la lingüística computacional, la traducción automática y la traducción asistida por ordenador son algunos de los primeros ámbitos en los que el corpus se consolidó como instrumento y materia prima por excelencia. También está siendo piedra angular en los estudios de historia de la lengua, cambio semántico y fenómenos de gramaticalización, de variación tipológico-textual, análisis crítico del discurso, la dialectología, la sociolingüística, la estilometría, la semántica cognitiva, la terminología, los estudios culturales y contrastivos, o la traducción y la interpretación, por citar los más relevantes. Y es que, en apenas unos años, la Lingüística de corpus se ha erigido como nuevo paradigma de investigación (cf. Corpas Pastor, 2008). En este contexto, la investigación en Fraseología tampoco se ha mantenido ajena a los avances del corpus. Es más, desde un principio los estudios con corpus pusieron de relieve dos aspectos que posteriormente se considerarían fundamentales. Por un lado, mostraron empíricamente la relativa baja frecuencia de uso de las unidades fraseológicas (UFs), en contraposición a la alta frecuencia de coaparición de sus elementos integrantes; y, por el otro, pusieron claramente de relieve la 'inestabilidad' real de la supuesta 'estabilidad' fraseológica, como se observa a partir del análisis de concordancias, donde se dan multiplicidad de variantes, variaciones ocasionales y modificaciones ostensibles con fines textuales, expresivos o lúdicos.

A pesar de que el uso de corpus ha abierto nuevas puertas y proporcionado bases reales para la investigación, su uso en Lingüística (y, por ende, en Fraseología) no está exento de problemas. Para empezar, es un recurso caro de desarrollar y actualizar, y aún sigue presentando numerosos escollos para el lingüista o el fraseólogo que pretenda analizar UFs mediante grandes volúmenes de datos, estudiar sus patrones de uso, sus valores pragmáticos o bien detectar usos lingüísticos recientes, cambiantes o neológicos. Además, no todas las lenguas disponen de corpus estables de referencia lo

suficientemente amplios o representativos como para poder realizar estudios fraseológicos con garantías. Incluso cuando estos existen, no siempre están a disposición de la comunidad académica, dado que suelen ser de acceso restringido o bien requieren suscripción o compra.

Dado que la investigación en fraseología necesita de grandes volúmenes textuales para poder obtener resultados relevantes, las limitaciones prácticas de los corpus actuales han llevado a muchos investigadores a «surfear» la Red mundial en busca de ejemplos de uso. Ello es posible hoy día debido al crecimiento exponencial y continuado de la red; y no sólo en cuanto al número de sitios en lengua inglesa, sino también con respecto al aumento de páginas en otras lenguas, especialmente germánicas, romances y asiáticas. Según el último estudio de Internet World Stat,¹ actualizado a 31 de mayo de 2011, el inglés encabeza el ránking de los 10 idiomas más utilizados en la red, con 565.004.126 usuarios (26,8% del total), seguido del chino (24,2%), el español (7,8%), el japonés (4,7%), y, a más distancia, del portugués (3,9%), el alemán (3,6%), el árabe (3,3%), el francés (3%), el ruso (3%) y el coreano (2%). A modo de ejemplo, citaremos el caso del español, que ocupa el tercer lugar del ránking, con un total de 164.968.742 usuarios de Internet —el 7.8 % de todos los internautas del mundo—, y cuya presencia en la red ha aumentado en un 807,4% en los últimos 11 años (2000-2011), frente al 301,4% de crecimiento del inglés.

La importante presencia de las lenguas en Internet hace posible hoy día utilizar la red como gigantesco repositorio de datos textuales para cualquier tipo de estudio lingüístico sobre (casi) cualquier lengua. De hecho, a día de hoy, la red es uno de los recursos electrónicos más utilizados en la investigación lingüística. En este trabajo nos vamos a centrar en el análisis de la UF contenida en el título mediante repertorios lexicográficos y recursos electrónicos. Nuestro objetivo es señalar las deficiencias de las que adolece el tratamiento lexicográfico de la fraseología para, a continuación, mostrar la manera en la que los recursos electrónicos y en línea actuales pueden contribuir a la mejora de investigación fraseológica y, por tanto, de la descripción lingüística resultante.

1. DESCRIPCIÓN LEXICOGRÁFICA

La red mundial, además de ser una valiosísima fuente de información, está considerada como uno de los mejores recursos electrónicos para la recuperación de información lingüística, en tanto constituye el repositorio textual más grande que existe, de acceso libre y amplísima cobertura temática, en aumento continuo y permanentemente actualizado. Entre ese maremágnum de sitios es fácil localizar muchos ejemplos de uso real y contextualizado de todo tipo de unidades fraseológicas. En contraposición a ello, los repertorios lexicográficos no suelen ofrecer información fraseológica fidedigna ni actualizada. Así las cosas, no es de extrañar que lingüistas, fraseólogos, lexicógrafos y traductores se decanten por la búsqueda de ejemplos y datos reales de uso en Internet.

¹ <http://www.internetworldstats.com/stats7.htm> [consulta: 17-09-2011].

Cuando se trata de estudiar la fraseología del español, de poco sirven los diccionarios actuales. Las principales deficiencias atañen a la propia selección de las unidades, a la ordenación del lema, a la categorización de la unidad en sí, o bien a la microestructura de la entrada correspondiente. Tomemos como ejemplo la unidad *pasarlo/pasárselo pipa* que hemos utilizado en el título de este trabajo. En la 22.^a edición del DRAE no viene siquiera recogida como UF, sino descompuesta como 9.^a y última acepción del lema *pipa*, categorizada como adverbio, y luego «recompuesta» como ejemplo de uso de dicha acepción:

pipa¹. [...] **9.** *adv. m. coloq.* Muy bien, estupendamente. *Pasarlo, pasárselo pipa.*

El artículo enmendado de la 23.^a edición del DRAE no presenta variaciones a este respecto, salvo por el número de la acepción, que pasa a ser la décima. Lo mismo ocurre en el DUE, donde la locución aparece como ejemplo de uso del lema *pipa*, en su octava acepción, como adverbio:

pipa [...] **8.** (inf.) *adv.* Muy *bien: *Los niños lo pasan pipa en el zoo. En este lugar se está pipa.*

Idéntico tratamiento recibe la UF en el diccionario CLAVE, donde aparece como quinta acepción del sustantivo, que sirve de palabra ordenatriz:

pipa. [...] **adv. 5 col.** Muy bien o estupendamente: *En el circo lo pasamos pipa.*

En el DEA se mantiene el mismo criterio que en los diccionarios anteriores, con la salvedad de que se considera claramente un caso de colocación de verbo + adverbio con verbos como *pasar* y *estar*. En los ejemplos de uso de la entrada se incluyen otros verbos posibles, como *montar*²:

pipa¹ [...]

II adv 9 (col) Muy bien. *Frec con vs como PASAR o ESTAR.* | Marsé Tardes 254: *Calla, hombre, lo vamos a pasar pipa, ya verás.* SSolís Blanca 139: *¡Ay, la conciencia, la dichosa conciencia! ¡Qué fácil es todo para los que no la tienen!, se lo montan pipa, que diría Mabel.*

Sólo dos diccionarios recogen la secuencia como UF. El LEMA incluye *pasarlo pipa* en negrita y como subtrada de *pipa*:

pipa1 n. f. [...]

pasarlo pipa coloquial. Divertirse mucho: *nos lo pasamos pipa en la fiesta.*

² A modo de anécdota mencionaremos que el diccionario REDES no recoge la secuencia *pasarlo/pasárselo pipa*, pero sí otras «combinaciones» del lema *pasar(lo)* [sic] con *bárbaro, bien, de fábula, de lo lindo, de maravilla, de miedo, estupendamente, mal, maravillosamente y regular.*

Y el DFEA, que siendo de los mismos autores que el DEA, pero algo posterior (2004), presenta la secuencia en negrita, como subentrada de *pipa*, pero ilustra la variante con *pasárselo* en el ejemplo:

PIPA¹

[...]

pasarlo pipa *v (col)* Pasarlo muy bien || Mañas *Kronen*: Tal y como lo contaba, se lo estaba pasando pipa metiéndole algodones por la nariz.

Es evidente que no estamos ante una colocación de verbo + adverbio [PASAR, etc. + PIPA] pues al descomponer la secuencia, esta pierde su unidad semántica y las partes constituyentes no conservan su propia integridad; además, desde el punto de vista formal, la unidad presenta fijación del componente *lo* (*la* en el español de América), pronombre en función de objeto directo, y fijación en singular del sustantivo, que no adverbio; y, finalmente, la secuencia no admite modificación o adición alguna (**pasarlos pipa*, **pasarle pipa*, **pasarlo pipas*, **pasarlo pipita*, **pasarlo medio pipa*, etc.). Más bien se trata de una de serie fraseológica con variantes sinonímicas: *pasarlo/pasárselo pipa/bomba/bárbaro/cañón/teta.*; y no de supuestos casos de colocaciones restringidas de verbo + adverbio: PASAR/MONTAR/ESTAR, etc. + PIPA/BOMBA/CAÑÓN/TETA/PIRATA.

Si se tratara de colocaciones, todos los sustantivos que forman parte de la UF como componentes serían categorizados como adverbios también. Y esto no es siempre así, como se deduce de la información recogida en el DRAE para dichos sustantivos, tanto en la edición actual como en el avance de la 23.^a edición.³ Bajo *teta*⁴ y *pirata* no se recogen acepciones como adverbios ni tampoco se incluye la combinación con *pasarlo* o *pasárselo*. Para *cañón*, en cambio, se indican sendas acepciones como adjetivo (en colocación con el verbo *estar*) y como adverbio (en colocación con el verbo *pasar*), dentro de la cual se inserta la UF como simple ejemplo de uso:

cañón. [...]

17. adj. Estupendo, fenomenal, muy bueno. *Está cañón.*

18. adv. m. coloq. Muy bien. *Nos lo pasamos cañón.*

El caso de *bomba* es algo diferente, pues aunque en el DRAE tampoco se categoriza como adverbio en ningún momento, la secuencia sí se recoge como subentrada, pero como dos locuciones verbales sinónimas («locs. verbs.»):

bomba. [...]

pasarlo, o pasárselo, ~.

1. locs. verbs. coloqs. Divertirse mucho. *Nos lo pasamos bomba en la fiesta.*

³ El avance de la 23.^a edición del DRAE está disponible en la URL: <http://buscon.rae.es/draeI/>. [Consulta: 16-09-2011].

⁴ En el DUE, sin embargo, sí se recoge una acepción similar, como adjetivo y adverbio, de la palabra *teta*: «4 (vulg.) adj. Estupendo, muy bueno. (vulg.) adv. Muy bien, estupendamente: “Pasarlo teta”».

Los ejemplos anteriores muestran una tremenda asistematicidad por parte de los repertorios lexicográficos a la hora de tratar la fraseología, pero también nos recuerdan la enorme dificultad que supone intentar clasificar y caracterizar este tipo de unidades. De las entradas analizadas se depende, en relación a la secuencia *pasarlo* o *pasárselo pipa*, lo siguiente:

- a) desde el punto de vista de su categoría fraseológica, se trataría de una colocación de verbo más adverbio o bien de una locución verbal;
- b) desde el punto de vista semántico se enfatiza el carácter elativo de la base (*pipa* = ‘muy bien’) o de toda la unidad (‘disfrutar mucho’), según se haya categorizado la unidad;
- c) desde el punto de vista de las restricciones diasistemáticas, se indica que diastráticamente pertenece a los niveles coloquial e informal; y, finalmente,
- d) desde un punto de vista puramente formal, el componente verbal de la secuencia parece admitir cierta variación paradigmática de carácter sinónimo (*pasarlo/pasárselo/montárselo/estar*).

Nada se dice, en cambio, sobre la fijación del componente *lo* (*la* en español de América), ni tampoco de la posibilidad de variación paradigmática del componente *pipa* (*bomba/cañón/teta/pirata*). No se mencionan otras restricciones diasistemáticas, ni aspectos pragmáticos y los ejemplos que se ofrecen son escasos, repetitivos y poco ilustrativos.

2. DESCRIPCIÓN LINGÜÍSTICA CON CORPUS

Otra opción para completar, actualizar y precisar la información lexicográfica sobre unidades fraseológicas es consultar bancos de datos o corpus estables de grandes dimensiones. De esta manera se accede a ejemplos que permiten determinar la frecuencia de las variantes, estudiar el uso de la unidad y ofrecer una información más precisa. Para el español se dispone de dos que reunirían las características necesarias: el *Corpus de Referencia del Español Actual* (CREA)⁵ de la RAE, que contiene unos 200 millones de palabras, y el *Corpus del Español* (CDE), compilado por Mark Davies (Brigham Young University, EE. UU.)⁶, de 100 millones de palabras. No obstante, el CDE contiene documentos desde el siglo XIII al XX, de los cuales solo 20 millones de palabras están fechadas entre los siglos XIX y XX. El CDE es, pues, un corpus eminentemente histórico y diacrónico, por lo cual no resultaría útil para investigar el uso de la fraseología actual. Por otra parte, el acceso a este corpus está muy limitado, incluso para los usuarios registrados. Aunque se accede bien a los ejemplos y funciones preparados a modo de guía de usuario, no es fácil consultar los datos del corpus, ya que continuamente aparecen mensajes de error del tipo «*You have exceeded the allowable number of queries per day.*»

⁵ <http://corpus.rae.es/>. [Consulta: 18-09-2011].

⁶ <http://www.corpusdelespanol.org/x.asp>. [Consulta: 18-09-2011].

Please try back later».⁷ Sin embargo, en el sitio *web* del corpus se pueden adquirir diversos lotes del corpus (por lemas, n-gramas, sinónimos, etc.), a distintos precios, según se trate de licencias de tipo académico o comercial, y siempre que no suponga tener acceso a los documentos completos, lo cual nos hace pensar que este corpus, en realidad, no es de acceso libre, sino que está pensado para darle un uso comercial.

El CREA, en cambio, permite acceso libre a las concordancias del corpus, dentro del rango textual, diacrónico y diatópico establecido en cada búsqueda. Sin embargo, a diferencia del CDE, no cuenta con un sistema de gestión versátil y flexible. Lo único que permite, en realidad, es poder presentar las concordancias ordenadas de la posición 1 a la 5, a la izquierda o a la derecha, del nodo (o *pivote*, en su terminología). Pero no permite realizar búsquedas truncadas mediante asteriscos o expresiones regulares, lo cual sería muy recomendable para una lengua tan derivativa como el español. Por ejemplo, al no estar lematizado, no permite localizar automáticamente las formas flexivas de la secuencia **[pasarlo + pipa]**. Así pues la localización de ejemplos para esta unidad tendría que realizarse mediante el análisis manual de las concordancias para *pipa*, alineadas, en este caso, a la izquierda. Sin embargo, ni siquiera de esta forma se consiguen ejemplos de uso de la UF en cuestión. Limitando las búsquedas a la variedad del español peninsular, en las 833 líneas de concordancias obtenidas como resultado no se registra la secuencia de búsqueda. Tampoco aparece entre los n-gramas (o *agrupaciones*, en su terminología) que el sistema recupera para el nodo *pipa*, ordenados por frecuencia:

- 2-gramas: *la pipa, una pipa, su pipa, en pipa, de pipa, La pipa, buena pipa, Una pipa, mi pipa;*
 3-gramas: *la pipa de, la pipa, una pipa de, su pipa, en pipa y, la pipa en, La pipa de, la pipa se, una pipa y;*
 5-gramas: *la pipa de la paz, la pipa de la boca, una pipa de espuma de, la pipa en la mano, la pipa apagada en la, la pipa de la dirección, la pipa entre los dientes, buena pipa que consiste en, una pipa en la boca.*

En cambio, sí se obtienen algunas concordancias si las búsquedas se realizan por secuencias concretas. Se trata de un problema de la arquitectura del sistema que si bien no impide la investigación, ralentiza considerablemente todo el proceso. En los ejemplos numerados que figuran a continuación se indica la secuencia concreta entre corchetes y el n.º de concordancias obtenidas (18 en total) entre paréntesis redondos.

[paso pipa] (1)

- 1 ones son de lo mejor que hay en televisión, me lo paso pipa rodándola, y encima, gustamos. ¿Qué más se **
 2 lo paso mejor. Disfruto haciendo la función me lo paso pipa. Uno piensa que después de tantos años de t **

[pasa pipa] (2)

- 3 estadas, que él sí ha visto a Morrisey, que se lo pasa pipa en el Arrebatado de Bilbao y mil cosas más. E **
 4 tal, y además, que como lo disfruta o sea Se lo pasa pipa. Y el otro, claro, yo le dije: "Mira, yo **

[pasan pipa] (2)

- 5 lido por la sencilla razón de que los críos se lo pasan pipa es, en realidad, tan absurdo como afirmar **
 6 jodas, miserable. Llevan una tajada atroz y se lo pasan pipa, los cabrones. Han hecho un montaje teatra **

⁷ «Ha excedido el número de consultas permitidas por día. Por favor, vuelva a intentarlo más tarde» [La traducción es nuestra].

[pasé pipa] (2)

7 final no fue ni campo de trabajo ni nada, y me lo pasé pipa, pues fue en cou y se me ocurrió la idea po **
8 r de isla, no sé que. Pues mira, fui solo y me lo pasé pipa, ¿entiendes? Espabilas mucho espabilas much **

[pasó pipa] (1)

9 que va de Liga ACB. Laso, excepcional Laso se lo pasó pipa alimentando a su gente, hasta el extremo de **

[pasaron pipa] (1)

10 el capote por aquí, el paquete por allá y se lo pasaron pipa. Fue una bueno, un torero le dio un infarto**

[pasaba pipa] (1)

11 actores europeos. En definitiva, el público se lo pasaba pipa. Recuerdo que Rafael Sánchez Ferlosio se **

[pasábamos pipa] (1)

12 simo porque tenía una cuñada maravillosa y nos lo pasábamos pipa. Llevo a Diana en el recuerdo". A la i **

[pasado pipa] (1)

13 s inéditas de Mójate Tenemos romance Y me lo he pasado pipa allí. Hasta tuve tiempo de enamorarme de **

[pasaré pipa] (1)

14 ta yo puedo meterme a crítica mañana mismo, me lo pasaré pipa. F. DÍAZ-FAES.- Eva, si son trapos..., ¿po **

[pasará pipa] (1)

15 te de lo que suceda en el terreno de juego, se lo pasará pipa. Canal +, 18.00 h. BALONCESTO Caja San Fe **

[pasaría pipa] (1)

16 obre todo si aún fuese un niño, seguramente me lo pasaría pipa. El reunir reproducciones arquitectónica **

[pasando pipa] (1)

17 r muy mal. - Tal y como lo contaba, se lo estaba pasando pipa metiéndole algodones por la nariz... - A **

[pasándolo pipa] (1)

18 en lasbotas," pasándolo la mar de bien, pasándolo pipa en la renfe. Entonces sería esto de cambiar imag **

Casi todos los ejemplos proceden del subcorpus oral y del componente periodístico del escrito, y solo el antepenúltimo procede de *Todo es comparable*, ensayo del arquitecto y pintor O. Tusquets Blanca (1998). Ello indica que la UF es típica de contextos informales o «pretendidamente» coloquiales, como son las entrevistas y los intercambios conversacionales; la prensa escrita y radiofónica, especialmente los artículos de opinión, la crónica social y las noticias deportivas; con alguna excepción, como en el ejemplo 16, al que nos acabamos de referir, donde funciona como recurso expresivo para acercar al lector de una forma amable a un tema un tanto árido, como es la arquitectura moderna.

Llama la atención que la UF tenga preferencias pragmáticas en cuanto a los tiempos y modos verbales. Así, sólo se conjuga en indicativo, preferentemente en presente (1-6), salvo en 2.^a persona de singular y plural y 1.^a de plural, y en pasado (7-13), concretamente en pretérito indefinido, salvo la 2.^a persona de singular y plural y la 1.^a de plural, y, en menor medida, en imperfecto (sólo en 3.^a persona de singular y 1.^a de plural), aunque se registra un ejemplo 1.^a persona singular del pretérito perfecto compuesto; también hay algún ejemplo en futuro (14-15), en 1.^a y 3.^a persona de singular únicamente, y solo un caso en 1.^a persona del condicional simple (16). En cuanto a sus formas no flexivas, no se registran usos en infinitivo (solo o formando perífrasis), y sólo hay un caso de participio presente o gerundio (17), que forma una perífrasis con *estar*, el cual ha sido utilizado como ejemplo de la UF en uno de los diccionarios mencionados más arriba (concretamente, el DFEA). Sin embargo, no hay ejemplos de uso en infinitivo.

En cuanto a las variantes léxicas de la UF, todos los ejemplos del CREA (salvo un caso) confirman la opción con el pronombre reflexivo *se* (*pasárselo bien*) y no hay rastro de usos con otros verbos recogidos por los diccionarios, como *estar* o *montar*, por ejemplo. Sí se atestiguan numerosas ocurrencias con *bomba* (*pasárselo/pasarlo bomba*), algunas con *teta* (*pasárselo teta*), como se ilustra seguidamente, pero ninguna con *cañón* o *pirata*.

[pasárselo bomba]

an felices. "El mío es absolutamente feliz. Se lo pasa bomba", afirma un padre anónimo, como piensan ot ** la raqueta mejor que casi nadie y, además, se lo pasa bomba. ¿Con qué cara podemos exigirle sacrificio ** s dudas sobre su grado de convencimiento. - Se lo pasa bomba el Ciegaliébres este. - No se cambia por e ** n mi despacho. ¡Se divertía el muy cabrito, se lo pasaba bomba haciéndose el mártir y el ofendido! ¡Ah, ** la menor vena de celos, sino al contrario, se lo pasaba bomba comentando a gente de su confianza, en e ** ual". Y vinieron a por mí, con los coches a . Te lo pasas bomba. Yo me lo paso bomba, chica. Lo que pasa que **

menos, del cadáver. En los cafés, Azucena, me lo paso bomba observando al personal, tratando de adivin **

[pasarlo bomba]

sino que me parece un sueño poder hacerlo y lo paso bomba hablando con los lectores y compartir opin ** iste ¿no? Ya la cla no existe. Yo en Sevilla lo pasaba bomba en la Antiguamente . Vamos, en todas ** el fondo de la estantería. Es muy triste pero lo paso bomba. -¿Qué lees? -Corazón. Oiga, he leído un c **

[pasárselo teta]

el de la víctima propiciatoria. Nos lo pasábamos teta. El tal Henri era un poco julandrón y lo que iba del maître observa la jugada y se lo está pasando teta. A lo mejor el muy hijoputa, al ver mi pasividad ** uejar. Con unas cosas y otras me lo estoy pasando teta. Muchos de ustedes quizá quisieran estar en mi l **

Por limitaciones de espacio no podemos analizar cada uno de los casos anteriores. No obstante, merece la pena señalar que presentan importantes diferencias entre sí en relación a su forma, su uso y su distribución:

- i. la variante con *teta* se construye en todos los casos con el reflexivo *se*, mientras que las variantes con *pipa* y *bomba*, aunque abundan en esa construcción, también se encuentran actualizadas en el discurso sin *se*;
- ii. las UF con *pipa* y *bomba* son mucho más frecuentes que la UF con *teta*, mientras que las variantes con *cañón* y *pirata* simplemente no aparecen en el corpus;
- iii. como hemos dicho más arriba, las variantes con *pipa* son propias de los textos orales y dentro del subcorpus escrito, de los periodísticos, mientras que la variante con *bomba* se encuentra tanto en el subcorpus oral como en el escrito, y dentro de este, no solo en textos periodísticos, sino también en todo tipo de literarios, y, finalmente, la variante con *teta* predomina en el componente narrativo —concretamente, los tres ejemplos anteriores proceden de la misma obra, *Días de guardar* (1981), de Carlos Pérez Merinero.

Está claro que la relación de sinonimia que todas estas UF parecen mantener en los repertorios lexicográficos resultaría, cuanto menos, discutible, a tenor de los resultados que arrojan las distintas consultas realizadas al corpus CREA, las cuales resumimos en la tabla que figura a continuación.

UF	FREC.		ORAL (col.)	ESCRITO (inf.)		PREFERENCIAS VERBALES	
				PER.	NARR.	+ frec.	- frec.
<i>pasárselo pipa</i>	17	18	6	10	1	Pres. Indic. (*2ª y 1ª pl) Pret. Indef. (*2ª y 1ª pl) Pret. Imp. (1ª y 3ª s)	Pret. perf. comp. (1º s) Futuro (1ª y 3ª s) Condicional (1ª s) Gerundio
<i>pasarle pipa</i>	1				1		
<i>pasárselo bomba</i>	9	13	3	3	3	Pres. Indic. (*1ª y 3ª pl) Pret. Imp. (solo 3ª s)	Infinitivo Gerundio
<i>pasarle bomba</i>	4		1	2			
<i>pasárselo teta</i>	3	3			3	Gerundio	Pret. Imp. (1ª pl)
<i>pasarle teta</i>	0						
<i>estar/montar + cañón/pirata</i>	0	0					

3. DESCRIPCIÓN LINGÜÍSTICA CON SRGI

El uso de corpus como mero repositorio textual o como metodología de análisis del nuevo paradigma reporta muchas e importantes ventajas para la fraseología y sus estudios aplicados. No hay más que releer la sección 2 destinada a los repertorios lexicográficos y comparar. Ahora bien, el uso de corpus también conlleva ciertos inconvenientes, algunos de los cuales hemos ido señalando en diversos lugares del apartado anterior (sección 3). Ante esta situación, muchos investigadores optan por utilizar diversos sistemas de recuperación y gestión de la información (SRGI), esto es, sistemas de recuperación de la información (SRI) y gestores de hipertextos. En primer lugar, dentro de los SRI se encuentran los motores de búsqueda generales, como los archiconocidos *Google*, *Yahoo*, *Altavista*, etc. A estos se unen los multibuscadores (*Clusty*, *Turbo10*, *Ixquick*, etc.) y los metabuscadores (*Metacrawler*, *Metasearch*, *All the Web*). Los metabuscadores permiten al usuario consultar varios motores de búsqueda e índices temáticos de forma simultánea a la vez, mediante una única consulta. Su ventaja es la comodidad de acceder a diversos buscadores en una misma página. Los resultados son muy similares, si no idénticos, a los que obtendríamos yendo directamente a cada buscador en cuestión. En esencia, los multibuscadores son similares a los metabuscadores con la diferencia de que los primeros no suman todos los resultados. Los metabuscadores, además, localizan otros buscadores.

En cualquier caso, estos S.R.I permiten realizar búsquedas simples de determinadas secuencias, y muchas de las veces, se obtienen resultados que, aunque básicos, son muy superiores en número y variedad a los que se podrían obtener mediante los corpus al uso actuales. Además son más fáciles y, sobre todo, muchísimo más rápidos de manejar. Piénsese que una consulta de diccionarios, además de frustrante por el tratamiento que

«propinan» a la fraseología, es extremadamente tediosa. La consulta al CREA, aunque mucho más fructífera, no está exenta de problemas y resulta mucho menos automatizada de lo que cabría esperar, por lo cual lleva bastante tiempo también. Un análisis de concordancias como el realizado en el apartado anterior puede llevar entre una y tres horas, dependiendo de la potencia del ordenador y la velocidad de conexión a la red, sin contar con el buen o mal funcionamiento puntual del servidor de la RAE o la mayor o menor pericia del que realiza las búsquedas.

Con el fin de ejemplificar el uso de SRI para el estudio de la fraseografía hemos elegido Google por ser uno de los buscadores más utilizados por los internautas, porque ofrece opciones de búsqueda avanzada y simple que permiten flexibilizar y adaptar las consultas a las necesidades de los usuarios y porque, como veremos más abajo, es el motor de búsqueda preferente del sistema gestor WebCorp.

Tomemos como punto de partida para nuestra consulta la secuencia [**pasa + pipa**]. Mediante la opción de búsqueda avanzada, donde se restringe a documentos escritos en nuestro idioma y publicados en España durante el último año, recupera aproximadamente 409.000 resultados en 0,08 segundos. Y sin la restricción temporal el buscador es capaz de recuperar unos 1.320.000 resultados (0,20 segundos). A continuación mostramos algunos de esos ejemplos:

- ...Va un KIKO y se lo pasa PIPA | Facebook...
- ...Pippa Middleton se lo pasa pipa en Madrid...
- ...Madonna se lo pasa pipa con su boy-toy...
- ...Romina Belluscio se lo pasa pipa - un vídeo de Gente...
- ...Miley Cyrus se lo pasa 'pipa', fumando hierba y fuera de control...
- ...Lady Gaga se lo pasa pipa en la piscina comiéndose a besos a su novio...
- ...Rihanna se lo pasa 'pipa' en Portofino. ¿En la playa o en la montaña? ...
- ...Rosa Diez se lo pasa pipa en la cabalgata Gay...
- ...Bejoncé se lo pasa pipa con su embarazo...
- ...Harry lo pasa 'pipa' en la costa de Croacia...

Google también permite automatizar las consultas mediante operadores booleanos en la búsqueda estándar o hasta tres alternativas directas en la búsqueda avanzada. Por ejemplo, una secuencia de búsqueda como [**paso OR pasas OR pasa OR pasamos OR pasais OR pasan AND pipa**] nos permitiría rastrear en la red ejemplos para todas las formas flexivas en presente de la UF, lo cual es ya todo un logro. Concretamente el buscador es capaz de recuperar hasta 2.270.000 resultados en 0,42 segundos. Además de todos los ejemplos anteriores, se pueden encontrar muchas más formas flexivas de la UF. Así, es posible recuperar ejemplos que contengan la UF todas las personas del presente, incluidas las de 2ª persona, singular y plural, y la 1ª de plural:

- ...Me lo paso pipa leyendo a estos tres idiotas. Estos tres imbéciles son los...
- ...tengo amigos canis que te lo pasas pipa con ellos XD eso sí, a veces salen para liarse...
- ...Pipa la Hermana de Kate Middleton en topless se lo pasa pipa...
- ...Gran noche de boda ayer, nos lo pasamos pipa, para los que no estuvisteis, os...

...jjeje lo pasais pipa, un saludo colegas riojanos, por cierto carmagedon el nombre...
 ...Dices que os lo pasais pipa follando. Coño, a ver si crees que yo no follo y que...
 ...Piqué y Puyol se lo pasan 'pipa'. Esta simple broma ya es publicada en Marca...

Cambiando las secuencias de búsqueda se obtienen resultados de esta UF conjugada en tiempos poco comunes, como el futuro compuesto, formando perífrasis de infinitivo, en modo subjuntivo, en imperativo, etc. Incluso, aplicando el operador booleano OR a los componentes sustantivos de las variantes se recuperan numerosas ocurrencias con *cañón*, *teta* y *pirata*, para las cuales el CREA no arroja apenas resultados:

...y algo saldrá de todo esto y si no pues nos lo habremos pasado pipa volando...
 ...Como se nos forme el anticiclón escandinavo vamos a pasarlo pipa...
 ...Enhorabuena de nuevo y pásalo pipa! Ya nos contarás :)...
 ... A un parque de bolas, ojalá se lo pase pipa, ya nos contará y mañana le llevamos a...
 ...jugando una de esas partidazas históricas con las que nos lo pasamos cañón...
 ...Y no te apures se lo estara pasando cañón con sus profes, amiguitos y un día sin...
 ...Este campeonato funciona bien, y nos lo pasamos teta...
 ...Saludos yo también soy nuevo y vamos a pasarlo teta con este juego...
 ...Nos lo pasamos pirata en Bilbao, os dejamos un vídeo con los mejores momentos...
 ...horas de sueño todos los días con vuestro programa, pero es que me lo paso pirata...

Digamos que se consigue reducir el silencio documental, pero, como contrapartida, se aumenta considerablemente el ruido documental⁸. Esto es, se consigue una mayor cantidad de ejemplos para cada una de las unidades estudiadas, pero, al mismo tiempo, también se recuperan todo tipo de páginas web, documentos en diversos formatos, archivos de audio y vídeo que contienen la secuencia como título o en otros lugares,⁹ etc., así como muchos resultados que nada tienen que ver con las UF, pues son ocurrencias de secuencias literales homónimas que contienen las palabras de búsqueda, como, por ejemplo [**pasas + pipa**]: *¿Me pasas una pipa, por favor? Y no, no me refiero a las pipas de comer, no. Me refiero a las que hacen pum*. Además, se encuentran ocurrencias con el pronombre *la* (en lugar de *lo*), probablemente por influencia de la población inmigrante

⁸ Se denomina silencio documental a aquella información que no ha sido encontrada por los motores de búsqueda, bien porque la estrategia de búsqueda que se ha definido resulta demasiado específica o bien porque las palabras claves utilizadas no son las adecuadas. El ruido documental, en cambio, hace referencia a aquella información que sí ha sido encontrada por el buscador pero que no es relevante, como ocurre cuando la estrategia de búsqueda ha resultado ser demasiado genérica.

⁹ Consúltense, por ejemplo, la URL: <http://poemitas.com/canciondialibro.htm> [Consulta: 20-09-2011], donde se puede escuchar una canción infantil con ese título e, incluso, bajarse la letra de la canción, que curiosamente, incluye la UF *pasárselo bomba* como sinónima; o el vídeo doméstico sobre un perro que se divierte tirándose a una piscina, y que lleva por título «El perro se lo pasa pipa en el tobogán» (URL: <http://www.notengotele.com/animales/el-perro-se-lo-pasa-pipa-en-el-tobogan>) [Consulta: 17-09-2011].

latinoamericana en España: *Así es que nada cogimos el bus y nos fuimos a una mega piscina que hay acá en Nimes, nos la pasamos pipa!!*

La combinación de silencio más ruido documental, unida a la democratización de la red, puede resultar en una auténtica *infoxicación* o sobrecarga informativa (*information overload*) que resta capacidad de análisis al investigador, el cual se ve desbordado por la ingente cantidad de ejemplos encontrados, en los más diversos sitios y géneros web (blogs, redes sociales, anuncios, wikis, webquests, etc.). Esta superabundancia de información hace que el investigador pierda la perspectiva que sí le proporcionaría un buen corpus, representativo, extenso y equilibrado. Por otra parte, las estrategias de búsqueda empleadas no son fácilmente automatizables u optimizables, por lo cual resulta muy difícil localizar UF de esta manera o determinar sus patrones combinatorios o sus variantes.

Una manera de superar algunos de los escollos anteriores es utilizar un sistema de gestión de hipertextos, que permita al usuario seguir utilizando la red como repositorio textual y base de datos, pero reduciendo la infoxicación, simplificando la clasificación de resultados, optimizando las estrategias de búsqueda mediante secuencias truncadas o expresiones regulares y ofreciendo las funcionalidades de cualquier programa de gestión de corpus al uso mediante operaciones como la lematización, el alineamiento, la extracción de colocaciones y n-gramas, etc. Uno de los sistemas más utilizados en la actualidad es WebCorp, que ha sido desarrollado por el RDUES (Research and Development Unit for English Studies) de la Universidad Central de Inglaterra, en Birmingham. Se trata de un programa¹⁰ gratuito, de libre acceso, que funciona como metabuscador y, a la vez, como generador de concordancias para documentos en .HTML y .TXT.

Este sistema localiza motores de búsqueda que conduzcan a páginas *web* relevantes para el término o secuencia de búsqueda, accede a dichas páginas, realiza un análisis del documento (*parsing*) en su formato original y extrae líneas de concordancia según la estrategia de búsqueda y los parámetros de presentación definidos por el usuario (Renouf, 2003). Como secuencia de salida (*output*), WebCorp utiliza el formato KWIC (*Key Word in Context*), esto es, líneas de concordancia con el nodo centrado y en contexto (entre 1 y 50 palabras a derecha e izquierda). El usuario puede obtener los resultados en .HTML (normal y tablas) o en .TXT. Puede especificar también si desea que en la búsqueda se distinga entre mayúsculas y minúsculas (Case Sensitive/Case Insensitive), o si desea que las líneas de concordancia obtenidas muestren la URL de la página de origen o bien la dirección vaya oculta, pero con un hipervínculo al nodo. En ambos casos, se ofrece información sobre la longitud de página, así como el porcentaje entre palabras tipo y formas, es decir, ocurrencias totales (*type/token*).

La opción de Búsqueda Avanzada (Advanced Search Options) permite seleccionar el universo textual, definir las estrategias de búsqueda y refinar los parámetros de análisis

¹⁰ El programa se encuentra disponible en la URL: <http://www.webcorp.org.uk/live/>.

que permitan extraer los resultados adecuados a las necesidades de información de los usuarios. Así pues, esta opción permite refinar las búsquedas mediante secuencias truncadas (asteriscos), secuencias discontinuas (expresiones regulares), filtrados léxicos y listas de *stop-words*; decidir qué motor de búsqueda se va a utilizar (Google, Yahoo o Altavista, Bing y Open Directory); establecer restricciones en cuanto a los ámbitos y tipos textuales, la variación diatópica y el rango temporal de los documentos que se van a seleccionar; y, finalmente, extraer directamente también colocaciones (externas e internas, con exclusión opcional de las mencionadas *stop-words*), establecer el número deseado de concordancias e, incluso, restringir el número de concordancias que se extraerán de cada documento analizado.

WebCorp resulta particularmente útil a la hora de estudiar unidades que presenten una frecuencia de aparición relativamente baja en el discurso, sean creaciones neológicas (formales o semánticas), ilustren usos raros o atípicos, o puedan adoptar formas cambiantes y fragmentadas. De ahí que este sistema resulte especialmente útil para el estudio de unas unidades como las fraseológicas, que encajan perfectamente en ese perfil: secuencias estables pero que, al mismo tiempo, resultan relativamente inestables; unidades que, por su propia institucionalización, son proclives a la variación y a las modificaciones discursivas, y que, en muchos casos, presentan una baja frecuencia de aparición. A continuación vamos a utilizar WebCorp para extraer concordancias para la UF seleccionada¹¹. Nos valdremos para ello de la opción Pattern Matching, que permite restringir las secuencias de búsqueda mediante corchetes, barras y asteriscos, además de utilizar la selección (positiva o negativa) de palabras. Seguiremos una metodología de análisis que contempla el establecimiento de secuencias de búsqueda generales y avanzadas; y la extracción de colocaciones externas y de frases clave.

En primer lugar, realizamos una búsqueda genérica mediante la secuencia truncada [**pas* pipa**], en la cual hemos empleado el asterisco en sustitución de las posibles desinencias verbales. Sin embargo, sólo se han obtenido 6 líneas de concordancias, 4 de ellas, procedentes del mismo blog de cine:

<http://www.blogfeed.es/horton-dos-mundos-para-pas-rselo-pipa/>
Plain Text Word List¹².

- Películas >> Horton dos mundos para pasárselo pipa Horton dos mundos para pasárselo
- pipa Horton dos mundos para pasárselo pipa Google : Horton dos mundos para
- Google : pipa Horton dos mundos para pasárselo pipa Horton dos mundos para pasárselo
- pipa Horton dos mundos para pasárselo pipa Viene tarde mi crítica de '
- Horton'. Horton dos mundos para pasárselo pipa Comment(0) Share on Facebook

¿Quiere esto decir que WebCorp solo ha sido capaz de recuperar de la red 3 ejemplos (1 de ellos repetidos) de la UF *pasárselo pipa* contenidos en un único documento? Hemos

¹¹ Las concordancias para el presente estudio fueron extraídas el 24 de septiembre de 2011.

¹² WebCorp permite el acceso al documento completo en formato plano o .TXT y, al mismo tiempo, ofrece la lista de palabras del documento por orden de frecuencia.

intentado dar respuesta a esta pregunta realizando una búsqueda para el sustantivo [**pipa**] y restringiendo diatópicamente la búsqueda a sitios web españoles (.es). WebCorp ha generado un total de 175 líneas de concordancia (por defecto el programa emplea Google y recupera concordancias procedentes de 64 sitios como máximo), pero en las concordancias obtenidas no hay rastro de la UF, como tampoco se encuentra en la lista de colocaciones externas de *pipa* ni en el listado de frases clave (Key Phrases)¹³, esto es, secuencias candidatas a unidades fraseológicas o representativas de los patrones sintagmáticos característicos del nodo analizado. El mismo resultado negativo se consigue cuando en la secuencia se sustituye el sustantivo por sus variantes con *bomba*, *cañón* y *pirata*. Para la variante con *teta*, en cambio, mediante el alineamiento a la posición 1 a la izquierda se obtienen 4 líneas de concordancia, 3 de ellas repetidas:

1. ojo a estos avances, ¡**pasarlo teteta** es vital! Más información | Facebook.
2. Cómics Horóscopo Robertus Fanta: **pasándolo teteta** en Nueva York Matías S.
3. divertidos (I) Robertus Fanta: **pasándolo teteta** en Nueva York El ascensor
4. Agenda happing Robertus Fanta: **pasándolo teteta** en Nueva York Conoce el

La opción Búsqueda avanzada permite filtrar las secuencias de palabras clave o truncadas con listas de palabras que deben (o no) aparecer con estas para seleccionar los documentos de los cuales se extraen las concordancias. Por ejemplo, si se combina *pipa* como palabra de búsqueda primaria con la secuencia [**pas[o|as|a|amos|áis|an]**] se localizarán UF con el verbo en presente de indicativo; si, por el contrario, se emplea [**pas[é|aste|ó|amos|ásteis|aron]**], las formas verbales estarán en pasado indefinido; una secuencia de filtrado como [**pas[ar|arselo|arlo|ando|ándolo|ado]**] recuperaría las formas verbales no personales, las perífrases de gerundo e infinitivo y, por supuesto, las formas compuestas con el participio, y así sucesivamente:

- entorno merengue, te lo **pasas pipa**. Sergio Ramos aplasta la única
- las opiniones me lo **paso pipa** 3 de Ago de 2009
- Semana Santa me lo **paso 'pipa'** Imagen de archivo de una
- chistes gráficos. Me lo **paso pipa** buscando por internet. Era seguidora
- si gateaba se lo **paso pipa** siguiendo a los otros niños
- diferente. Yo me lo **paso pipa** chinchando a los merengues cuando
- es que me lo **paso pipa** y me flipa esa descarga
- Yujuuu! Miravete se lo **paso pipa** escalando en su tierra! Juanito
- recuerdo (y, de paso, **pasártelo pipa** haciendo fotos ;). Pero antes de
- José Manuel Calderón se lo **pasó pipa** en Londres, haciendo de anfitrión como buen
- juventud, para reunirnos y **pasarlo... pipa!!!!** Inmaculada D se ha unido
- yo igualmente me lo **pase pipa** y al menos no cae
- segunda. Yo me lo **pase pipa**, que es lo que cuenta
- en Bilbao. Me lo **pasé pipa**, me rei mucho pero vamos

¹³ Esta opción es similar a las Agrupaciones ofrecidas por el CREA.

- es que se lo **pasaron pipa**!! #58 Vota Vota Denunciar 6
- Se lo pasarían **pipa** al principio. #23 Vota Vota
- Orientalia Lo vamos a **pasar pipa** 07:35 Lo vamos a pasar
- 07:35 Lo vamos a **pasar pipa** RAMÓN DÍAZ Pues da la
- cayendo, lo vamos a **pasar pipa** en los próximos meses. La
- subj=Lo+vamos+a+**pasar+pipa+** target="_blank" title="Comparte esta
- C) Se lo estaban **pasando pipa**. #95 Vota Vota Denunciar 5
- bien de qué significa **pasarlo pipa**', cuando lo dije, no me

La opción de filtrado secundario reduce considerablemente el ruido documental y automatiza el proceso de la investigación, pues permite localizar diversas formas flexivas de un determinado lema e, incluso, posibilita el almacenamiento de desinencias y afijos para investigaciones posteriores. Así, una secuencia como **[*[o]as[a]amos[áis]an]]** podría ser empleada para localizar formas en presente de cualquier verbo de la primera conjugación. Pero, como contrapartida, aumenta exponencialmente el silencio documental, hasta el punto de que son muy pocos los ejemplos que se puede obtener de esta manera. Además, el filtrado secundario no garantiza que sólo se recuperen las formas verbales deseadas, puesto que, por ejemplo, no se distinguen las vocales con tilde de las que no la llevan, o no se puede definir más la búsqueda indicando la distancia colocacional máxima permitida. De ahí que en la primera tanda de ejemplos que acabamos de mostrar aparezca uno con la forma *pasó* (*se lo pasó pipa*), en pasado, dado que el sistema no es capaz de diferenciarla de la forma presente (*paso*), mientras que en la segunda tanda, entre los ejemplos con el verbo en pasado indefinido, aparezca alguna concordancia con el verbo de la UF en condicional simple (*se lo pasarían* [sic] *pipa*), dado que es posible que en el resto del documento sí aparezca el verbo *pasar* conjugado en ese tiempo.

Otras estrategias de búsqueda interesantes para recuperar información fraseológica es la opción de utilizar el asterisco para indicar casillas vacías. Una secuencia como **[se lo pasa *]** muestra las palabras que aparecen en la posición 1 a la derecha de la expresión, así como la palabra que figura en la segunda posición, dado que trata a toda la secuencia como un bloque para la búsqueda y por eso indica en otro color y subrayado las palabras que ocupan esa casilla, así como la palabra con la cual toda la secuencia se coloca en la posición 2 a la derecha. Un ejemplo de secuencia completa con la casilla actualizada sería *se lo pasa pipa* y esta, a su vez, se coloca (o mejor, se coliga) con la preposición *en*. Por otro lado, esa casilla vacía se puede actualizar también mediante las siguientes palabras: sustantivos como *pipa*, *bomba* y *teta*, adverbios como *bien* (*muy bien*, *mejor*), adjetivos como *fenomenal* y *genial*, y preposiciones como *a*, *de*, *en*. Cuando la casilla está ocupada por las preposiciones *a* y *en*, la secuencia completa sería otra UF parcialmente sinónima con variación interna: *pasárselo a lo/en grande*, al igual que cuando aparece con la preposición *de* forma otra UF sinónima: *pasárselo de miedo*.

Ahora bien, cuando la casilla está ocupada por los sustantivos, la secuencia se coliga con la preposición *en*, que introduce un complemento circunstancial de lugar (*se lo pasa pipa en el concierto*), con la preposición *con*, que introduce un complemento circunstancial de instrumento o de compañía (*se lo pasa bomba con Wario Ware*); y con gerundios (*se lo pasa pipa contando sus anécdotas*), que introduce cláusulas subordinadas de modo. Resultados análogos, pero más simplificados, arroja el análisis automático de frases clave: *se lo pasa pipa*, *se lo pasa en grande*, *se lo pasa en*, *se lo pasa a*, *se lo pasa bien* y *se lo pasa pipa en*.

1. El equipo técnico del Barça se lo pasa a lo grande con el 'Xarlatan'
2. VÍDEO Última hora La Selección se lo pasa a lo grande de vuelta a
3. Activado - Desactivado La selección española se lo pasa a lo grande de vuelta a
4. vídeos Explorar Vídeo: La Selección se lo pasa a lo grande de vuelta a
5. Activado - Desactivado La selección española se lo pasa a lo grande de vuelta a
6. campaña El alcalde asegura que se lo pasa «bien» en los debates y no
7. Disfruto viendo que la gente se lo pasa bien mientras yo trabajo» Imprimir Enviar
8. Radio Bloddec Reset Satoru Iwata se lo pasa bomba con Wario Ware: Snapped! Por
9. Sus grupos Datos Ayuda miley se lo pasa de miedo con una de sus
10. login. Acceder Registrarse [E3] R2-D2 se lo pasa en grande jugando a 'Star Wars
11. Trayectoria Vídeos Fotos Foro Verdasco se lo pasa en grande en El Hormiguero Publicado
12. Buscar Carga de fotos Piqué se lo pasa en grande Siguiente >> Ver imagen en
13. Quiniela diariodemallorca.es » Deportes Colombia se lo pasa en grande con Nadal Quince mil
14. Tienda adn » cultura Gabriel Salvadó se lo pasa en grande dibujando y no se
15. Galería Lne.es » Galería Sasha se lo pasa en grande La hija menor de
16. Ingresar al sitio {header-content} Como se lo pasa este "Jubileta" en la Última Fiesta
17. cocina" La actriz reconoce que se lo pasa fenomenal cada día encima del escenario,
18. que en su ausencia, l se lo pasa fenomenal. La web 'RadarOnline' ha publicado
19. cuando vamos donde los abuelos se lo pasa genial pero si que me reclama`,
20. te confundas. ¿Quieres ver quien se lo pasa pipa contando sus anécdotas del 11S?
21. que será pero el niño se lo pasa pipa e intenta por todos los
22. Tweet Lleno de VIPS Letizia se lo pasa pipa en el concierto de Alejandro
23. S guenos en: Gente Pippa se lo pasa pipa La hermana de la princesa
24. 9 > Últimas contribuciones 6 Ronaldinho se lo pasa 'teta' en Río de Janeiro solofutboltv -
25. Fotos VÁdeos Actividad Quien se lo pasa mejor en el campo??? AÑ±adido
26. italianpaparazzi.com 0 Michelle Rodriguez se lo pasa pipa en Formentera cotibluemos -
27. Vota Resultados 0 votos Rihanna se lo pasa 'pipa' en Portofino En la playa
28. artista, soy una artesana que se lo pasa pipa haciendo lo que hace", confiesa
29. un bote de FANTA! Ramiro se lo pasa teta, pero teta, teta, con sus

Igualmente, se puede emplear el asterisco para recuperar los verbos que se combinarían con *pipa* mediante la secuencia [**lo * pipa**]. En este caso hemos optado por marcar la casilla de colocaciones internas (*Internal Collocates*) para que el sistema nos devuelva de forma automática las formas verbales, ordenadas por frecuencias. Todas ellas son formas flexivas del verbo *pasar*, a excepción de la palabra *imaginas*, que es

claramente un *hapax legomenon* que nada tiene que ver con la UF que analizamos: *pasas* (5), *pasaron* (3), *pasaba* (3), *pasarás* (3), *pasó* (3), *pasen* (2), *pase* (2), *paso* (1), *pasaremos* (1), *pasaríamos* (1), *pasé* (1), *pasan* (1), *pasamos* (1), *pasarán* (1) y *pasaban* (1). Cuando se sustituye el sustantivo por *bomba*, *teta* y *pirata* la opción de colocaciones internas indica que la casilla intermedia sigue estando ocupada por formas flexivas del verbo *pasar*, también en construcción reflexiva con pronombres personales (*me*, *te*, *se*, etc.).¹⁴ El sistema, en cambio, no muestra resultados para la secuencia [lo * **cañón**].

CONCLUSIONES

Las UF son conglomerados semántico-pragmáticos complejos, formados por una invariante o núcleo y varios planos adicionales, que se actualizan continuamente en el discurso. A esto se une la inestabilidad formal de tales unidades y las divergencias entre la relevancia psicológica de determinadas unidades y su frecuencia real de uso. Tales circunstancias hacen que estas unidades resulten bastante escurridizas para la descripción lingüística. No es de extrañar que el tratamiento de la fraseología siga siendo una de las asignaturas pendientes de la lexicografía actual. Tomemos como ejemplo el DRAE, uno de los repertorios más representativos para el español. Por un lado, no incluye *pasarlo pipa* como UF, sino como un ejemplo del «adverbio» *pipa*, sin ofrecer mayor información al respecto; mientras que, por el otro, no parece tener problemas en recoger otras obsoletas, como *tomar pipa*. «1. *loc. verb. coloq. desus.* Marcharse, irse, huir».

Son muchas las ocasiones en las que las páginas de los diccionarios se llenan de «cadáveres» fraseológicos, esto es, unidades ya caídas en desuso, que son recogidas (o, más bien, enterradas) en dichos repertorios. Otras veces, más que ante cadáveres, estamos ante casos de unidades fraseológicas «maltratadas». La falta de criterio, los errores de categorización, el forzar determinadas acepciones, la selección dudosa de las unidades que finalmente se recogen en los diccionarios, y que hacen que muchas UF vivas, actuales y de uso común no encuentren su sitio o, si lo tienen, este no sea el apropiado, son algunos de los signos de «violencia fraseológica» más evidentes.

Una buena alternativa a la «violencia contra la fraseología» que exhiben los repertorios lexicográficos al uso es consultar corpus estables o bancos de datos de grandes dimensiones. De esta manera se accede a ejemplos que permiten determinar la frecuencia de las variantes, estudiar el uso de la unidad y ofrecer una información más precisa. En las páginas anteriores hemos mostrado cómo el CREA, por ejemplo, permite extraer datos generales sobre la UF *pasárselo pipa* y sus variantes en relación a su variación paradigmática interna, su frecuencia de aparición, sus patrones sintácticos y sintagmáticos, así como su diferente distribución en cuanto a géneros y tipos textuales,

¹⁴ En el caso de la secuencia con *pirata*, se recuperan otros verbos, como *vender* o *hacer*, con los cuales se combina la unidad para formar colocaciones con otros significados: *vender pirata* es vender productos que no son originales, sino imitaciones, o bien que se han copiado del original de forma ilegal; *hacer algo pirata* es hacerlo de forma ilegal o sin respetar las normas establecidas.

que ponen en tela de juicio la pretendida relación de sinonimia que estas unidades supuestamente mantienen entre sí.

Es evidente que la consulta de corpus estables arroja datos enriquecedores para cualquier tipo de estudio fraseológico serio, pero tampoco hay que ignorar otras desventajas. Mencionaremos, por ejemplo, el hecho de que no siempre es posible encontrar ejemplos de UF, dada la baja frecuencia de aparición de estas en el discurso y el tamaño relativamente limitado de los corpus disponibles. O, incluso, aunque se encuentren ejemplos, podrían proceder de una única obra o de un único autor, con lo cual más bien serían representativos de un determinado idiolecto o estilo literario. En el caso concreto del CREA, otras dificultades propias se deben a la imposibilidad del sistema para ejecutar búsquedas truncadas que permitan localizar de forma automática las distintas formas flexivas de un mismo lema o para interrogar el corpus mediante expresiones regulares y opciones de búsqueda que permitan localizar variantes de forma automática. Incluso, como hemos visto, el sistema presenta errores consistentes en que cuando se piden líneas de concordancia para un determinado nodo, alineadas a izquierda o derecha (lo cual permitirían detectar, aunque fuese manualmente, UF), no siempre se obtienen todas, por lo que podría parecer que no hay ocurrencias de la UF objeto de estudio. Sin embargo, en esos casos podrían obtenerse resultados si se interroga el corpus mediante las cadenas de palabras exactas, una por una, de forma manual. Esto es un problema que se puede deber a la propia arquitectura del sistema, pero que sin duda dificulta y ralentiza cualquier análisis.

Por su parte, los sistemas de recuperación y gestión de información permiten al fraseólogo aumentar de forma exponencial los datos empíricos disponibles para el análisis. Mediante el empleo de diversos motores de búsqueda, y con las estrategias adecuadas, frecuentemente es posible conseguir resultados mucho más abundantes de una forma más rápida y sencilla. No obstante, aunque estos sistemas reducen el silencio documental, al mismo tiempo aumentan el ruido, por lo cual, a menudo, este exceso de información se torna en infoxicación e incapacidad del investigador en distinguir lo relevante de lo que no lo es, o lo anecdótico de las restricciones y preferencias reales. Todas estas dificultades se superarían si contásemos con auténticos sistemas gestores de datos en línea. Bien es cierto que existen algunos prototipos bastante efectivos, como WebCorp. Dichos prototipos admiten secuencias de búsqueda truncadas y alternativas, filtrados léxicos y, además, permiten realizar muchas de las operaciones básicas de cualquier programa de gestión de corpus (alineamiento del nodo, establecimiento de patrones sintácticos, extracción de n-gramas y frases clave, etc.). Sin embargo, esta aplicación en concreto adolece de una serie de problemas que lo sitúan lejos de ser la panacea para el estudio fraseológico. Algunas de las desventajas de WebCorp son, por ejemplo, la lentitud del programa en ofrecer resultados frente a cualquier SRI; la cantidad de ruido documental que genera, especialmente mediante las opciones de colocaciones externas y frases clave; las inconsistencias que se detectan cuando la búsqueda se realiza en el modo simple o en el avanzado, mediante secuencias monoverbales o pluriverbales, y

el menor número de concordancias que genera en comparación con un buscador como Google, al cual accede por defecto.

A pesar de las consideraciones anteriores, el uso del corpus como paradigma y como banco de datos (ya sea estable o virtual y *ad hoc*) ofrece un potencial enorme para el estudio fraseológico, ya que de forma inmediata se amplían y validan los datos disponibles para cualquier UF que se esté analizando. El fraseólogo internauta no solo debería poder disfrutar de la red como un divertimento, sino que también debería poder contar con una herramienta que le permitiese explotar las posibilidades que brinda la red de forma efectiva. O lo que es lo mismo, para poder pasarlo pipa de verdad el fraseólogo necesita tener lo mejor de ambos mundos: acceso a la ingente cantidad de información electrónica en línea accesible hoy día, pero tamizada mediante un potente gestor de corpus. Pero este es un tema complejo, y complicado a nivel técnico, el cual esperamos poder abordar en futuros trabajos.

REFERENCIAS BIBLIOGRÁFICAS

- BOSQUE MUÑOZ, I. (Dir.). (2004): REDES. *Diccionario combinatorio del español contemporáneo*. Madrid: Ediciones SM. [REDES].
- CORPAS PASTOR, G. (2008): *Investigar con corpus en traducción: los retos de un nuevo paradigma*. (Studien zur romanischen Sprachwissenschaft und interkulturellen Kommunikation, 49). Fránkfurt: Peter Lang.
- DAVIES, M. (2002-): *Corpus del Español: 100 million words, 1200s-1900s* [en línea]. <http://www.corpusdelespanol.org>.
- MALDONADO GONZÁLEZ, C. (dir.). (2006): *Diccionario CLAVE de uso del español*. 8.^a ed. (con CD-ROM). Madrid: Ediciones SM.[En línea] <http://clave.librosvos.net>. [CLAVE].
- MOLINER, M. (2007): *Diccionario de uso del español*. 2 vols. 3.^a ed. Madrid: Gredos. [En línea]. <http://www.diclib.com/>. [DUE].
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. <http://www.rae.es>.
- REAL ACADEMIA ESPAÑOLA (2001): *Diccionario de la lengua española*. 22.^a edición. Madrid: Espasa Calpe. [En línea]. <http://www.rae.es>. [DRAE].
- SECO, M.; ANDRÉS, O.; RAMOS, G. (1999): *Diccionario del español actual*. 2 vols. Madrid: Aguilar. [DEA].
- SECO, M.; ANDRÉS, O.; RAMOS, G. (2006²): *Diccionario fraseológico documentado del español: actual: locuciones y modismos españoles*. Madrid, Santillana.[DFEA].