

Aproximación al análisis de la fraseología española en situaciones comunicativas informales: el caso de *Gran Hermano*

An Approach to the Analysis of the Spanish Phraseology in Spontaneous Informal Language Production: The Case of *Big Brother*

ENRIQUE GUTIÉRREZ RUBIO [enrique.gutierrez@upol.cz]
Univerzita Palackého v Olomouci, República Checa

RESUMEN

Este artículo analiza las características de la fraseología española empleada en situaciones comunicativas informales. Para un estudio de esta naturaleza resulta imprescindible tener acceso a conversaciones espontáneas. Dado que los corpus orales existentes del español no son adecuados para los propósitos específicos de esta investigación, la fuente de las unidades fraseológicas (UF) analizadas es *Gran Hermano*, la versión española del *reality show Big Brother*. En este trabajo se presentan los resultados de un análisis de 120 minutos de conversaciones extraídas de este concurso televisivo. Dicho estudio se centrará en las siguientes dimensiones de análisis: a) Tipología de las UF registradas en contextos conversacionales espontáneos. b) Su frecuencia relativa de uso, tanto en función de los distintos tipos como de ocurrencias. c) Su presencia o ausencia en diccionarios, así como las marcas de registro asociadas a las UF.

PALABRAS CLAVE

fraseología española; español coloquial; corpus orales; *Gran Hermano*; lexicografía

ABSTRACT

This paper investigates the characteristics of the Spanish phraseology as it is used in natural communicative situations. For this kind of analysis, it is necessary to have access to spontaneous, free spoken conversations. Since oral corpora of Spanish are not sufficiently reliable for the specific purposes of this research study, the source of the phraseological units (PhU) analysed in this research is *Gran Hermano*, the Spanish version of the reality show *Big Brother*. In this paper the results of an analysis of 120 minutes of conversations will be presented. The focus will be on the following research dimensions: a) Typology of the PhU documented in spontaneous conversational contexts. b) Their frequency in terms of types and occurrences. c) Their presence or absence in dictionaries, together with the stylistic registers associated to the PhU.

KEYWORDS

Spanish phraseology; spoken Spanish language; oral corpora; *Big Brother*; lexicography

RECIBIDO 2020-05-11; ACEPTADO 2020-07-18

La investigación presentada en este artículo ha sido posible gracias a las actividades del proyecto de investigación FPVČ2017/04, financiado por el Fondo para el Fomento de la Actividad Científica de la Facultad de Filosofía y Letras de la Universidad Palacký de Olomouc.

1. Registro coloquial y situaciones comunicativas informales

No cabe duda de la dificultad que entraña recopilar material lingüístico válido para el análisis del habla en situaciones informales de comunicación, es decir, en las que los hablantes se expresen de forma natural, improvisada, sin que exista una voluntad de usar un español estándar o normativo, de acuerdo siempre con sus propias características individuales y capacidades como hablantes. En este trabajo no nos referimos tanto a un tipo de registro concreto, como al contexto en que las conversaciones objeto de análisis tienen lugar, es decir, *situaciones comunicativas informales*. Y remarcamos aquí que se trata de conversaciones, dado que, tal y como afirma Briz: “[...] la conversación es el uso más auténtico del lenguaje”; si bien, matiza enseguida: “[...] conversar no es hablar coloquialmente, puesto que hay conversaciones no coloquiales” (2010a: 10). En este sentido, no está de más recordar que incluso dentro de la lengua coloquial habría distintos registros, tal y como nos recuerda Vigarra Tauste (1987: 12): “El error de enfoque en que seguimos incurriendo con frecuencia es el de identificar sin más ‘lengua coloquial’ con ‘lengua popular informal’, olvidando que también hay una vertiente coloquial en la lengua ‘media’”.

Así, no juzgamos *a priori* el registro de las conversaciones objeto de análisis de coloquial, popular, informal o vulgar; tan solo indicamos que tienen lugar en situaciones comunicativas informales y espontáneas. En el marco de estas conversaciones, por tanto, aparecerán distintos tipos de registros o subregistros dependiendo de los propios hablantes y de la situación comunicativa, que, aun siendo de naturaleza espontánea, puede presentar multitud de matices. Este punto de partida metodológico, además de parecernos el más objetivo, permite evitar discusiones terminológicas, ya que son numerosos (y en ocasiones confusos) los adjetivos empleados para describir este fenómeno, tal y como recoge Cascón Martín (1995: 9): “popular, familiar, coloquial, informal, conversacional, hablada, oral, común, corriente, diaria, estándar, etc.”. Por su parte, Beinhauer, en su pionero estudio sobre el tema, escrito originalmente en alemán, afirma que la característica principal de este “registro” reside primordialmente en su carácter espontáneo, y lo describía así: “el habla tal como brota, natural y espontáneamente en la conversación diaria, a diferencia de las manifestaciones lingüísticas conscientemente formuladas, y por tanto más cerebrales, de oradores, predicadores, abogados, conferenciantes, etc.” (Beinhauer 1985: 9). En este sentido, coincide en lo fundamental con Seco (1973), quien caracteriza el objeto de nuestro estudio como *registro informal del habla*, entendido este como producto lingüístico de una actitud espontánea y que, generalmente, se viene denominando *registro coloquial*. De forma similar, Lázaro Carreter (1980) propone el término *lengua oral estándar*, mientras que Casares (1950/1992) prefiere *familiar* sobre *conversacional*, ya que este último, en su opinión, parece referirse a “[...] toda clase de temas, desde los más graves y elevados hasta los triviales e intrascendentes. Los interlocutores pueden hallarse en las situaciones de nivel o de desnivel” (Casares 1950/1992: 229). Sin duda, la aproximación más actual, y la más cercana al espíritu de nuestro trabajo, es la visión dinámica y contextualizada defendida por Briz, de modo que se trata, ante todo, de una cuestión contextual y gradual, sin caer en parámetros absolutos; así, “cabe distinguir entre *situaciones de mayor o menor coloquialidad* y *situaciones de mayor o menor formalidad* (entre la *constelación comunicativa +/–coloquial* y la *+/–formal* [o, en otros términos, de] situaciones de más o menos inmediatez comunicativa” (Briz y Albelda 2013: 293).

En este sentido, consideramos un acierto su propuesta de una serie de características propias de la conversación coloquial prototípica, los llamados rasgos coloquializadores, en función de los cuales una conversación tendrá una naturaleza más o menos coloquial/formal:

- + *relación social o funcional de igualdad entre los interlocutores*: acercamiento social o de los papeles comunicativos en un momento dado
- + *relación vivencial de proximidad entre estos*: saberes, experiencias y contextos compartidos
- + *marco interaccional familiar*: relación de cotidianidad de los participantes con el marco espacial en el que se sitúa la interacción
- + *cotidianidad temática de la interacción*: temas de la vida cotidiana, no especializados (Briz 2010b: 126).

A pesar de todo lo expuesto en este apartado, y por cuestiones prácticas, a lo largo del trabajo nos referiremos, en ocasiones, a nuestro objeto de estudio con la etiqueta general de *registro coloquial/informal*.

2. Corpus orales del español

De lo que no cabe duda es de que el material para un análisis como el nuestro ha de ser de naturaleza oral y conversacional, para cuyo estudio destacan distintos corpus del español, entre los que el más conocido y extenso sería el CREA oral, que en la versión 3.2 (2008) incluye casi nueve millones de entradas reunidas en más de 1 600 documentos. La metodología empleada por la Real Academia para la creación de este corpus es la transcripción y codificación de documentos sonoros, principalmente procedentes de grabaciones de radio o de televisión. A estos se les han añadido otros corpus orales previamente existentes, procedentes de “discursos políticos, conversaciones telefónicas, mensajes en contestadores, diálogos informales”¹. Además, los documentos van acompañados de algunas informaciones relevantes, como el grado de formalidad o el carácter libre o dirigido de la grabación; sin embargo, se trata de unos datos que no parecen seguir una metodología sistemática y que, de hecho, se incluyen tan solo en algunos de los subcorpus, no en todos. Si nos centramos, primero, en las grabaciones de radio y de televisión, creemos que, aun siendo orales, nadie se atrevería a calificarlas, salvo en contadas excepciones, de registro informal, coloquial, natural, familiar o bajo alguna de las variadas etiquetas con las que cabe denominar el objeto principal de nuestra investigación, ya que se presupone que los participantes tratan, en general, de mantener un nivel formal y normativo propio de estos medios. A este respecto, Briz parece confirmar nuestra premisa, ya que, en su opinión, el habla de la televisión está “planificada”. Aun así, esta afirmación parece estar más dirigida a los informativos que a otro tipo de programas, donde observamos un “[...] proceso de coloquialización creciente en TV y radio, favorecido, entre otras razones, por la tendencia a la comunicación interactiva y por la constante lucha por la audiencia” (Briz 2010a: 11). En cuanto al resto de materiales del CREA oral, se trata de un cajón de sastre con el que no parece posible trabajar

1 <http://www.rae.es/recursos/banco-de-datos/crea-oral>.

de forma sistemática a causa de sus heterogéneas características y por haber sido obtenidos mediante diversas metodologías.

Entre los demás corpus orales del español queremos destacar algunos que, por distintas causas, han llamado nuestra atención. Primero, el conjunto de entrevistas y conversaciones semidirigidas, grabadas en el marco del *Proyecto para el Estudio del Español de España y de América* (PRESEEA). Algunos de sus subcorpus (Alcalá de Henares y Barrio de Salamanca) nos resultan especialmente interesantes porque su fraseología ha sido objeto de análisis por parte de Penadés Martínez (2004; 2012; 2015a; 2015b). A pesar de que se trataron temas cotidianos, los participantes se hallaban en una relación de relativa igualdad y las grabaciones tenían una duración relativamente extensa –de entre 45 y 60 minutos ininterrumpidos–, resulta evidente que este tipo de material no es la fuente ideal de registro informal, dado que no se cumplen en rigor, al menos, dos de los cuatro rasgos coloquializadores propuestos por Briz: *relación vivencial de proximidad* y *marco interaccional familiar*². Por otra parte, para nuestra propia propuesta metodológica, resulta enormemente interesante el hecho apuntado por esta autora de que la formalidad del registro disminuía con el tiempo, de modo que algunos participantes acababan usando palabras propias de un tono informal e incluso llegaban a comentar “[...] no recordar que la conversación estaba siendo grabada” (Penadés Martínez 2004: 2227). Aun así, la propia autora reconoce explícitamente que esta metodología proporciona material de discursos de estilo formal o semiformal, es decir, estamos ante conversaciones coloquiales periféricas, si bien cabe hablar, al final de las grabaciones, de “un estilo algo más espontáneo” (Penadés Martínez 2015b: 252). Por todo ello, cabe afirmar que esta metodología basada en grabaciones semidirigidas y con el magnetofón a la vista no parece la más adecuada, dado que, de saberse grabados, resulta casi imposible que “[...] los interlocutores se presten sin prejuicios al coloquio” (Vigara Tauste 1987: 30). En definitiva, estamos ante lo que Labov denominó la “paradoja del observador”: “the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain these data by systematic observation” (Labov 1972: 209). Otro de los subcorpus del proyecto PRESEEA, concretamente el de Granada, también le ha servido a Sosiński (2010, 2011, 2012) para un estudio fraseológico de características relativamente similares a los publicados por Penadés Martínez, tal y como detallaremos más adelante.

Otro corpus que, sin duda, destaca entre los del español oral, sería el creado por el grupo Val. Es.Co. (*Valencia, Español Coloquial*) de la Universidad de Valencia y dirigido, en sus orígenes, por el propio Briz. El corpus está formado por 46 conversaciones con un total de 12 976 intervenciones y 120 246 palabras. Su tamaño es significativamente menor que el corpus oral de la RAE; sin embargo, tiene la ventaja de que, en su mayoría, evita la ya comentada “paradoja del observador” de Labov, dado que el 90% de las conversaciones se realizaron mediante grabación secreta, de modo que apenas en un 10% los hablantes sabían que estaban siendo monitorizados.³ Este corpus no ha sido escogido como fuente del material lingüístico de esta investigación, ante todo, porque para algunas de las dimensiones de nuestro estudio complejo de la fraseología en situaciones comunicativas informales necesitamos tener acceso a conversaciones mantenidas por un mismo hablante en distintas situaciones comunicativas y con personas tanto de su mismo sexo como del

2 En este sentido, Penadés Martínez opina que este último rasgo sí se cumple en el caso de las grabaciones de la capital de España (Barrio de Salamanca), ya que “[...] las conversaciones se llevaron a cabo en las casas de los informantes o en aulas de diversos centros públicos situados en el propio distrito de Salamanca” (2015b: 252).

3 *cf.* <http://valesco.es/?q=preguntas>.

sexo opuesto. Además, uno de los objetivos de nuestra investigación es comparar los datos obtenidos de *Gran Hermano* con los del corpus Val.Es.Co. De este modo, podremos arrojar luz sobre si *Gran Hermano*, a pesar de ser un producto televisivo (ver Sección 3), es o no una fuente adecuada para el estudio del registro coloquial en general, es decir, más allá del propio campo de estudio de esta investigación, la fraseología.⁴

Un último corpus oral del español que queremos mencionar en este breve repaso sería el Corpus Oral de Lenguaje Adolescente (COLA), creado por la Universidad de Bergen (Noruega) y que recoge 76 horas de conversaciones informales grabadas en tres capitales del mundo hispanohablante: Madrid, Buenos Aires y Santiago de Chile.⁵ En este caso, los hablantes eran conscientes en todo momento de que estaban siendo monitorizados. Sin embargo, a diferencia de la metodología empleada en el proyecto PRESEEA, en el marco del COLA las grabaciones se realizaron con micrófonos que llevaban los propios participantes, en muchas ocasiones en contextos alejados del centro de estudio, como parques, en casa, en el metro, etc. Así, al menos de acuerdo con Annette Myre Jørgensen, del grupo de investigación del que depende el COLA, la presencia de los micrófonos no habría influido en las conversaciones de los jóvenes, lo que se observaría en los “diferentes temas de conversación y el modo casual de hablar”. Además, añade, “si hacen referencia al micrófono a lo largo de sus conversaciones, solamente es para ver si funciona”.⁶ Nuevamente, las características propias de nuestra investigación –y, muy especialmente, del análisis realizado desde la perspectiva de los estudios de género– nos impiden hacer uso del COLA, si bien, al igual que en el caso ya comentado del corpus del grupo Val.Es.Co., realizaremos en un futuro cercano un estudio comparativo que aclare el grado de *coloquialidad* (y *vulgaridad*) de las expresiones recogidas en este corpus en relación con las obtenidas de las grabaciones de *Gran Hermano*.

Por último, tenemos que referirnos aquí, aunque sea brevemente, al intento más destacado de llevar a cabo un análisis de una naturaleza similar al nuestro, el propuesto en la obra *La fraseología del español coloquial* por Leonor Ruiz Gurillo (1998). Sin duda, se trata de un referente de indudable valía para la conformación de nuestro propio estudio de la fraseología en situaciones comunicativas informales. Sin embargo, el trabajo de Ruiz Gurillo presenta unos principios metodológicos distintos a los nuestros, principalmente respecto a dos aspectos. Primero, su análisis se centra en los sintagmas prepositivos fraseológicos. Segundo, para su investigación la autora hace uso de un corpus oral mixto, con conversaciones coloquiales y programadas de radio y televisión, es decir, no cabe hablar de auténticas situaciones comunicativas informales.

3. *Gran Hermano* como fuente de material lingüístico espontáneo

Como ya ha sido adelantado anteriormente en varias ocasiones, nuestra opción metodológica no será ninguno de los corpus brevemente apuntados en la sección 2, sino las conversaciones que tienen lugar en el marco del *reality show Gran Hermano*⁷.

4 Los primeros resultados aún provisionales pueden encontrarse en Gutiérrez Rubio (2020).

5 *cf.* <http://colam.org/publikasjoner/corpuslenguajeadoles.htm>.

6 *cf.* <http://colam.org/publikasjoner/corpuslenguajeadoles.htm>.

7 No consideramos que sea necesario explicar aquí en detalle el funcionamiento del concurso; para más información, recomendamos consultar la obra de Castañares (2003), en español, o de Mathijs y Jones (2004), en inglés.

A pesar de la enorme popularidad de este concurso televisivo en buena parte del mundo y de la propia naturaleza de este, basada en la idea de que reproduce fidedignamente la realidad cotidiana (incluida la lengua), el número de análisis lingüísticos realizados sobre la base de este programa es, hasta el momento, sorprendentemente escaso, muy especialmente en el caso concreto de la lengua española. Así, existe un trabajo –de reciente publicación, además– en el que sus autores, Haugh y Sinkevičiute (2018), presentan un análisis del discurso de hablantes españoles y argentinos, más concretamente de acusaciones y conflictos interpersonales, con base en los diálogos mantenidos por los concursantes de las respectivas ediciones nacionales de este *reality show*. Además, la propia Sinkevičiute ha publicado varios trabajos sobre las variantes británica y australiana de la lengua inglesa tal y como es empleada por los concursantes de *Big Brother*, tanto sobre la construcción de la identidad de grupo (2017a) como acerca de la descortesía (2017b).

Mucho más interesantes nos resultan los recientes trabajos publicados por Bertomeu Pi (2019; 2020) con base en las ediciones española y alemana del concurso. Tras realizar un análisis cualitativo de siete rasgos coloquializadores, este especialista acaba afirmando que *Gran Hermano* es “a highly suitable corpus to study colloquial register in German and Spanish” (Bertomeu Pi 2020: 142).

Resulta significativo y, en nuestra opinión, ciertamente inexplicable que, a pesar de las características específicas de este concurso, en la única obra que aborda de forma monográfica un estudio de carácter lingüístico del fenómeno de los *reality shows* en general, más concretamente desde la perspectiva del análisis del discurso, apenas se le preste atención a *Gran Hermano* (Lorenzo-Dus y Garcés-Conejos Blitvich 2013).

Por otra parte, resulta innegable que el material lingüístico obtenido de las conversaciones mantenidas por los concursantes de *Gran Hermano* no es completamente natural, dado que estamos ante un producto televisivo y, en mayor o menor grado, dirigido: un espacio que no es real, en el que la realidad, más bien, se recrea o que, como señala Castañares (2003), se halla a medio camino entre el mundo de la vida real y el de la ficción. Partiendo, por tanto, de que no estamos ante una producción de habla absolutamente natural o espontánea, sí entendemos que puede considerarse una variante muy cercana a esta. En este sentido, Castañares propone que uno de los motivos por los que los concursantes de los *reality shows* (en general, no en referencia específica a *Gran Hermano*) son escogidos entre decenas de miles de candidatos tiene que ver con el hecho de que representan “el paradigma del ciudadano medio, que poseen una gran espontaneidad, que utilizan un lenguaje llano, popular, a veces incluso incorrecto o vulgar [...]” (Castañares 1995: 86).

Además, creemos no exagerar al afirmar que los cuatro principales rasgos coloquializadores propuestos por Briz (relación social o funcional de igualdad, relación vivencial de proximidad, marco interaccional familiar y cotidianidad temática) se cumplen a la perfección en el entorno de confinamiento y convivencia *forzada* de la casa de *Gran Hermano*. Por supuesto, los hablantes son en todo momento perfectamente conscientes de que están hablando ante las cámaras a través de micrófonos, lo que indudablemente podría influir en su registro. Sin embargo, si Penadés Martínez afirmaba que entre algunos de los participantes del corpus de Alcalá en menos de una hora de conversación, y con la grabadora siempre presente, era perceptible un cambio significativo desde un registro más formal hacia uno claramente más informal, no parece descabellado pensar que, a pesar de las cámaras y los micrófonos omnipresentes, después de no horas ni días, sino de semanas e incluso meses de grabación y emisión ininterrumpida 24 horas al día, los participantes se acostumbren a la presencia de estos aparatos hasta el punto de que ya afecten muy poco, e incluso

no lo hagan en absoluto, a su registro. Por tanto, si nuestra premisa fuera cierta, pasadas algunas semanas de concurso, es muy probable que lo que escuchemos sean conversaciones de registro informal auténticamente naturales o, al menos, algo muy cercano a esto. Además, contamos con una calidad de sonido relativamente elevada, lo que evita los problemas propios de las grabaciones secretas como los padecidos por el Grupo Val.Es.Co. (Briz 2002).

Nuestra premisa parece verse sustentada, además, por algunos de los numerosos trabajos que, desde la perspectiva sociológica y de los estudios de los medios de comunicación, han abordado el fenómeno de *Big Brother*. Nadie pone en duda que somos testigos de una construcción en cierta medida dirigida por la producción televisiva a cargo; hecho este observable, por ejemplo, en algunas de las novedades de la quinta edición del *Big Brother* británico, claramente enfocadas a fomentar la actividad sexual entre los concursantes (Bignell 2005: 78). Sin embargo, autores como Dovey ponen de manifiesto que, aun tratándose de una *simulación*, *Gran Hermano* produce conocimiento *real*:

[Although we are watching] simulated situations that only exist because of the intervention of the TV production [...] the simulation, in my understanding, is not about unreal states [...] the simulation as a material practice produces real knowledge about real things in the real world and has real effects upon real lives (Dovey 2004: 233).

Por todo ello, partimos de la premisa de que este material lingüístico es más válido para nuestro objetivo que el obtenido de entrevistas o diálogos semidirigidos, especialmente pasadas varias semanas de concurso.

4. Delimitación y clasificación de la fraseología española

Antes de pasar a presentar los principios metodológicos que rigen nuestro análisis, hay que destacar la problemática misma de la delimitación del objeto de estudio de la fraseología, las ya famosas concepciones ancha y estrecha de esta disciplina, y posicionarse metodológicamente para nuestra investigación.

En principio, nos decantamos por una concepción relativamente amplia, dado que incluimos en nuestro análisis cualquier combinación de palabras⁸ cuyo significado o uso, como conjunto, no se corresponda con la mera suma de los significados o usos asociados a sus constituyentes. Esto nos haría coincidir, en gran medida, con la clasificación de Corpas Pastor (1996) –que goza de una aceptación mayoritaria, si bien no absoluta, entre los especialistas españoles–, pero de la que en nuestra investigación, por la necesidad de acotarla a los principales tipos de UF, descartamos las colocaciones. Así, en nuestro estudio clasificamos tres tipos principales de UF: locuciones, fórmulas oracionales y paremias, entendidos estos tal y como se detalla en los siguientes tres apartados

Por último, creemos que los vocativos pluriverbales, especialmente los de carácter metafórico –como los propuestos por Miranda (1998: 87): “mi vida, vida mía, mi cielo, mi salvación”– también

8 Descartamos, por tanto, todas las expresiones univerbales, a pesar de que, tal y como señala Zuluaga (1980: 193), *Adiós* presenta el mismo funcionamiento como UF que *Buenos días*.

deberían formar parte del conjunto de unidades fraseológicas y, serán, lógicamente, incluidos entre las locuciones nominales a pesar de que no suelen ser recogidos en las fuentes fraseográficas.

4.1. Locuciones

Aquí nos referimos a locuciones en el sentido más o menos tradicional del término. En general, podemos decir que coincidimos con la definición propuesta por Corpas Pastor (1996: 88): “unidades fraseológicas del sistema de la lengua con los siguientes rasgos distintivos: fijación interna, unidad de significado y fijación externa pasemática. Estas unidades no constituyen enunciados completos y, generalmente, funcionan como elementos oracionales”. Sin embargo, no compartimos a pie juntillas su categorización, que reformulamos siguiendo aquello que nos parece más adecuado, sistemático y coherente de entre las múltiples clasificaciones publicadas para el español (Casares 1950/1992, Zuluaga 1980, Montoro del Arco 2006, García-Page Sánchez 2009, entre otros); así, en nuestro trabajo distinguimos siete tipos de locuciones, según la función que estas desempeñan en la oración: adjetiva, adverbial, conjuntiva, nominal, marca-dora, prepositiva y verbal.

4.2. Fórmulas oracionales

La principal característica de las fórmulas oracionales frente a las locuciones (de cualquier tipo) es su capacidad para construir enunciados por sí mismas. Sin embargo, podemos sumar un segundo elemento delimitador: mientras que las locuciones funcionan en el marco de la oración (excepcionalmente a nivel supraoracional, como en el caso de las locuciones marcadoras), las fórmulas lo hacen claramente en el nivel del discurso, poseyendo una función eminentemente comunicativa. Dicho de otro modo, no se les asocia tanto un significado léxico (o gramatical), como una función dentro del discurso. En palabras de López Simó: “son enunciados, productos lingüísticos de un acto de habla, y, por lo tanto, no son unidades de la gramática sino del discurso-en-interacción (dirigido a otro que también actúa) cuyo ejemplo más prototípico es la conversación” (López Simó 2016: 77). En este sentido habría que advertir de que no todas las fórmulas oracionales son actos de habla por sí mismas ni necesariamente son reacciones a una interacción, ya que existen otros tipos de expresiones que, a modo de ejemplo, expresan más una emoción o sensación del hablante como reacción a una información nueva, que puede ser un acto de habla, pero también una situación extralingüística.

Para la categorización de las fórmulas oracionales nos basamos precisamente en la reciente y detallada propuesta de López Simó. Así, de acuerdo con esta especialista española, distinguimos cuatro categorías principales: a) *fórmulas de relación interpersonal* (sin distinguir entre activas, reactivas, bi-activas y activo-activas), que son “enunciados del discurso-en-interacción mediante los cuales los sujetos interactantes llevan a cabo [...] ciertos actos de habla interactivos en los que hablante y oyente están implicados y que manifiestan una relación entre ellos” (López Simó 2016: 179); entre los numerosos ejemplos propuestas, mencionamos *Mi más sentido pésame*, *Que te den morcillas o ¿No te parece?*; b) *fórmulas personales* (ya sean sintomáticas o cognitivas),

donde el centro de la atención se sitúa en el hablante, cuyos sentimientos o pensamientos son expresados mediante la UF (*¡Qué guay!*, *¡Maldita sea!* o *¡A otro perro con ese hueso!*). Estos enunciados pueden deberse a la interacción con el oyente (aunque no siempre), pero no expresan ningún tipo de beneficio o perjuicio para este, algo que sí resulta fundamental en las fórmulas de relación interpersonal; c) *fórmulas impersonales*, similares en gran medida a las fórmulas personales, pero donde el foco se sitúa en el mensaje. Así, el hablante las usa con la intención de expresar certeza, probabilidad o duda (*A la vista está*, *Dios dirá* o *Mentira cochina*); y d) *fórmulas metainteractivas*, que serían aquellas UF centradas en el código, “pues relacionan al hablante (yo) u oyente (tú) con su propio decir y que tienen que ver con la gestión u organización de la conversación, [de modo que] los interlocutores comentan, sitúan o califican el acto de enunciación” (López Simó 2016: 181). Sería el caso de *Al grano*, *Corto y cambio* o *Eso es todo*.

4.3. Paremias

En este tercer grupo se incluirían refranes, proverbios, citas y otras expresiones que son, sin lugar a duda, enunciados independientes y que poseen, además, una naturaleza marcadamente distinta a la de las locuciones. El Centro Virtual Cervantes⁹ diferencia en su *Refranero multilingüe* los siguientes subtipos: aforismo, apotegma, cliché, dialogismo, frase hecha, frase proverbial, locución proverbial, máxima, proverbio, refrán, refrán dialogado, refrán meteorológico, refrán temporal y sentencia. Estas expresiones carecerían de una integración real en el discurso, siendo, de algún modo, supradiscursivas, ya que generalmente se pronuncian tras una pausa o incluso deben ser presentadas por el hablante (*Como decía mi abuela...*, *Ya se sabe que...*, *No en vano se dice...*, etc.), lo que no sucede con las fórmulas oracionales, perfectamente integradas en el discurso. A pesar de que hoy en día se da por hecho que la paremiología es la disciplina encargada de su estudio y que, por tanto, la fraseología debería centrarse en las locuciones, en las fórmulas oracionales y, como mucho, en las colocaciones, no cabe duda de que la tradición española (Casares, Zuluaga, Corpas Pastor) las ha considerado parte integrante de la concepción nuclear de la fraseología y, como tal, las vamos a recoger en nuestro estudio.

5. Metodología del análisis

Una vez aclarada y justificada la elección de *Gran Hermano* como fuente de material lingüístico frente a otras más convencionales y presentado, a grandes rasgos, el estado de la cuestión, pasamos a detallar el modo en que se extrajeron las UF del *reality show*. En primer lugar, se realizaron grabaciones de varias horas en distintos días del canal 24 horas de la 16ª edición de *Gran Hermano* España, entre noviembre y diciembre de 2015, es decir, cuando los participantes ya habían convivido durante más de dos meses en la casa y, por tanto, habían mantenido conversaciones entre sí durante un periodo de tiempo potencial de unas 1 500 horas. Recalcamos el hecho de que se trata de fragmentos obtenidos al azar del canal 24 horas y no del resumen diario donde se

9 <https://cvc.cervantes.es/lengua/refranero/>.

recogen algunos de los momentos más polémicos y que, por tanto, podrían estar especialmente dirigidos por la producción del programa. Para el estudio se analizaron diálogos en contextos de comunicación espontánea sobre temáticas generales, descartando, por tanto, los monólogos en el “confesionario” y las conexiones en directo con el programa diario o semanal.

Para el estudio de carácter exploratorio presentado en este artículo se han analizado exactamente 120 minutos de grabación (que se corresponden a 10 021 palabras), concretamente un fragmento de 85 minutos del día 19 de noviembre y otro de 35 minutos del día 28 de noviembre. Debe aclararse que los 120 minutos no son de conversación continua, puesto que se dan varios momentos de silencio entre las conversaciones, además de que, a causa del ruido ambiental o de conversaciones cruzadas, algunas partes no han podido ser analizadas, dada la imposibilidad de una recogida efectiva de datos.

A continuación, realizamos personalmente varios visionados de los vídeos y anotamos en una hoja de cálculo todas las expresiones institucionalizadas, junto a otras informaciones que podrían tener importancia en el ulterior análisis de los datos, como quién habla (y a quién lo hace) o el contexto comunicativo (relax, discusión, juego, etc.). Después, se evaluó de la forma más sistemática posible si las expresiones anotadas eran o no UF. En muchos casos tuvimos dudas al respecto, especialmente cuando las expresiones no se hallaban recogidas en ninguno de nuestros diccionarios de referencia (*DELE*, *DFDEA*, *DFEM*, *DiLEA* y *DRAE*). Por último, se contó el número de palabras pronunciadas por cada uno de los participantes para, de este modo, poder calcular la frecuencia relativa de UF por cada mil palabras (ver sección 6.2.).

En cuanto a los participantes, al comienzo de esta edición de *Gran Hermano* su número era de dieciocho. Sin embargo, en el periodo analizado para este estudio vivían en la casa tan solo diez concursantes: seis hombres (Carlos, Aritz, Suso, Han, Ricky y Vera) y cuatro mujeres (Marta, Niedziela, Marina y Sofía). Su rango de edad oscilaba entre los 19 y los 28 años –es decir, claramente pertenecientes a la Generación 1 a la que se refiere el grupo PRESEEA (Moreno Fernández *et al.* 2005), que abarca hasta los 34 años– y tres de ellos cursaban en ese momento estudios superiores (Marta, Sofía y Han). En general cabe hablar de que se trata de un grupo formado por individuos de clase media o media alta con estudios medios o superiores. Ocho de ellos son españoles. Han es de origen chino, aunque se mudó con doce años junto a su madre a Granada, concretamente a la localidad de Armilla¹⁰, y muestra un gran dominio del español, como demostraría el hecho de que cursara estudios de Ingeniería Eléctrica en la Universidad de Granada. El otro concursante extranjero es Vera, nacido en Argentina, pero que, en el momento de su participación en el concurso, llevaba ya dieciséis años residiendo en Gran Canaria. Su acento es, a primera vista, canario, pero su origen argentino será un elemento para tener en cuenta. Más importante aún nos parece el hecho de que entrara en la casa a finales de octubre, es decir, que en el momento del análisis el número de horas de conversaciones coloquiales entabladas era considerablemente menor que el de sus compañeros. Por último, habría que comentar que Niedziela es de origen polaco, si bien ya nació en España y, por lo tanto, la consideramos una hablante nativa de español más.

10 cfr. https://www.lecturas.com/actualidad/desvelamos-el-duro-pasado-de-han_17606.

6. Presentación e interpretación de los datos obtenidos del análisis

En este sexto apartado vamos a presentar los datos obtenidos del análisis de las conversaciones de *Gran Hermano* desde tres perspectivas: la cantidad y tipología de las UF documentadas, su frecuencia de uso y, por último, un estudio de carácter lexicográfico que incluye los registros asociados a las UF recogidas en estas conversaciones.

6. 1. Cantidad y tipología de las unidades fraseológicas documentadas

En las 10 021 palabras documentadas a lo largo de los 120 minutos de grabación analizados, hemos recogido 267 UF distintas (tipos) y 553 ocurrencias, lo que evidencia que, de media, cada UF ha sido empleada poco más de dos veces.

En la Tabla 1 presentamos los datos pormenorizados relativos a la distribución de las UF registradas, tanto en cuanto al inventario de UF (tipos), como a números absolutos, es decir, teniendo en cuenta las 553 entradas totales (ocurrencias).¹¹

Clase de UF	Subclase	UF (tipos)	UF (ocurrencias)
Locución		169 (63,3%)	379 (68,5%)
	<i>adjetiva</i>	7 (2,6%)	7 (1,3%)
	<i>adverbial</i>	64 (24,0%)	113 (20,4%)
	<i>conjuntiva</i>	17 (6,4%)	24 (4,3%)
	<i>nominal</i>	12 (4,5%)	26 (4,7%)
	<i>marcadora</i>	23 (8,6%)	135 (24,4%)
	<i>prepositiva</i>	3 (1,1%)	3 (0,5%)
	<i>verbal</i>	43 (16,1%)	71 (12,8%)
Fórmula oracional		97 (36,3%)	173 (31,3%)
	<i>FO relación interpersonal</i>	46 (17,2%)	97 (17,5%)
	<i>FO personal</i>	42 (15,7%)	63 (11,4%)
	<i>FO impersonal</i>	7 (2,6%)	11 (2,0%)
	<i>FO metainteractiva</i>	2 (0,7%)	2 (0,4%)
Paremia		1 (0,4%)	1 (0,2%)
TOTAL		267 (100%)	553 (100%)

Tabla 1. Distribución de las UF documentadas en *Gran Hermano*.

De los datos presentados en la Tabla 1 cabe extraer las siguientes conclusiones. Primero, el uso de las paremias en las conversaciones llevadas a cabo en situaciones comunicativas informales por

¹¹ El listado completo de las locuciones y fórmulas oracionales documentadas puede ser consultado en los anexos situados al final del artículo.

miembros de la Generación 1 es enormemente infrecuente, con apenas el 0,2% del total de ocurrencias. Se trataría de una única entrada a lo largo de los 120 minutos analizados, concretamente *Nunca cantes victoria antes de tiempo* que, de acuerdo con el *Refranero multilingüe* del Centro Virtual Cervantes, sería una variante que “se utiliza más” de *No cantes gloria hasta el fin de la victoria*¹². Se trata, además, de un caso algo dudoso, ya que algunas de nuestras fuentes lexicográficas de referencia documentan la locución verbal *Cantar victoria*. A pesar de ello, consideramos que la variante con “antes de tiempo” y el imperativo negativo del verbo *cantar* se acerca más a la forma de la paremia que a la de la locución.

Respecto a las paremias (y refranes), por tanto, no podemos, sino confirmar la apreciación de Ruiz Gurillo, quien afirmara que en su propio estudio de la fraseología del español coloquial “las paremias, y en concreto los refranes, son escasos” (Ruiz Gurillo 1998: 47). Sin embargo, la autora opina que este tipo de expresión se vería sustituido con frecuencia por enunciados fraseológicos bimembres del tipo *De eso nada, monada*. En contra de esta apreciación, en nuestro estudio no se registra ni una sola ocurrencia de este tipo de enunciado bimembre.

Por otra parte, en nuestro corpus las locuciones son más frecuentes que las fórmulas oracionales (el 63,3% de los tipos frente al 36,3%), diferencia esta que se amplía aún más si lo que tenemos en cuenta es el número total de ocurrencias (68,5% frente a 31,3%, es decir, más del doble de frecuencia). Estos datos dejarían claro que las locuciones se muestran como una herramienta considerablemente más común incluso en los contextos comunicativos supuestamente más propicios para el uso de las fórmulas oracionales. Sin embargo, sí podría llamar la atención que sean precisamente las locuciones adverbiales, en cuanto a tipos, y las marcadoras, en cuanto a ocurrencias, las subclases más comunes en nuestras dos estadísticas.

Un dato fuera de toda duda se refiere a que las locuciones marcadoras son, de entre todas las subclases, las que con un menor número de tipos distintos acumulan un mayor número de entradas, con una media de casi seis ocurrencias por cada tipo. En el extremo contrario se situarían las locuciones prepositivas y adjetivas, que, además de ser poco frecuentes en cuanto a tipos (tres y siete, respectivamente), presentan siempre una única ocurrencia.

En cuanto a las fórmulas oracionales, resulta evidente que hay dos subclases que, al menos en este contexto comunicativo, se emplean con mucha más frecuencia que el resto: las fórmulas interpersonales y las personales, siendo las otras dos subclases (impersonales y metainteractivas) claramente minoritarias.

Por otra parte, y en consonancia con lo expuesto en los párrafos anteriores, entre las UF más repetidas, aquellas con diez o más ocurrencias en nuestro corpus, destacan dos locuciones marcadoras: *es que*, con cuarenta ocurrencias, y *o sea*, con 29. Además, la fórmula oracional de relación interpersonal *venga va/vamos* se documenta en 17 ocasiones, mientras que la locución adverbial *a lo mejor* lo hace en once casos.

De cara a poder interpretar adecuadamente estos datos, queremos compararlos con el único análisis estadístico general de la fraseología española por número de unidades léxicas llevado a cabo hasta el momento, el propuesto por Sosiński (2010, 2011, 2012)¹³. En su análisis de

12 <https://cvc.cervantes.es/lengua/refranero/Busqueda.aspx>. Entrada “victoria”.

13 Decimos “general”, ya que otros autores se han centrado en un único subtipo de locución, como en el caso de los trabajos ya arriba comentados de Penadés Martínez sobre locuciones verbales o de Ruiz Gurillo respecto a sintagmas prepositivos fraseológicos.

la fraseología presente en el corpus semiformal del habla de Granada perteneciente al proyecto PRESEEA y compuesto de dieciocho entrevistas, Sosiński documenta 646 tipos de UF y 2 816 ocurrencias de un total de 82 141 unidades léxicas (Sosiński 2010: 140–141). En principio, los datos obtenidos de ambos corpus (el de PRESEEA Granada y el de *Gran Hermano*), a pesar de su diferente extensión, deberían poder ser comparados de forma satisfactoria, ya que el vaciado de las UF sigue criterios metodológicos similares, siendo la divergencia más llamativa la aceptación en nuestro estudio de las paremias (un elemento claramente insignificante a tenor de su escasísima frecuencia). En cuanto a las locuciones y las fórmulas oracionales (denominadas *fórmulas rutinarias* en los trabajos de Sosiński), la única aparente diferencia metodológica es la inclusión en el estudio de Sosiński de las locuciones clausales –que nosotros enmarcamos dentro de las verbales–, algo que no habría de desvirtuar los datos en absoluto, dado que, de acuerdo con los números aportados por Sosiński, es un tipo de locución que no documenta ni una única ocurrencia en las 82 141 palabras analizadas. En cualquier caso, las enormes diferencias en el tamaño de los corpus no permiten realizar un análisis comparativo de tipos, sino tan solo de ocurrencias.

En la Tabla 2 presentamos los datos obtenidos en ambos estudios por cada mil unidades léxicas.

Tipo de UF	<i>Gran Hermano</i>	PRESEEA Granada	Diferencia
Locución adjetiva	0,7	0,4	+0,3
Locución adverbial	11,3	11,1	+0,2
Locución conjuntiva	2,4	1,2	+1,2
Locución nominal	2,6	3,5	-0,9
Locución marcadora	13,5	7,1	+6,4
Locución prepositiva	0,3	1,7	-1,4
Locución verbal	7,1	2,5	+4,6
Fórmula oracional	17,3	6,7	+10,6

Tabla. 2. Distribución comparativa de las UF (ocurrencias) documentadas en *Gran Hermano* y PRESEEA (Granada) por cada mil unidades léxicas.

De los números expuestos en la Tabla 2 cabe extraer algunas tendencias que podrían marcar las principales diferencias respecto a la distribución fraseológica en contextos semiformales e informales como son, a priori, los aquí analizados. Quizá lo más llamativo sea que en seis de las ocho categorías las UF son más frecuentes en nuestro corpus que en el analizado por Sosiński (todas excepto las locuciones nominales y prepositivas). Sin embargo, estos datos deben ser tomados con gran precaución, dadas las muy diversas formas de entender, limitar y clasificar las UF. De hecho, al revisar más en detalle el trabajo de Sosiński centrado en el estudio de las locuciones nominales (2010), resulta evidente que existe un contraste enorme respecto al propio concepto de UF, ya que un buen número de las locuciones nominales que registra son lo que el DFDEA etiqueta

denominaciones normales, como, por ejemplo, *Fin de semana*, y que nosotros no consideramos UF. De hecho, de las diez locuciones nominales más frecuentes recogidas por Sosiński, nosotros descartaríamos, al menos, siete: *Fin de semana* (30 ocurrencias en el corpus de Granada), *Primera comunión* (13), *Cuarto de baño* (11), *Semana Santa* (10), *Centro de salud* (7), *Tiempo libre* (6) y *Enseñanza secundaria* (5). Estas enormes discrepancias sobre la propia concepción de lo que es y no es el objeto de estudio de la fraseología hacen que, incluso en un caso como este en que la metodología es aparentemente compatible, cualquier tipo de estudio comparativo resulte muy arriesgado.¹⁴ A la vista de estos datos, hemos decidido centrar nuestro análisis tan solo en aquellos tipos de locuciones que podemos cotejar con ciertas garantías. Así, una vez descartadas las locuciones nominales, centramos nuestra atención en los otros dos tipos que Sosiński aborda en sus trabajos: las locuciones verbales (2011) y las locuciones adverbiales (2012). Por desgracia, en ninguno de estos estudios el autor presenta un listado completo de las locuciones registradas en el corpus de Granada, ya que excluye de estos todas las UF con una única ocurrencia. Sin embargo, y a diferencia del artículo sobre locuciones nominales (2010), en el caso de las locuciones verbales y adverbiales observamos pocas diferencias de relevancia respecto a nuestra forma de entender el objeto de estudio de la fraseología, por lo que, siempre con cautela, consideramos que, en este caso sí, los datos obtenidos de los dos corpus pueden ser confrontados con ciertas garantías.¹⁵

Si comenzamos con el análisis comparativo de las locuciones adverbiales –basado, como se comprobará, en frecuencia y registros y, por tanto, adelantando para locuciones verbales y adverbiales los datos que se presentan para el conjunto de las UF en 6.2. y 6.3.–, destaca que apenas se documentan diferencias entre los datos obtenidos de los dos corpus, siendo las locuciones adverbiales muy frecuentes, además, en ambos casos: 11,3 locuciones adverbiales por mil palabras en las conversaciones de *Gran Hermano* frente al 11,1 del corpus de Granada, una diferencia de apenas 0,2 ocurrencias por mil palabras.

<i>Gran Hermano</i>		PRESEEA Granada	
Locución	Ocurrencias	Locución	Ocurrencias
<i>a lo mejor</i>	11	<i>un poco</i>	154
<i>un poco</i>	8	<i>sobre todo</i>	83
<i>al final</i>	6	<i>a lo mejor</i>	59
<i>al revés</i>	5	<i>a veces</i>	33
<i>en plan</i>	5	<i>más o menos</i>	28
<i>a veces</i>	4	<i>al final</i>	26
<i>un montón</i>	4	<i>ahora mismo</i>	24
<i>por encima</i>	3	<i>nada más</i>	17

14 Debemos insistir en que consideramos errónea la metodología empleada por Sosiński, lo que explicaría, además, por qué las locuciones nominales presentadas en este estudio están mucho más frecuentemente recogidas por un diccionario general como el DRAE que por los diccionarios fraseológicos (Sosiński 2010: 143–144).

15 En este sentido la mayor divergencia sería la inclusión de *por qué*, considerada por Sosiński locución adverbial, tal y como se indica, de hecho, en el DRAE: “1. loc. adv. Por cuál razón, causa o motivo. ¿Por qué te agrada la compañía de un hombre como ese? No acierto a explicarme por qué le tengo tanto cariño”. Nosotros, en consonancia con las fuentes fraseográficas (DFDEA, DFEM y DELE) no consideramos que *por qué* sea una UF.

<i>Gran Hermano</i>		PRESEEA Granada	
Locución	Ocurrencias	Locución	Ocurrencias
<i>que te cagas</i>	3	<i>en general</i>	16
<i>un huevo</i>	3	<i>por qué</i>	15

Tabla 3. Locuciones adverbiales (ocurrencias) más frecuentemente documentadas en *Gran Hermano* y PRESEEA (Granada).

Tal y como podemos observar en la Tabla 3, cuatro locuciones verbales se hallan entre las diez más frecuentes en ambos corpus: *un poco*, *a lo mejor*, *a veces* y *al final*. Esto vendría a demostrar que son UF muy comunes en cualquier contexto discursivo oral. Otras dos locuciones verbales, concretamente *al revés* y *por encima*, se documentan con más frecuencia en las conversaciones de *Gran Hermano* que en el corpus de Granada, a pesar de contener este último un número de palabras más de ocho veces superior. La explicación creemos hallarla precisamente en el hecho de que el corpus de *Gran Hermano* sea relativamente pequeño y que, por tanto, una situación discursiva concreta puede colocar UF de un uso generalmente poco común entre las más frecuentes. De hecho, las cinco ocurrencias de *al revés* se documentan en apenas dos de los catorce contextos discursivos analizados en *Gran Hermano*. Algo muy parecido cabe decir de *por encima*, ya que sus tres ocurrencias se concentran en una única conversación. El dato más interesante, y lo que marca, sin lugar a dudas, la mayor discrepancia entre los datos obtenidos de ambos corpus, es que entre las diez locuciones adverbiales más comunes de *Gran Hermano* se documentan dos UF coloquiales (*un montón* y *en plan*) y dos vulgares (*que te cagas* y *un huevo*), todas ellos repartidas en, al menos, dos conversaciones distintas. Por tanto, los datos relativos a las locuciones adverbiales señalan que la diferencia entre los dos registros no es tanto una cuestión de frecuencia absoluta, como de grado de formalidad de las UF empleadas, siendo mucho más comunes las locuciones adverbiales coloquiales y vulgares en las conversaciones que cumplen los principales rasgos coloquializadores propuestos por Briz (ver apartado 1).

A diferencia de las locuciones adverbiales, sí documentamos divergencias de importancia en los datos relativos a la frecuencia de uso de las locuciones verbales, documentadas casi tres veces más frecuentemente en las conversaciones de *Gran Hermano* (7,1 ocurrencias por mil palabras) que en el corpus semiformal de Granada (2,5).

<i>Gran Hermano</i>		PRESEEA Granada	
Locución	Ocurrencias	Locución	Ocurrencias
<i>ser/dar igual</i>	8	<i>llamar la atención</i>	18
<i>hacer falta</i>	4	<i>darse cuenta</i>	15
<i>tener que ver</i>	3	<i>tener ganas</i>	8
<i>llevarse [+ valoración]</i>	3	<i>dar igual</i>	7
<i>tocar/chupar la polla</i>	3	<i>dársele bien</i>	6
<i>saber mal</i>	3	<i>hacer tiempo</i>	6

<i>Gran Hermano</i>		PRESEEA Granada	
Locución	Ocurrencias	Locución	Ocurrencias
-	-	<i>tener claro</i>	6
-	-	<i>dar tiempo</i>	6
-	-	<i>merecer la pena</i>	5
-	-	<i>tener que ver</i>	4

Tabla 4. Locuciones verbales (ocurrencias) más frecuentemente documentadas en *Gran Hermano* y PRESEEA (Granada).

La comparación en este caso resulta más compleja, dado que en *Gran Hermano* apenas se documentan seis locuciones verbales con tres o más ocurrencias y no consideramos oportuno estudiar aquellas locuciones que apenas se documentan en una o dos ocasiones. Aun así, nuevamente hay algunas UF que presentan, en ambos corpus, una frecuencia elevada (*ser/dar igual* y *tener que ver*) y que se suponen, por tanto, comunes a cualquier registro oral del español. De igual modo, destaca en las conversaciones de *Gran Hermano* una mayor presencia de UF coloquiales (*saber mal*, al menos según se etiqueta en el *DFEM*, no así en el *DFDEA*) y vulgares (*tocar/chupar la polla*).

Por último, el hecho de que se documente una mucha mayor presencia relativa de locuciones verbales en contextos coloquializadores parece indicar que este tipo de locución es mucho más característico de los registros informales que las locuciones adverbiales. Este hecho estaría relacionado con la mayor presencia relativa de locuciones verbales propias de registros informales en las conversaciones de *Gran Hermano*: el 55% de las locuciones verbales (tipos) registradas en nuestra fuente de referencia principal (*DFDEA*) está etiquetada como *coloquial*, *juvenil* o *vulgar*. Por el contrario, esta cifra se reduce prácticamente a la mitad (28%) en el caso de las locuciones adverbiales. La duda que surge es por qué serían precisamente las locuciones verbales, frente a las adverbiales en este caso, más propias del discurso coloquial. Una posible explicación, que habrá de ser estudiada en el futuro, es que en los registros coloquiales hay una mayor frecuencia de uso de UF expresivas y que, precisamente, esta expresividad es una característica más propia de las locuciones verbales que adverbiales del español.¹⁶

6. 2. Estudio de frecuencias

Al igual que ocurriera con las cifras aportadas en el apartado anterior, son muy escasos los trabajos publicados sobre la frecuencia de uso general de la fraseología. Es cierto que existen estudios que han investigado este aspecto en español, si bien generalmente se centran en la búsqueda de

16 Por otra parte, la expresividad parece estar asociada al grado de idiomatidad, aunque no existen estudios empíricos que aclaren de forma taxativa este punto. Curiosamente, Penadés Martínez (2015b) afirma que las locuciones verbales, dada su escasa frecuencia, no constituyen un rasgo característico del registro coloquial y que, además, la mayoría de las documentadas en el corpus del Barrio de Salamanca (Madrid) no son idiomáticas. Sus conclusiones, por tanto, irían en contra de las presentadas en esta investigación. Sin embargo, en nuestra opinión, podría ser precisamente la diferencia en los registros, coloquial en *Gran Hermano* frente a semiformal (o “coloquial periférico” como ella lo denomina) en el corpus PRESEEA, lo que podría generar estas diferencias.

una lista cerrada de UF en grandes corpus textuales (Corpas Pastor 1998; Castillo Carballo 2000; Boers, Stengers 2008; algo similar cabe decir de varios trabajos en inglés como, por ejemplo, los publicados por Moon 1997; 1998). Las investigaciones más interesantes en este sentido serían las ya reiteradamente mencionadas de Penadés Martínez (2004; 2012; 2015a; 2015b) y de Sosiński (2010; 2011; 2012), si bien, como ya hemos podido observar en 6.1., ni siquiera en estos casos resulta sencillo equiparar los datos obtenidos a causa de las diferentes metodologías empleadas.

Una cuestión relevante a este respecto, si bien de carácter orientativo, lo encontramos en las palabras de Penadés Martínez para quien, a pesar de haberse incrementado notablemente el número de corpus orales a la disposición de los investigadores:

[...] la búsqueda de unidades fraseológicas en ellos puede ser, si no infructuosa, sí un tanto descorazonadora: las locuciones, al menos, no abundan en ellos tanto como el investigador desearía ni como habitualmente se cree y se afirma; cosa distinta puede que sea la utilización, incluso abusiva, de fraseologismos en los guiones que reproducen situaciones de la vida diaria, como los que se hacen para series de televisión del tipo *Cámara Café*, *Aquí no hay quien viva* o *Aída*, pero ya digo, son *recreaciones*, no transcripciones de conversaciones reales (Penadés Martínez 2015c: 76).¹⁷

En las conversaciones de *Gran Hermano* registramos 26,7 tipos y 55,2 ocurrencias por cada mil palabras. Así, se observa una muy mayor frecuencia fraseológica que en los datos aportados para el corpus de Granada: 7,9 tipos y 34,3 ocurrencias por cada mil palabras. Si bien la comparación de la frecuencia relativa de tipos se ve sin duda seriamente afectada por el tamaño de los corpus con los que se ha trabajado (siendo el analizado por Sosiński hasta ocho veces más extenso que el nuestro), los datos relativos a ocurrencias sí pueden ser comparados de manera sistemática y deberían darnos una idea clara de la situación. Por tanto, cabría afirmar que las UF son aproximadamente 1,6 veces más frecuentes en las conversaciones llevadas a cabo en situaciones comunicativas informales (*Gran Hermano*) que en las semiformales del corpus de entrevistas y conversaciones de Granada. Sin embargo, como ya anunciáramos anteriormente, estas cifras deben ser tomadas con suma cautela.

Queremos aún comparar nuestros datos con los de otro corpus del proyecto PRESEEA, en este caso los propuestos por Penadés Martínez (2004 y 2015a). Sin embargo, nos encontramos con varias dificultades a la hora de realizar esta comparación de forma eficaz. Primero, Penadés Martínez trabaja solo con los datos relativos a las locuciones verbales, lo que, sin duda, reduce notablemente el alcance de las conclusiones a las que podemos llegar. Además, su trabajo más completo, basado en el corpus de Madrid (Barrio de Salamanca) no especifica el total de unidades léxicas analizadas, apenas se señala que se trata de un subcorpus de 36 entrevistas de las 54 que constituyen el corpus PRESEEA-MADRID (distrito de Salamanca), que comprenden, en total, aproximadamente 36 horas de grabación (2015a) y que el análisis queda restringido a las palabras pronunciadas por los informantes (excluyendo, por tanto, las intervenciones de los entrevistadores).

Así, debemos conformarnos con los datos aportados sobre locuciones verbales por la misma autora para el corpus PRESEEA-ALCALÁ. En este caso sí aporta el número de unidades léxicas

17 En Gutiérrez Rubio (2020) analizamos precisamente el uso de la fraseología en un guion televisivo (concretamente en las conversaciones de la serie de televisión *Cuéntame cómo pasó*) y comparamos los resultados obtenidos con los de diálogos reales recogidos en los corpus del grupo Val.Es.Co. y de *Gran Hermano*.

analizadas, dado que se supone que ha trabajado con el corpus completo, es decir, con 322 000 palabras. Aquí el problema surge a causa de que, al tratarse de un trabajo previo a la publicación de la primera edición del *DFDEA*, la consideración o no de una UF como locución verbal se basa en los datos aparecidos en el *Diccionario del español actual (DEA)*, obra que, por otra parte, será la base del propio *DFDEA*, pero que, al tratarse de un diccionario general, contiene un número significativamente menor de UF. Además, de entre las locuciones verbales documentadas en el *DEA* la autora descarta todas aquellas que, de acuerdo con la definición del diccionario (pero en contradicción con su propia clasificación), puedan ser analizadas como “una locución adverbial, adjetiva o de cualquier otra clase que no correspondiera a la de las locuciones verbales” (Penadés Martínez 2004: 2229). De acuerdo con estos parámetros, en el corpus de Alcalá se documentan 158 tipos de locuciones verbales y 515 ocurrencias en un total de aproximadamente 322 000 palabras. Por tanto, el resultado es de aproximadamente 0,5 tipos y 1,6 ocurrencias cada mil palabras. Si ajustáramos los datos extraídos de las conversaciones de *Gran Hermano* a la metodología empleada por Penadés Martínez, es decir, tomando en cuenta tan solo aquellas locuciones verbales que se documentan en el *DEA* y no pueden ser analizadas como otro tipo de locución, obtendríamos un total de 22 tipos (doce de los cuales se hallan, además, entre los 158 documentados por Penadés Martínez) y 44 ocurrencias en 10 021 palabras, es decir, aproximadamente 2,2 tipos y 4,4 ocurrencias por cada mil palabras. De acuerdo con estos datos, las diferencias en cuanto a la frecuencia de uso de las locuciones verbales son inesperadamente elevadas. Así, en *Gran Hermano* las locuciones verbales son 4,4 veces más frecuentes, si hablamos de tipos, y 2,75 veces más frecuentes, si hablamos de ocurrencias, que en las entrevistas y conversaciones del corpus de Alcalá. Es cierto que la ratio de tipos se ve sin duda afectada por el número de palabras analizadas, pero no debería ser así con las ocurrencias. Si lo comparamos, por otra parte, con las cifras relativas a las locuciones verbales presentadas por Sosiński (7,1 UF por cada mil palabras en *Gran Hermano* frente a 2,5 en el corpus de Granada, ver Tabla 2), observamos que los datos son prácticamente idénticos a los presentados por Penadés Martínez, y eso a pesar de que ambos especialistas emplean metodologías ciertamente distintas: las ocurrencias de locuciones verbales en *Gran Hermano* serían 2,84 veces más frecuentes que en el corpus de Granada y 2,75 que en el de Alcalá. La única explicación razonable que encontramos a estos resultados es que, tal y como señalan algunos especialistas, y recoge, aunque no comparte, Penadés Martínez (2015b: 251): “las unidades fraseológicas caracterizan el registro coloquial de la lengua por la frecuencia con la que se usan y por la expresividad que conllevan”.

6. 3. Estudio lexicográfico y de registros

En este apartado trataremos de mostrar si las UF empleadas por los concursantes de *Gran Hermano* se hallan o no recogidas en la segunda edición del *Diccionario fraseológico documentado del español actual: locuciones y modismos españoles (DFDEA)*, el más completo y actual de los publicados en español hasta el momento, y eso a pesar de las críticas de que (su primera edición) es objeto por parte de Penadés Martínez (2015c). Puesto que se trata de un diccionario del español actual y moderno, y no podemos descartar la aparición de expresiones arcaicas, hacemos uso asimismo del *DRAE* que, si bien es un diccionario de carácter general, incluye numerosísimas UF, incluidas

las arcaicas, con lo que lo consideramos un importante complemento al *DFDEA*. Por último, tal y como ya ha sido anteriormente comentado, la única paremia documentada se ha confirmado en el *Refranero multilingüe* del Centro Virtual Cervantes.

En este apartado no nos interesa tan solo conocer si el diccionario en cuestión recoge las UF, si no, además, se tiene en cuenta si le añade alguna marca de registro, como *coloquial*, *informal*, *grosera*, *arcaica*, *frecuente*, etc. Uno de los problemas básicos de este tipo de estudio lexicográfico es conocer en qué criterios se basan los autores de los diccionarios para distribuir estas etiquetas, tal y como advierte Penadés Martínez (2015c: 180) para el caso del *DELE*, y el uso sistemático, pero no justificado, de las marcas *poco frecuente (pf)*, *frecuente (f)* y *muy frecuente (mf)*. Aunque el centro de nuestro interés no sea la frecuencia, sino el carácter coloquial, informal o vulgar de las UF, no cabe duda de que la reflexión de Penadés Martínez puede extrapolarse perfectamente a la falta de criterios objetivos a la hora de etiquetar el registro propio de una UF.¹⁸ Así, opinamos que la solución más adecuada sería hacer el estudio desde el uso hacia el diccionario: aquellas expresiones que se documenten preferentemente en conversaciones realizadas en situaciones comunicativas informales deberían ser etiquetadas como coloquiales.

En cuanto a los datos obtenidos del estudio, de las 266 UF documentadas –una vez excluido el refrán, dado que el diccionario no recoge este tipo de expresiones– 179 (67%) están registradas en el *DFDEA* con una forma y un uso idénticos (o, al menos, muy similares) a los empleados en las conversaciones analizadas, mientras que 87 (33%) no se documentan en nuestra fuente fraseográfica principal. Por tanto, y a pesar de estar ante conversaciones de elevado carácter coloquial, algo más de dos de cada tres UF sí se registran en el diccionario, lo que, *a priori*, ya que no contamos con datos con los que comparar estas cifras, nos parece un porcentaje relativamente elevado que demostraría que el *DFDEA* es un diccionario útil incluso para usuarios interesados en la vertiente más informal del español. Y mucho más aún si tenemos en cuenta que la mayoría de las UF no documentadas en este diccionario son fórmulas oracionales, ya que, con 45 tipos no registrados en el *DFDEA*, estas suponen el 52% de los 87 casos, a pesar de formar tan solo el 36% del total de UF (ver Tabla 1). Este desequilibrio entre los datos documentados en las conversaciones y los relativos a la inclusión o no en los diccionarios se puede deber, probablemente, a que las fórmulas oracionales se hallan en los márgenes del objeto de los estudios fraseológicos y, por tanto, tienden a ocupar un lugar secundario en las investigaciones y diccionarios fraseológicos, que parecen recoger, como hemos podido observar, tan solo las fórmulas oracionales más representativas.

En el caso del *DRAE*, el número de UF recogidas es menor: 117 de las 266 UF documentadas en las conversaciones de *Gran Hermano*, es decir, algo menos de la mitad (44%). Si bien es cierto que una diferencia entre los dos diccionarios de 23 puntos puede parecer importante, teniendo en cuenta que el *DRAE* es un diccionario generalista y el *DFDEA*, uno fraseológico (y actual), realmente consideramos que son datos, hasta cierto punto, sorprendentes. Por otra parte, resulta indudable que la información que encontramos en el *DFDEA* suele ser mucho más completa, precisa y actualizada que en el *DRAE*.

En la Tabla 5 podemos observar los datos relativos a los dos diccionarios en función de la tipología de las UF.

18 El caso de las expresiones vulgares probablemente resulte menos problemático, por otra parte, ya que estas expresiones suelen incluir vocablos que cabe denominar vulgares, un elemento mejor estudiado desde la lexicología.

Clase	DFDEA		DRAE	
	Registrado	No registrado	Registrado	No registrado
Locuciones	75%	25%	52%	48%
Fórmulas rutinarias	54%	46%	28%	72%
Total	67%	33%	44%	56%

Tabla. 5. Distribución por tipologías de las UF registradas en los diccionarios.

Un dato curioso en este sentido es que la mayor presencia de las locuciones respecto a las fórmulas oracionales resulta relativamente similar en ambos diccionarios, dado que el porcentaje es un 21% más alto en las locuciones en el caso del *DFDEA* y un 24% en el caso del *DRAE*.

Uno de los principales elementos que pretende sacar a la luz esta investigación se refiere a la proporción de expresiones coloquiales y vulgares empleadas en situaciones comunicativas informales. A pesar de que podríamos enmarcar fácilmente muchas de las UF no registradas en los diccionarios dentro de uno de estos registros, hemos decidido, por una cuestión de sistematización, trabajar tan solo con aquellas UF recogidas en nuestros dos diccionarios de referencia. Así, de las 187 UF registradas entre las dos fuentes lexicográficas consultadas, 79 presentan alguna marca de registro: 55 coloquial, 19 vulgar, 2 popular, 2 juvenil y 1 regional. Sin prescindir de la última etiqueta, resulta que el 42% de las UF (tipos) documentadas en *Gran Hermano* y registradas en alguno de los diccionarios serían propias de registros no formales, es decir, a pesar de tratarse de conversaciones llevadas a cabo en situaciones comunicativas informales, estaríamos relativamente lejos de alcanzar siquiera la mitad del total de las UF documentadas. Si pasamos nuestra atención ahora del estudio de tipos al de ocurrencias, observamos que el grado de coloquialidad incluso disminuye: de las 423 ocurrencias de UF que se hallan registradas en nuestras dos fuentes lexicográficas, 149 son propias del registro informal: coloquial (118), popular (2), juvenil (2) o vulgar (27). En otras palabras, apenas el 35% del total. Estos datos vendrían a probar que las UF mayoritarias, incluso en los contextos que Briz propone como altamente coloquializadores, son las que cabe etiquetar de neutras. Por el contrario, las cifras que sí consideramos relativamente elevadas son las que hacen referencia a las UF vulgares, dado que significan el 10,2% de los tipos y el 6,4% del total de ocurrencias.

Si comparamos nuevamente estos datos con los obtenidos por Penadés Martínez (2004) del corpus semiformal de Alcalá,¹⁹ el porcentaje de expresiones propias del registro coloquial es muy superior en el *reality show*. Así, el 38% de las ocurrencias de locuciones verbales obtenidas de *Gran Hermano* estarían etiquetadas en el *DFDEA* como *informal* frente a apenas el 18% en el caso de las recogidas en el trabajo de Penadés Martínez (2004). En otras palabras, el grado de coloquialidad de las expresiones de *Gran Hermano* sería 2,1 veces mayor que el de las obtenidas del corpus oral de Alcalá. Esta cifra sería aún más elevada si nos centramos en las locuciones verbales vulgares: estas suponen el 5,6% de las ocurrencias en *Gran Hermano* frente a apenas el 0,2% en el corpus de Alcalá, es decir, 28 veces más frecuentes entre las locuciones verbales.

19 Un tema ya tratado en Gutiérrez Rubio (2019).

6. Conclusiones

A pesar de que este trabajo tiene un carácter eminentemente exploratorio y se basa en una cantidad aún relativamente escasa de material analizado (120 minutos de conversaciones y algo más de 10 000 palabras), consideramos que aporta unos datos muy prometedores que evidenciarían la importancia de emprender estudios que investiguen, desde distintas perspectivas, la fraseología tal como esta se utiliza en distintas situaciones comunicativas, incluyendo, claro está, las informales.

Si comenzamos con la dimensión relativa a la tipología de las UF documentadas, probablemente la principal conclusión sea que, tal y como afirma Ruiz Gurillo (1998: 85–97), las paremias son un recurso expresivo muy poco frecuente, casi insignificante, en la situación comunicativa objeto de estudio, dado que en nuestro corpus representan apenas el 0,2% del total de ocurrencias. Sin embargo, y en contra de la apreciación de Ruiz Gurillo, esta escasa frecuencia no se ve compensada por la presencia de otros recursos lingüísticos como los enunciados fraseológicos bimembres del tipo *De eso nada, monada*. Además, las locuciones son significativamente más frecuentes que las fórmulas oracionales, tanto en cuanto a tipos (63,3% frente a 36,3%), como a ocurrencias (68,5% frente a 31,3%). Entre las primeras destaca la frecuencia de las locuciones adverbiales, en cuanto a tipos, y de las marcadoras, en cuanto a ocurrencias. En el extremo opuesto se situarían las muy poco comunes locuciones prepositivas y adjetivas. Por su parte, las subclases más comunes entre las fórmulas oracionales son, con una marcada diferencia, las interpersonales y las personales.

Los datos más llamativos del estudio de frecuencias señalan que en las conversaciones de *Gran Hermano* se documentan 26,7 tipos y 55,2 ocurrencias por cada mil palabras. Resulta muy complejo afirmar si esta frecuencia de uso es o no elevada; sin embargo, los datos presentados por autores como Penadés Martínez o Sosiński mostrarían cifras significativamente menores, por lo que podría proponerse una frecuencia más elevada de UF en situaciones comunicativas informales que en otros tipos de situaciones orales más tendentes hacia el polo formal.

La dimensión lexicográfica, por su parte, muestra que el 67% de las UF documentadas en las conversaciones de *Gran Hermano* se hallan registradas en el *DFDEA* con una forma y un uso idénticos o muy semejantes, frente al 44% en el caso del *DRAE*. Esta diferencia de 23 puntos nos parece en cierta medida pequeña si tenemos en cuenta que el *DRAE* es un diccionario generalista y la segunda edición del *DFDEA* el diccionario fraseológico más representativo, extenso y actual de la lengua española. En cuanto al grado de coloquialidad de la fraseología empleada en situaciones comunicativas informales, el 42% de las UF (tipos) y el 35% de las ocurrencias documentadas en *Gran Hermano* y registradas en alguno de estos dos diccionarios son propias de registros que cabría denominar, de modo general, informales (coloquial, vulgar, popular o juvenil). Así, la mayoría de las UF que empleamos, incluso en los contextos altamente coloquializadores, son de carácter neutro. En cuanto a las UF etiquetadas como vulgares, se documenta un número que consideramos relativamente elevado y que podría ser uno de los rasgos característicos de la fraseología empleada en esta clase de conversaciones: un 10,2% entre los tipos y un 6,4% entre las ocurrencias. Por último, se ha comprobado que las UF coloquiales son, al menos, el doble de frecuentes en el corpus de *Gran Hermano* que en el semiformal de Alcalá. Sin embargo, la principal diferencia entre estos dos registros parece hallarse en la frecuencia de uso de las UF etiquetadas como vulgar, dado que en el *reality show* son 28 veces más frecuentes que en las conversaciones y entrevistas recogidas por el grupo PRESEEA.

Referencias bibliográficas

- Beinhauer, W. (1985). *El español coloquial* (3a ed. española). Madrid: Gredos.
- Bertomeu Pi, P. (2019). Las peticiones en la conversación coloquial española y alemana: un acercamiento a sus tipos y formas. *Círculo de Lingüística Aplicada a La Comunicación*, 79, 139–160. <<https://doi.org/10.5209/clac.65653>>.
- . (2020). Big Brother Germany and Spain: a Linguistic Characterization from a Theory of Register. *Revista de Filología Alemana*, 28, 125–146. <<https://doi.org/10.5209/rfal.70052>>.
- Bignell, J. (2005). *Big brother: reality TV in the twenty-first century*. Basingstoke: Palgrave Macmillan.
- Boers, F.; & H. Stengers (2008). A Quantitative Comparison of the English and Spanish Repertoires of Figurative Idioms. In F. Boers, & S. Lindstromberg (Eds.), *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology* (pp. 355–374). Berlin/New York: Mouton de Gruyter.
- Briz, A. (Coord.) (2002). *La conversación coloquial (Materiales para el estudio)*. Madrid: Arco/Libros.
- . (2010a). *El español coloquial: situación y uso*. 6a ed. Madrid: Arco/Libros.
- . (2010b). Lo coloquial y lo formal, el eje de la variedad lingüística. In R. M. Castañer Martín, & V. Lagüéns Gracia (Coord.), *De moneda nunca usada: Estudios dedicados a José M^a Enguita Utrilla* (pp. 125–133). Zaragoza: Instituto Fernando El Católico, CSIC.
- Briz, A.; & Albelda, M. (2013). Una propuesta teórica y metodológica para el análisis de la atenuación lingüística en español y portugués. La base de un proyecto en común (ES.POR.ATENUACIÓN). *Onomázein. Revista semestral de lingüística, filología y traducción*, 28, 288–319.
- Casares, J. (1950/1992). *Introducción a la lexicografía moderna*. Madrid: CSIC.
- Cascón Martín, E. (1995). *Español coloquial. Rasgos, formas y fraseología de la lengua diaria*. Madrid: Edinumen.
- Castañares, W. (2003). La vida en directo, el último espectáculo televisivo. In G. Gómez Rodríguez, & M. M. Renero Quintanar, *Tv Global y espectáculos locales* (pp. 127–145). Guadalajara (Méx.): Universidad de Guadalajara.
- . (1995). Géneros realistas en televisión: los reality shows. *CIC: Cuadernos de información y comunicación*, 1, 79–92.
- Castillo Carballo, M. A. (2000). Lingüística de corpus y fraseología: algunas consideraciones sobre las locuciones verbales. In G. Corpas Pastor (Ed.), *Las lenguas de Europa: estudios de fraseología, fraseografía y traducción* (pp. 95–126). Granada: Comares.
- Centro Virtual Cervantes (s.f.). *Refranero multilingüe* (<https://cvc.cervantes.es/lengua/refranero/>).
- Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid: Gredos.
- . (1998). El uso de paremias en un corpus del español peninsular actual. In G. Wotjak (Ed.), *Estudios de fraseología y fraseografía del español actual* (pp. 365–390). Frankfurt am Main/Madrid: Vervuert/Iberoamericana.
- DEA = Seco, M. et al. (1999). *Diccionario del español actual*. Madrid: Aguilar.
- DELE = Martínez López, J. A.; & Jørgensen, A. M. (2009). *Diccionario de expresiones y locuciones del español*. Madrid: Ediciones de la Torre.
- DFDEA = Seco, M. et al. (2017). *Diccionario fraseológico documentado del español actual: locuciones y modismos españoles*. 2ª edición. Madrid: Aguilar.
- DFEM = Varela, F.; & Kubarth, H. (1994). *Diccionario fraseológico del español moderno*. Madrid: Gredos.
- DiLEA = Penadés Martínez, I. (2019). *Diccionario de locuciones idiomáticas del español actual* (<http://www.diccionariodilea.es/diccionario>).

- DRAE = Real Academia Española (2014). *Diccionario de la lengua española* [versión 23.3 en línea] (<https://dle.rae.es>).
- Dovey, J. (2004). It's only a game show: Big Brother and the theatre of spontaneity. In E. Mathijs, & J. Jones (Eds.), *Big Brother International. Formats, Critics and Publics* (pp. 232–248). London & New York: Wallflower Press.
- García-Page Sánchez, M. (2009). *Introducción a la fraseología española. Estudio de las locuciones*. Barcelona: Anthropos.
- Gutiérrez Rubio, E. (2019). Typologies of phraseological units and their frequency in the reality show *Gran Hermano*. *Phrasis*, 3, 246–252.
- . (2020). Realidad, *reality* y ficción. Análisis de la fraseología española en tres contextos comunicativos informales. *Lingüística y Literatura*, 79, 321–344. <<https://doi.org/10.17533/udea.lyl.n77a15>>.
- Haugh, M.; & Sinkeviciute, V. (2018). Accusations and interpersonal conflict in televised multi-party interactions amongst speakers of (Argentinian and Peninsular) Spanish. *Journal of Language Aggression and Conflict*, 6 (2), 248–270. <https://doi.org/10.1075/jlac.00012.hau>.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lázaro Carreter, F. (1980). *Estudios de lingüística*. Barcelona: Crítica.
- López Simó, M. (2016). *Fórmulas de la conversación. Propuesta de definición y clasificación con vistas a su traducción español-francés, francés-español*. Tesis doctoral inédita. Universidad de Alicante.
- Lorenzo-Dus, N.; & Garcés-Conejos Blitvich, P. (2013). *Real Talk: Reality Television and Discourse Analysis in Action*. Basingstoke: Palgrave Macmillan.
- Mathijs, E.; & Jones, J. (2004). Introduction: Big Brother International. In E. Mathijs, & J. Jones (Eds.), *Big Brother International. Formats, Critics and Publics* (pp. 1–8). London & New York: Wallflower Press.
- Miranda, A. (1998). *Usos coloquiales del español*. 2a ed. corregida y aumentada. Salamanca: Editorial Colegio de España.
- Moon, R. (1997). Vocabulary Connections: Multi-word Items in English. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 40–63). Cambridge: Cambridge University Press.
- Moon, R. (1998). *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. New York: Clarendon Press.
- Montoro del Arco, E. T. (2006). *Teoría fraseológica de las locuciones particulares. Las locuciones prepositivas, conjuntivas y marcadoras en español*. Frankfurt am Main: Peter Lang.
- Moreno Fernández, F. et al. (2005). *La lengua hablada en Alcalá de Henares. Corpus PRESEEA-ALCALÁ. II. Hablantes de Instrucción Media*. Alcalá de Henares: Universidad de Alcalá.
- Penadés Martínez, I. (2004). ¿Caracterizan las locuciones el registro coloquial? In M. Villayandre Llamazares (Ed.), *Actas del V Congreso de Lingüística General, León, 5–8 de marzo de 2002*, III (pp. 2225–2235). Madrid.
- . (2012). La variación en las locuciones a partir de materiales del PRESEEA (Barrio de Salamanca, Madrid). In A. M. Cestero Mancera, I. Molina Martos, & F. Paredes García (Eds.), *Actas del XVI Congreso Internacional de La Asociación de Lingüística y Filología de la América Latina (Alcalá de Henares, 6–9 de junio de 2011)* (pp. 2081–2091). Alcalá de Henares: Universidad de Alcalá.
- . (2015a). Implicaciones de la frecuencia de uso de las locuciones en la elaboración de un diccionario. *Estudios de Lingüística Universidad de Alicante (ELUA)*, 29, 253–277.

- . (2015b). Las locuciones verbales en el habla de Madrid (distrito de Salamanca). In A. M. Cestero Mancera, I. Molina Martos, & F. Paredes García (Eds.), *Patrones sociolingüísticos de Madrid* (pp. 251–286). Bern: Peter Lang.
- . (2015c). *Para un diccionario de locuciones. De la lingüística teórica a la fraseología práctica*. Alcalá de Henares: Universidad de Alcalá.
- Ruiz Gurillo, L. (1998). *La fraseología del español coloquial*. Madrid: Ariel.
- Seco, M. (1973). La lengua coloquial: “Entre visillos”, de Carmen Martín Gaité. In E. Alarcos *et al.*, *El comentario de textos* (pp. 361–379). Madrid: Castalia.
- Sinkeviciute, V. (2017a). Variability in group identity construction: A case study of the Australian and British *Big Brother* houses. *Discourse, Context & Media*, 20, 70–82.
- . (2017b). What makes teasing impolite in Australian and British English? “Step[ping] over those lines [...] you shouldn’t be crossing”. *Journal of Politeness Research*, 13(2), 175–207.
- Sosiński, M. (2010). Datos sobre las locuciones nominales en el corpus PRESEEA-Granada. In *Actas del VII Congreso Internacional de la Asociación Asiática de Hispanistas* (pp. 137–148). Pekin: Foreign Language Teaching and Research Press.
- . (2011). La norma fraseológica: locuciones verbales en el corpus PRESEEA-Granada. In E. Waluch de la Torre (Ed.), *Encuentros 2010. Vol. 1: La norma lingüística del español* (pp. 119–129). Varsovia: Editorial Instituto de Estudios Ibéricos e Iberoamericanos.
- . (2012). Las locuciones adverbiales en el corpus PRESEEA-Granada: análisis estadístico, lexicográfico y comparativo. In E. Waluch de la Torre, & J. A. Moya Corral (Coords.), *Español hablado. Estudios sobre el corpus: PRESEEA-Granada* (pp. 217–238). Varsovia: Instituto de Estudios Ibéricos e Iberoamericanos de la Universidad de Varsovia.
- Vigara Tauste, A. M. (1987). *Aspectos del español hablado*. 2a ed. Madrid: SGEL.
- Zuluaga, A. (1980). *Introducción al estudio de las expresiones fijas*. Frankfurt a.M./Bern: Verlag Peter D. Lang.

Anexo I. Locuciones registradas en el corpus de *Gran Hermano*

a bocajarro; a la altura; a la defensiva; a lo mejor; a lo/la que; a pulso; a que [+ pregunta con respuesta afirmativa]; a su bola; a última hora; a veces; a ver si [+ oración subordinada]; a ver; abrir los ojos; ahora mismo/mismito; al final; al lado; al principio; al revés; alguna vez; antes de; aun (y) así/ incluso/siquiera así; aunque sea; cada vez que; coger el truco; comer la polla [a alguien]; comerse la cabeza; como (muy) tarde; como el perro y el gato; como mucho; como que; como si fuera una mierda; como un loco; con todo mi amor; correr riesgo; dar (o meter) caña; dar guerra; dar la razón; dar por hecho; dar/pegar para el pelo; dar/pegar un corte; darle [a alguien]; dársele algo [+ valoración]; de cojones; de frente; de golpe; de lujo; de marca; de mierda; de momento; de puta madre; de siempre; de todo; dejar tirado; dejarlo para última hora; después de que; echar de menos; el tema es que...; en cabeza; en contra; en directo; en el fondo; en frente; en las mismas; en opinión; en plan; en realidad; en serio; en teoría; en vez de; entre horas; es decir; es que; es verdad que; está claro que; estar ahí ahí; fundírsele los plomos; hacer falta; hacerse el loco; hasta el gorro; hasta la saciedad; ir/salir de marcha; jugársela; llevar la contraria; llevar(lo) [+ valoración]; llevarse [+ valoración]; lo mismo; lo peor; lo que pasa/ocurre es que; mala leche; mi vida; ni nada; ni por el forro; no por nada; no sé; no sé qué [Ladj.]; no sé qué [Lmarc.]; no sé qué [Lnom.]; no tener (ni) (puta,

*pajolera) idea; o sea / o séase; oler que alimenta; otra vez; para nada; para que; pasarlo [+ valoración]; pero bueno; perro/ito faldero; pillarla; pisarle la cara; ponerse burro; por A por B; por ahí; por ciento; por debajo; por dentro; por ejemplo; por encima; por la cara / su cara bonita; por lo menos; por los pelos / un pelo; por mucho/más que; por si (acaso); pues eso; pues mira; pues sí; (pues) sí que; que es un gusto / da gusto; ¡qué sé yo!; que si [+ repetir el verbo anterior]; que te cagas; quedarse frito; saber mal; según qué; ser el/lo/la que es; ser/dar igual; ser/dar lo mismo; sí, ya; siempre que; sin más; sin que; sin querer; soltarse la melena; tal vez; tanto... como...; te guste o no; (te) mentiría si (te) dijera que; tener (un par de) cojones; tener algo con alguien; tener que ver; tener razón; tener/ser *tu día; tocar/chupar la polla [a alg.]; todas las cosas; todo el día; todo el día; todo el mundo; un euro; un huevo; un montón [Ladv.]; un montón [Lnom.]; un poco; una mierda; una vez que; y mira que; y nada; y tal; y todo; (y) ya está.*

Anexo II. Fórmulas oracionales registradas en el corpus de *Gran Hermano*

a la una, a las dos, (y) a las tres; a lo mejor; a ver; ¿a que sí?; ¿ah sí?; buenos días; como lo oyes / lo que oyes; cómo que no; de nada; de verdad; déjame en paz; déjate de historias; ¿en serio?; es lo que digo; es verdad [fórmula de relación interpersonal]; es verdad [fórmula personal]; eso es; está claro / claro está; estoy hasta la polla; hostia puta, la has cagado; la madre que le parió/trajo; let's go; ¿lo entiendes?; lo más seguro; lo siento; ¿lo ves?; me cago en la leche; me cago en la puta; me importa un bledo; me la pela; me la suda; mira a ver; ni que...; ¿no crees?; no (me) jodas; no es justo; no lo sé; no mientas; no pasa nada; no sé yo; no sé; no seas putas; no veas; no volveré a hacerlo; ¿o qué?; otra vez; pero qué va a ser [alguien algo]; por eso / ya, por eso te digo; por favor [fórmula personal]; por favor [fórmula de relación interpersonal]; puede ser; ¡qué bien!; ¡qué cabrón!; ¡qué fuerte!; ¡qué guay!; que le* den por (el) culo; ¡qué mal (me parece)!; ¡que no; ¡qué pesa(d)ol!; ¡qué peste!; ¿qué quieres que (le) haga?; ¡que sí!; ¡qué sé yo!; ¡qué suerte!; ¿qué tal?; ¡que te den por (el) culo!; ¡qué va!; ¿sabes lo que te digo / quiero decir?; ¡será cabrón!; ¿sí o no?; ¿sí? ¿No?; si tú supieras; sí, hombre; sí, pero no; ¿te das cuenta?; te digo una cosa; te lo digo (yo); (te) lo juro/prometo; te pido perdón; tres, dos, uno, ya; tú a lo tuyo; ¿tú crees?; ¿tú te crees (que es normal/posible)?; (vamos) digo yo; ¡vaya mierda!; ¡vaya tela!; venga va/vamos; ¡vete a saber!; ¿y eso?; y lo sabes; ¿y qué?; ya está; ya lo sé; ya te digo; ya ves; zas en toda la boca.*



This work can be used in accordance with the Creative Commons BY-SA 4.0 International license terms and conditions (<https://creativecommons.org/licenses/by-sa/4.0/legalcode>). This does not apply to works or elements (such as images or photographs) that are used in the work under a contractual license or exception or limitation to relevant rights.

