# Análisis de contenido para la elaboración del mínimo paremiológico portugués

Ana María DÍAZ FERRERO
*Universidad de Granada (ESPAÑA)*
anadiaz@ugr.es || https://orcid.org/0000-0002-0916-5552

Recibido: 26/2/2024 || Aceptado: 17/3/2024

**Resumen:** En este artículo presentamos los criterios seguidos para llevar a cabo un análisis de contenido (AC) como fase inicial en la elaboración del mínimo paremiológico portugués de la variedad de Portugal. El objetivo de este trabajo es elaborar un corpus con las paremias que, en mayor o menor medida, están presentes en el día a día de la lengua portuguesa. Se trata de crear un corpus manejable con un número limitado de paremias de modo que pueda ser testada posteriormente la frecuencia de uso de cada una ellas y determinar cuáles son las más usadas. Tras definir los objetivos de la investigación y realizar un estudio previo, el primer paso para realizar el AC es seleccionar los documentos en formato digital que constituyen la muestra de análisis. A continuación, se eligen los segmentos que contienen paremias y se procede al tratamiento informático de los datos con una plantilla de códigos previamente establecida. Los resultados obtenidos constatan que la aplicación de la técnica de AC es eficaz para conseguir los objetivos planteados en la presente investigación.
**Palabras clave:** Paremiología. Paremia. Mínimo paremiológico. Portugués.

**Titre : «Analyse de contenu pour l'élaboration du minimum parémiologique portugais ».**
**Résumé :** Dans cet article, nous présentons les critères que nous avons suivis pour effectuer une analyse de contenu (AC) dans la phase initiale de l'élaboration du minimum parémiologique du portugais européen. L'objectif de ce travail est de créer un corpus de parémies qui sont`présentes, dans des degrés divers, dans la langue portugaise courante. L'objectif est de créer un corpus gérable avec un nombre limité de parémies afin de tester la fréquence d'utilisation de chacune d'entre elles et de déterminer celles qui sont les plus utilisées. Après avoir défini les objectifs de la recherche et réalisé une étude préliminaire, la première étape de l'AC a consisté à sélectionner les documents en format électronique qui constituent l'échantillon d'analyse. Ensuite, les segments contenant des parémies ont été sélectionnés et les données ont été traitées à l'aide d'un modèle de code préalablement établi. Les résultats obtenus montrent que l'application de la technique de l'AC est efficace pour atteindre les objectifs fixés dans cette recherche.
**Mots-clés :** Parémiologie. Parémie. Minimum parémiologique. Portugais.

**Title: "Content analysis for the elaboration of the Portuguese paremiological minimum."**
**Abstract:** In this article we set out the criteria to draw up a content analysis (CA) as the initial stage in the elaboration of the Portuguese minimum paremiological of the variety of Portugal. The object of this work is to create a corpus of paremias that, in one way or another, are present in the day to day of the Portuguese language. The aim is to create a manageable corpus with a limited number of paremias so that it is possible to analyse the frequency of use of each of them and determine which are the most used.
After defining the research objectives and carrying out a preliminary study, the first step in carrying out a CA is to select the documents in digital format that make up the analysis sample. Next, the segments containing paremias are chosen and the data is computerised using a pre-established code plan. The results obtained showed that the application of the CA technique is effective in achieving the objectives set out in this research.
**Keywords:** Paremiology. Pareamia. Paremiological minimum. Portuguese.

## INTRODUCCIÓN

El término «mínimo paremiológico» se emplea para designar «el conjunto de enunciados sentenciosos estables más conocidos por todos o, al menos, por una mayoría significativa de los hablantes de una comunidad sociocultural concreta» (Zurdo Ruiz-Ayúcar y Sevilla Muñoz, 2016: 21). Consecuentemente, las investigaciones para elaborar el mínimo paremiológico de una determinada lengua recogen las paremias más comunes de esa comunidad lingüística, a saber, enunciados breves y sentenciosos como proverbios, aforismos, refranes, frases proverbiales, locuciones proverbiales o dialogismos. Los estudios conducentes a obtener el mínimo paremiológico en distintas lenguas (Mieder, 1994; Baur; Chlosta; Grzybek, 1996; Ĉermak, 2003; Tarnovska, 2004; Haas, 2008; Sardelli, 2011; Reis; Baptista, 2020; Korshunova, et al 2021; Crida Álvarez, 2021; Huang; Mohd Nawi, 2024) han ido creciendo de forma notable desde los años 70 del siglo XX, a tenor de los trabajos publicados por Permiakov (1970, 1973, 1985). Estos estudios se han elaborado con métodos de indagación diversos y, por consiguiente, con resultados también distintos, por lo que su análisis ha posibilitado proponer mejoras para las futuras investigaciones. Para la elaboración de un mínimo paremiológico, los dos métodos más frecuentes empleados son el demoscópico con encuestas realizadas a informantes, y la lingüística de corpus, si bien, los criterios para la elaboración de los corpus (cantidad, calidad y tipología de textos...), así como para el diseño y administración de los cuestionarios (tipo de preguntas, tamaño de la muestra, selección de informantes...) varían considerablemente (Forgács, 2014; Ďurčo; 2014). Para muchos investigadores (Ďurčo 2014; Zurdo Ruiz-Ayúcar; Sevilla Muñoz, 2016), ambos métodos son complementarios, ya que, aunque el enfoque metodológico introducido por la lingüística de corpus permite obtener resultados más precisos para cuantificar el uso de las paremias, la consulta a los informantes sigue siendo en muchos casos un método necesario para conocer determinados elementos sobre la forma y el uso de las paremias. En opinión de Zurdo Ruiz-Ayúcar y Sevilla Muñoz (2016: 18):

> El método tradicional, basado en las encuestas a informantes, se sigue considerando necesario como complemento del análisis de corpus por su capacidad para establecer el término medio del conocimiento y uso de paremias entre los hablantes y por su utilidad para descubrir la incidencia de los factores diatópicos, diastráticos, diafásicos o estilísticos en las respuestas.

El presente estudio, orientado a determinar el mínimo paremiológico de la lengua portuguesa, se enmarca en un proyecto de paremiología cuyos parámetros teóricos y metodológicos fundamentales se encuentran en el número 1 de la serie «Mínimo paremiológico» de la Biblioteca fraseológica y paremiológica del Centro Virtual Cervantes (Zurdo Ruiz-Ayúcar y Sevilla Muñoz, 2016).

El acervo paremiológico de la lengua portuguesa es inmenso y sería prácticamente imposible testar el conocimiento de todas las paremias de la lengua portuguesa mediante encuestas o a través de un análisis de todos los documentos existentes en portugués. Además, debido a la heterogeneidad del ámbito geográfico del portugués y las consecuentes diferencias que manifiestan sus hablantes, es necesario diferenciar en la investigación las distintas variedades de la lengua portuguesa: portugués europeo, portugués de Brasil, portugués de los diferentes países africanos de lengua oficial portuguesa (Cabo Verde, Guinea-Bisáu, Santo Tomé y Príncipe, Angola y Mozambique) y portugués de Timor Oriental. Por consiguiente, hemos diseñado un modelo de investigación para establecer el mínimo paremiológico de cada una de las variedades de la lengua portuguesa.

Después de realizar un estudio piloto para la elaboración del mínimo paremiológico, decidimos elaborar un plan de trabajo dividido en tres fases de estudio y basado en una estrategia de triangulación metodológica (Denzin, 1970; Morse, 1991; Rodríguez Sabiote, *et al*.

2006) en la que se combinan técnicas de observación y análisis de datos documentales con técnicas de conversación basadas en cuestionarios y entrevistas a informantes nativos. Dos métodos complementarios en los que el cotejo y verificación de los datos obtenidos en ambos casos permite establecer con más precisión el mínimo paremiológico final. En la primera fase, mediante un análisis de contenido y un estudio demoscópico, se crea un corpus manejable con las paremias que, en mayor o menor medida, están presentes en el día a día de la lengua portuguesa. En la segunda fase se comprueba, mediante análisis cuantitativo en la Web y a través de encuestas a informantes nativos, la frecuencia de uso de las paremias obtenidas en la primera fase, y en la tercera fase se efectúa un proceso de validación final de los resultados, se realizan encuestas en los casos de paremias con resultados confusos y finalmente se eligen las paremias base y las variantes que constituyen el mínimo paremiológico.

En este artículo presentamos la metodología empleada para llevar a cabo el análisis de contenido (AC) de documentos en la primera fase del estudio del *Mínimo paremiológico del portugués europeo* (Díaz Ferrero y Sabio Pinilla, 2023). Explicamos los pasos seguidos con este análisis y proporcionamos algunos ejemplos de los resultados obtenidos.

## 1. ANÁLISIS DE CONTENIDO

El AC ha sido estudiado y definido desde diferentes perspectivas (Bardin, 1977; Krippendorf, 1990; López Noguero, 2002; Andréu Abela, 2002; Espín López, 2002; Piñuel Raigada, 2002). Según la definición de Bardin (1977), el término AC se emplea para designar:

> Um conjunto de técnicas de análise das comunicações visando obter, por procedimentos, sistemáticos e objectivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis inferidas) destas mensagens (Bardin, 1977: 42).

Los orígenes de los estudios de AC se remontan al siglo XVII «cuando la Iglesia estaba inquieta por la difusión de los temas de índole no religiosa a través de los periódicos» (Krippendorff, 1990: 15), pero los rasgos definitorios de esta técnica y la metodología de análisis comenzaron a sistematizarse en la primera mitad del siglo XX con las publicaciones de Lasswell (1927) sobre las técnicas de propaganda política, y principalmente a raíz de los estudios teóricos de Berelson y Lazarsfeld (1948) y Berelson (1952), quienes desarrollaron en profundidad esta técnica de investigación. Los primeros análisis eran de carácter cuantitativo e iban destinados primordialmente al estudio de discursos políticos y textos periodísticos, si bien desde mediados del siglo pasado esta técnica se ha ido extendiendo a otras disciplinas. Actualmente el AC se emplea tanto en el ámbito cuantitativo como en el cualitativo y se aplica a diferentes áreas de estudio con resultados más precisos gracias al avance del tratamiento informático de los datos. Existen tres tipos de análisis de contenido (Palmquist *et al*. 1997; Colle, 2011; Gil Pascual, 2011): análisis temático, también llamado conceptual, cuyo objetivo es examinar documentos para comprobar la presencia o ausencia de determinados términos o conceptos sin tener en cuenta la relación entre ellos; análisis semántico, denominado también relacional, que se centra no solo en localizar determinados conceptos o términos en el documento, sino también en las relaciones de significado que se establecen entre ellos, y por último, el análisis de redes, que estudia la copresencia de pares de componentes o la ubicación relativa de ciertos componentes en el texto.

Como hemos mencionado, la primera fase de la investigación para establecer el mínimo paremiológico está destinada a crear un corpus de paremias en uso que sirva de punto de partida en la investigación. Para ello empleamos dos técnicas de recogida de información: por un lado, un estudio demoscópico, y por otro, un AC conceptual para hallar las paremias que han sido

usadas en diferentes textos (periodísticos, literarios, publicitarios, académicos, humorísticos...) en soporte digital. Los pasos o etapas que se siguen en la aplicación de un AC (Bardin, 1977; Krippendorf, K. 1990; Espín, 2002; Andréu Abela, 2002; Tinto Arandes, 2013) son los siguientes:

- Definir los objetivos de la investigación.
- Estudio previo.
- Seleccionar la muestra o corpus documental de análisis, es decir, el conjunto de documentos (orales, escritos, visuales) que serán analizados y su tamaño.
- Elegir las unidades de registro y de contexto. Los segmentos de texto que se van a analizar (pueden ser palabras, frases, temas, conceptos, símbolos... ), así como la unidad de contexto o porción de texto que será analizado.
- Determinar el sistema de codificación, es decir, asignar un código numérico o alfabético a los datos para su clasificación y estudio.
- Analizar e interpretar los datos obtenidos.
- Presentar los resultados.

### 1.1. Objetivos de la investigación

El objetivo final de todo el proyecto es identificar cuáles son las paremias que están en uso actualmente y son conocidas por una amplia mayoría de los hablantes nativos de Portugal. Con este AC se pretende obtener un corpus con un amplio registro de paremias que hayan sido citadas en algún documento por los portugueses cuando hablan o escriben en portugués. De este modo se excluyen las paremias en desuso o muy poco usadas y se facilita el análisis posterior de frecuencia de uso. Se trata de crear un corpus que contenga, no las paremias existentes en portugués y que están recogidas en los repertorios, sino aquellas que han sido usadas actualmente por los hablantes. Además, con este estudio pretendemos conocer posibles variantes, conseguir contextos de uso de las paremias y obtener ejemplos en diferentes formatos para comprobar el sentido con el que se emplean.

### 1.2. Estudio previo

Para llevar a cabo la presente investigación realizamos en primer lugar un estudio previo con el objetivo de evaluar la viabilidad de la metodología del proyecto. Se realizaron distintas pruebas para obtener y analizar los datos y, una vez examinados los resultados, se hicieron los ajustes necesarios para diseñar la estrategia específica de trabajo. Gracias a este estudio previo pudimos determinar los criterios para garantizar la validez, representatividad y exhaustividad de la recogida de datos, es decir, el tamaño de la muestra, las normas para elegir documentos y la eficacia del programa informático elegido para procesar los datos (codificación y análisis). En suma, el estudio previo permitió verificar que el diseño del AC permite conseguir los objetivos planteados.

### 1.3. Muestra de análisis

Para conseguir resultados relevantes hay que manejar un gran número de textos que sean representativos del uso de la lengua. «La función principal de un corpus, tanto textual como oral, es establecer la relación entre la teoría y los datos; el corpus tiene que mostrar a pequeña escala cómo funciona una lengua natural» (Torruella y Llisterri, 1999: 46). En esta investigación recurrimos a la web para crear el corpus de documentos que constituirían la muestra objeto de estudio del AC. A pesar de las mejoras que deben realizarse para afinar las búsquedas y optimizar los resultados, el uso de la web como fuente de datos es una opción muy usada desde hace décadas y cada vez más extendida (De Schryver, 2002; Kilgarriff; Grefenstette, 2003; Lapata; Keller, 2005; Xatara, 2008; Corpas Pastor, 2014; Valença; Sabino,

2016; González, 2017; 2018). La cantidad de datos presentes en la web es tan variada y extensa que permite obtener ejemplos fiables y representativos del uso de la lengua. Como afirma Corpas Pastor (2014: 135):

> las limitaciones prácticas de los corpus actuales han llevado a muchos investigadores a «surfear» la Red mundial en busca de ejemplos de uso. Ello es posible hoy día debido al crecimiento exponencial y continuado de la red; y no sólo en cuanto al número de sitios en lengua inglesa, sino también con respecto al aumento de páginas en otras lenguas, especialmente germánicas, romances y asiáticas.

Para escoger los documentos hemos tenido en cuenta las reglas establecidas por Bardin (1977) de pertinencia (los documentos contienen información adecuada para los objetivos planteados); exhaustividad (se seleccionan todos los documentos que cumplen los criterios establecidos para el estudio); representatividad (los documentos son representativos del universo, en este caso, de Portugal continental e insular) y homogeneidad (se emplean los mismos criterios para seleccionar los documentos). No obstante, somos conscientes de que los resultados de una búsqueda en Internet varían en función de diferentes factores como el motor de búsqueda utilizado o la ubicación. Además, el número máximo de resultados visibles es limitado. En cualquier caso, nuestro objetivo no es analizar todas las páginas de Internet que contienen paremias, sino conseguir una extensa relación de documentos procedentes de distintas regiones y de diferente autoría para poder elaborar una muestra homogénea y representativa del portugués europeo. Así pues, la muestra para el AC está formada por textos escritos, orales, visuales y audiovisuales de portugués europeo principalmente en formato digital[1]. Para seleccionar estos documentos se realizó en primer lugar un rastreo en Internet que nos permitió comprobar cuáles son las frases más frecuentes usadas en portugués europeo como *presentadores paremiológicos* (Pamies Bertrán, 2017) o *marcadores mediativos* (Anscombre, 2010), es decir, las frases o términos usados para introducir las paremias en el discurso. Obtuvimos 20 fórmulas diferentes y seleccionamos las ocho más frecuentes: (*como) diz o ditado*, (*o) ditado que diz*, (*como*) *diz o provérbio*, (*o*) *provérbio que diz*, *(como) diz o adágio*, (*o*) *adágio que diz*, *como diz o povo* y *diz a minha avó*. A continuación se realizaron diferentes búsquedas en Internet con el motor de búsqueda de Google ─en la sección general y en la de Imágenes, la de Vídeos y la de Libros─ en páginas de Portugal que incluyesen las palabras «*o ditado*» «*o proverbio*», «*o adagio*», «*diz o povo*» o «*diz a minha avó*», así como búsquedas específicas en páginas de trabajos académicos, discursos parlamentarios, periódicos portugueses, emisoras de radio y canales de televisión[2] empleando los presentadores paremiológicos más usados en Portugal. A continuación, se analizó la información para seleccionar los textos cuyas características se ajustaban a la investigación. De esta manera, los textos de la muestra debían cumplir los siguientes requisitos: contener una o más paremias y pertenecer a cualquier sección de un medio de comunicación tradicional (prensa, radio y televisión), a medios sociales (blogs, redes sociales...) u otro tipo de texto (literario, divulgativo,

[1] Aunque la mayoría de las paremias se han obtenido mediante el análisis de contenido de documentos en soporte digital, también se han incorporado paremias que hemos localizado en documentos en papel (publicidad, periódicos, textos literarios...) durante el proceso de investigación, como los refranes encontrados en los azucarillos del café Delta con motivo de la campaña publicitaria «Sociedade Ponto Verde».

[2] Los periódicos consultados fueron: Diário de Notícias, Jornal de Notícias, Correio da Manhã, Expresso, Observador, Público, Jornal i, Jornal de Negócios, Jornal Económico, Record, O Minho, O Mirante, Jornal da Madeira, Diário dos Açores, Jornal A voz de Trás-os-Montes, Jornal O Interior, Diário do Sul y Jornal do Algarve. Las emisoras de radio: TSF, Rádio comercial, Rádio renascença, y los canales de TV: RTP, RTP Madeira y RTP Açores.

expositivo, publicitario...) de Portugal continental, de las Azores o de Madeira; haber sido publicado o editado entre el año 2000[3] y el año 2020 y ser un documento de autoría portuguesa. Se descartan las páginas duplicadas, los diccionarios, glosarios o recopilaciones de paremias, las traducciones al portugués de paremias extranjeras, los textos que, a pesar de estar escritos en portugués, no pertenecen al portugués europeo, así como los textos con unidades fraseológicas que, pese a ser introducidas con un presentador paremiológico, no son paremias (colocaciones, locuciones verbales, fórmulas oracionales, frases célebres o paremias inventadas), en definitiva todas aquellas que no son pertinentes o relevantes para el estudio. Ejemplos de casos descartados:

- Os árabes têm um provérbio que diz: "O nosso inimigo, não é aquele que nos odeia, mas aquele que nós odiamos." (Blog Recuperar das dependências, 21-02-08).
- Em Belém do Pará, no norte do Brasil, há um ditado que diz: "Chegou ao Pará parou, tomou açaí ficou". (*Expresso*, 05-02-2011).
- E o resultado, segundo Buffett, tenderá a ser o do adágio que diz "quando uma pessoa com dinheiro conhece uma pessoa com experiência, a que tem experiência fica com o dinheiro e a que tem o dinheiro fica com a experiência". (*Jornal de Negócios*, 27-02-2017).
- PROVÉRBIO DA SEMANA: "No mercado de Alcobaça só o preço é que massa".
  Episódio 13. Alcobaça a Punho. Super Mercados. Sapo Vídeos (25-06-2009).

Después de analizar los resultados y descartar los textos que no cumplían los criterios de la investigación, la muestra quedo configurada con un total de 4237 documentos.

### 1.4. Unidad de registro

El paso siguiente consiste en extraer de cada texto la unidad de registro que nos interesa para la investigación, es decir, las paremias, entendiendo por ello los enunciados breves y sentenciosos que corresponden «a una oración simple o compuesta, que se ha fijado en el habla y que forma parte del acervo sociocultural de una comunidad hablante» (Sevilla Muñoz; Crida Álvarez, 2013: 106). Siguiendo esta definición de paremia, seleccionamos por un lado, paremias de origen anónimo y popular como refranes (*Quem semeia ventos colhe tempestades*), frases proverbiales (*Um dia não são dias*), locuciones proverbiales (*Ir por lã e sair tosquiado*) y dialogismos (*Diz o roto ao nu: porque não te vestes tu?*), y, por otro, paremias de origen conocido y culto como proverbios (*Faça o bem sem olhar a quem*) y aforismos (*Mente sã em corpo são*). No obstante, cuando hemos tenido dificultad en distinguir las paremias de otras unidades fraseológicas se ha optado por incluir aquellas que en los textos analizados aparecen citadas como paremias (*provérbios*, *ditados*, *adágios*, *máximas*, *aforismos*, *ditos*, o *sentenças*) y aparecen registradas como tal en los diccionarios especializados consultados[4].

Siguiendo estos parámetros, se examinan todos los textos seleccionados, se extraen las unidades que cumplen los requisitos establecidos y se organizan en fichas para su posterior codificación. Dado que en esta fase inicial se trabaja con un volumen muy extenso de paremias, la unidad de contexto que se selecciona es la que aparece como resultado en la búsqueda de Internet, es decir, una mínima parte de la frase o párrafo en la que se encuentra la paremia. Los datos incluidos en la ficha permitirán posteriormente localizar la paremia y ampliar el contexto u otros datos, si fuera necesario.

[3] Se incluyen algunos textos literarios contemporáneos cuya primera edición es anterior al 2000.

[4] Las obras consultadas en la elaboración del mínimo paremiológico del portugués europeo se pueden consultar en el capítulo 2.2. «Caracterización de las paremias objeto de estudio» (Díaz Ferrero y Sabio Pinilla, 2023).

| | | | |
|---|---|---|---|
| **Paremia** | Só se lembra(m) de Santa Bárbara quando troveja | | |
| **Unidad contextual** | *Diz o ditado* popular que só nos lembramos de Santa Bárbara quando faz trovoada. E é muitas vezes no Inverno que as tempestades e catástrofes naturais fazem... | | |
| **Fuente documental** | *Jornal O Mirante* | | |
| **Forma de la paremia** | forma canónica | | |
| | Desautomatizada | X | |
| | Truncada | | |
| **Tipo de paremia** | Refrán | X | |
| | frase proverbial | | |
| | locución proverbial | | |
| | Proverbio | | |
| | Aoforismo | | |
| | Dialogismo | | |
| **Tipo de texto** | Periodístico o divulgativo (periódicos, radio, televisión, blogs...) | | X |
| | Narrativo o literario (novelas, teatro, comics, canciones, películas...) | | |
| | Publicitario o comercial (anuncios o carteles publicitarios, folletos comerciales y turísticos ...) | | |
| | Argumentativo (monografías, ensayos, artículos científicos, debates...) | | |
| | Epistolar (cartas, correos electrónicos, comentarios de lectores...) | | |
| **Formato** | Escrito | X | |
| | Oral | | |
| | Imagen | | |
| **Región** | No se especifica | | |
| | Norte | | |
| | Centro y Beiras | X | |
| | Lisboa y Vale do Tejo | | |
| | Alentejo | | |
| | Algarve | | |
| | Azores | | |
| | Madeira | | |
| **Fecha** | 14-01-2016 | | |

Ejemplo de ficha de unidades de registro (Fuente: elaboración propia)

### 1.5. Codificación

Los datos obtenidos mediante el AC se gestionan con la ayuda del software estadístico SPSS lo que permite sistematizar la información y poder analizar y evaluar los datos posteriormente. A cada paremia y sus categorías de análisis se les asigna un código siguiendo una plantilla previamente establecida. El procesamiento de los datos mediante la codificación y tabulación de las unidades de registro permite obtener no solo una relación de todas las paremias en uso en los documentos consultados junto con tablas de frecuencias de las paremias más citadas, sino también información relativa a las paremias, clasificarlas por regiones, tipo de texto u otro componente contemplado en la ficha. Así mismo, se pueden obtener datos sobre la forma canónica de las paremias, sus variantes o posibles formas desautomatizadas. En suma, el tratamiento informático de las paremias recopiladas nos proporciona una base de datos que sirve como corpus de trabajo para testar la frecuencia de uso en las siguientes fases del estudio, pero también nos ofrece información complementaria de gran utilidad para el estudio de las paremias.

**1.6. Resultados**

Después de analizar y evaluar la información recuperada en Internet, con una muestra formada por 4237 documentos, se obtuvo una base de datos con 867 paremias diferentes que, junto con las paremias obtenidas mediante la encuesta de respuesta espontánea nos permitió crear un corpus inicial de 961 paremias.

De las 867 paremias seleccionadas, 61 % corresponden a refranes, 24 % a frases proverbiales, 8 % a proverbios, 4 % a aforismos y 3 % a locuciones proverbiales. Muchas de ellas aparecen tan solo una o dos veces como el caso del refrán *Quem não governa a lenha, não governa a casa que tenha:*

> Nessa noite, o pai não se deitou. Ficou a matutar na conversa dos filhos. Lembrava-se de um ditado antigo" Quem não governa a lenha, não governa casa que tenha".Afinal, quantos braços haveriam na aldeia? pensou. (Palavra de rei tem que voltar atrás! En: Conta a história, querida avó. https://conta-a-historia-querida-avo.webnode.pt/, 04-01-2013).

Otras paremias, sin embargo, aparecen en diferentes textos y en diferentes formatos; por ejemplo, el refrán *De pequenino é que se torce o pepino* se localiza, entre otros, en un informe elaborado por el diputado del PSD Mendes Bota:

> Até então, o que se aceitava era algo que a sabedoria popular traduzia no provérbio "de pequenino se torce o pepino": o cérebro não mudava grande coisa desde a infância. (Relatório elaborado pelo deputado do PSD Mendes Bota, 20-05-2010).

En una noticia sobre la influencia que Internet ejerce en nuestro cerebro publicada en el suplemento Pública del periódico *Público*:

> Até então, o que se aceitava era algo que a sabedoria popular traduzia no provérbio "de pequenino se torce o pepino": o cérebro não mudava grande coisa desde a infância. (*Público*, 31-10-2010).

En el título de una imagen de la sección “Imagem do dia” de la radiotelevisión portuguesa RTP desporto: De pequenino se torce o “menino”

> A competição é real, mas apesar do duelo ser numa arena de Sumo, o que está em jogo é saber qual das crianças chora mais e faz a cara mais feia. O "baby crying contest" decorre no templo de Sensoji, em Tokyo, no Japão. (https://www.rtp.pt/noticias/imagem-do-dia/desporto/, 28 Abril 2018).

O en un texto divulgativo sobre la importancia de tomar calcio desde pequeños:

> Lá diz o ditado: “de pequenino é que se torce o pepino”. É na fase inicial da vida que se dita a saúde para os anos futuros! (https://www.blissnatura.pt/, 13-11-2017).

Respecto al tipo de texto en el que se insertan, la mayoría (54 %) de las paremias obtenidas se localizaron en textos de carácter periodístico o divulgativo, principalmente en periódicos y blogs, seguido de textos epistolares (13 %), fundamentalmente mensajes o comentarios de los lectores en medios sociales o medios de comunicación; texto literarios o narrativos (12 %); publicitarios o comerciales (11 %), muchos de ellos en imágenes, y textos argumentativos de carácter académico o político (10 %). Ejemplos:

**Prensa**

- "Achado não é roubado": o ditado popular que não é para seguir à letra (*Diário de Notícias*, 23-08-2010).
- Realmente, como diz o povo, "mais vale cair em graça que ser engraçado", e o estranho caso do treinador José Peseiro ilustra bem a asserção popular. (*Correio da manhã*, 27-09-2005).
- Costa sugere banir o ditado "entre marido e mulher não se mete a colher" (*Jornal de Negócios*, 08-03-2017).
- Entrámos no ano de 2018 com o pé direito! Mas, bem diz o ditado que "até ao lavar dos cestos é vindima" e quem anda neste meio sabe isso melhor que ninguém. (*Visão*, 24-08-2018).

**Radio y televisión**

- "Diz-me com quem andas e dir-te-ei quem és" é o provérbio que nos aconselha a escolher bem as nossas companhias para que não sejamos mal influenciados. (*Rádio Renascença*, 30-04-2018).
- Diz o provérbio que quem corre por gosto não cansa. É o caso dos nossos convidados que fizeram do atletismo, uma carreira. O Sociedade Civil propõe falar dos treinos, das provas, das vitórias e de muito mais. (*RTP*. Programa Sociedade Civil. Atletismo. Ep. 16, 16-02 2016).
- Diz o provérbio popular que pelo São Martinho come-se castanhas e bebe-se vinho e o Verão de São Martinho são três dias e mais um bocadinho. A valsousa.tv foi para a rua descobrir um pouco mais sobre a nossa cultura popular. (Vale do Sousa TV; 2017).

**Blogs**

- Uma desgraça nunca vem só. Diz o ditado é bem verdade (eu e a minha mania dos ditados populares, que me ficou da minha avó, que não se cansava de os usar). (Blog Nada acontece por acaso,19-09-2013).
- Quase todos conhecemos o ditado popular "E*m abril, águas mil"*. Já o interiorizamos e ninguém o questiona. (Blog Narrativa Diária, 03-04-2014).
- Teres fama sem proveito é uma "coisa lixada"... Às vezes, só queremos ser divertidos! Mas lá diz o ditado, onde há fumo há fogo. O facebook é um sítio que, para algumas pessoas, vale tudo. Eu prefiro usar o meu blogue para as minhas "confidências" de momento. (Blog, Tudo avulso, 09-10-2015).
- Já diz o ditado "até ao lavar dos cestos é vindima"! Este ditado tem diversas aplicações, em várias áreas, mas todas elas têm o mesmo significado, as coisas só acabam no final! (Blog A bola não é quadrada!!, 29-04-2019).

**Narrativo**

- *E se da parte do comum dos ocidentais reina o ditado* longe da vista, longe do coração, Rafael nem isso espelha, é como se tivesse feito o correcto e, desse ponto de vista, o que está feito, feito está, e não se fala mais nisso. (Luís Miguel Rocha, *O último Papa*, Porto: Porto editora, 2011, p. 185).
- Quem sai aos seus... (*Cartoon* de Henrique Monteiro, 07/03/2013).
- Está certo o ditado que diz que a esperança é a última a morrer. Mas também é igualmente certo que «só há fumo, se existir lume». É desejável que alguém o acenda e provoque uma revolução de Amor (António Pinto da Silva, Caducos & Ginetos, *Escritas de Penamacor. Antologia de textos literários*. Penamacor: Câmara Municipal de Penamacor, p. 202).

**Publicitario o comercial**

- É bem conhecido o provérbio que *"não se pode agradar e gregos e troianos"*; e nem sempre é possível alcançar um estímulo global de boa disposição e aceitação das decisões gerais. (Agencia Dual Up consulting).
- Publicidad de la cerveza Super Bock: "O que vem à rede é super" A Vida é Super.
- A Nossa Empresa.
  Como diz o ditado popular "Roma e Pavia não se fizeram num dia", e a MADECA-Madeiras de Caxarias, S.A. também não; foi fundada em 1952 com sede em Caxarias Norte, e a actividade inicial constituiu na compra de matéria... (Madeca.pt).
- NOVIDADES • GUIMARÃESHOPPING
  Celebre o Mês Saudável no seu Centro com dois showcookings

Abril, receitas mil. Mudamos o provérbio para lhe apresentar os eventos que irão fazer de abril, o Mês Saudável no nosso Centro. (guimarãeshopping 10-04-2018).

**Epistolar**

- A mim dá-me vontade de rir. Porque, como diz o adágio popular: ladrão que rouba a ladrão, tem cem anos de perdão. (Comentário. Anónimo. https://sporting.blogs.sapo.pt/ 30-04-2014).
- É! Tal como diz o provérbio, por vezes o "feitiço vira-se contra o feitiçeiro". Só se realizam quando amesquinham? Ai, que falta de paciência para gente dessa! (Comentário sobre Grande Ego?! Blog Uma carta fora do baralho, 26-07-2017).
- Porque efetivamente a internet nos coloca numa aldeia global, foi finalmente criado o site da Freguesia de Coutos de Viseu. Lá diz o ditado "mais vale tarde que nunca"! (Mensagem do presidente da Junta da Freguesia de Coutos de Viseu, 27-09-2018).

**Argumentativo**

- Como diz o adágio popular "ser avó é ser mãe duas vezes". Significa dar duas vezes carinho e amor. Ser avó é um exemplo de paz, carinho e amor Ana Teresa Pais Mordono Campos, A Família e o Cuidado dos Avós. *Relatório Final da Prática de Ensino*, Coimbra: Universidade Católica Portuguesa, 2018).
- Qualquer que fosse a opção feita, haveria sempre quem a criticasse. Mas como diz o provérbio "a frases loucas, orelhas moucas" (Ata da sessão ordinária da Assembleia municipal de Viseu, 29-06-2015).
- Por isso mesmo, Srª Secretária de Estado, queria citar-lhe um provérbio popular: «Diz-me com quem andas, dir-te-ei quem és». Neste caso, o Partido Socialista anda de mão dada com aqueles que, lá fora, representam o pior que existe do ponto de vista do protecionismo e do isolacionismo (Debates parlamentares, nº 065, 18-03-2017).

## CONCLUSIONES

En este artículo se ha presentado la metodología llevada a cabo para realizar un AC con el objetivo de obtener una relación de paremias usadas actualmente en mayor o menor grado en el portugués europeo. Hemos trabajado con un corpus construido principalmente con documentos en formato digital (texto, imagen y audio) por lo que se ha podido constatar que la web es una gran fuente de datos con ejemplos reales del uso que los nativos hacen de la lengua.

Para gestionar y clasificar los datos se ha elaborado una ficha en la que se recogen no solo las paremias, sino también las categorías o factores de análisis asociados a cada paremia (tipo de texto, formato, región...). Esta ficha junto con los códigos creados para registrarla en el programa SPSS ha resultado ser de gran utilidad para conseguir una relación de paremias en uso, pero también para obtener información complementaria muy útil en todo el proceso de investigación y para la presentación final de las paremias en fichas descriptivas.

Cabe señalar también que la comparación de la información conseguida mediante encuestas y la obtenida mediante el AC nos ofrecen datos próximos, pero dispares. Por ejemplo, algunas paremias, cuyo uso ha quedado claramente constatado mediante el AC, apenas fueron mencionadas por los hablantes en las encuestas realizadas. Es el caso del refrán *ano novo, vida nova*, que tiene una elevada presencia en documentos periodísticos, comerciales, así como en programas de radio y televisión, pero solamente fue mencionado por una persona en las encuestas. Ello puede deberse a que las paremias normalmente se usan cuando se produce un determinado estímulo contextual, circunstancial o temporal que propicia la alusión a la paremia, y las encuestas no se realizaron ni a finales de año ni durante el mes de enero. En definitiva, para garantizar la representatividad y fiabilidad del corpus de trabajo es necesario emplear diferentes técnicas de investigación. En nuestro caso, la combinación del AC y las encuestas nos ha permitido obtener datos fiables para elaborar el mínimo paremiológico por lo que consideramos que este modelo de análisis se puede aplicar a otros estudios exploratorios en otras lenguas.

## REFERENCIAS BIBLIOGRÁFICAS


Andréu Abela, J. (2002). *Las técnicas de Análisis de Contenido: Una revisión actualizada*. Sevilla: Fundación Centro de Estudios Andaluces. https://abacoenred.com/wp-content/uploads/2019/02/Las-t%C3%A9cnicas-de-an%C3%A1lisis-de-contenido-una-revisi%C3%B3n-actualizada.pdf [consultado: 10/2/2024]

Anscombre, J. C. (2010). Las formas sentenciosas: un fenómeno lingüístico. *Revista de Investigación Lingüística*, 13: 17-43. https://revistas.um.es/ril/article/view/114111 [consultado: 10/2/2024]

Bardin, L. (1977). *Análise de conteúdo*. Lisboa: edições, 70.

Baur, R. S.; Chlosta, CH.; Grzybek, P. (1996). Das Projekt Sprichwörter-Minima im Deutschen und Kroatischen. What is worth doing – do it well. *Muttersprache*, 2: 162-179.

Berelson, B. (1952). *Content Analysis in Communications Research*, Nueva York: Free Press.

Berelson, B. y Lazarsfeld P. F. (1948). *The Analysis of Communication Conten*t, Chicago y Nueva York: University of Chicago and Columbia University.

Ĉermak, F. (2003). Paremiological Minimum of Czech: The Corpus Evidence. Burger, H; Häcki Buhofer, A.; Gréciano, G. (eds.), *Flut von Texten - Vielfalt der Kulturen: Ascona 2001 zur Methodologie und Kulturspezifik der Phraseologie*. Baltmannsweiler: Schneider Verlag Hohengehren, 15-31.

Colle, R. (2011). *El análisis de contenido de las comunicaciones*. La Laguna: Sociedad Latina de Comunicación Social.

Corpas Pastor, G. (2014). El fraseólogo internauta: cómo pasarlo pipa en la red. Durante, V.; Tschann, C. (Coords.), *Fraseología y paremiología: enfoques y aplicaciones*. Madrid: Centro Virtual Cervantes, 133-152.

Crida Álvarez, C. A. (2021). *El mínimo paremiológico griego con correspondencias en español*. Madrid: Centro Virtual Cervantes (Instituto Cervantes). https://cvc.cervantes.es/lengua/biblioteca_fraseologica/m3_crida/ [consultado: 10/1/2024]

De Schryver, G. M. (2002). Web for / as corpus: a perspective for the African languages, *Nordic Journal of African Studies*, 11(2): 266-282. https://tshwanedje.com/publications/webtocorpus.pdf [consultado: 23/1/2024]

Denzin, N. K. (1970). *The Research Act: A Theoretical Introduction to Sociological Methods*. New Jersey: Transaction Publishers.

Díaz Ferrero, A. M.; Sabio Pinilla, J. A. (2023). *El mínimo paremiológico portugués. El portugués europeo*. Madrid: Centro Virtual Cervantes (Instituto Cervantes).

Ďurčo, P. (2014). Empirical research and paremiological minimum. Hrisztalina Hrisztova-Gotthardt, Melita Aleksa Varga (eds.), *Introduction to Paremiology: A Comprehensive Guide to Proverb Studies*. Berlin/Munich/Boston: Walter de Gruyter, 183-205.

Espín López, J. V. (2002). El análisis de contenido una técnica para explorar y sistematizar información, XXI. *Revista de educación*, n.º 4: 95-106. https://rabida.uhu.es/dspace/bitstream/handle/10272/1913/b15141895.pdf?sequence=1 [consultado: 23/1/2024]

Forgács, T. (2014). Über das parömische Minimum des Ungarischen. *Proverbium*, 31: 255-278. https://hrcak.srce.hr/file/403067 [consultado: 16/1/2024]

González Fernández, A. (2017). La web como corpus: un esbozo. *Lengua y habla*, 21: 126-150.

González Fernández, A. (2018). Aproximación a la Lingüística Computacional y sus herramientas para el trabajo con corpus. *Futhark* , 13: 37-52. https://revistascientificas.us.es/index.php/futhark/article/view/15938 [consultado: 23/1/2024]

Gil Pascual, J. A. (2011). *Técnicas e instrumentos para la recogida de información*. Madrid: UNED. https://www.calameo.com/read/0019555911a566c110989 [consultado: 2/2/2024]

Haas, H. A. (2008). Proverb Familiarity in the United States: Cross-Regional Comparisons of the Paremiological Minimum. *Journal of American Folklore*, vol. 121, n.º 481: 319-347. https://www.jstor.org/stable/20487611 [consultado: 13/2/2024]

Huang, X.; Mohd Nawi, A. (2024). Establishment of a Chinese Paremiological Minimum Based on International English Media, *Theory and Practice in Language Studies*, vol. 14, n.º 1: 167-176: https://tpls.academypublication.com/index.php/tpls/article/view/7235/5913 [consultado: 15/1/2024]

Kilgarriff, A.; Grefenstette, G. (2003). Introduction to the Special Issue on the Web as Corpus, Computational Linguistics, v. 29, 3: 333-347. https://aclanthology.org/J03-3001.pdf [consultado: 15/1/2024]

Korshunova, A. A.; Bochina, T. G.; Zharkynbekova, S. K. (2021). Um algoritmo para a formação do mínimo paremiológico russo. *Revista EntreLinguas*, Araraquara, v. 7, n. esp. 7. https://periodicos.fclar.unesp.br/entrelinguas/article/view/16301 [consultado:14/2/2024].

Krippendorf, K. (1990). *Metodología de análisis de contenido: Teoría y práctica*. Barcelona: Paidós.

Lapata M.; Keller, F. (2005). Web-based Models for Natural Language Processing. *ACM Transactions on Speech and Language Processing*, vol. 2, n.º 1: 1-30. https://homepages.inf.ed.ac.uk/mlap/Papers/tslp05.pdf [consultado: 4/2/2024]

Lasswell, H. D. (1927). *Propaganda Technique in the Worl War*. Nueva York: A. A. Knopf.

López Noguero, F. (2002). El Análisis de contenido como método de investigación, XXI. *Revista de educación*, n.º 4: 167-180.

Mieder, W. (1994). Paremiological Minimum and Cultural Literacy. W. Mieder (ed.) *Wise Words. Essays on the Proverb*. New York: Garland Publishing, 297–316.

Morse, J. M. (1991). Approaches to Qualitative-Quantitative Methodological Triangulation. Method ology Corner. *Nursing Research*, 40(2): 120-123.

Palmquist, M. E.: Carley, K.M.; Dale, T. A. (1997). Two applications of automated text analysis: Analyzing literary and non-literary texts. En C. Roberts (Ed.), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Tanscripts*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pamies Bertrán, A. (2017). Fraseología y variación diatópica en español. *Verba hispanica:* nº. 25: 55-81

Permiakov, G. L. (1970). *Ot pogovorki do skazki*, Moskva: Nauka.

Permiakov, G. L. (1973). On the Paremiological Level of Language and the Russian Paremiological Minimum. *Proverbium*, 22: 862-863.

Permiakov, G. L. (1985). *300 obscheupotrebitel´nykh russkikh poslovits i pogovorok dlia govoriaschikh na nemetskom yiazyke*,. Moskva: Nauka.

Piñuel Raigada, J. L. (2002). Epistemología, metodología y técnicas del análisis de contenido. *Estudios de Sociolingüística*, 3(1): 1-42. https://www.ucm.es/data/cont/docs/268-2013-07-29-Pinuel_Raigada_AnalisisContenido_2002_EstudiosSociolinguisticaUVigo.pdf [consultado: 21/1/2024]

Rodríguez Sabiote, C.; Pozo Llorente, T.; Gutiérrez Pérez, J. (2006). La triangulación analítica como recurso para la validación de estudios de encuesta recurrentes e investigaciones de réplica en Educación Superior. *RELIEVE, Revista electrónica de investigación y evaluación educativa*, vol. 12, n.º 2: 289-305. http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_6.htm [consultado: 21/1/2024]

Reis, S.; Baptista, J. (2020). Determinação de um mínimo paremiológico do português europeu. *Acta Scientiarum. Language and Culture*, vol. 42, n.º 2. https://www.redalyc.org/journal/3074/307466046014/html/ [consultado: 21/1/2024]

Sardelli, M. A. (2011). Recopilación y sistematización del mínimo paremiológico italiano. Pamies Bertrán, A.; Luque Durán, J.D.; Fernández Martín, P. (eds.), *Paremiología y herencia cultural*, Granada: Educatori, 87-95.

Sevilla Muñoz, J.; Crida Álvarez, C. A. (2013). Las paremias y su clasificación, *Paremia*, 22: 105-114.

Tarnovska, O. (2004). *Refranero básico español con correspondencias en ruso y ucraniano*. Kiev: Logos.

Tinto Arandes, J. A. (2013). El análisis de contenido como herramienta de utilidad para la realización de una investigación descriptiva. *Provincia*, 29: 135-173. https://www.redalyc.org/pdf/555/55530465007.pdf [consultado: 13/2/2024]

Torruella, J.; Llisterri, J. (1999). Diseño de corpus textuales y orales. Blecua, J. M. Clavería, G.; Sánchez, C.; Torruella, J. (eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Editorial Milenio, 45-77.

Xatara, C. M. (2008). A web para um levantamento de freqüência. Magalhães, J. S.; Travaglia, L. C. (orgs.), *Múltiplas perspectivas em lingüística*. Uberlândia: EDUFU, 770-777. http://www.filologia.org.br/ileel/artigos/artigo_398.pdf. [consultado: 22/1/2024]

Zurdo Ruiz-Ayúcar, M.ª T.; Sevilla Muñoz, J. (2016). *El mínimo paremiológico: aspectos teóricos y metodológicos*. Madrid: Centro Virtual Cervantes (Instituto Cervantes). http://cvc.cervantes.es/lengua/biblioteca_fraseologica/m1_zurdo/ [consultado:13/2/2024]