Estudio de corpus de los presentadores paremiológicos en español[1]

Enrique GUTIÉRREZ RUBIO
University Palacký Olomouc (REPÚBLICA CHECA)
enrique.gutierrez@upol.cz || http://orcid.org/0000-0001-8877-4446

Recibido: 13/9/2023 || Aceptado: 12/10/2023

Resumen: Este estudio empírico de los presentadores paremiológicos pretende sacar a la luz nueva información sobre su uso en el español contemporáneo. Con este fin, se han analizado 859 apariciones de nueve paremias documentadas en webs españolas recogidas en el *gigacorpus* esTenTen18. Entre las conclusiones más significativas destaca la disparidad en la frecuencia de uso de los presentadores respecto a las nueve paremias estudiadas, así como que la presencia o no de presentadores depende, al menos parcialmente, de la frecuencia de uso de las paremias y del tipo de registro, siendo más comunes los presentadores en textos formales y junto a las paremias más frecuentes. Además, entre los presentadores destaca la escasa presencia de los referentes personales del tipo «como decía mi abuela». Por el contrario, el verbo *decir* sobresale como presentador paremiológico, ya sea junto a un término paremiológico del tipo *refrán/dicho* o sin él.
Palabras clave: Paremiología. Pragmática. Paremia. Español.

Titre : « Étude de corpus des introducteurs parémiologiques en espagnol ».
Résumé : Cette étude empirique des introducteurs parémiologiques a pour but de fournir de nouveaux renseignements quant à leur emploi en espagnol contemporain. Pour ce faire, nous avons analysé 859 résultats de la rechercher sur des sites webs espagnols inclus dans le giga-corpus « esTenTen18 » de neuf parémies documentées. Parmi les conclusions les plus significatives, il convient de relever la disparité que l'on observe dans la fréquence d'utilisation des introducteurs par rapport aux neuf parémies étudiées, ainsi que le fait que la présence ou l'absence de ceux-ci dépende, au moins en partie, de la fréquence d'utilisation des parémies et du type de registre, les introducteurs étant plus fréquents dans les textes formels et avec les parémies les plus courants. En outre, parmi les introducteurs, il faut souligner la faible présence des références personnelles du type « comme disait ma grand-mère ». D'autre part, le verbe *dire* se démarque en tant qu'introducteur parémiologique, qu'il soit juxtaposé ou non à un terme parémiologique du type *proverbe/dicton*.
Mots-clés : Parémiologie. Pragmatique. Parémie. Espagnol.

Title: "Corpus study of paremiologically introductors."
Abstract: This empirical study of paremiological introductors aims to bring to light new information about their actual use in contemporary Spanish. To this end, 859 occurrences of nine paremias documented in Spanish web texts collected in the gigacorpus esTenTen18 have been analysed. Among the most significant conclusions, the disparity in the frequency of appearance of the introductors regarding the nine paremias studied stands out, as does the fact that their occurrence depends, at least partially, on the frequency of use of the paremias and the register, with the introductors appearing more frequently in formal texts and when they go with the most frequently used paremias. In addition, among the introductors, the scarce presence of personal referents of the type "as my grandmother used to say" is remarkable. On the contrary, the verb *to say* stands out as a paremiological introductor, either with or without the use of a paremiological term such as *proverb* or *saying*.
Keywords: Paremiology. Pragmatics. Paremia. Spanish.

[1] La preparación y publicación de este artículo ha sido posible gracias a la ayuda financiera del Fondo para el Apoyo de Actividades Científicas de la Facultad de Filosofía y Letras de la Universidad Palacký de Olomouc (2023-2025) (Proyecto FPVC2023/03).

INTRODUCCIÓN

Hasta el momento, el estudio empírico de las paremias mediante análisis de corpus ha sido un campo de investigación relativamente inexplorado en español, y eso a pesar de la contundente afirmación que Corpas Pastor realizara hace ya veinticinco años: «Una reflexión seria acerca del emplazamiento y función de las paremias en la lengua debe comenzar por el análisis empírico de estas unidades fraseológicas en contexto, teniendo en cuenta los datos fiables que sólo un estudio a gran escala es capaz de aportar» (1998: 368). Una de las causas más probables para la escasez de este tipo de análisis podría hallarse en el hecho de que los datos que se obtenían de los corpus «clásicos» de fácil acceso y manejo –la única opción viable para muchos investigadores hasta hace algunos años– eran claramente insuficientes para realizar análisis concluyentes; resulta ilustrativo el caso de Norrick al que alude Corpas Pastor (1998: 370), quien documentó tan solo dos paremias en dos corpus orales. Sin embargo, la reciente aparición de *gigacorpus* con miles de millones de palabras y eficaces herramientas de búsqueda y, muy especialmente, los de la familia TenTen accesibles a través de la herramienta Sketch Engine, supone una auténtica revolución metodológica respecto al análisis paremiológico. A modo de ejemplo, la paremia *A Dios rogando y con el mazo dando* cuenta con apenas doce entradas válidas en el corpus CREA (España) de la RAE. Por el contrario, en el subcorpus Spanish Web 2018 del corpus esTenTen18, compuesto por más de tres mil millones de palabras[2] extraídas de todo tipo de páginas webs con el dominio «.es», se documentan 175 ocurrencias válidas de esta misma paremia. Además, a diferencia de otros corpus, esTenTen18 recoge textos de distintos registros, tal y como se detallará a lo largo de este trabajo.

El interés concreto de este artículo son los presentadores paremiológicos, un elemento que con frecuencia acompaña a las paremias de toda clase, pero que, con excepción de algunos (muy pocos) trabajos, no ha sido hasta el momento objeto de un estudio detallado que, a través de un análisis sistemático, aclare sus principales características. Este artículo arroja nueva luz sobre el tema gracias a los datos obtenidos del corpus esTenTen18.

1. PRESENTADORES PAREMIOLÓGICOS

Al abordar el tema de los presentadores paremiológicos –o *fórmulas introductorias de refranes* en la terminología empleada por Fernández-Sevilla (1983)– partimos de la definición de estos «elementos deícticos contextuales» aportada por Corpas Pastor en su *Manual de fraseología española* (1996: 137)*:*

> Se trata de frases o palabras, acompañadas alternativamente de indicaciones formales o prosódicas, que cumplen dos funciones principales. Por un lado, distancian al emisor de lo enunciado, librándole de parte de su responsabilidad; y, por otro, introducen estas unidades en el discurso resaltando su carácter paremiológico.

En este sentido, Fernández-Sevilla, en su trabajo sobre las fórmulas introductorias de refranes en la obra *La Celestina*,[3] profundiza en el tipo de función ejercida, al afirmar que esta «puede ser simplemente la de destacar el carácter paremiológico de la unidad que sigue, o bien

[2] Pensemos que, en el marco de sus investigaciones precursoras en el campo de la lingüística de corpus aplicada a la paremiología, Corpas Pastor (1998), para el español peninsular, y Čermák (2005), para el inglés y el checo, emplearon corpus considerablemente menores, de diez y cien millones de palabras, respectivamente. Para otras lenguas sí se han publicado más recientemente análisis paremiológicos basados en corpus de un tamaño significativamente mayor (*cfr.* Meterc, 2016).

[3] Por su parte, Cantera Ortiz de Urbina (1998) aporta en su artículo listados de los presentadores paremiológicos empleados en esta y otras obras de la literatura medieval española.

establecer relaciones gramaticales de causalidad, ilación, etc., con el contexto» (1983: 211). En cuanto a qué palabras o combinaciones de estas se emplean como presentadores, Corpas Pastor propone que, entre las más frecuentes, «figuran términos como *refrán, dicho, frase, eslogan...* así como frases del tipo *como dice el refranero, como asegura el refrán, se dice eso de, ya lo dice el refrán, como decía el poeta (mi abuela, mi padre, etc.), dicen que, ya se dice, como ellos dicen* y similares» (1996: 137). Fernández-Sevilla, por su parte, añade que, en la obra atribuida a Fernando de Rojas, «abundan las fórmulas introductorias de refranes en las que interviene el verbo *decir* con sentido impersonal o suprapersonal e intemporal, tales como *(y) como dizen, porque, como dizen, por esto dizen, pues dizen*» (1983: 211). Además del verbo *decir*, en *La Celestina* se documentan frases construidas con el verbo *saber*, como «pues sabe que, sabete que, pero ya sabes que, porque ya se sabe que, ya sabes que» (1983: 214). Este autor, al igual que Corpas Pastor y Čermák (2005), se refiere también a los presentadores que emplean explícitamente términos paremiológicos como *refrán* y *proverbio*[4].

En cuanto al grado de presencia o ausencia de los presentadores, Veyrat Rigat, al abordar las ideas de Hernando Cuadrado, apunta a su obligatoriedad cuando el refrán no va «engastado en el fluir del discurso», ya que, en estos casos, «*tiene que* venir precedido de un presentador de diversa índole» (2008: 11; la cursiva es nuestra). El único dato numérico con el que contamos en este sentido para el español se refiere exclusivamente a los presentadores documentados en *La Celestina*, obra en la que más del 60 % de las paremias *carece* de presentador o nexo gramatical como *que, así que, así, porque, por tanto, de manera que, y, y que, y también, mas* o *pero* (Fernández-Sevilla, 1983). Además, estos presentadores y nexos habrían de ser más frecuentes con paremias que han sufrido una ruptura de la forma estándar, ya que su presencia «ayuda a reconocer la unidad paremiológica, aunque ésta haya sido desarticulada con el fin de mejor situarla y anclarla en el contexto» (Fernández-Sevilla, 1983: 218). En el caso del inglés y el checo, se han documentado presentadores en aproximadamente el 20 % y el 27 % de las paremias, respectivamente (Čermák, 2005: 64, 66), si bien hay que tener en cuenta que la metodología de este especialista checo es, al igual que en el caso de Fernández-Sevilla, menos restrictiva que la nuestra al etiquetar palabras o frases como presentadores (ver sección 2). Por último, los presentadores escasearían cuando las paremias «pueden ser reconocidas, gracias, sobre todo, a que el autor ha puesto especial cuidado en situarlas en lugares prominentes, destacándolas por medio de pausas y de otros recursos» (Fernández-Sevilla, 1983: 218).

2. METODOLOGÍA

Los datos presentados en este artículo se basan parcialmente en los obtenidos para un trabajo anterior (Gutiérrez Rubio, 2022) en el que se analizó la frecuencia de uso en el *gigacorpus* esTenTen18 de 27 locuciones y 27 paremias españolas extraídas de *Hablar por los codos* (Vranic, 2004), una de las pocas obras ELE diseñadas específicamente para el proceso de enseñanza/aprendizaje de la fraseología y la paremiología. Para centrar el análisis en el español peninsular, las búsquedas se realizaron exclusivamente en el subcorpus Spanish Web 2018, que contiene 3 421 734 353 formas *(tokens)* extraídas de páginas web con dominio «.es». Se empleó el motor de búsqueda «concordance» y el tipo de consulta *(query type)* «phrase». Se buscaron siempre combinaciones de tres palabras consecutivas invariables e impropias *a priori* del «discurso libre», tratando así de reducir la inclusión en los resultados de combinaciones libres de palabras que coincidieran formalmente con parte de la locución o paremia en cuestión.

[4] En este sentido, Nicolás López, en su estudio de la *metarrefraneidad*, es decir, «la capacidad metalingüística propia de los refranes» (1996: 153), denomina *metarrefraneidad enmarcadora* al uso del término *refrán* «para poder circunscribir una parte del texto paremiológico dentro del todo que supone su concepción global» (1996: 154).

A modo de ejemplo, la búsqueda para *A cada cerdo le llega su San Martín* fue «su san martín», combinación está muy improbable en el discurso libre. Al reducir las búsquedas a tres palabras se persiguió evitar los problemas asociados a la localización de paremias en corpus a causa de la elevada presencia de lo que Corpas Pastor (1998: 381) denomina *manipulación creativa*. De hecho, en lo relativo a esta paremia en concreto, se documentó un gran número de variantes no canónicas (el 14 % de las 172 ocurrencias válidas de esta paremia documentadas en el corpus), con formas como *a todo teléfono / sistema económico / ídolo teen / gente como usted / golfo / mujeriego / Jaume Matas...* En el caso de la segunda paremia con mayor número de variantes (9,1 %), *A Dios rogando y con el mazo dando,* se localizaron, entre otras, las siguientes formas creativas: *A Dios rogando y con la sotana renegando / bajo el paso trianeando / con la derecha dando / con el Rabo dando / el cazo de oro llenando...*

De entre las 27 paremias analizadas en el primer estudio, se escogieron nueve repartidas en tres grupos en función de su frecuencia de aparición en las búsquedas:

• Frecuencia baja: *A enemigo que huye, puente de plata* (27 ocurrencias válidas en el corpus), *Unos tienen la fama y otros cardan la lana* (32) y *No se hizo la miel para la boca del asno* (39).

• Frecuencia media: *A la cama no te irás sin saber una cosa más* (81), *A falta de pan, buenas son tortas* (95) y *A caballo regalado, no le mires el diente* (103).

• Frecuencia alta: *A mal tiempo, buena cara* (135), *A cada cerdo le llega su San Martín* (172) y *A Dios rogando y con el mazo dando* (175).

Justificamos la selección de paremias en función de su frecuencia de uso por la primera de las hipótesis de esta investigación: las paremias con mayor presencia en los documentos que componen el corpus y, por tanto, probablemente más (re)conocidas por los hablantes, tenderán a emplear menos presentadores, ya que «se trata, precisamente, de unidades acuñadas y pertenecientes al acervo común» (Fernández-Sevilla, 1983: 218) y, consecuentemente, no necesitan que su carácter paremiológico sea destacado. Esto, sin embargo, parece contradecir los resultados del trabajo de Čermák (2005), en el que las paremias checas, más frecuentes en el corpus que las inglesas[5], documentan un mayor porcentaje de presentadores.

Nuestra segunda hipótesis se basa nuevamente en las apreciaciones de Fernández-Sevilla, quien afirma que los presentadores habrían de ser más frecuentes con paremias que han sufrido una ruptura de la forma estándar (ver sección 1).

Una tercera variable que consideramos que podría afectar a la frecuencia de aparición de presentadores sería el registro, ya que partimos de la hipótesis de que en el español formal los presentadores serán más frecuentes, a causa de la mayor *planificación* que se les presupone a los textos de este registro. En este sentido, resulta necesario señalar que en algunos textos y, muy especialmente, en los publicados en páginas webs: «a veces se escribe como si se hablara (en cuanto a la construcción y progresión del discurso), es decir, puede ser escrito, pero menos planificado» (Briz Gómez, 2010: 127). Por desgracia, la herramienta *Sketch Engine*, mediante la que se realizan las búsquedas en el corpus esTenTen18, no permite filtrar los resultados en función del registro o del tipo de web en que se ha publicado, lo que dificulta la comprobación de esta hipótesis. Así, nos valdremos del análisis de las características del contexto en que se enmarca la paremia en cada uno de los resultados del corpus para clasificarlos de forma manual en dos registros:

• Registro formal: textos publicados en periódicos y revistas, transcripciones de discursos de carácter formal, blogs, webs de productos y servicios (incluidas aquellas que muestran un estilo cercano al cliente y emplean tuteo), textos de términos y condiciones, pero también comentarios

[5] Este hecho se debe a que la selección de las paremias checas se hizo a partir de un trabajo de frecuencia anterior, algo que no ocurrió en el caso de las inglesas.

bajo estos tipos de webs o testimonios de clientes escritos con clara intención de mantener el registro formal. Por ejemplo: «Que iba destinado, insisto, a atender a eso que usted llama 'los pobres inmigrantes'. Señoría, a Dios rogando y con el mazo dando. No se puede venir aquí a esta tribuna a hacer un discurso demagógico».

• Registro semiformal o informal: comentarios referidos a webs o testimonios de clientes escritos sin una clara intención de mantener el registro formal. Estos textos suelen incluir interjecciones coloquiales *(hombre, vaya tela)*, emoticones, faltas de ortografía, puntos suspensivos, abreviaturas (Bcn) y ausencia de tildes: «Impresionante esta ciudad y muy buenas fotos. Lastima que cosas tan bonitas como esta se vayan a perder. al mal tiempo buena cara… o como era? jeje bueno, al menos a algunos les sirvió de diversión». Los textos más marcadamente informales, además, tienden a emplear vulgarismos, interjecciones muy informales *(juer, uffff, OMG)*, elementos subjetivos *(qué estafa!!)* o abreviaturas no estandarizadas: «ellos pueden quejarse lo q quieran de sus respectivas madres,pero si in día te quejas tú...uffff.....en fin...q al mal tiempo buena cara..a estar bien con nuesttros chicos y nada más importaa... amanda...jajajajaj....».

El resto de las variables presentadas en la sección 1, es decir, si las paremias van engastadas «en el fluir del discurso» o destacadas «en lugares prominentes», así como si se emplean para establecer relaciones gramaticales de diversa índole, no serán tenidas en cuenta para este primer acercamiento al tema de los presentadores.

Además de investigar la cuestión de la frecuencia de aparición de los presentadores, vamos a dedicar una parte de esta investigación a catalogar los distintos tipos de presentadores extraídos del corpus, objetivo este para el que partiremos nuevamente de las publicaciones de Fernández-Sevilla (1983) y Corpas Pastor (1996) presentadas arriba[6].

En cuanto a los problemas metodológicos asociados al trabajo con corpus computarizados, el más relevante se refiere a la presencia de *ruido*, en otras palabras, a la aparición en los resultados de datos irrelevantes que pueden interferir en la precisión y fiabilidad del análisis. Para minimizar el ruido, hemos realizado una revisión manual de todos los resultados de las búsquedas. Así, tras excluir los fragmentos textuales repetidos en el corpus (a causa, aparentemente, de un error en el procesamiento de las webs que lo conforman); los listados de paremias presentes en glosarios y manuales; los títulos de libros, películas y programas de radio o televisión[7]; las citas de obras literarias; y los fragmentos poco claros como consecuencia de la falta de información contextual, se ha descartado el 15 % de los resultados obtenidos de las búsquedas automáticas. Además, hemos eliminado las variantes de paremias que, a causa de un proceso de institucionalización, han dado lugar a locuciones. Este fenómeno, brevemente discutido por Corpas Pastor (1998: 375), afectaría a dos de las paremias analizadas. Por una parte, *A mal tiempo, buena cara*, origen de la locución verbal *Poner a(l) mal tiempo, buena cara*, recogida como tal en los diccionarios *DiLEA* (Penadés Martínez, 2019) y *DFDEA* (Seco *et al.*, 2017: 830), y que cuenta con una presencia muy elevada en el corpus: 137 de las 272 entradas válidas localizadas con la combinación «mal tiempo buena». También hemos excluido las seis combinaciones de palabras que incluyen «su san Martín» cuando, como elemento aislado del resto de la paremia, denotan 'muerte', como en «soñando con eludir su san Martín» y, por tanto, convertidas en una locución nominal, si bien, esta no se halla recogida en ninguna de las fuentes fraseográficas arriba citadas.

Por último, queremos hacer constar que no incluiremos en nuestra investigación los presentadores que Fernández-Sevilla (1983) denomina *nexos gramaticales* –como *que, porque, de manera que*–, ya que no son exclusivos de las paremias, y nos centraremos en los

[6] Čermák (2005), por su parte, incluye entre los *presentadores paremiológicos estándares* en checo e inglés términos y verbos muy similares a los ya señalados aquí para el español (ver sección 1).

[7] No se trata de una metodología unánime. Así, especialistas como Pamies y Pazos (2023) consideran que estos títulos no son *ruido* y, consecuentemente, nos los excluyen de sus datos.

presentadores *específicos*, tanto en forma de frases (*como se dice, ya se sabe que*), como de términos que denominan tipos de paremias: *refrán, proverbio, aforismo*... También hemos descartado de nuestros datos el presentador (o *semipresentador* para ser más exactos) *aquello/eso de (que)*, a pesar de haberse mostrado relativamente frecuente en el corpus analizado: «conviene mucho aplicar *aquello de que* a Dios rogando y con el mazo dando».

3. RESULTADOS DEL ANÁLISIS

3.1.Frecuencia de los presentadores paremiológicos en el corpus

En la Tabla 1 presentamos el porcentaje relativo de presentadores documentado para cada una de las nueve paremias analizadas.

Paremias	Presentador (%)
A enemigo que huye, puente de plata	40,7
Unos tienen la fama, y otros cardan la lana	31,3
A la cama no te irás sin saber una cosa más	30,9
A caballo regalado, no le mires el diente	27,2
A Dios rogando y con el mazo dando	26,9
A mal tiempo, buena cara	20,7
A cada cerdo le llega su San Martín	20,3
A falta de pan, buenas son tortas	20,0
No se hizo la miel para la boca del asno	17,9
MEDIA	24,4

Tabla 1: Frecuencia de aparición de presentadores para cada una de las paremias analizadas

El primer dato que llama poderosamente nuestra atención se refiere a la sustancial divergencia entre el porcentaje de presentadores documentado para las distintas paremias analizadas. Frente al más del 40 % de paremias acompañadas por presentadores en el caso de *A enemigo que huye, puente de plata*, se documentan cuatro paremias con un porcentaje de presentadores que ronda la mitad de esta cifra: desde el 20,7% de *A mal tiempo, buena cara* hasta el apenas 17,9 % de *No se hizo la miel para la boca del asno*.

Por otra parte, la primera de nuestras hipótesis preveía que las paremias más frecuentemente localizadas en el corpus tenderían a ir menos acompañadas de presentadores a causa de la menor necesidad de destacar su carácter paremiológico. Los datos expuestos en la Tabla 2 parecen apuntar a la confirmación de esta hipótesis.

Paremias	Suma de *tokens*	Presentador (%)
Paremias POCO FRECUENTES	98	28,6
Paremias FRECUENTES	279	25,8
Paremias MUY FRECUENTES	482	22,8
MEDIA		24,4

Tabla 2: Presencia de presentadores por grupos de frecuencia

Sin embargo, estos datos son, hasta cierto punto, engañosos. La Tabla 3 muestra que se trata, en parte, de un fenómeno individual. Así, aunque la paremia menos frecuente *–A enemigo que huye, puente de plata–* es la que más comúnmente va acompañada de presentadores (40,7 %), la tercera menos frecuente *–No se hizo la miel para la boca del asno–* es, en contra de nuestra hipótesis, la que menos aparece presentada en el corpus por una fórmula introductoria (17,9 %). Por otra parte, aunque dos de las tres paremias más comunes *–A mal tiempo, buena cara* y *A Dios rogando y con el mazo dando–* se hallan entre las paremias con menor frecuencia de presentadores, la segunda paremia más frecuente *–A cada cerdo le llega su San Martín–* documenta un porcentaje de presentadores relativamente elevado y ligeramente superior a la

media (26,9 %). Por tanto, aunque cabe afirmar que la frecuencia de uso de las paremias tiene influencia sobre el número de presentadores, no se trata de una variable concluyente.

Paremias	*Tokens*	Presentador (%)
A enemigo que huye, puente de plata	27	40,7
Unos tienen la fama, y otros cardan la lana	32	31,3
No se hizo la miel para la boca del asno	39	17,9
A la cama no te irás sin saber una cosa más	81	30,9
A falta de pan, buenas son tortas	95	20,0
A caballo regalado, no le mires el diente	103	27,2
A mal tiempo, buena cara	135	20,3
A cada cerdo le llega su San Martín	172	26,9
A Dios rogando y con el mazo dando	175	20,7
TOTAL/MEDIA	859	24,4

Tabla 3: Presencia de presentadores de acuerdo con la frecuencia individual

Nuestra segunda hipótesis proponía una mayor presencia de presentadores junto a paremias que han sufrido una ruptura de su forma estándar. En este caso, los datos muestran con claridad que se trata de una hipótesis falsa: de las 49 variantes creativas documentadas –como «A caballo regalado, no preguntes que hace con tus datos»– tan solo seis incluyen presentador, lo que equivale al 12,2 % del total (debemos aclarar que hemos excluido de esta lista las paremias con sinónimos más o menos canónicos, como «a caballo regalado no se le revisan los colmillos»). Si tenemos en cuenta que, de media, el 24,4 % del total de paremias va acompañado de un presentador (el doble exactamente), la hipótesis no solo queda sin evidencia empírica, sino que incluso cabría afirmarse que las paremias que han sufrido modificaciones tienden a emplearse sin fórmulas introductorias.

La tercera hipótesis se refería al registro del texto en que se documentan las paremias, de modo que, en contextos formales, se espera una mayor presencia de presentadores respecto a los semiformales e informales.

Paremias	Presentador formal (%)	Presentador informal (%)	Diferencia > (%)
A enemigo que huye, puente de plata	50,0	22,2	27,8
A caballo regalado, no le mires el diente	41,0	18,8	22,3
A falta de pan, buenas son tortas	29,2	10,6	18,5
A la cama no te irás sin saber una cosa más	40,5	22,7	17,8
A Dios rogando y con el mazo dando	30,6	17,6	13,0
Unos tienen la fama, y otros cardan la lana	33,3	25,0	8,3
A cada cerdo le llega su San Martín	21,1	19,8	1,3
A mal tiempo, buena cara	20,3	21,3	-1,0
No se hizo la miel para la boca del asno	13,6	23,5	-9,9
MEDIA	34,6	17,1	16,9

Tabla 4: Presencia de presentadores en función del registro

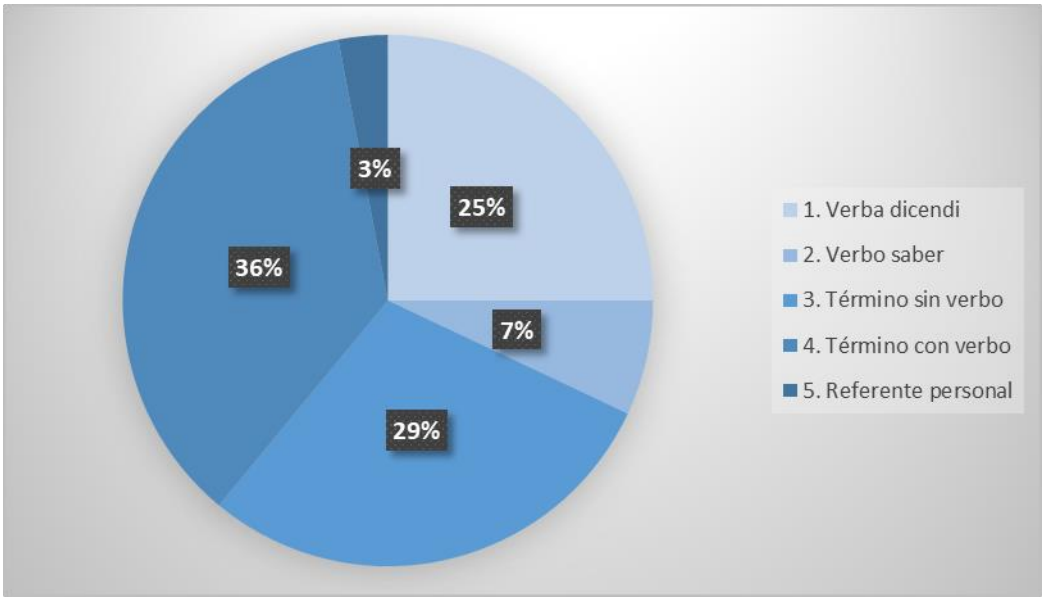
La Tabla 4 muestra que en seis de las nueve paremias analizadas se documenta una clara tendencia a una mayor aparición de presentadores en los contextos formales. En dos casos –*A cada cerdo le llega su San Martín* y *A mal tiempo, buena cara*– las diferencias son muy escasas, de entorno al 1 %. La única paremia que presenta unos datos claramente contrarios a nuestra suposición sería, como en el caso de la primera hipótesis, *No se hizo la miel para la boca del asno*, paremia esta que, por tanto, muestras características específicas.

3.2. Clasificación cuantitativa de los presentadores

Otro objeto de nuestro interés se refiere a la frecuencia de aparición en las páginas web analizadas de los diversos tipos de presentadores. Tras analizar los 210 presentadores localizados en el corpus y partiendo de los datos aportados por los especialistas (ver sección 1), los presentadores se pueden dividir en cinco categorías:

1. *Verba dicendi*: *como se suele decir/ se dice / afirman...*
2. Verbo *saber*: *todo el mundo sabe / ya se sabe…*
3. Mención explícita a un término paremiológico sin *verba dicendi* o el verbo *saber*: *el refranero es claro / ojalá se cumpla ese dicho...*
4. Mención explícita a un término paremiológico en combinación con *verba dicendi* o el verbo *saber*: *como dice el dicho / dice el refranero que / como reza el dicho / aunque ya se sabe el dicho...*
5. Referente personal (con o sin *verba dicendi*): *mi padre siempre decía / como dice tu madre / hagan caso al sabio consejo de mi abuela…*

La frecuencia de aparición de estas cinco categorías puede observarse en la Gráfica 1.



Gráfica 1: Frecuencia de aparición de los cinco grupos de presentadores

Entre los cinco tipos de presentadores, el más frecuentes es, con un 36 % del total, el grupo 4, es decir, *verba dicendi* o el verbo *saber* con mención explícita de un término paremiológico. Destaca sobre todos los demás el verbo *decir*, con 68 de las 75 apariciones (*rezar* cuenta con seis y *saber*, apenas con una). En cuanto a los términos documentados, sobresalen *refrán* (45 *tokens) y dicho* (22), a pesar de que este último no sea un término aceptado por la paremiología. Además, se documentan las palabras *refranero* (4), *proverbio* (2), *adagio* (1) y *aforismo* (1). El segundo grupo más frecuente (29 %) se refiere a estos mismos términos, pero sin acompañamiento de *verba dicendi* (o *saber*). Aunque los tres primeros coinciden con los del grupo 4, se recoge un total de diez términos en lugar de los seis anteriores: *refrán* (23 *tokens*), *dicho* (17), *refranero* (7), *máxima* (5), *frase* (2), *expresión* (2), *divisa* (1), *adagio* (1), *jaculatoria* (1) y *frase hecha* (1). El tercer grupo en cuanto a frecuencia son *verba dicendi* sin término paremiológico (25 %); en realidad, se limita al verbo *decir* y a una única aparición de *afirmar*.

Por su parte, el grupo 2 (verbo *saber* sin término paremiológico) representa tan solo el 7 % del total de los presentadores documentados. En último lugar, aunque pueda resultar ligeramente sorprendente, dado que Corpas Pastor (1996) los incluye entre los principales presentadores, los referentes personales suponen apenas el 3 % de las entradas. En todos los casos se trata, además, de familiares: *abuela* (3), *abuelo* (2), *madre* (1), *padre* (1), junto a una única entrada de carácter humorístico: «Como dijo el sobrino-nieto de Confucio».

Por último, en cuanto a la distribución de estos cinco grupos por registros, en el contexto formal destaca claramente el grupo 3 (término sin *verba dicendi*), que documenta 49 ocurrencias en textos formales y apenas once en aquellos categorizados como informales.

CONCLUSIONES

Los datos presentados en este artículo arrojan nueva luz sobre el uso de los presentadores paremiológicos en páginas webs españolas actuales. En cuanto a la relación entre la presencia de presentadores, la frecuencia de uso de las paremias y el tipo de registro en que se documentan (primera y tercera hipótesis), ambas han sido respaldadas, si bien solo parcialmente. Por una parte, se ha demostrado empíricamente que, en general, a menor presencia de una paremia en el corpus, mayor su tendencia a ir acompañada por un presentador en el texto. Sin embargo, debemos también reconocer que algunas paremias aportan datos contrarios a esta tendencia, especialmente el proverbio *No se hizo la miel para la boca del asno.* Algo parecido cabe afirmar sobre nuestra tercera hipótesis, en que se analiza la presencia de fórmulas introductorias en función del registro: ha sido comprobado que, a mayor grado de formalidad, mayor presencia de presentadores. Aun así, se trata tan solo de una tendencia general, ya que en dos casos las paremias no muestran grandes diferencias entre ambos registros y una tercera paremia, nuevamente *No se hizo la miel para la boca del asno*, evidencia la tendencia contraria, ya que viene acompañada de presentadores con mayor frecuencia en contextos informales.

A pesar de estas divergencias, las dos hipótesis quedan demostradas, al menos como tendencia, de modo que sí parece verificado que el uso de los presentadores depende en gran medida de la frecuencia de uso de las paremias a las que acompañan, así como del registro del texto en que se emplean. Por el contrario, se ha demostrado sin evidencia empírica la segunda hipótesis, la cual propondría que los presentadores serían más frecuentes con paremias que han sufrido una ruptura de la forma estándar.

Por último, se ha aclarado qué presentadores paremiológicos son los más frecuentes en las webs españolas y cuáles podrían catalogarse de marginales. Destacan por su escasa presencia (3 % del total) los referentes personales. Además, el verbo *decir* sería el presentador prototípico, ya sea junto a un término paremiológico (36 %) o sin él (25 %). Sobresalen también por su frecuencia los términos sin acompañamiento de *verba dicendi* y, muy especialmente, *refrán* y *dicho.* Se han documentado, con mucha menor frecuencia, otros términos especializados (*proverbio, refranero, adagio, jaculatoria, divisa, aforismo, máxima* e incluso *frase hecha),* además de formas más generales como *expresión* o *frase.*

No queremos cerrar este artículo sin hacer un llamamiento a los especialistas en paremiología a que empleen *gigacorpus* en sus futuras investigaciones, un recurso que consideramos imprescindible para hacer avanzar de forma significativa el conocimiento que tenemos sobre la naturaleza y el uso auténticos de las paremias en contextos reales.

REFERENCIAS BIBLIOGRÁFICAS

Briz Gómez, A. (2010). Lo coloquial y lo formal, el eje de la variedad lingüística. R. M. Castañer Martín, y V. Lagüéns Gracia (coords.) *De moneda nunca usada.* Zaragoza: Instituto Fernando El Católico, CSIC, 125-133.

Cantera Ortiz de Urbina, J. (1998). Refranes y sentencias en la literatura medieval española. *Paremia*, 7: 11-26. https://cvc.cervantes.es/lengua/paremia/pdf/007/002_cantera.pdf [consultado: 4/8/2023]

Čermák, F. (2005). Text introducers of proverbs and other idioms. *Jezikoslovlje*, 6.1: 57-77.

Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid: Gredos.

Corpas Pastor, G. (1998). El uso de paremias en un corpus del español peninsular actual. Wotjak, G. (coord). *Estudios de fraseología y fraseografía del español actual*, Madrid: Iberoamericana Vervuert, 365-390.

Fernández-Sevilla, J. (1983). Presentadores de refranes en el texto de *La Celestina. Sexta Philologica F. Lázaro Carreter*, vol. 1. Madrid: Cátedra, 209-218.

Gutiérrez Rubio, E. (2022). Frecuencia de uso de locuciones y paremias en el corpus Spanish Web 2018 (esTenTen18): implicaciones didácticas y lexicográficas. Corpas Pastor, G. *et al.* (eds.). *Proceedings of the International Conference EUROPHRAS 2022, Malaga, Spain, 28-30 September*, 2022, Bulgaria: INCOMA, 26-33.

Meterc, M. (2016). The Slovene paremiological optimum: new empirical research tools and the augmentation of the field of minimum-oriented research. *Proverbium*, 33: 319-338.

Nicolás López, R. (1996). Acercamiento metalingüístico y metadiscursivo a lo paremiológico. *Paremia*, 5: 153-159. https://cvc.cervantes.es/lengua/paremia/pdf/005/019_lopez.pdf [consultado: 4/8/2023]

Pamies, A.; Pazos, J. M. (2023). Frasemas combinados 'a pedir de boca'. *Romanica Olomucensia*, 35.1:139-156.

Penadés Martínez, I. (2019). *Diccionario de locuciones idiomáticas del español actual (DiLEA):* http://www.diccionariodilea.es/diccionario [consultado 4/8/2023].

Seco, M.; Andrés, O.; Ramos, G. (2018). *Diccionario fraseológico documentado del español actual.* 2.ª ed. Madrid: JdeJ Editores.

Veyrat Rigat, M. (2008). Aproximación lingüística al estudio del refrán como unidad comunicativa. *Dialogía: revista de lingüística, literatura y cultura*, 3: 5-31.

Vranic, G. (2004). *Hablar por los codos: Frases para un español cotidiano*. Madrid: Edelsa.