

## A TODO CHANCHO LE TOCA SU NAVIDAD Estudio de corpus sobre la variación geoparémica en el español actual

VITO ANTONIO MICCOLI  
UNIVERSIDAD COMPLUTENSE DE MADRID

**Abstract** - Although limited, research on paremiological variation constitutes a valuable contribution to achieving a more balanced representation of the various voices within the Hispanophony. An even smaller number of studies have explored the connection between paremiology and corpus linguistics, partly attributable to the structural and semantic complexity of paremias. This work aims to shed light on the advantages of combining these two disciplines. This article is the result of a corpus-based study on geoparemiological variation. We have worked with two lexical variants of a proverb: the first, traditionally considered canonical in the European context; the second, which, according to our findings, has attained the status of reference form on the American continent. After identifying global patterns of dispersion for the lexical variants in question, an interterritorial comparison was conducted on their diffusion and frequency in two macro-varieties of contemporary Spanish: Rioplatense Spanish and European Spanish. This methodological approach to paremiology opens new lines of investigation, providing an enriching perspective on the proverb as a unit of linguistic and cultural analysis. Additionally, this methodology for data consultation and collection is of significant relevance to other fields of knowledge, such as language acquisition and translation, insofar as it enables the identification of functional and translation equivalents across the various geolinguistic varieties of Spanish.

**Keywords:** geoparemiology; corpus linguistics; lexical variation; varieties of Spanish; translation.

### 1. Introducción

Las investigaciones desarrolladas desde una perspectiva variacionista han sentado las bases para una reorientación teórica y metodológica en el estudio de la fraseología y la paremiología. En una fase inicial, el rasgo de fijación fue considerado como un criterio inamovible para la caracterización de las unidades fraseológicas (UF), pese a que la variabilidad comenzaba ya a despertar el interés de la comunidad investigadora, dada su relativa omnipresencia (García-Page 1999; 2008; Mogorrón 2020). Tradicionalmente, de los rasgos mencionados, únicamente la fijación fue admitida como definitorio, mientras que la variación quedaba relegada a una mera constatación de manifestaciones sintópicas (López 2020, p. 145). Posteriormente, los estudios centrados en la variación fraseológica impulsaron un cambio de paradigma, propiciando una aceptación casi unánime de un modelo teórico más flexible, en el que la fijación se concibe como un fenómeno gradual (García-Page 2008; Mellado 2004; Penadés 1996). De este modo, estabilidad y variación pasan a entenderse como dos dimensiones complementarias e indisociables del fenómeno fraseológico y paremiológico, hasta el punto de ser consideradas las dos caras de la misma moneda (Penadés 2022, pp. 23-24).

El concepto de variación, entendido como la capacidad de todo sistema lingüístico natural de producir diferentes realizaciones en el habla (Sinner y Tabares 2016, p. 15), se presenta como una noción compleja y polisémica, en parte por la heterogeneidad terminológica utilizada para describirlo. En el caso del español, esta complejidad se acentúa

por el alto grado de difusión geográfica de la lengua, que da lugar a una rica variación diatópica. Es incuestionable que la variación afecta a todos los niveles del análisis lingüístico, incluido el nivel paremiológico tal como evidencian con claridad Scutto y Asensio (2022), editoras de un número especial de la *Revista Nebrja*, parcialmente dedicado a la interdependencia entre diatopía y fenómenos de variación y cambio en el panorama fraseológico y paremiológico. Paralelamente, proyectos de envergadura tales como el *Refranero multilingüe* (Sevilla y Zurdo 2009) o *ParemiRom* (Gargallo y Álvarez 2014) abordan la variación parémica respaldándola con muestras de sólido rigor científico. Efectivamente, en un panorama global, las paremias no parecen someterse a limitaciones rígidas. En particular, el refrán ha sido descrito como una entidad textual que “no respeta fronteras ni, en sentido estricto, tiene patria: va, viene, arraiga, se adapta en dondequiera que es adoptado y, cuando ya no es útil, simplemente desaparece o emigra en busca de alguien más que se interese en él” (Pérez 2007, p. 45).

A partir de esta premisa, el presente estudio se propone analizar la variación léxica de uno de los refranes más emblemáticos del repertorio peninsular: *A cada cerdo le llega su San Martín*, el cual ha traspasado las fronteras geográficas y culturales hasta adquirir nuevas formas y matices en diferentes territorios hispanohablantes. Para ello, se ha optado por una aproximación metodológica que combina la paremiología y la lingüística de corpus, cuya convergencia, según se ha podido constatar, permite obtener resultados significativos en cuanto a la caracterización de la difusión panhispánica de las paremias.

La investigación, de corte empírico, se basa en el análisis de corpus, lo que implica la utilización de datos reales y exhaustivos que reflejan el uso efectivo de la lengua a través de fragmentos textuales representativos de actos de habla espontáneos en contextos comunicativos concretos (Rojo 2014, 2021). Al estar constituidos por datos representativos de comunidades lingüísticas específicas, los corpus pueden ser legítimamente considerados como sustitutos o reflejos del uso lingüístico real (Pérez 2020, p. 1). Este planteamiento se ve reforzado por el interés sostenido que la lingüística de corpus ha mostrado por la investigación lexicográfica, así como por el papel destacado que ha desempeñado en la centralización del léxico y de las UF en los estudios lingüísticos (Liendo *et al.* 2022). Una de las principales ventajas de esta metodología reside en la posibilidad de visualizar, recopilar e investigar una cantidad considerable de datos cualitativos, lo que permite superar las limitaciones de los estudios basados en registros restringidos. En este sentido, los corpus digitales han proporcionado a los fraseólogos y paremiólogos una base de datos más sólida y fiable, haciendo posible realizar investigaciones “a gran escala” que arrojan resultados representativos y estadísticamente relevantes (Corpas 1998, p. 368).

El uso combinado de corpus de referencia y gigacorpus web permite abordar el fenómeno de la variación parémica desde una perspectiva documentada y cuantitativa. Mediante este enfoque digital, el presente estudio persigue tres objetivos fundamentales. En primer lugar, se propone un modelo de aproximación al estudio de la variación paremiológica, área aun escasamente explorada, como reconocen Sevilla y Ugarte (2022, p. 94). En este marco, se pretende demostrar la amplia diversidad formal que puede presentar una sola unidad paremiológica al ser sometida a una revisión sistemática a partir de múltiples fuentes, tanto computacionales como tradicionales. Con frecuencia, dicha variación se relaciona con la diversidad geográfica en que se produce su uso.

En segundo lugar, se busca establecer patrones de difusión en la hispanoesfera de la variante léxica americana *A cada chancho le llega su San Martín*, aplicando una metodología específica de localización de paremias en corpus electrónicos. La identificación de dichos patrones se fundamenta en parámetros de frecuencia relativa y absoluta, procedentes de la observación de ocurrencias asociadas a determinadas entidades geográficas.

Por último, se analiza comparativamente la difusión y frecuencia de dos variantes léxicas de la paremia en cuestión —la forma europea y la americana— mediante un enfoque inicialmente lexicográfico, que se complementa con un análisis paremiográfico y contrastivo. La comparación se efectúa entre dos macrovariedades del español: la variedad estándar peninsular y la variedad estándar rioplatense. La considerable abundancia de datos documentados y su segmentación en torno a estas dos regiones justifican el enfoque contrastivo adoptado, lo cual favorece desentrañar el particularismo léxico de ambas variedades.

La adopción de los corpus digitales para el estudio de la variación geoparémica no solo resulta clave en la descripción de los patrones de uso, sino que también aporta beneficios en otros ámbitos del saber, tales como la didáctica de lenguas y la traducción. En efecto, la ausencia de recopilaciones paremiográficas exhaustivas y contrastivas que documentan las variantes léxicas de un refrán en todas las variedades del español, genera obstáculos a la hora de localizar equivalentes funcionales o traductológicos. Por ello, la metodología aquí propuesta permite el acceso a registros detallados sobre la variación geoparémica de un refrán, facilitando así la toma de decisiones fundamentadas en contextos traductológicos y pedagógicos.

## 2. Apuntes sobre la variación léxica geoparemiológica

Los refranes, en tanto UF de carácter sentencioso, presentan determinados rasgos lingüísticos que favorecen su estabilidad y memorización, como la marcada tendencia al bimembrismo o el empleo de recursos mnemotécnicos, entre ellos la rima interna. Estas propiedades estructurales facilitan su transmisión intergeneracional. No obstante, tales características no impiden que estas unidades experimenten procesos de variación. En esta línea, y siguiendo las directrices de Jarilla (2024, p. 278), para enmarcar un concepto tan amplio como el de variación —difícil de delimitar teórica y empíricamente— resulta pertinente integrar tanto estudios teóricos aplicados a las UF como investigaciones centradas en la variación paremiológica, dado que ambas comparten características estructurales fundamentales como la fijación.

Existe consenso entre fraseólogos y paremiólogos en identificar la sustitución de los componentes léxicos como uno de los mecanismos primordiales responsables de la variación fraseológica. Según lo plantea Zurdo (2002, p. 131):

Una de las constantes que se pueden postular en relación con el estudio científico y aplicado de la fraseología es, justamente, la relativa a la activación de mecanismos que tienen- o pueden tener- como consecuencia la alteración morfosintáctica, semántica o funcional de la lexía que constituye la base invariante imprescindible en toda unidad fraseológica.

Este mecanismo de variación, que incide directamente sobre los componentes de una estructura fija se ha denominado variación léxica. García-Page (1999, p. 225) reconoce que algunas UF pueden modificar su estructura léxica —en relación más o menos directa con el contexto— hasta el punto de que cabe hablar de variantes léxicas. En esta misma línea, Montoro del Arco (2005, p. 138; 2006, p. 60) respalda dicha postura y explica que las variantes léxicas afectan, precisamente, al componente léxico proporcionando una alternativa por sustitución. También Mogorron (2020, pp. 18-20) en su estudio sobre las construcciones verbales fijas (CVF), determina que uno de los procesos de variación de las UF es la sustitución léxica, clasificando la variación léxica nominal y la verbal como las más frecuentes. Rigurosa, en su exposición, es la postura de Alvarado (2008, p. 18) quien

considera “que las variantes se dan en las UF<sup>s</sup> cuando presenta un solo cambio léxico en su estructura, bien sea por adición o reducción, bien por sustitución, sin alterar su componente principal, su significado, en los contextos en los que puede aparecer”. Más recientemente, Penadés (2022, pp. 25-27) introduce el concepto de “variación formal léxica” entendido como el proceso que afecta a una o varias unidades léxicas constitutivas de una UF.

Tal y como lo exponen Corpas y Mena (2003, p. 186), mediante criterios formales es posible identificar los cambios léxicos responsables de la emergencia de variantes, particularmente en dos de las tres esferas fraseológicas establecidas por Corpas Pastor (1996): las locuciones y los enunciados fraseológicos. Las autoras defienden el carácter institucionalizado de dichas variantes, en muchos casos motivadas por relaciones semánticas como la sinonimia. A modo ilustrativo, se citan *A caballo regalado no le mires el diente* y su variante léxica *A caballo regalo no le mires el dentado* (Corpas, Mena 2003, pp. 186-187).

Para el presente estudio, resulta crucial enfatizar la codependencia entre variación léxica y factores geográficos. Al respecto, Penadés (2022, pp. 29-30) distingue entre variantes fraseológicas *no marcadas* —de uso neutro— y *marcadas*, cuyo empleo está condicionado por factores contextuales, como la diatopía, las variables sociales del hablante o el eje temporal. Desde esta perspectiva dialectológica, la variación constituye el resultado directo de los desplazamientos territoriales y culturales que afectan al comportamiento de las unidades fraseológicas en español. En el ámbito geofraseológico y geopolémico, la difusión del español en vastos dominios, a veces muy alejados entre sí, como causa natural, ha dado lugar a variantes geolectales diversas que designan una misma realidad (García-Page 2008, p. 217). En términos similares, Montoro del Arco (2005, pp. 137-143) distingue entre *variantes internas* y *variantes externas*. Las segundas se producen como consecuencia de factores diatópicos, diastráticos o diafásicos, lo cual facilita la comparación intervarietal. A modo de síntesis, puede acogerse la propuesta de Koike (2003, pp. 57-58), quien afirma que la variación léxica se manifiesta a través de la sustitución de los elementos léxicos de la UF, siendo la diatopía una de sus motivaciones más recurrentes.

A nivel paremiológico, se observan igualmente las tendencias variacionistas mencionadas anteriormente. También en este tipo de UF se ha documentado “la potencial existencia de variantes” como fenómeno generalizado, motivado por factores diafásicos, diatópicos o diacrónicos que inciden tanto en su estructura como en su componente léxico (Sevilla y Crida 2013, p. 107). Según Sevilla y Ugarte (2022, p. 94) las variantes son aquellas paremias que conviven de forma paralela con otra canónica o de referencia. En esta línea, Grzybek (2012) propone una clasificación gradual de la variación paremiológica, a partir de su forma canónica,<sup>1</sup> punto de partida con el cual se comparan las variantes, en la que el primer grado de variación consiste en cambios mínimos de la estructura paremiológica, entre ellos la sustitución léxica. Desde una perspectiva filológica, Almela y Sevilla (2000, pp. 14-15) identifican dos tipos de variantes: a) variantes léxicas, que afectan “bien las formas propiamente léxicas, bien las formas léxicas junto con las llamadas formas gramaticales” y b) variantes diatópicas, entendidas como realizaciones geográficamente condicionadas. Este tipo de modificación suele obedecer a un proceso de adaptación cultural en función de las particularidades del territorio en el que se emplea la paremia, fenómeno ampliamente documentado en la hispanofonía. En este sentido, resulta ilustrativa la afirmación de Shirley Arora quien, ya hace más de dos décadas, advertía que:

<sup>1</sup> La forma que registra un uso predominante según los diccionarios actuales o, en el caso de formas caídas en desuso, la realización que ha evidenciado un desarrollo diacrónico más significativo (Klein 2006, p. 163; 2017, p. 38).

Se ha calculado que alrededor de 80% de los refranes hispanoamericanos son de origen español o por lo menos han llegado al Nuevo Mundo por medio de los españoles. Esto significa automáticamente que hay mucha semejanza entre los refraneros de distintos países hispanoamericanos. Entre las diferencias que existen, las más relevantes resultan de la sustitución de alguna palabra por otra de tipo regional, o de la invención de refranes nuevos, generalmente mediante fórmulas o estructuras ya conocidas, por ejemplo, en el refrán mexicano «El que con coyote anda, a aullar se enseña», el lobo queda reemplazado por un animal netamente americano. (en Sevilla 2020, p. 10)

De esa misma época data la reflexión de Soto (2000, p. 44) en torno al concepto de *americanismo*, referido tanto a palabras españolas con acepciones nuevas en América como a voces autóctonas o de origen americano que no tuvieron su equivalente en el español peninsular y fueron castellanizadas. Según el autor, con las paremias sucede un proceso análogo, observándose la presencia de refranes que representan puros y auténticos americanismos, autóctonos y vernáculos, reflejo del mestizaje y de la identidad cultural regional (Soto 2000, p. 45). Consolidando esta posición, Torres (2013, p. 102) acuña el término “americanismo parémico” por analogía con el tradicional americanismo léxico, para designar aquellas paremias propias del ámbito americano, desconocidas o en desuso en el español peninsular, que conviven con formas heredadas de la tradición hispánica. Con la misma connotación, Messina (2011, p. 164) habla de “americanismos fraseparemiológicos”, evocando su “auténtico modo de ser, un genuino pensamiento y un verdadero sentimiento criollo, típica y exclusivamente hispanoamericano”. Por su parte, Guerra (1997, p. 299) destaca el contenido sentencioso y moralizante similar a la base de los que llama “paralelismos paremiológicos de las dos orillas de nuestro idioma”, haciendo hincapié en que, con relativa frecuencia, la filosofía popular de estos enunciados sentenciosos se forja en procesos metafóricos análogos, aunque expresados con vocablos locales.

Este fenómeno de variación geográfica ha permitido a varios autores clasificar las UF de acuerdo con su distribución territorial. Una primera categorización distingue entre formas de alcance general, o panhispánico, y formas de alcance local, estas últimas subdivididas en: i) expresiones de uso preferente en España; ii) expresiones mayoritariamente americanas; y, iii) expresiones propias de un país hispanohablante concreto (Koike 2003, p. 47). Un esquema similar es adoptado por Nilo (en Pamies 2017, p. 60) quien argumenta sobre la necesidad de distinguir entre: a) UF usadas a nivel panhispánico; b) UF empleadas prevalentemente en el espacio europeo; c) UF compartidas por varios países americanos; y, d) UF prototípicas de un país americano específico de habla hispana. En el ámbito paremiológico, Torres (2013, p. 96) plantea una taxonomía próxima: i) paremias que presentan un fondo común panhispánico; ii) paremias con alguna variación formal, por lo general léxica; y, iii) paremias de origen netamente autóctono.

Para concluir este breve recorrido teórico sobre la variación léxica paremiológica marcada por diatopía, resulta imprescindible destacar la notable capacidad de adaptación del refranero hispánico, cuya flexibilidad ha motivado que se lo considere, con razón, “un refranero verdaderamente mundial” (Arora 1997, p. 76). Esta calificación pone de manifiesto no solo su vitalidad cultural, sino también su aptitud para integrarse y resignificarse en los contextos más diversos.

### 3. Notas sobre la variación diatópica del refrán

El punto de partida del presente análisis lo constituye la forma canónica o de referencia del refrán, también conocida como grado cero (Grzybek 2012). Para este fin, se ha seleccionado

como forma de referencia<sup>2</sup> la que aparece en la primera ficha del *Refranero multilingüe*, a saber, *A cada cerdo le llega su San Martín*. Este refrán se interpreta como una advertencia moralizante que señala que: “no queda impune el comportamiento del malvado, pues, antes o después, quien ha obrado mal recibe su merecido” (Sevilla y Zurdo 2009). En tanto unidad de uso frecuente y alto grado de fijación, esta paremia posee también un notable potencial de variabilidad. De hecho, la misma base de datos recoge diversas variantes, acompañadas en varios casos por indicaciones sobre su procedencia geográfica. A continuación, se listan algunas de las variantes documentadas:

*Enunciado:* A cada puerco le llega su San Martín

*Enunciado:* A cada puerco le viene su San Martín

*Enunciado:* A cada puerco le llega su sábado / Nochebuena (Puerto Rico)

*Enunciado:* A todo cochino gordo le llega su sábado (Venezuela)

*Enunciado:* A cada marrano le llega su Nochebuena (Colombia, fuente oral)

Estos enunciados ejemplifican la capacidad de adaptación del refrán original, que no solo ha logrado trascender el espacio peninsular, sino que también ha evidenciado una marcada propensión a la variación léxica en función de los territorios de uso. Sevilla (2020, p. 81), en su estudio comparativo sobre la variación en los refraneros español y mexicano, documenta la variante léxica *A cada puerco le llega su San Martín*, asignándola al español de México. Por su parte, Arora (1997, p. 76) ya había registrado una forma relacionada —*A cada chancho le llega su sábado*—, descrita por la paremióloga como “variante común centroamericana de un refrán español muy conocido”. En esta misma dirección apuntan otros autores, quienes refuerzan la vinculación de la paremia en cuestión con el espacio hispanoamericano. Así, Torres (2013, p. 97) consigna la forma *A cada chancho le llega su San Martín*, mientras que Mántica (1999) la recoge en el refranero nicaragüense como *A cada chancho le llega su sábado*. Asimismo, Messina (2011, p. 175) incluye en su estudio sobre los americanismos fraseparemiológicos la variante *A todo gocho le llega su San Martín*, asociada al español de Venezuela. Por su parte, el *Diccionario de americanismos* (DA) documenta otras formas de esta paremia: *A todo chancho gordo le llega su día (de San Juan)*, atribuida a Costa Rica, así como dos variantes pertenecientes al español de Chile: *A todo chancho le llega su mazorca* y *A todo chancho le llega su Navidad*. De forma complementaria, Navarro-Brotóns (2020, p. 122), basándose en encuestas realizadas a hispanohablantes, confirma la existencia de formas que difieren respecto de la variante europea, entre ellas *A cada chancho le llega su San Martín*, documentada como forma de uso frecuente en Argentina.

La asignación de una variante léxica de un refrán a una variedad funcional específica del español plantea no pocas dificultades, tal como se desprende del repertorio presentado. Tal como se apuntaba al finalizar el apartado anterior, existen paremias que comparten un trasfondo semántico común de carácter panhispánico, aunque difieren en el plano formal, especialmente por la sustitución de uno o más de sus componentes léxicos. Conforme a la clasificación propuesta por Mogorron (2020, pp. 22-25), la variación de la lexía en las UF puede deberse a tres tipos de procedencias: a) el español diatópico; b) las lenguas indígenas autóctonas; y, c) las lenguas extranjeras en contacto. En el marco del presente estudio, la

<sup>2</sup> El criterio pragmático que rige el *Refranero multilingüe* pretende buscar la forma más usual de la paremia hoy día o la más utilizada en los textos. El resto de las formas ocupan su lugar entre las variantes (Sevilla 2012).

variación analizada responde al primer tipo, lo cual justifica su abordaje desde una perspectiva geoparémica.

Partiendo de los registros anteriormente expuestos, el presente estudio se propone establecer patrones de difusión de la variante americana *A cada chancho le llega su San Martín* y sus correspondientes formas alternantes. Para ello, se ha recurrido a técnicas específicas de extracción y filtrado de datos, propias del análisis basado en corpus. La configuración de estas técnicas y la descripción detallada de los corpus utilizados se abordarán en los apartados metodológicos subsiguientes.

## 4. Aproximación metodológica al estudio de corpus

Con el objetivo de analizar la paremia *A cada cerdo le llega su San Martín* y examinar su comportamiento léxico condicionado por la diatopía, se ha optado por una aproximación empírica basada en el análisis de corpus. Concretamente, se trata de una investigación *corpus-based*, cuyo propósito principal es la verificación de fenómenos lingüísticos mediante parámetros consultables y sistematizados (Rojo 2021, p. 154; Novodvorski 2017, pp. 1635-1636). En este sentido, tal como plantean Vázquez y Blanco (2022, pp. 347-348), los corpus pueden concebirse como una fuente fiable de información lingüística, en tanto permiten la extracción de muestras auténticas del uso real de la lengua. Esta característica se ve potenciada cuando, como en el presente estudio, se puede condicionar la búsqueda con criterios geográficos específicos. Por esta razón, se establecen criterios de delimitación tanto lingüísticos como espaciales, a fin de acotar el universo de análisis y garantizar la fiabilidad de los datos extraídos.

En lo que respecta al tratamiento digital del refrán, se adopta la metodología de búsqueda para la localización de refranes en corpus lingüísticos digitales propuesta por Miccoli (2025), cuyas directrices permiten optimizar la precisión en la recuperación de datos paremiológicos.

### 4.1. Criterios lingüísticos y espaciales

Con el propósito de limitar la amplitud, potencialmente inabarcable, de los registros disponibles en los corpus, se han definido criterios específicos para la selección de las paremias objeto de estudio. Esta delimitación no solo responde a razones de viabilidad metodológica, sino también a la necesidad de garantizar un umbral de ocurrencias suficiente y representativo. Desde un punto de vista lingüístico, se ha optado por analizar una unidad que, por sus características formales y estructurales, se adscribe inequívocamente a la categoría del refrán, esto es, una paremia de origen anónimo y uso marcadamente popular (Sevilla y Crida 2013, p. 111; Sevilla y Crida 2017, p. 124). En cuanto al criterio espacial, este permite observar cómo se manifiesta la variación en las variedades del español, con especial atención a la peninsular y a la rioplatense. En los estudios basados en corpus digitales, la asignación de una paremia a una determinada variedad geográfica del español se efectúa mediante la identificación del dominio web del fragmento analizado. Así, para contrastar el uso entre el español peninsular y el rioplatense, se seleccionan únicamente registros provenientes de los dominios<sup>3</sup> .es, .ar y .uy. Se excluyen, por tanto, los dominios genéricos como .com, .org o .net, cuya procedencia geográfica resulta ambigua o

<sup>3</sup> Los dominios web están vinculados, respectivamente, con las entidades territoriales de España, Argentina y Uruguay. Dado que cada corpus utiliza una denominación diferente para ellos, en el desarrollo del trabajo se especifica con detalle el nombre asignado a los subcorpus dentro de cada corpus general.

indeterminada. En lo que respecta a la variante austral, se trabaja exclusivamente con los dominios asociados a Argentina y Uruguay<sup>4</sup>.

#### **4.2. Los corpus empleados y el filtrado por geolocalización**

En este apartado se describen los corpus utilizados y las estrategias aplicadas para asignar las paremias a regiones geográficas específicas. En algunos corpus, la información geográfica se encuentra integrada en la interfaz misma; en otros, es necesario aplicar técnicas adicionales de filtrado para obtener dicha información.

Los dos primeros corpus empleados pertenecen al sistema de corpus de la *Real Academia Española* (RAE): i) el *Corpus de Referencia del Español Actual* (CREA), versión anotada 1.0 disponible desde diciembre de 2023; y, ii) el *Corpus del Español del Siglo XXI* (CORPES XXI), versión 1.2 disponible desde noviembre de 2024. En ambos casos, la variable geográfica se presenta junto al fragmento textual consultado, en la sección correspondiente al país de procedencia. A estos, se añade el *Corpus del Español* (Mark Davies 2016), en su sección Web/Dialectos (CdEweb). En este corpus, el país de origen del documento consultado aparece al inicio de cada fragmento textual, lo que facilita la categorización territorial de las ocurrencias.

Sin embargo, la consulta exclusiva de estos tres corpus de referencia habría arrojado un total aproximado de cincuenta ejemplos, cifra insuficiente para sostener un análisis empírico con garantías de representatividad. Por esta razón, se ha optado por complementar los datos mediante la triangulación con corpus de tipo web. Pese al escepticismo que suscita el uso de corpus de este tipo, investigaciones previas han demostrado que los corpus web permiten acceder a muestras lingüísticas tan válidas y auténticas como las ofrecidas por corpus de referencia (Kilgarriff, Renau 2013, p. 14). Son numerosos los trabajos fraseológicos que, partiendo de este supuesto, han logrado formular conclusiones sólidas y relevantes (Cuadrado 2020; Gutiérrez 2024; Mellado 2020; Pamies y Pazos 2023).

En este marco, se ha recurrido al sistema de consulta de corpus *Sketch Engine*, seleccionando tres corpus pertenecientes a la familia de los gigacorpus *TenTen*. Inicialmente, estos corpus fueron utilizados para observar la difusión global de las variantes paremiológicas. En una etapa sucesiva, se les han asignado funciones más específicas: dos de ellos se han empleado para la comparación entre el español rioplatense y el español peninsular; el tercero, para validar los datos obtenidos en el ámbito americano.

El primero de estos corpus es *Spanish Web 2023 (esTenTen23)*, el más reciente de la familia *TenTen* para el español. Se trata de un corpus panhispánico que ofrece representatividad equilibrada de las distintas variedades del español. Para la localización de las variantes europeas y australes, se han utilizado técnicas de filtrado. La herramienta de gestión permite seleccionar subcorpus específicos, facilitando así la extracción exclusiva de los registros correspondientes a los dominios: .ar (Argentina), .es (España) y .uy (Uruguay). El segundo corpus empleado es *Spanish Web 2018 (esTenTen18)*, de la misma familia. En este caso, se han aplicado los mismos criterios de segmentación, recurriendo al trinomio de dominios ya mencionado.

Para finalizar, se incorpora al estudio el corpus *American Spanish Web 2011 (esamTenTen11)*, representativo del español de América. En este corpus, se han consultado únicamente los dominios .ar y .uy. Aunque este corpus no permite una comparación directa con el español peninsular, su inclusión resulta útil para validar los datos americanos

<sup>4</sup> Formarían parte de la variedad austral también el español hablado en algunas zonas de Chile y Paraguay. Sin embargo, las notables diferencias geolingüísticas que presenta el español chileno le suelen otorgar el estatus de variedad dialectal autónoma. Por su parte, Paraguay también ofrece relevantes diferencias lingüísticas, si se compara con el español de los países limítrofes (Moreno 2020, p. 118).

previamente obtenidos y fortalecer la triangulación metodológica. Cabe destacar que *Sketch Engine* ofrece dos procedimientos para vincular las ocurrencias a una región o dominio web determinado. En el primero, la búsqueda se inicia directamente en un subcorpus previamente definido por el usuario. En el segundo, la búsqueda parte del corpus general, pero luego se acota a través de la secuencia *Filter > Text types > Top-level domain*, introduciendo posteriormente el dominio geográfico de interés. Ambos métodos conducen a resultados significativos desde el punto de vista cuantitativo y cualitativo. En virtud de lo expuesto, a lo largo del estudio se adopta el término *técnica de filtrado por geolocalización* para designar de manera precisa estos procedimientos de segmentación destinados a relacionar las paremias documentadas con una variedad funcional específica del español.

#### **4.3. La localización de los refranes en los corpus**

Para la localización del refrán *A cada cerdo le llega su San Martín* y de sus variantes en los corpus digitales, se ha adoptado un método de búsqueda específicamente delineado para estos tipos de enunciados sentenciosos. Como advertencia metodológica preliminar, cabe señalar que en el contexto de los corpus electrónicos se desaconseja la introducción del refrán en su forma completa (Miccoli 2025, p. 75). Pese a que este sea el método tradicional para la búsqueda de signos lingüísticos en corpus digitales, dicha aproximación obliga al corpus a proporcionar exclusivamente aquellos registros en los que se repite la expresión tal y como ha sido digitada, resultando en la depuración de cualquier tipo de variación. Por esta razón, en el caso de la localización de paremias, se recomienda recurrir a la búsqueda mediante combinaciones de palabras consecutivas invariables y típicas del discurso repetido (Gutiérrez 2024, p. 47).

La formulación del método de búsqueda empleado parte de una observación previa sobre los componentes léxicos del refrán y su comportamiento variable. Esta observación permitió identificar el componente que presenta mayor estabilidad a lo largo de las variantes registradas: el pronombre personal *le*. Este elemento aparece sistemáticamente en posición central, justo entre las dos mitades de la estructura bimembre del refrán, lo que refuerza su valor como ancla estructural.

A partir de ello, y con el objetivo de detectar variantes léxicas que incluyan los términos *chancho* y *cerdo*, se formulan las secuencias de búsqueda “chancho le” y “cerdo le”. La elección de estas combinaciones presenta, al menos, tres ventajas metodológicas: i) se trata de secuencias atípicas del discurso libre, lo que reduce significativamente la aparición de “fragmentos incidentales (FI)” (Miccoli 2025, p. 73); ii) la secuencia se sitúa en la parte central del refrán, permitiendo captar elementos léxicos de ambas mitades de una estructura bimembre; iii) la combinación integra un elemento léxico estable (*le*) y otro variable (*cerdo* o *chancho*), lo que posibilita el estudio de la variación sin perder la referencia a la forma canónica. Hemos podido constatar que, usando esta secuencia central, la mayoría de los datos que proporcionan los corpus son realizaciones del refrán en cuestión. Sin embargo, pese a las ventajas que se alcanzan utilizando esta secuencia de búsqueda, cabe considerar también las limitaciones que conlleva este procedimiento.

Una de las principales dificultades proviene de la abundante proliferación de modificaciones que pueden derivarse de una estructura canónica que un estudio de corpus supone. Como ya ha sido señalado por Corpas (2014, p. 134), las UF, incluso cuando poseen un alto grado de fijación, presentan en los corpus digitales una frecuencia de aparición inferior respecto de sus variantes modificadas. Este fenómeno, si bien es de gran interés para los estudios sobre variación, puede entorpecer la localización de las ocurrencias pertinentes, dada la abundancia de realizaciones marginales o desautomatizadas.

Ante esta situación, se ha considerado imprescindible delimitar qué formas se han considerado válidas para el análisis y cuáles han sido excluidas, conforme a los objetivos específicos de este estudio centrado en la variación diatópica. A continuación, se enumeran los tipos de registros descartados: i) refranes desautomatizados (Mena 2003) o creativamente manipulados (Corpas 1996, pp. 233-234), como *A cada cerdo le toca su cárcel en Marruecos*, *A cada cerdo le llega su indulto* o *A todo cerdo le llega su 15M*, que alteran el sentido proverbial mediante modificaciones pragmáticas o ideológicas; ii) formas truncadas (Sevilla y Barbadillo 2021, p. 58), esto es, fragmentos que reproducen solo una parte del refrán, como *A cada cerdo...* o *A cada chancho...*; iii) formas interrumpidas por un comentario o aclaraciones del hablante, tales como *A cada chancho (puerco, cerdo, cochino, o lechón) le llega su San Martín* o *A cada cerdo le llegaba el 11 de noviembre su San Martín*; iv) refranes con inversión sintáctica, como *El San Martín que a cada chancho le llega*, que alteran el orden prototípico; v) refranes donde aparece un referente del santoral<sup>5</sup> distinto de *San Martín* o *San Juan*, como *A cada chancho le llega su San Quintín* o *A cada cerdo le llega su San Benito*. Por el contrario, sí se han incluido todas aquellas variantes<sup>6</sup> cuya variación responda a patrones de naturaleza léxica o diatópica. A modo de ejemplo, se han registrado y analizado formas en las que se sustituye el sintagma “su San Martín” por referentes polilexémáticos como: *su Nochebuena*, *su Navidad*, *su día*, *su hora*, *su sábado* (prevalentemente en Nicaragua, 36 ocurrencias, en *esamTenTen11*) y *su* 31 (prevalentemente en Cuba, con 8 ocurrencias, en el mismo corpus).

En lo que respecta a la configuración técnica, todas las búsquedas se han realizado utilizando la opción *avanzada* por *Concordancia* seleccionando *query type: phrase*. Dado que el número de ocurrencias registradas en los distintos corpus no ha superado el umbral de las mil instancias en ningún caso, la revisión y exclusión de los fragmentos colaterales ha sido llevada a cabo manualmente, salvo en los casos en que se especifica una metodología de filtrado diferente.

## 5. Discusión de los resultados

### 5.1. Estadísticas distribucionales de chancho y de su sinónimo peninsular

La localización geográfica precisa y detallada de términos relacionados por sinonimia requiere de un procedimiento metodológico riguroso, compuesto por varias etapas. La primera de ellas, ineludiblemente, consiste en la consulta de repertorios lexicográficos normativos y especializados. En este sentido, el *Diccionario de la lengua española* (DLE) de la Real Academia Española apostilla *chancho* como un americanismo y lo presenta como sinónimo de *cerdo*, sin añadir información adicional sobre su distribución geográfica o usos específicos. En cambio, el *Diccionario de americanismos* (DA) proporciona una caracterización más detallada, añadiendo etiquetas geográficas específicas, que abarcan una gran parte de los países hispanoamericanos en los que se documenta este vocablo. Asimismo, el DA incluye datos relativos no solo a las diferentes acepciones que adopta el sustantivo en cada zona, sino también a nuevas formas paremiológicas, algunas de las cuales han sido ya referidas en el apartado 2. La adscripción del sustantivo *chancho* al

<sup>5</sup> Hemos decidido excluir estos fragmentos, ya que representan apariciones esporádicas en el corpus. En el mejor de los casos, solo presentan dos ocurrencias entre las miles consultadas, producto de deturpaciones (Mena 2003; Jarilla 2024) o alteraciones de la forma original. Por ello, resulta imposible establecer un patrón de frecuencia/distribución.

<sup>6</sup> Nos referimos a variantes estructurales (Corpas 1996, p. 28) e hipotéticas variantes léxicas producto de diatopía, diastratía o diafasía.

español americano se ve además respaldada por su inclusión en repertorios dialectales como el *Diccionario del español de México* (DEM) o el *Diccionario del habla de los argentinos* (DIHA). Pese a que tanto cerdo como chancho pueden considerarse formas panhispánicas, el presente estudio se enfoca en el contraste entre dos macrovariedades del español: la variedad peninsular y la variedad rioplatense.

Las observaciones léxicas preliminares permiten anticipar una notable concentración del uso de chancho en el español rioplatense. Esta previsión se confirma mediante la consulta de fuentes lexicográficas nacionales como el DIHA y el *Diccionario del español de la Argentina* (DEArg), en los que el término figura como un argentinismo consolidado. Su vitalidad como regionalismo rioplatense ha sido objeto de análisis por parte de varios estudios. Entre ellos destacan los trabajos de Sciutto (2006, p. 143; 2017, pp. 422-423), centrados en los somatismos fraseológicos en el español de Argentina.

Estos patrones de distribución pueden apreciarse igualmente a través de un enfoque de corpus. Por razones de claridad expositiva, se presenta a continuación la Tabla 1, en la que se detallan los porcentajes de frecuencia absoluta de los sustantivos *cerdo* y *chancho*<sup>7</sup> en distintas áreas dialectales del español, según dos corpus digitales de referencia de la RAE: el CORPES XXI y el CREA. Las zonas dialectales consideradas corresponden a las categorías establecidas por los propios corpus.

Chancho						
	Rioplatense	Europea	Caribeña	Chilena	Andina	Centroam <sup>8</sup>
CREA	41,1 %	5,73 %	1,04 %	19,3 %	30,2 %	2,6 %
CORPES XXI	33,1 %	1,33 %	3,09 %	23,6 %	20,6 %	18,1 %
Cerdo						
CREA	8,14 %	57,4 %	2,34 %	1,79 %	6,22 %	19,2 %
CORPES XXI	9,73 %	36,3 %	12,5 %	4,18 %	7,82 %	21,5 %

Tabla 1  
Porcentajes de difusión geográfica de los sinónimos *cerdo* y *chancho*.  
Fuentes: CORPES XXI y CREA. Elaboración propia.

Gracias a las funcionalidades estadísticas ofrecidas por los corpus de la RAE, los datos recogidos proporcionan una visión comprehensiva de la difusión de ambos términos, tanto a nivel intraterritorial como interterritorial. Si se atiende a los porcentajes correspondientes a las variedades en análisis, se observa que ambas formas coexisten en los mismos espacios, aunque se observan ciertas asimetrías. Dicha convivencia es desequilibrada. En particular, *chancho* presenta su frecuencia mínima en la variedad peninsular y alcanza su máximo en español rioplatense. La distribución de *cerdo*, por el contrario, refleja una tendencia opuesta: su uso se concentra especialmente en el ámbito peninsular, mientras que su frecuencia disminuye significativamente en el área rioplatense, sin llegar a ser marginal. Estas observaciones terminológicas preliminares establecen hipótesis fundadas sobre la dispersión de las variantes léxicas del refrán analizado. Dado que los refranes constituyen estructuras de alta fijación formal, pero sujetas a adaptación contextual, cabe esperar que las unidades léxicas que los componen reproduzcan los patrones distribucionales ya observados en los sustantivos *cerdo* y *chancho*. Esta continuidad en la distribución léxica constituye el punto de partida para el análisis de la variación geoparémica que se desarrollará en los apartados siguientes.

<sup>7</sup> Configuración de búsqueda: *Elementos Gramaticales > Lema; Resultado > Estadísticas*.

<sup>8</sup> México y Centroamérica.

## 5.2. La variante léxica y su distribución en la hispanoesfera

La consulta de varios corpus de distinta índole permite acceder a una representación más amplia y matizada del comportamiento léxico de una paremia determinada. En cuanto a la distribución de las dos variantes léxicas del refrán objeto de estudio, se presentan a continuación los resultados cuantitativos obtenidos sobre su frecuencia y distribución en el ámbito de la hispanofonía.

En los corpus de referencia de la RAE —Corpus del Español del Siglo XXI (CORPES XXI) y Corpus de Referencia del Español Actual (CREA)— la información relativa al refrán es limitada. Solo se han documentado dos ocurrencias con el componente *chancho*: una en CORPES XXI, adscrita a Argentina, y otra en el CREA, atribuida a Uruguay. Esta escasez de datos impide establecer comparaciones concluyentes, aunque su consulta ha sido clave para el análisis de la distribución léxica presentado en el apartado 5.1.

El panorama cambia de forma notable al incorporar el Corpus del Español Web/Dialectos (CdEweb), en el que se evidencia una frecuencia considerablemente mayor de la paremia con el actante *chancho*. Se han identificado 37 fragmentos que contienen la secuencia “*chancho le*”, de los cuales 9 fueron descartados mediante revisión manual por no corresponder el refrán en estudio. Las restantes 28 ocurrencias, adscribibles a diversos países hispanohablantes, se representan en el Gráfico 1:

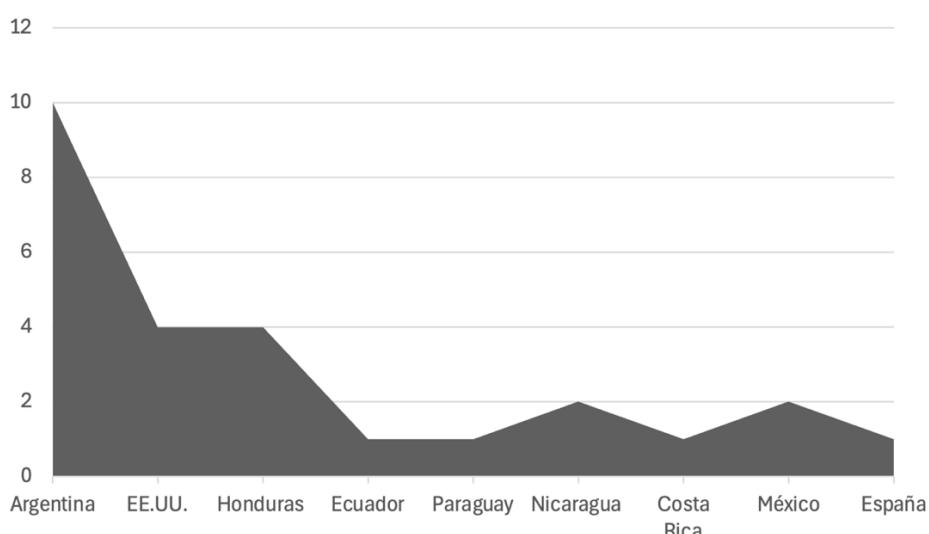


Gráfico 1  
Distribución de la variante léxica con *chancho* según CdEweb.  
Fuente: CdEweb. Elaboración propia.

De acuerdo con los datos recogidos, el refrán alcanza su mayor frecuencia en Argentina, mientras que registra su mínimo en países como Paraguay y España. Este primer gráfico, respeta los datos mencionados hasta el momento y es una anticipación coherente de los resultados sucesivos. No obstante, la cantidad de datos sigue siendo escasa para un análisis empírico sólido, lo que hace necesaria la búsqueda en otros corpus.

La incorporación de corpus web, pese a la complejidad que conlleva el tratamiento de sus abundantes registros, permite establecer patrones de distribución de las dos variantes léxicas más exactos. En los tres corpus de la familia *TenTen* se han utilizado las secuencias “cerdo le” y “chancho le”, las cuales han demostrado ser altamente eficaces en la localización del refrán, reduciendo drásticamente los fragmentos incidentales (FI). A modo de ejemplo, en *esTenTen23*, la búsqueda con la secuencia “cerdo le” arrojó un total de 1249

ocurrencias, de las cuales aproximadamente un 31,92 % fueron excluidas tras filtrado manual. De forma similar, en *esTenTen18*, la secuencia “chancho le” produjo 170 resultados, con una tasa de exclusión del 27,65 %. En ambos casos, los datos conservados superaron el 68 % de las instancias recuperadas, lo que confirma la eficacia del método delineado en el apartado 4.3. En lo que se refiere a la distribución geográfica de los enunciados sentenciosos y los actantes relativos, se observa el mismo patrón de distribución supuesto hasta el momento. Los datos recogidos en los tres corpus de la familia *TenTen* han sido fundamentales para la elaboración del Mapa 1, que muestra la difusión en América de la variante léxica que alberga *chancho*.



Mapa 1  
Distribución de la variante léxica americana.  
Fuentes: *esTenTen23*, *esTenTen18* y *esamTenTen11*.  
Elaboración propia.

El mapa se ha elaborado sumando las ocurrencias obtenidas en régimen de frecuencia absoluta. Dicho proceso se ha reiterado por cada país hispanoamericano<sup>9</sup> en los tres corpus web consultados en *Sketch Engine*, filtrando por cada uno el dominio web de pertenencia. El gradiente cromático indica la densidad de frecuencia. En total, se registraron unas 350 ocurrencias: más de 200 en dominios web argentinos, alrededor de 100 en Nicaragua, y ausencia total en países como Guatemala o Venezuela. De este modo, se observa una dispersión amplia en la hispanoesfera americana, con un claro epicentro en Argentina, donde el refrán no solo alcanza su mayor frecuencia, sino que presenta además mayor variabilidad estructural.

Al comparar esta distribución con la de la variante europea, se advierte una diferencia significativa. En España, las paremias que albergan *cerdo* reflejan una frecuencia superior a 500 ocurrencias (294 *esTenTen23* y 210 *esTenTen18*). Por el contrario, en el

<sup>9</sup> Con excepción de Puerto Rico, cuyo dominio no es localizable en los tres corpus.

ámbito rioplatense, la forma canónica se encuentra apenas representada: solo seis ocurrencias válidas (tres en cada uno de los corpus mencionados), todas ellas localizadas en dominios web argentinos. Esta asimetría se mantiene si se invierten los términos de comparación. Los refranes con sustituto léxico *chancho* circulan libremente en América, pero apenas tienen presencia en España, donde solo se documenta un caso aislado. Este fragmento, al ser examinado en detalle, revela que su autoría corresponde a una escritora uruguaya<sup>10</sup>:

[...] la piel de gallina, poner ojos de carnero, ver en alguien a una dulce palomita, esperar que las vacas vuelen o a que cada chancho le llegue su San Martín? Más bien escuchando la canción infinita con la piel que habito y los ojos atentos de una lechuza [...]. (esTenTen23, 47140017)

Ahora bien, a nivel cuantitativo, es posible trazar un patrón de difusión comparando los parámetros de frecuencia absoluta en los dominios de referencia. Los gráficos 2 y 3 muestran la distribución de las variantes léxicas en relación con los dominios web de los tres países analizados: *.es*, *.ar* y *.uy*.

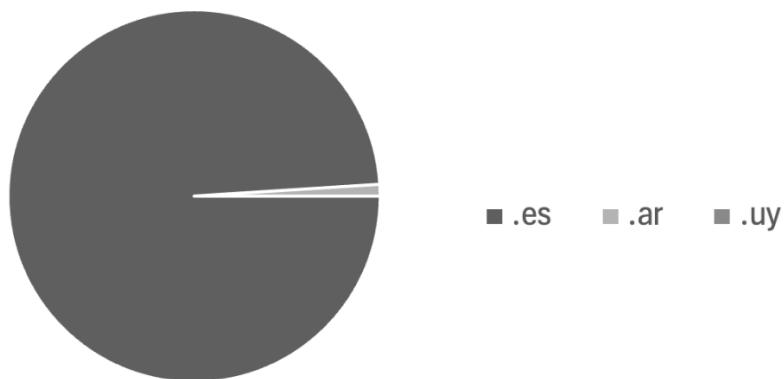


Gráfico 2  
Difusión del refrán con actante Cердо por dominios web.  
*Fuente:* esTenTen23. Elaboración propia.

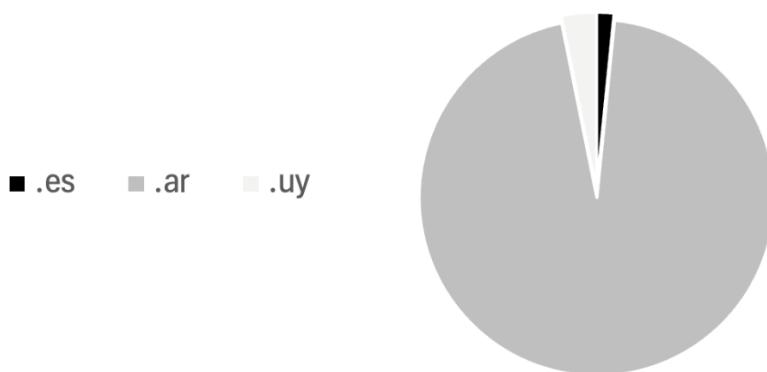


Gráfico 3  
Difusión del refrán con actante Chancho por dominios web.  
*Fuente:* esTenTen23. Elaboración propia.

<sup>10</sup> Es frecuente que el corpus recoja obras escritas y publicadas en un determinado país. Es de igual modo frecuente que estas obras sean escritas por hablantes con nacionalidad distinta a la del país donde se publican. Por ello, se requiere una investigación particular, que estos recursos normalmente facilitan (Rojo 2021, p. 166).

Ambos gráficos circulares proporcionan una distribución casi dicotómica entre las dos variantes léxicas en las dos variedades comparadas. Un dato que merece la pena subrayar es la escasa representación del dominio *.uy*, con menos de diez registros en total para ambas formas. En cambio, los porcentajes registrados para Argentina encuentran una posible explicación en el testimonio de un hablante recogido en *esTenTen23*:

[...] El refrán en sí es español, y carece de sentido en Sudamérica donde se mata al cerdo para hacer facturas o embutidos en mayo o junio, y su forma en Argentina es siempre "a cada chancho", nunca "puerco" y sólo a veces (muy pocas) "cerdo" [...]. (*esTenTen23*, 59046030)

Más allá de la frecuencia, el análisis de corpus permite también evaluar la variabilidad interna del refrán. Desde el punto de vista estadístico, Nicaragua sigue a Argentina en número de ocurrencias del refrán con *chancho* como actante. En los registros nicaragüenses, se observa además una sustitución significativa de la expresión *su San Martín* por *su sábado*. Este fenómeno confirma que la variación no se limita al sustantivo, sino que puede afectar a otros componentes del enunciado. Se han registrado variantes que incorporan sinónimos actantes de *chancho* como *gocho*, *gorrino*, *cochino*, *marrano* y *puerco*, entre otros. La sustitución motivada por sinonimia parece afectar a todos los elementos del refrán, a excepción del pronombre *le* y la preposición inicial *a*, que se mantienen estables en todos los casos observados. Entre los elementos particularmente propensos a la variación se encuentran el verbo principal (*llegar* o *tocar*) y el determinante (*cada* o *todo*). El Gráfico 4 resume la frecuencia de estos cambios en los casos en que el actante principal es *chancho*.

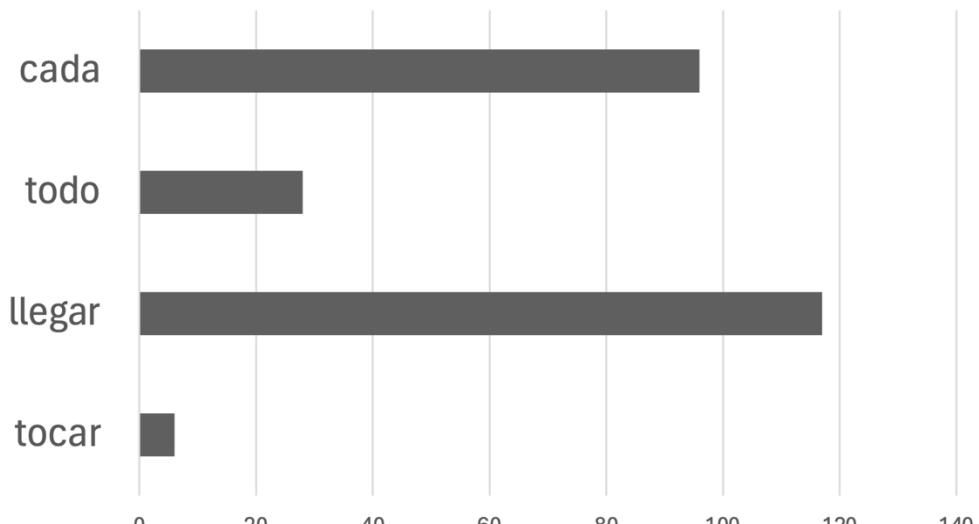


Gráfico 4  
Variabilidad de lexía en los refranes con actante *chancho*.  
Fuente: *esTenTen23*. Elaboración propia.

En definitiva, aunque los componentes léxicos tradicionales siguen manteniendo su peso en la modulación del refrán, los datos permiten constatar una plasticidad significativa. Esta se manifiesta en procesos de sustitución léxica motivados por sinonimia, que afectan tanto a los núcleos léxicos como a los constituyentes funcionales del enunciado paremiológico.

## 6. Conclusiones

El presente estudio ha tenido como objetivo principal demostrar la aplicabilidad de los corpus lingüísticos al análisis de la variación paremiológica, en particular, a la observación de la variación léxica motivada por factores diatópicos. Esta dimensión, propia de la expansión del español en áreas, a menudo alejadas entre sí, se ha abordado mediante el estudio de un refrán muy popular, ampliamente difundido a nivel panhispánico: *A cada cerdo le llega su San Martín*, cuya frecuencia de uso permitió pronosticar su extensa gama de cambios. Tomando como punto de partida los valiosos datos previamente recopilados por otros estudiosos, se ha seleccionado como objeto de análisis la variante léxica rioplatense: *A cada chancho le llega su San Martín*. La recopilación y sistematización de ejemplos reales ha permitido identificar los componentes estructurales del refrán que presentan mayor tendencia a la variación, así como aquellos que se mantienen estables. Esta observación ha sido clave para formular una estrategia de búsqueda optimizada para la localización de paremias en corpus digitales.

Entre las principales ventajas de la secuencia de búsqueda propuesta —basada en combinaciones léxicas de carácter mixto, con un elemento estable (*le*) y otro variable (*chancho / cerdo*)— cabe destacar dos aspectos fundamentales: por un lado, su eficacia para recuperar fragmentos en los que aparece el refrán en sus distintas variantes; por otro, su capacidad de depurar significativamente los resultados espurios o incidentales, cuyo filtrado se ha realizado mediante revisión manual. A efectos comparativos, se han aplicado las secuencias “cerdo le” y “chancho le” en múltiples corpus digitales, utilizando técnicas de filtrado por geolocalización que permiten restringir los resultados a dominios web específicos. Esta metodología ha facilitado el establecimiento de un patrón de distribución geográfica de la variante americana, visualizado en el Mapa 1, y ha permitido, además, realizar un análisis contrastivo entre dos macrovariedades del español: la europea y la rioplatense. Los resultados obtenidos en esta comparación validan las hipótesis de partida. Las dos variantes léxicas del refrán reproducen patrones de distribución similares a los de sus respectivos actantes, *cerdo* y *chancho* (véase Tabla 1), aunque de forma más acentuada. La forma canónica con *cerdo* muestra su hegemonía en España, mientras que en el ámbito rioplatense aparece de forma esporádica y marginal. De manera inversa, la variante con *chancho* se consolida como forma de referencia en América —especialmente en Argentina— y carece de vitalidad en el español peninsular, donde su aparición es prácticamente inexistente. A nivel panhispánico, la variante *A cada chancho le llega su San Martín* se perfila como la forma preferente del refrán en América. Su uso se documenta en la mayoría de los países hispanohablantes del continente, aunque en ciertos territorios coexiste con otras variantes que presentan modificaciones léxicas adicionales. Esta tendencia a la diversificación afecta a distintos componentes del enunciado paremiológico y no se limita al actante. En cambio, elementos como la preposición inicial *a* y el pronombre *le* se mantienen como elementos estructurales constantes.

Particularmente destacable es la situación observada en Argentina, donde la variante con *chancho* ha adquirido un estatus de estandarización funcional: es la forma más utilizada, la que presenta mayor número de ocurrencias y la que muestra una mayor capacidad para generar nuevas variantes (véase Gráfico 6).

De todo lo anteriormente expuesto, se infiere que la forma con “chancho” constituye un claro indicador de autonomización regional, con un arraigo especialmente notable en Argentina. Paralelamente, estas variantes evidencian una circulación asimétrica: los refranes peninsulares migran con mayor facilidad hacia América y no a la inversa, lo cual subraya la coexistencia funcional parcial de formas canónica y locales. Finalmente, el refrán se consolida como unidad heurística privilegiada para la detección y representación de la

variación dialectal, resultado de la naturaleza polisistémica y no unificada del refranero panhispánico.

Desde un punto de vista metodológico, esta aproximación al estudio de la variación geoparemiológica presenta múltiples ventajas. La principal reside en su capacidad para visibilizar la diversidad funcional y territorial de las paremias, lo que permite acceder a una representación más fiel de su dimensión cultural, histórica e identitaria. Asimismo, los resultados obtenidos trascienden el plano descriptivo y aportan herramientas útiles para otros ámbitos del conocimiento lingüístico. De manera puntual, los hallazgos de esta investigación parémico-contrastiva, resultan de especial interés para la didáctica de lenguas y la traducción. En efecto, el conocimiento de la distribución real de las variantes paremiológicas en las distintas variedades del español permite adoptar decisiones informadas y culturalmente pertinentes al seleccionar equivalentes funcionales en contextos educativos o traductológicos. En suma, se reivindica aquí el valor del análisis corporalingüístico no solo como herramienta descriptiva, sino como un instrumento clave para alcanzar una mayor representatividad y equidad en el tratamiento de la fraseología panhispánica.

**Nota biográfica:** Vito Antonio Miccoli es doctorando en el programa de *Lingüística Teórica y Aplicada* de la Universidad Complutense de Madrid. Su investigación se centra en la paremiología desde una perspectiva contrastiva y aplicada, con especial atención a los estudios basados en corpus y a las implicaciones traductológicas de las unidades fraseológicas. Es coautor del *Refranero multilingüe* (lengua italiana, Centro Virtual Cervantes) y miembro-colaborador del grupo de investigación UCM 930235 *Fraseología y Paremiología* (PAREFRAS). Asimismo, es *Cultore della materia* en el SSD SPAN-01 por la Universidad del Salento y por la Universidad de Foggia. Líneas de investigación: paremiología, lingüística de corpus y traducción.

**Dirección del autor:** [vmiccoli@ucm.es](mailto:vmiccoli@ucm.es)

## Referencias bibliográficas

- Almela Pérez R. and Sevilla Muñoz J. 2000, *Paremiología contrastiva: propuesta de análisis lingüístico*, en "Revista de Investigación Lingüística" III, pp. 7-47.
- Alvarado Ortega B. 2008, *Sobre el concepto de variación fraseológica*, en "ELUA" 22, pp. 9-21.
- Arora S.L. 1997, *El refranero español en Los Ángeles (California)*, en "Paremia" 6, pp. 67-76.
- Corpas Pastor G. and Mena Martínez F. 2003, *Aproximación a la variabilidad fraseológica de las lenguas alemana, inglesa y española*, en "ELUA" 17, pp. 181-201.
- Corpas Pastor G. 1996, *Manual de fraseología española*, Gredos.
- Corpas Pastor G. 1998, *El uso de paremias en un corpus del español peninsular actual*, en Wotjak G. (ed.), *Estudios de fraseología y fraseografía del español actual*, Iberoamericana Vervuert, pp. 365-390.
- Corpas Pastor G. 2014, *El fraseólogo internauta: cómo pasarlo pipa en la red*, en Durante V. (ed.), *Fraseología y paremiología: enfoques y aplicaciones*, Instituto Cervantes, pp. 133-152.
- Cuadrado Rey A. 2020, *Neologismos fraseológicos en el español hablado en Argentina: los procedimientos de creación*, en MogorronHuerta P. and CuadradoRey A. (ed.), *ELUA: Fraseología y variaciones (socio)lingüísticas y diatópicas*, Anexo VII, pp. 129-143.
- García-Page M. 1999, *Variantes morfológicas y unidades fraseológicas*, en "Paremia" 8, pp. 225-230.
- García-Page M. 2008, *Introducción a la fraseología española. Estudio de locuciones*, Anthropos.
- Gargallo Gil J.E. and Álvarez Pérez X.A. 2014, *El proyecto Paremio-Rom. Refranes meteorológicos y geoparemiología romance*, en "Estudis Romànics" 36, pp. 313-324.
- Grzybek P. 2012, *Proverb Variants and Variations: A New Old Problem?*, en Laukalangas O. and Soares R. (ed.), *Proceedings of the Fifth Interdisciplinary Colloquium on Proverbs*, IAP, pp. 136-152.
- Guerra Garrido R. 1997, *Cachicamo llamándole a morrocoy conchudo. Una aproximación a la Paremiología hispanoamericana*, en "Paremia" 6, pp. 297-300.
- Gutiérrez Rubio E. 2024, *Estudio de corpus de los presentadores paremiológicos en español*, en "Paremia" 34, pp. 45-54.
- Jarilla Bravo S.M. 2024, *Estudio de variantes y vigencia de los refranes recogidos por Fernán Caballero*, en "Romanica Olomucensis" 36 [2], pp. 275-288.
- Kilgarriff A. and Renau I. 2013, *EsTenTen, a Vast Web Corpus of Peninsular and American Spanish*, en "Procedia - Social and Behavioral Sciences" 95, pp. 12-19.
- Klein J.R. 2006, *Problemas relacionados coa determinación da forma 'canónica' nunha base de datos de refráns franceses (DicAuPro)*, en "Cadernos de fraseoloxía galega" 8, pp. 147-163.
- Klein J.R. 2017, *Portrait d'un dictionnaire automatique et philologique des proverbes français*, en "Paremia" 26, pp. 35-40.
- Koike K. 2003, *Las unidades fraseológicas del español: su distribución geográfica y variantes diatópicas*, en "Epos" 19, pp. 47-65.
- Liendo P.J. et al. 2022, *Introducción a la Lingüística de Corpus: sus usos y aplicaciones* (1 ed.), General Roca.
- López Meirama B. 2020, *Variación diatópica y análisis de corpus: algunos casos en la fraseología del español*, en Mogorron Huerta P. and Cuadrado Rey A. (ed.), *ELUA: Fraseología y variaciones (socio)lingüísticas y diatópicas*, Anexo VII, pp. 145-159.
- Mántica C. 1999, *Refranero nicaragüense* (2 ed.), Editorial Hispamer.
- Mellado Blanco C. 2004, *Fraseologismos somáticos del alemán*, Peter Lang.
- Mellado Blanco C. 2020, *(No) me importa un comino y sus variantes diatópicas. Estudio de corpus desde la Gramática de Construcciones*, en Mogorron Huerta P. and Cuadrado Rey A. (ed.), *ELUA: Fraseología y variaciones (socio)lingüísticas y diatópicas*, Anexo VII, pp. 89-111.
- Mena Martínez F. 2003, *En torno al concepto de desautomatización fraseológica: aspectos básicos*, en "Tonos Digital" 5.
- Messina Fajardo L.A. 2010, *Como avispa, que cigarrón atorar: paremiología venezolana y "comparancias"*, en Pamies Bertrán A., Luque Durán J. and Fernández Martín P. (ed.), *Paremiología y herencia cultural*, Granada Lingüística, pp. 169-179.
- Miccoli V.A. 2025, *La localización de refranes en corpus lingüísticos digitales: métodos y aplicaciones*, en "Paremia" 35, pp. 69-80.
- Mogorron Huerta P. 2020, *Locuciones verbales, variación fraseológica y diatopía*, en MogorronHuerta P. and Cuadrado Rey A. (ed.), *ELUA: Fraseología y variaciones (socio)lingüísticas y diatópicas*, Anexo VII, pp. 11-31.
- Montoro DelArco E.T. 2005, *Hacia una sistematización de la variabilidad fraseológica*, en Pastor Milán M.A. (ed.), *Estudios lingüísticos en recuerdo del profesor Juan Martínez Marín*, Universidad de Granada, pp. 125-152.
- Montoro Del Arco E.T. 2006, *Teoría fraseológica de las "locuciones particulares": las locuciones prepositivas, conjuntivas y marcadoras en español*, Peter Lang.

- Moreno Fernández F. 2020, *Variedades de la lengua española*, Routledge.
- Navarro-Brotóns L. 2020, *Aproximación mediante encuestas a las paremias usuales comunes entre España e Hispanoamérica aplicables a las clases de traducción general*, en MogorronHuerta P. and Cuadrado Rey A. (ed.), *ELUA: Fraseología y variaciones (socio)lingüísticas y diatópicas*, Anexo VII, pp. 113-128.
- Novodvorski A. 2017, *A equivalência tradutória de argentinismos: um estudo contrastivo léxico-fraseológico em corpus jornalístico de matérias políticas*, en "Domínios De Lingua@gem" 11 [5], pp. 1628-1648.
- Pamies A. 2017, *Frageología y variación diatópica en español*, en "Verba Hispánica" XXV, pp. 55-81.
- Pamies A. and Pazos J.M. 2023, *Frasemas combinados "a pedir de boca"*, en "Romanica Olomucensis" 35 [1], pp. 139-156.
- Penadés Martínez I. 1996, *Las expresiones fijas desde los conceptos centro y periferia de los lingüistas praguenses*, en CasasGómez M. (ed.), pp. 91-134.
- Penadés Martínez I. 2022, *La variación en las unidades fraseológicas*, en "Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas" 16 [32], pp. 21-39.
- Pérez Pérez H. 2007, *La fortuna del refranero mexicano*, en "Paremia" 16, pp. 45-50.
- Pérez-Paredes P. 2020, *Corpus Linguistics for Education: A Guide for Research*, Routledge.
- Rojo G. 2014, *Hispanic Corpus Linguistics*, en Lacorte M. (ed.), *The Routledge Handbook of Hispanic Applied Linguistics*, Routledge, pp. 371-387.
- Rojo G. 2021, *Introducción a la lingüística de corpus en español*, Routledge.
- Sciutto V. 2006, *Elementos somáticos en la fraseología del español de Argentina*, Aracne.
- Sciutto V. 2017, *Locuciones denominativas en un corpus de la variedad argentina del español*, en "Lingue e Linguaggi" 23, pp. 291-305.
- Sciutto V. and Asensio Ferreiro M.D. 2022, *Frageología y Paremiología: diatopía, variación, fraseodidáctica y paremioididáctica*, en "Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas" 16 [32], pp. 10-19.
- Sevilla Muñoz J. and Barbadillo M.T. 2021, *El mínimo paremiológico español*, Biblioteca fraseológica y paremiológica, serie "El mínimo paremiológico" 2, Centro Virtual Cervantes (Instituto Cervantes).
- Sevilla Muñoz J. 2000, *Shirley L. Arora y la paremiología hispanoamericana*, en "Paremia" 9, pp. 7-14.
- Sevilla Muñoz J. 2012, *Aplicación de las TIC a la paremiología y la paremiografía: el Refranero multilingüe*, en González Rey M.I. (ed.), *Unidades fraseológicas y TIC*, Instituto Cervantes, pp. 51-62.
- Sevilla Muñoz J. 2020, *La variación en las paremias del español de España y de México*, en Mogorron Huerta P. and Cuadrado Rey A. (ed.), *ELUA: Fraseología y variaciones (socio)lingüísticas y diatópicas*, Anexo VII, pp. 73-88.
- Sevilla Muñoz J. and Crida Álvarez C.A. 2013, *Las paremias y su clasificación*, en "Paremia" 22, pp. 105-114.
- Sevilla Muñoz J. and Crida Álvarez C.A. 2017, *Taxonomía de las paremias en lengua española*, en "Phrasis" 1, pp. 117-129.
- Sevilla Muñoz J. and Ugarte García M.ª 2022, *La variación en los refranes y frases proverbiales en español: un caso práctico*, en "Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas" 16 [32], pp. 92-108.
- Sevilla Muñoz J. and Zurdo Ruiz-Ayúcar M.I.T. 2009, *Refranero Multilingüe*, Instituto Cervantes (Centro Virtual Cervantes).
- Sinner C. and Tabares Plasencia E. 2016, *El problema de las variantes fraseológicas desde la perspectiva de la lingüística de variedades*, en "Revista de Lingüística Teórica y Aplicada" 54 [2], pp. 13-41.
- SotoPosada G. 2000, *Aculturación e identidad del hombre latinoamericano: una aproximación paremiológica*, en "Paremia" 9, pp. 43-48.
- TorresTorres A. 2013, *Paremiología española e hispanoamericana. Los americanismos parémicos*, en "Dialectología" 10, pp. 87-105.
- VázquezRozas V. and Blanco M. 2022, *Corpus y enseñanza del español*, en Parodi G., Lewis C.H. and Pascual C. (ed.), *Lingüística de corpus en español / The Routledge Handbook of Spanish Corpus Linguistics*, Routledge, pp. 342-356.
- Zurdo Ruiz-Ayúcar M.I.T. 2002, *En torno a la modificación por expansión en fraseologismos de núcleo verbal alemanes y españoles*, en "Estudios Filológicos Alemanes" 1, pp. 131-149.

## Diccionarios

DA = Asociación de Academias de la Lengua Española 2010, *Diccionario de americanismos*, Santillana.  
 DEArg = Haensch, G., Chuchuy C. y Werner, R. 2000, *Diccionario del español de Argentina: español de Argentina - español de España*, Gredos.  
 DEM = Lara L. F. (dir.) 2010, *Diccionario del español de México*, El Colegio de México.  
 DIHA = Academia Argentina de Letras 2003, *Diccionario del habla de los argentinos* (2. ed.), Espasa Calpe.  
 DLE = Real Academia Española 2014, *Diccionario de la lengua española* (23. ed.)

## Corpus empleados

CdEweb: *Corpus del Español (Web/Dialectos)*. Dir. Mark Davies, <[www.corpusdelespanol.org/web-dial/](http://www.corpusdelespanol.org/web-dial/)>.  
 CORPES XXI: Real Academia Española. *Corpus del Español del Siglo XXI*,  
 <<http://rae.es/recursos/bancode-datos/corpes-xxi>>.  
 CREA: Real Academia Española. *Corpus de Referencia del Español Actual*, <<http://rae.es/recursos/banco-de-datos/crea>>.  
 esamTenTen11: The Sketch Engine. <<http://www.sketchengine.co.uk>>.  
 esTenTen18: The Sketch Engine. <<http://www.sketchengine.co.uk>>.  
 esTenTen23: The Sketch Engine. <<http://www.sketchengine.co.uk>>.