



PONTIFICIA UNIVERSIDAD  
**CATOLICA**  
**DE VALPARAISO**

Instituto de Literatura y Ciencias del Lenguaje

Facultad de Filosofía y Educación

Pontificia Universidad Católica de Valparaíso

**Detección automática**  
**de verbos del español en un corpus general y su**  
**aplicación para la detección de neología léxica**

Tesis para optar al grado académico de

Licenciada en Lengua y Literatura Hispánica

**Tesista: Ana Castro Páez**

**Directores de tesis: Irene Renau – Rogelio Nazar**

Viña del Mar, enero de 2019

## Índice

1. Introducción .....	4
2. Marco teórico .....	8
2.1. El estudio del léxico .....	8
2.2. Neología y neologismos .....	13
2.3. Neología léxica.....	16
2.4. Detección automática de neologismos .....	23
2.5. Verbos: morfología, modelos de conjugación y neología verbal .....	25
2.5.1. Participios.....	32
3. Marco metodológico .....	33
3.1. Tipo de investigación .....	33
3.2. Preguntas de investigación .....	33
3.3. Objetivos de investigación .....	34
3.3.1. Objetivo general .....	34
3.3.2. Objetivos específicos.....	34
3.4. Materiales y métodos .....	34
3.4.1. Materiales .....	34
3.4.2. Fases metodológicas de la investigación.....	35
3.4.2.1. Conformación de un listado de las desinencias verbales del español.....	35
3.4.2.2. Conformación de un listado de formas.....	37
3.4.2.3. Procesamiento del listado de formas .....	37
3.4.2.4. Comparación con fuente lexicográfica.....	40
3.4.2.5. Realización de prueba de confiabilidad.....	41
3.4.2.6. Análisis de verbos no encontrados en el DLE.....	41
3.5. Criterios de análisis .....	43
4. Resultados y discusión .....	46
4.1. Resultados generales .....	46
4.1.1. Prueba de confiabilidad.....	49
4.2. Verbos detectados en el corpus que se encuentran registrados en el DLE.....	49
4.3. Verbos detectados en el corpus que no se encuentran registrados en el DLE.....	50
4.3.1. Neologismos.....	50
4.3.1.1. Derivación .....	51
4.3.1.1.1. Prefijación (FPRE) .....	51
4.3.1.1.2. Sufijación (FSUF) .....	54

4.3.1.1.3. Parasíntesis (FPRSU) .....	57
4.3.1.2. Préstamos adaptados (AM) .....	58
4.3.1.3. Otros mecanismos de formación de palabras .....	59
4.3.2. Errores ortográficos y fallas del método .....	61
4.4. Participios.....	62
5. Conclusiones .....	63
Bibliografía .....	66

## 1. Introducción

El presente trabajo consiste en una propuesta de monitoreo del estado actual del léxico utilizado en castellano, particularmente enfocado en los verbos, mediante análisis de corpus. Su objetivo fundamental es identificar, por medio del procesamiento automático de datos, los verbos del castellano efectivamente utilizados por los hablantes, de manera que sea posible detectar aquellos que aún se encuentran vigentes en el uso a pesar del paso del tiempo, así como casos de neologismos. Junto con ello, se ofrece una propuesta metodológica para la detección de dichos verbos en un corpus general de lengua castellana, con especial foco en verbos neológicos.

El interés de este estudio radica en que se propone extraer por procedimientos automáticos las unidades estudiadas a partir de datos de corpus, datos que por tanto corresponden al uso real de dichos verbos, a diferencia de lo que ocurre comúnmente en los diccionarios, en los que persisten registros de verbos ya anticuados o en franco desuso. En este sentido, resulta interesante determinar cuáles de los verbos que se encuentran registrados en los diccionarios, particularmente en el *Diccionario de la lengua española*, DLE (RAE, 2014), efectivamente son utilizados por los hablantes. Junto con ello, se requieren datos empíricos obtenidos de un análisis de corpus que releve dicha información por medio de un procedimiento automático que atienda no solo a los verbos ya establecidos en el sistema de la lengua, sino también a aquellos que se van incorporando producto de nuevas necesidades comunicativas y expresivas y que por su reciente uso no son abordados por los diccionarios.

Para efectos de la presente investigación, se utiliza un listado de formas (es decir, de segmentos separados por un espacio en el corpus, lo que las identifica como posibles palabras) extraído del corpus EsTenTen (Kilgarriff y Renau, 2013), el cual es procesado automáticamente a través de un script en código Perl, para detectar verbos neológicos y no neológicos a partir de su estructura morfológica. Esta estructura resulta particularmente útil para dicha detección en tanto que uno de sus componentes, las desinencias, representa un conjunto finito y estable que aplica a los verbos en general, particularmente a los verbos regulares, que constituye el grupo más abundante y productivo del español, en desmedro de las conjugaciones irregulares (RAE, 2009). Por el contrario, las raíces corresponden al elemento inestable en la producción de nuevas palabras, que puede variar en relación con la

creatividad de los hablantes o las necesidades comunicativas que motivan la incorporación de una nueva forma verbal. Posteriormente, los verbos detectados en el EsTenTen son comparados con aquellos registrados en el DLE para realizar la medición cuantitativa ya mencionada.

Como producto de la investigación, se genera una base de datos léxica representativa de los verbos existentes en el español actual por medio de un proceso de detección automatizada. A pesar de los avances en el procesamiento del lenguaje natural, la detección automática de neologismos resulta ser particularmente problemática, debido a que incluso los etiquetadores más utilizados, como TreeTagger (Schmid, 1994), no son capaces de detectar verbos neológicos. Por ejemplo, al ingresar a TreeTagger<sup>1</sup>, la oración contenida en (1) extraída del corpus EsTenTen (Kilgarriff y Renau, 2013) que contiene el verbo *aflamencar* (uno de los neologismos detectados en este estudio) se obtiene el resultado que muestra la Figura 1.

(1) *su privilegiada garganta se convirtió en un crisol en el que se aflamencaron estilos de origen popular andaluz.*

---

<sup>1</sup> Se utiliza la interfaz Porcus, de R. Nazar, D. Sánchez y N. Acosta (<http://www.tecling.com/porcus/index.php>, [Consulta: 4/1/2019]).

Su	PPO	suyo
privilegiada	ULadj	privilegiar
garganta	NC	garganta
se	SE	se
convirtió	ULfin	convertir
en	PREP	en
un	ART	un
crisol	NC	crisol
en	PREP	en
el	ART	el
que	CQUE	que
se	SE	se
aflamencaron	ULfin	<b>UNKNOWN</b>
estilos	NC	estilo
de	PREP	de
origen	NC	origen
popular	ADJ	popular
andaluz	ADJ	andaluz
.	FS	.

**Figura 1.** Resultado de TreeTagger al ingresar oración con el verbo neológico *aflamencar*.

En esta figura, se puede observar que el analizador morfosintáctico reconoce la unidad neológica como un verbo (VLfin), mas no es capaz de reconstruir su lema (UNKNOWN), al no encontrarse este integrado en su diccionario y por enfocarse únicamente en lo que concierne al aspecto sintáctico, en desmedro del aspecto semántico. Por ende, resulta necesaria la propuesta de un método que prescinda de un diccionario previo y, en cambio, realice la detección automática a partir de la estructura morfológica que es común a todos los verbos independiente del carácter neológico.

Esta tesis se organiza en los siguientes apartados: primero, el Marco teórico (2), donde se detallarán los lineamientos teóricos que sustentan el estudio, desde las disciplinas en que se inserta hasta sus conceptos y variables fundamentales; segundo, el Marco metodológico (3) que ofrece en detalle la descripción del tipo de estudio, materiales, métodos y criterios de análisis que se siguieron para la realización de la investigación; tercero, se presenta el

apartado de Resultados y discusión (4), donde se exponen los resultados obtenidos y su correspondiente interpretación; y, por último, las Conclusiones (5) que derivan de la totalidad de la investigación, donde se incluyen, además, las proyecciones y el trabajo futuro.

## **2. Marco teórico**

En este apartado se expondrán los principios teóricos que rigen la presente investigación. En primer lugar, se presentan las características generales que sustentan el estudio del léxico desde diferentes disciplinas (2.1); en segundo lugar, se expone, en términos generales, el fenómeno de la neología y el surgimiento de neologismos (2.2); en tercer lugar, se detallan los mecanismos de formación de palabras para caracterizar los neologismos de tipo formal (2.3); en cuarto lugar, se exponen las diferentes perspectivas desde las que se aborda la detección automática de neologismos para puntualizar en el criterio lexicográfico en que se basa la presente investigación (2.4); y, por último, se ofrecen las características morfológicas de los verbos con el fin de describir cómo operan los neologismos verbales (2.5).

### **2.1. El estudio del léxico**

El vocabulario se concibe como el conjunto de unidades léxicas que conforman una lengua y que se caracteriza por ser “estructurado, dinámico y creativo” (Feliu, 2009, p. 51). Esto implica que es un componente de la lengua sujeto a múltiples cambios, lo que lo convierte en un complejo objeto de estudio que ha sido abordado desde diferentes perspectivas. Las disciplinas de la lingüística que han enfocado sus estudios en el componente léxico son la lexicología y la lexicografía, cuyo desarrollo se presenta brevemente a continuación.

La lexicología, por su parte, es una subdisciplina de la lingüística que, como tal, hereda su carácter científico y enfoca sus estudios en “el significado de las unidades léxicas de una lengua y las relaciones sistemáticas que se establecen entre ellas en virtud de su significado” (De Miguel, 2015, p. 1). En este sentido, la lexicología aborda el léxico desde una perspectiva mayormente analítica, cuyo propósito principal es la formulación de principios generales que rigen su funcionamiento (Porto Dapena, 2002), dando lugar a un saber fundamentalmente teórico sobre este.

Por otra parte, la lexicografía constituye una técnica o metodológica enfocada en la elaboración de diccionarios, que estudia el léxico desde un punto de vista sintético para dar énfasis a las particularidades de cada unidad léxica y, con ello, a su correspondiente abordaje lexicográfico (Porto Dapena, 2002). Desde esta perspectiva se entiende que la lexicografía no implica solo un saber práctico o metodológico, sino también un saber teórico que sustenta

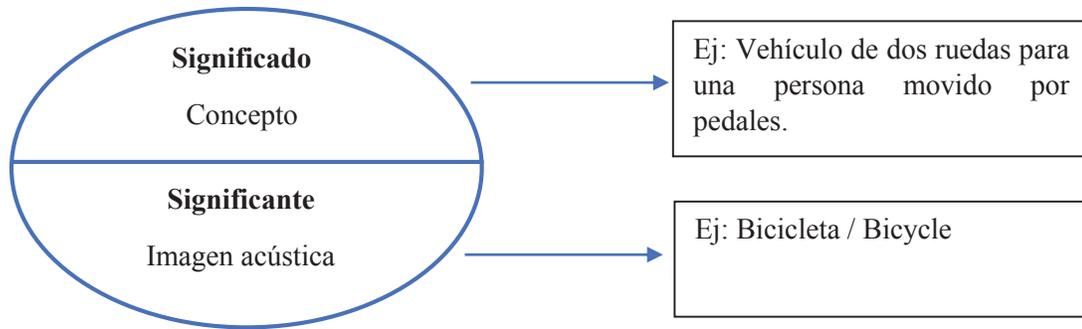
la descripción del léxico en los diccionarios, además de la propia concepción de estas fuentes lexicográficas. En este sentido, la lexicografía se definiría como aquella disciplina “que se ocupa de todo lo concerniente a los diccionarios, tanto en lo que se refiere a su contenido científico (estudio del léxico) como a su elaboración material y a las técnicas adoptadas en su realización” (Porto Dapena, 2002, p. 24).

Como se puede observar, la lexicología y la lexicografía se proponen objetivos diferentes respecto del léxico, como señala Batiukova (2009):

la lexicología como parte de la lingüística ha de dar respuestas sobre la estructura y el funcionamiento del léxico; la meta de la lexicografía es, en cambio, de tipo aplicado: consiste en recoger el vocabulario y describirlo en forma de repertorios (p. 487).

En consideración de lo anterior, resulta útil definir la unidad de análisis de la lexicología, disciplina en la que se enmarca el presente estudio. En este sentido, a partir de la noción de que el léxico es el conjunto de palabras de una lengua, se diría que la lexicología estudia, en concreto, las palabras. El concepto de palabra en el área de la lingüística moderna se comienza a configurar a partir de la noción de signo propuesta por Saussure (1916[2007]). En principio, respecto de los estudios del lenguaje, Saussure plantea que existe, por un lado, la lengua, que corresponde básicamente a un constructo teórico definido como un sistema de signos y, por otro lado, el habla, que corresponde a las realizaciones concretas que permite tal sistema de signos.

En este sentido, un signo es aquel elemento de la lengua que refiere a un objeto o fenómeno de la realidad y que se compone de un significado y un significante, tal como se representa en la Figura 2.



**Figura 2.** Diagrama del signo lingüístico según Saussure (1916[2007]).

Como se observa en la Figura 2, el significado corresponde al concepto o representación mental del referente, mientras que el significante es la imagen acústica que se corresponde con dicha representación mental. Cabe destacar que, desde el punto de vista de Saussure (1916[2007]), la relación entre ambos componentes es totalmente arbitraria, en tanto que nada motiva que el concepto correspondiente a *bicicleta* sea designado a través de la secuencia *b-i-c-i-c-l-e-t-a* en español, *b-i-c-y-c-l-e* en inglés, etc.

El signo, además de ser arbitrario, se caracteriza por ser inmutable y mutable a la vez, lo que quiere decir que un signo, en tanto que depende de un sistema de reglas establecido de manera convencional por los hablantes, goza de cierta estabilidad que permite su conservación en el tiempo tanto a nivel del significado como del significante. Sin embargo, en un espectro temporal más amplio, el mismo signo puede ir sufriendo cambios motivados por nuevas necesidades comunicativas o expresivas de los hablantes, lo que puede modificar alguno de sus componentes. Por ejemplo, la palabra *nube* actualmente ya no solo designa al “Agregado visible de minúsculas gotitas de agua, de cristales de hielo o de ambos, suspendido en la atmósfera y producido por la condensación de vapor de agua” (RAE, 2014, acep. 1), sino que con el avance tecnológico de Internet también refiere al “Espacio de almacenamiento y procesamiento de datos y archivos ubicado en internet, al que puede acceder el usuario desde cualquier dispositivo” (RAE, 2014, acep. 8), experimentando así un cambio en su significado, lo que, como se verá después (v. apartado 2.2), recibe el nombre de *neologismo semántico*.

En consideración de lo anterior, una palabra es, entonces, un signo, particularmente un signo lingüístico que, como tal, es uno de los objetos de estudio de la lingüística. No obstante, para definir qué es una palabra no basta con señalar que es un signo lingüístico, sino que hace falta considerar una serie de aspectos que delimitan con mayor claridad este concepto.

En cuanto a la determinación de la noción de palabra se han ofrecido diversas propuestas desde distintas perspectivas, todas ellas con sus aportes, pero también con claras limitaciones que deben tenerse en cuenta. Siguiendo a Escandell (2007), se reconocen, al menos, cuatro criterios que han sido propuestos para distinguir una palabra: primero, que una palabra se diferencia de otra a nivel escritural por estar estas separadas entre sí por un espacio. No obstante, es evidente que el modo como se escriben las palabras responde a normas ortográficas que frecuentemente son arbitrarias en cada lengua o, dentro de la misma lengua, varían en el tiempo. Por ejemplo, la forma *dámelo* está constituida por una sola palabra, mientras que *me lo das* por tres palabras: aunque son las mismas unidades su distribución en el espacio escritural es arbitraria en cada caso. En segundo lugar, según Bloomfield (citado en Escandell, 2007), “la palabra es la mínima forma libre” (p. 21), es decir, es la unidad mínima del discurso y es independiente. Sin embargo, hay otras unidades también consideradas palabras, como los artículos, las preposiciones o las conjunciones, que no gozan de tal independencia. En tercer lugar, Lyons (citado en Escandell, 2007) propone que lo que define una palabra es que esta está dotada de coherencia interna, es decir, que los elementos que la componen construyen su significado y este se erige como independiente: “una palabra puede tener otros componentes menores, pero estos no pueden reordenarse ni admiten la interpolación de otras palabras” (Escandell, 2007, p. 21). Por ejemplo, una palabra como *aerotransportadas* no admite la reubicación de sus componentes menores para formarse como *\*portransaeradas*. No obstante, esta característica no es suficiente para delimitar el concepto. Ante ello, Cruse (citado en Escandell, 2007) complementa la propuesta de Lyons y plantea que, además de coherencia interna, la palabra goza de movilidad sintáctica, limitada únicamente por las posibilidades combinatorias de las clases de palabras. Por ejemplo, las palabras contenidas en (2) pueden combinarse también como figura en (3) por la movilidad sintáctica de la que gozan, pero no de la manera en que aparece en (4).

(2) *Juan contestó rápidamente el teléfono*

(3) *Rápidamente contestó Juan el teléfono*

(4) *Juan el rápidamente contestó teléfono*

En concreto, las palabras por sí solas o combinadas pueden constituir unidades léxicas, las cuales poseen un significado léxico independiente, tal como ocurre con los sustantivos, verbos, adjetivos y adverbios, o bien palabras funcionales, que no poseen significado léxico, sino que cumplen determinada función en el discurso. En consecuencia, cobra especial relevancia para este estudio el concepto de unidad léxica. Al respecto, Cabré (2015), a partir de los postulados de Alain Rey, define unidad léxica como “toda palabra, lexía o sintagma (sintagma lexicalizado)” (p. 10). En este sentido, una unidad léxica se caracteriza, principalmente, por estar dotada de un significado independiente (significado léxico) y por gozar de independencia sintáctica.

La noción de unidad léxica planteada hasta ahora responde mayoritariamente a la lexicología inicial, la que estaba en estrecha relación con el estructuralismo preponderante durante la primera mitad del siglo XX y que se basaba en un análisis componencial de las unidades del léxico. Particularmente, de acuerdo a este tipo de análisis estructural, la unidad del léxico es el *lexema*, que se materializa en el habla como *palabra* y que está constituido por elementos mínimos de contenido que son los *semas*. Así, los semas permiten poner en oposición las unidades léxicas como, por ejemplo, *sillón* y *silla* se oponen por el sema [±CON BRAZOS], aunque ambos pertenecen al mismo *campo léxico*. Evidentemente, esta oposición binaria suscita una serie de dificultades, en tanto para distinguir, por ejemplo, *silla*, *sillón*, *taburete*, *asiento* y *sofá* haría falta recurrir a rasgos extralingüísticos y propios de la realidad, lo que atenta contra la supuesta independencia de la lengua respecto del mundo. Ante esta problemática, la semántica cognitiva propone una categorización por prototipos, planteando que la pertenencia de un miembro a una categoría no siempre está determinada por una oposición binaria, sino por su aproximación a un *prototipo* que reúne las principales características del grupo, como, por ejemplo, lo es *gorrión* en la categoría de *aves*, relevando los aspectos extralingüísticos en su definición. Esta última postura responde, más que a la lexicología, al desarrollo de ciertos procesos cognitivos (De Miguel, 2015).

Posteriormente, modelos funcionalistas y generativistas heredaron algunos principios de la lexicología estructural, principalmente el análisis componencial y la idea de que las unidades

léxicas se definen según las relaciones que establecen unas con otras. De allí, pervive la idea de contemplar la unidad léxica en su contexto de uso, surgiendo así la necesidad de recurrir a la información semántica y, de esta manera, estableciendo un vínculo indisoluble entre la sintaxis y la semántica hasta los estudios actuales en lexicología (De Miguel, 2015). Para tales efectos, surge la necesidad de obtener datos reales de uso que permitan observar el comportamiento de las unidades léxicas en relación con otras y, junto con ello, las relaciones entre sus significados, por lo que una metodología basada en corpus se vuelve imprescindible. En este contexto, han surgido propuestas como la Theory of Norms and Exploitations (Hanks, 2013), que plantea un modelo de análisis de corpus para observar el comportamiento sintáctico y semántico de las palabras en contextos reales de uso.

## **2.2. Neología y neologismos**

El dinamismo que caracteriza a las lenguas se hace evidente principalmente a nivel léxico, a través de, entre otras cosas, la pervivencia, pérdida e incorporación de palabras. Este dinamismo se vincula directamente con la mutabilidad del signo que se abordó en el apartado anterior y ha sido objeto de interés por parte de la lexicología y motivo de constante actualización de fuentes lexicográficas que pretenden describir el estado de una lengua particular en un momento dado.

A partir del origen de las palabras o unidades léxicas que son utilizadas por los hablantes de una lengua, se pueden reconocer tres categorías (Álvarez de Miranda, 2009): el *léxico heredado*, correspondiente a las palabras que “no nacen”, sino que vienen heredadas de lenguas precedentes, como la palabra *casa* que viene de la voz del latín *casa* que refería a una ‘choza’; el *léxico adquirido*, compuesto por todas aquellas palabras que una lengua toma prestadas de otra, tales como *football* o *happy hour*, provenientes del inglés; y el *léxico multiplicado*, que consta de palabras que surgen dentro de una misma lengua producto de sus propios mecanismos de formación, como, por ejemplo, *predecir*, que consiste en la base léxica *decir* a la que se le ha añadido el prefijo *pre-*. A estas categorías se suman también las *creaciones inducidas* o *calcos*, que comparten características del léxico adquirido y del multiplicado y tienen lugar cuando una palabra surge motivada por la existencia de otra palabra en otra lengua, a partir de la cual se aplican mecanismos de derivación o composición. Tal es el caso de la palabra *contenedor*, tomada del inglés *container* y adaptada a los procesos

de creación de palabras del español. Tanto el léxico adquirido como el multiplicado y las creaciones inducidas, junto con otros mecanismos de menor relevancia, contribuyen en el aumento del volumen del léxico de una lengua como parte de lo que se conoce como *lexicogénesis* (Álvarez de Miranda, 2009). Por el contrario, la pérdida léxica, que ocurre cuando una palabra deja de ser utilizada por los hablantes de una lengua, implica su disminución. Ambos fenómenos constituyen lo que se denomina *cambio léxico*, el que junto con el *cambio semántico* contribuye al dinamismo de la lengua (De Miguel, 2015).

En este sentido, la neología corresponde a aquella parte de la lexicología que se centra en los procesos de lexicogénesis, es decir, en la incorporación de nuevas unidades léxicas, llamados *neologismos*, al caudal léxico de una lengua y que constituye el eje central del presente estudio. En cuanto a su desarrollo disciplinar, Cabré (2015) señala que desde fines del siglo XX que la neología ha ido cobrando relevancia, mas no ha logrado un desarrollo teórico apropiado debido a, entre otras cosas, la insistencia de los académicos e investigadores en centrar sus estudios en la vertiente aplicada de esta subdisciplina. Además, desde la aparición del término *neología* en 1758 este estaba asociado, primero, con la creación de nuevas palabras o significados, pero posteriormente, durante los siglos XVIII y XIX, estuvo asociado con la transgresión de la norma, lo que relegó su estudio.

Por esta razón, se manejan escasos antecedentes teóricos sobre neología y, evidentemente, sobre el concepto de neologismo. Entre estos antecedentes, Cabré (2015) destaca los aportes de Louis Guilbert y Alain Rey. Por una parte, Guilbert (citado en Cabré, 2015) entre los años 1973 y 1975 establece que la neología se rige por cinco principios, que se podrían sintetizar en que, primero, esta depende del código en que funciona la lengua; segundo, la creación neológica modifica tanto el significado como el significante del signo lingüístico; tercero, un neologismo no es una unidad de significación mínima, sino construida; cuarto, la creación neológica siempre depende de la situación comunicativa o discursiva en que tiene lugar; y, por último, el neologismo tiene un carácter tanto escrito como oral. Asimismo, Guilbert retoma las nociones de mutabilidad e inmutabilidad del signo (v. apartado 2.1) para explicar el fenómeno de la neología. Al respecto, sostiene que las mutaciones en el léxico se encuentran en estrecho vínculo con las nociones de mutabilidad e inmutabilidad del signo planteadas por Saussure (1916[2007]). En este sentido, los cambios producidos responden a

necesidades comunicativas y expresivas diversas, como, por ejemplo, mejorar la economía lingüística (Cabré, 2015).

Por otra parte, Rey (citado en Cabré, 2015) en 1973 sitúa directamente el neologismo en el nivel léxico y describe su complejidad a partir de tres aspectos: en primer lugar, señala que se trata de una unidad léxica, siguiendo la definición abordada anteriormente (v. apartado 2.2), que es producida individualmente por los hablantes y cuyo paso al uso colectivo constituye un hecho psicosocial más que lingüístico. En segundo lugar, plantea que un neologismo siempre está determinado por la *novedad*, pero que esta puede afectar diferentes aspectos de esta unidad: su forma, su contenido o su uso. En tercer lugar, propone que un neologismo solo es tal en determinado contexto o en relación con determinado subcódigo, es decir, la novedad es relativa en tanto que lo que puede ser nuevo para un grupo social puede no serlo para otro.

En definitiva, el concepto de neologismo resulta difícil de abordar principalmente en relación con el aspecto relativo que se menciona anteriormente, la novedad, puesto que no hay claridad respecto de los criterios para determinar cuándo una palabra deja de ser considerada un neologismo y pasa a formar parte del léxico estable. Al respecto, Álvarez de Miranda (2009) señala:

Es obvio, por lo demás, que la pérdida (la “superación”, podríamos decir) de la condición de neologismo algo tendrá que ver, más que con el paso del tiempo, con cierto grado de extensión y aun generalización, es decir, con datos vinculados a la frecuencia de empleo (el cual, naturalmente, puede estar limitado a un territorio, un estrato social o un determinado registro). Y también, con la conciencia lingüística de los hablantes, con su “conciencia neológica” (p. 137).

A partir de esto, se puede observar, por un lado, que los neologismos de una lengua pueden presentarse con una frecuencia lo suficientemente alta como para advertir su circulación en el habla y que los hablantes puedan reconocerlos como palabras nuevas, pero lo suficientemente baja como para determinar que no se ha equiparado a la frecuencia de las palabras establecidas en el sistema de la lengua. Por otro lado, en tanto que el uso colectivo de un neologismo constituye un proceso psicosocial, se puede observar que se encuentra en

estrecha relación con cómo los hablantes perciben esta nueva unidad léxica (Álvarez de Miranda, 2009), lo que difícilmente puede ser abordado de manera exhaustiva y sistemática.

Por último, cabe destacar que la incorporación de nuevas unidades léxicas es bastante recurrente: se registran más entradas al léxico, es decir, neologismos, que salidas o pérdidas léxicas (Álvarez de Miranda, 2009). Esto conlleva la imperiosa necesidad de monitorear el proceso de lexicogénesis de una lengua tanto para promover el avance teórico de los estudios sobre neología como para la actualización de las bases de datos lexicográficas: “La neología ha sido y es necesaria en lexicografía por cuanto los diccionarios generales deben actualizar su nomenclatura al ritmo de los nuevos usos, siempre constantes por la evolución social incesante” (Cabré, 2015, p. 3).

### **2.3. Neología léxica**

Como se menciona anteriormente (apartado 2.2), los neologismos pueden ser producto del cambio léxico o del cambio semántico. En otros términos, puede incorporarse al caudal léxico de una lengua una palabra totalmente nueva por medio de un préstamo, mecanismos de formación de palabras o por creación inducida, lo que corresponde a la neología léxica; o bien, puede incorporarse únicamente un nuevo significado asociado a una palabra preexistente en la lengua, lo que corresponde a la neología semántica.

Para efectos de la presente investigación, se hará especial énfasis en la neología léxica o formal, que, en términos generales, se define como el surgimiento de una unidad léxica que no existía previamente en el lexicón (Varo, 2012) o, lo que es lo mismo, de un nuevo signo (significado y significante nuevo) (Álvarez de Miranda, 2009). Este proceso depende de recursos morfológicos, sintácticos, fonológicos y semánticos previamente existentes en la lengua (Cabré, Bayà, Bernal, Freixa, Solé y Vallès, 2002), cuyo funcionamiento los hablantes conocen por la competencia lingüística que poseen. Sin embargo, cabe recordar, a partir de la distinción anterior, que no todo proceso neológico está asociado a estas reglas de carácter formal o gramatical:

La creatividad lingüística es una propiedad ligada a las lenguas. Forma parte de la competencia del hablante en general la capacidad de crear nuevas unidades a partir de las reglas de la gramática. El léxico es el componente gramatical en el que se

evidencia mejor y más explícitamente esta capacidad. Pero la creatividad léxica como propiedad de las lenguas no puede confundirse con la neología. Los neologismos no siempre se producen utilizando los mecanismos de formación de palabras que ofrece el sistema de una lengua. La neología es un ámbito de conocimiento centrado en el análisis de las unidades consideradas nuevas en una lengua. La creatividad, tal como se ha estudiado en la lingüística, es un proceso regulado totalmente o parcialmente por reglas de la gramática. La neología, a diferencia de la creatividad que es una capacidad inherente a las personas hablantes y una propiedad constitutiva de las lenguas, es una opción que a menudo, pero no únicamente, aprovecha los mecanismos gramaticales para realizarse (Cabré, 2015, p. 14)

Como se puede advertir en la cita anterior, la creatividad lingüística es un fenómeno mucho mayor a la neología formal que aborda este estudio puesto que los hablantes, si bien aprovechan los recursos preexistentes en la lengua, muchas transgreden sus normas para producir nuevos elementos que respondan a sus necesidades comunicativas y expresivas. En consecuencia, el presente estudio aborda parcialmente el fenómeno de la lexicogénesis, centrándose únicamente en aquellos neologismos, particularmente neologismos verbales, que surgen producto de los mecanismos de creación de palabras del español, especialmente mediante la derivación. En otras palabras, el foco se encuentra en parte del *léxico multiplicado* en tanto mantiene directa relación con uno de los principales procesos morfológicos de formación de palabras.

Particularmente, el estudio de estos mecanismos corresponde a la morfología léxica, que se define como aquella que “comprende el conjunto de procedimientos formales empleados en una lengua como el español para crear palabras nuevas a partir de unidades léxicas ya existentes” (Feliu, 2009, p. 53). Esta se diferencia de la morfología flexiva en dos aspectos: la morfología léxica, por su parte, tiene capacidad creativa, pero un nivel de productividad restringido, ya que los procedimientos de formación de palabras no aplican a todas las palabras que conforman el léxico de una lengua. Por otra parte, la morfología flexiva, que es aquella que “da lugar a distintas formas gramaticales de una misma palabra” (Feliu, 2009, p. 53), no comparte dicha capacidad creativa, pero sí posee un alto nivel de productividad en tanto que la mayoría de las palabras pueden flexionarse (Feliu, 2009). En este sentido, el

presente estudio aprovecha los recursos teóricos ofrecidos por ambas, en tanto, por una parte, la morfología léxica ofrece una concepción y caracterización acabada de los mecanismos de formación de palabras y, por otra parte, la morfología flexiva permite determinar qué es un verbo y cómo debe ser lematizado (v. apartado 3.4.2.3).

Antes de exponer en detalle cómo funcionan tales mecanismos de formación de palabras, resulta conveniente abordar algunos conceptos que permiten comprender la estructura morfológica de las palabras y, junto con ello, sus posibilidades de modificación y creación. En primer lugar, el morfema es la unidad mínima de segmentación de las palabras. Según Torner (2006), siguiendo un criterio distribucional, es decir, basado en la posición que ocupan los morfemas, se pueden distinguir raíces y afijos: la *raíz* o *lexema* es aquel morfema que aporta el significado a la palabra y que se mantiene tanto en las realizaciones flexivas como en las familias de palabras; mientras que los *afijos* son aquellos morfemas que se adhieren a la raíz para flexionar una palabra o formar una nueva por derivación. Por ejemplo, en las palabras *asustaba*, *asusté* y *asustaste*, la raíz común a todas ellas es *asust-*, mientras que los afijos son *-aba*, *-é* y *-aste* que permiten flexionar el verbo *asustar* en los distintos tiempos y personas gramaticales. El elemento al que se adhiere el afijo puede ser una raíz o un segmento formado por una raíz y un afijo y recibe el nombre de *base*. Por ejemplo, en la palabra *véndeme*, *-me* es el sufijo que se adhiere a la base *vende-*; luego, si se agrega el sufijo *-lo* para formar *véndemelo*, la base se considera *véndeme-*.

Particularmente, los afijos se clasifican de acuerdo con dos criterios. En primer lugar, según la posición que ocupan, se distinguen *prefijos*, que se adhieren al principio de la palabra, como *multi-* en el caso de *multifuncional*, *multiuso*, etc.; *sufijos*, que se sitúan al final de la palabra, como lo es *-ble* en *deseable*, *habitabile*, etc.; *circunfijos*, que se adjuntan simultáneamente al principio y al final de la palabra, dando lugar a la parasíntesis, como ocurre en las palabras *en-tern-ecer* o *en-trist-ecer* donde el circunfijo es *en...ecer*; e *interfijos*, que se sitúan en medio de dos bases que forman una palabra compuesta, como el afijo *-i-* en *altibajo*. En segundo lugar, según la función que cumplen, los afijos pueden ser *derivativos*, cuando permiten formar una nueva palabra a partir de un recurso preexistente, como el prefijo *in-* que permite derivar *inútil* o *inusual* de las palabras *útil* y *usual* respectivamente; o *flexivos*, cuando permiten formar diferentes variantes flexivas de una palabra como los afijos *ía*, *iste*

e *ían* que permiten flexionar el verbo *comer* para formar *comía*, *comiste* y *comían* respectivamente (Torner, 2006). Evidentemente, entre estos, los que suponen mayor relevancia para esta investigación son los afijos derivativos.

La derivación es aquel proceso que consiste en “añadir un afijo derivativo a una base léxica, ya sea a una raíz o al conjunto formado por una raíz más otro(s) afijo(s). Según el tipo de afijo utilizado en la derivación, se distinguen la prefijación, la sufijación y la parasíntesis” (Torner, 2006, p. 39).

En concreto, la prefijación, como es de suponer, es un proceso de derivación que consiste en la adhesión de un prefijo a una base, lo que da lugar a una palabra derivada por prefijación. Ejemplos de verbos sometidos a este proceso son *reabrir* y *reabsorber*, a cuyas bases, *abrir* y *absorber* respectivamente, se les ha agregado el prefijo *re-* para indicar que la acción se repite. En general, este mecanismo no cambia la categoría gramatical de la base. Así, para efectos de este estudio, según el protocolo de vaciado de textos escritos (espontáneo) propuesto por OBNEO (2004) para la detección de neologismos distintos a los textos de prensa, son palabras neológicas derivadas por prefijación todas aquellas que contienen los prefijos que muestra la Tabla 1.

a-, an-	ceto-	extra-	multi-	re-
ab-	circun-, circum-	ex-, exo-	neo-	semi-
ana-	cis-	fuera-	nitr-, nitro-	tras-
anfi-	con-, com-, co-	hemi-	nitros-, nitroso-	sin-, sim-, sil-
ante-	contra-, contra	hidroxi-	ob-	sobre-
anti-	des-	hiper-	oxo-	sub-
apo-	di-	in-, im-, i-	pan-	sulf-, sulfo
archi-	di-, dia-	infra-	oxi-	supra-
azoxi-	diazo-	inter-	para-	super-
bajo-	e-	intra-	peri-	trans-
benz-, benzo-	dis-	mal-	platini-	tri-

bi-, bis-	eu-	menos-	pluri-	ultra-, ultra
bien-	endo-	meta-	poli-	uni-
carboxi-	epi-	mono-	post-, pos-	vice-
casi-	eso-	metoxi-	pre-	yuxta-
cata-	ex-	no-, no	pro-	

**Tabla 1.** Tabla de prefijos para la detección de neologismos (OBNEO, 2004, pp. 43-44)

La sufijación, por su parte, consiste en añadir un sufijo derivativo a una base, lo que da lugar a una palabra derivada por sufijación. De acuerdo con lo planteado por Torner (2006), “muchos sufijos seleccionan la categoría gramatical de la base a la cual se pueden adjuntar. Algunos de ellos no cambian la categoría de dicha base, pero otros la pueden modificar” (p. 36). Particularmente, en el caso de los verbos, se distinguen sufijos verbalizadores tales como *-izar*, *-ear*, *-ecer*, *-ificar*, etc. Estos sufijos crean, por ejemplo, verbos a partir de nombres, como en el caso *golear*, proveniente del sustantivo *gol*, o bien, a partir de adjetivos, como en el caso de *palidecer*, proveniente del adjetivo *pálido* (Torner, 2006).

En cuanto a la parasíntesis, cabe recordar, como se menciona anteriormente, que consiste en la adhesión simultánea de un prefijo y un sufijo o, lo que es lo mismo, de un circunfijo a una base. Particularmente, en el caso de los neologismos en que se distingue una base afectada tanto por un prefijo como por un sufijo, según lo planteado por OBNEO (2004), resultaría pertinente detectar el proceso más reciente entre ambos. Si no fuera posible determinar esto, entonces se trataría de interferencias entre prefijación y sufijación. Sin embargo, para efectos de este estudio, bastará con señalar que se trata de una palabra derivada por parasíntesis.

En el caso de los verbos, la parasíntesis opera a partir de bases nominales y adjetivas mayoritariamente y en escasas ocasiones a partir de bases adverbiales. De acuerdo con lo planteado por Feliu (2009), los prefijos más frecuentes en su formación son *a-* y *en-*, y en menor medida *des-*. La Tabla 2 muestra a modo de ejemplo los esquemas de formación de estos verbos con los prefijos *a-* y *en-*.

	<b>Estructura</b>	<b>Ejemplo</b>
<b>Verbos parasintéticos denominales</b>	<i>a-N-ar</i>	<i>Acaramelar</i>
	<i>a-N-ecer</i>	<i>Atardecer</i>
	<i>a-N-izar</i>	<i>Aterrorizar</i>
	<i>en-N-ar</i>	<i>Empolvar</i>
	<i>en-N-ecer</i>	<i>Enmohecer</i>
	<i>en-N-izar</i>	<i>Encolerizar</i>
<b>Verbos parasintéticos deadjetivos</b>	<i>a-A-ar</i>	<i>Agrandar</i>
	<i>en-A-ar</i>	<i>Endulzar</i>
	<i>n-A-ecer</i>	<i>Ensordecen</i>

**Tabla 2.** Esquema de formación de verbos parasintéticos con prefijos *a-* y *en-* (Feliu, 2009, p. 70). N = nombre; A = adjetivo.

Otro de los mecanismos de formación de palabras más recurrentes es la composición. En términos generales, Feliu (2009) define la composición como un mecanismo de formación de palabras “mediante la unión de dos o más palabras, raíces españolas o temas grecolatinos, que poseen un significado unitario y un referente único, independiente de que formen una unidad gráfica o no” (p. 71). Ejemplos de palabras compuestas son *bocacalle*, *agridulce*, *lavamanos*, etc. Además, como parte de este mecanismo se distingue también la composición culta que, según OBNEO (2004), es un procedimiento de formación de palabras cercano a la prefijación y la sufijación. En términos generales, se trata de neologismos que contienen como forma sufijada o forma prefijada un formante culto proveniente del griego o del latín, tal como *fito-* (proveniente del griego) en *fitocosmética*, o un formante obtenido mediante el acortamiento de una palabra, como *tele-* en la palabra *telepromoción*. Al respecto, Feliu (2009) señala que, si bien a simple vista la derivación y la composición culta pueden parecer semejantes, presentan diferencias notorias, principalmente en el hecho de que, por una parte, los formantes cultos pueden combinarse entre sí y formar palabras nuevas y, por otra, las palabras acortadas no necesariamente deben ir ligadas a una base. En particular, la composición y la composición culta no constituyen un foco de interés para el presente estudio, en tanto la atención está centrada en la neología derivativa, lo que no implica que no puedan detectarse a través del método propuesto por su estructura morfológica.

Las palabras obtenidas a través de los mecanismos anteriormente descritos, derivación y composición, forman parte, tal como se indicó anteriormente, de lo que Álvarez de Miranda (2009) denomina *léxico multiplicado*, es decir, aquel que se obtiene por medio de los mecanismos de formación de palabras de una lengua en particular, en este caso del español. Asimismo, dentro de tales categorías del léxico, se encuentra el *léxico adquirido*, que corresponde a aquellas palabras que se obtienen por préstamos de otras lenguas. Según este autor, “se produce un *préstamo* cuando una lengua toma de otra un elemento lingüístico, con mayor o menor adaptación al carácter y fisonomía de la lengua receptora (...). Los préstamos son la más importante y visible consecuencia del contacto entre lenguas” (p. 142). Ejemplos de préstamos establecidos en el español son *sushi*, proveniente del japonés; *hardware*, proveniente del inglés; *mousse*, del francés, etc.

La incorporación de estas palabras al caudal léxico de una lengua, como se observa en las palabras de Álvarez de Miranda (2009), puede en ciertos casos requerir de un proceso de adaptación; en tal caso se habla de un *préstamo adaptado* pues este se acomoda a las reglas ortográficas o a la fonética de la lengua receptora. Ejemplo de ello son las palabras *chequear*, proveniente del inglés *to check* (‘comprobar’) y que ha sido adaptada a la ortografía del español (además de a su sistema flexivo); o *chalé*, proveniente del francés *chalet* (con el mismo significado de tipo de ‘vivienda’) y que ha experimentado una adaptación ortográfica. En definitiva, un préstamo adaptado comparte características morfológicas con las palabras de la lengua en la que se inserta. Esto es posible en tanto el préstamo es sometido a los mecanismos de derivación correspondientes para que pertenezca, en términos de su estructura morfológica, a una de las clases gramaticales preestablecidas en la lengua. Por ejemplo, *to check* pasa a ser un verbo del español cuando se toma su base y se le aplica el sufijo *-ear*, terminación propia de los verbos de nuestra lengua. En consideración de esto, los préstamos adaptados también constituyen un foco de interés en el presente estudio.

En síntesis, para esta investigación se considerarán como centro de atención los neologismos obtenidos a partir de procesos de derivación, ya sea por medio de sufijación, prefijación o parasíntesis, o bien a partir de préstamos adaptados.

## 2.4. Detección automática de neologismos

Todo lo anteriormente expuesto en este apartado pretende delimitar los parámetros en que se concibe la neología y, particularmente, los neologismos formales. Sin embargo, estos lineamientos responden, evidentemente, a la dimensión teórica del fenómeno. En este apartado se revisarán algunos aspectos relativos a los métodos computacionales para el tratamiento automático de los neologismos, especialmente su detección, área en la que se inscribe esta tesis.

En general, según lo planteado por Janssen (2009), la neología se ha visto beneficiada por el avance tecnológico, lo que ha implicado un cambio en los diferentes objetivos que se propone con respecto a los neologismos:

La neología originalmente era solamente una práctica lexicográfica, recopilando palabras nuevas que entran en la lengua para la creación de diccionarios de neologismos y para la actualización de diccionarios de lengua existentes. Pero ahora, la neología se ha transformado en una disciplina extensa, que tiene un número de finalidades diversas, entre las cuales se encuentra el perfeccionamiento de las herramientas para el procesamiento del lenguaje, la regularización de préstamos y el estudio de la productividad morfológica. (p. 2)

A pesar de este avance, el autor señala que la funcionalidad de las herramientas para la detección de neologismos se ha mantenido estable. En este sentido, distingue tres abordajes diferentes para esta tarea. En primer lugar, se encuentra la utilización de una lista de exclusión compuesta por palabras conocidas para contrastar con el corpus de estudio, de modo que las palabras que están en el corpus de estudio, pero no en el de exclusión, se consideran neologismos. En segundo lugar, cuando se quiere descartar la comparación anterior entre un corpus de estudio y uno de exclusión, se recurre a la extracción de patrones léxico-sintácticos que se basa en la relación del neologismo con las palabras de su contexto y “se funda en la idea de que los neologismos son palabras poco conocidas por los lectores y, por lo tanto, son introducidos de una manera especial” (Janssen, 2009, p. 3) a través de elementos como, por ejemplo, comillas o cursiva. Por último, se encuentra el análisis estadístico, que consiste en “contar las ocurrencias de palabras en el corpus de estudio, normalmente en comparación con el corpus de referencia” (Janssen, 2009, p. 3).

A partir de lo anterior, es posible observar que cada abordaje se rige por sus propios criterios. Al respecto, Cabré (1993) señala que para el abordaje metodológico de la detección de neologismos se han establecido los siguientes parámetros:

- a. *La diacronía*: una unidad es neológica si ha aparecido en un período reciente;
  - b. *La lexicografía*: una unidad es neológica si no aparece en los diccionarios;
  - c. *La inestabilidad sistemática*: una unidad es neológica si presenta signos de inestabilidad formal (morfológica, gráfica, fonética) o semántica;
  - d. *La psicología*: una unidad es neológica si los hablantes la perciben como una unidad nueva.
- (p. 445)

En la presente investigación, para efectos del análisis de los datos obtenidos, se siguen los principios del primer abordaje presentado por Janssen (2009), es decir, el de la utilización de una lista de exclusión. En consecuencia, el parámetro utilizado para determinar si un verbo es neológico es el de la lexicografía o criterio lexicográfico: “Con este criterio, un término es un neologismo si no aparece en el corpus lexicográfico seleccionado como corpus de referencia (llamado también corpus de exclusión)” (Cabré, 1993, p. 446). En este sentido, se considera que toda aquella palabra que se encuentre en uso, pero que no aparezca en una fuente lexicográfica representativa, tiene un carácter neológico. Con ello, se concibe el diccionario como una fuente confiable de información en tanto registra las palabras que poseen un alto grado de estabilidad en el léxico de una lengua.

No obstante, la aplicación de este criterio no se encuentra exento de dificultades en tanto no es posible asegurar que toda palabra del corpus de estudio y que no se encuentre en el corpus de exclusión sea efectivamente un neologismo, percibido como tal por los hablantes. En ocasiones, un neologismo puede permanecer largo tiempo presente en el uso, incluso hasta perder su carácter neológico, sin que con ello pase necesariamente a formar parte de un diccionario. Asimismo, que un neologismo haya sido efectivamente incorporado a un diccionario por su alta frecuencia en determinado momento, no asegura que su uso se mantenga estable posteriormente. Al respecto, Renau, Nazar y Lecaros (en prensa) señalan que la principal limitación de este criterio responde la naturaleza de los elementos que están en cuestión: la neología es un fenómeno de naturaleza lingüística, mientras que los diccionarios son objetos de naturaleza cultural. Junto con ello, “la mayoría de ellos son

prescriptivos y sancionan los usos antinormativos, están sometidos a restricciones económicas o a la presión social que hace aceptar antes una palabra que otra, etc.” (p. 20).

En definitiva, ciertas obras lexicográficas, como el *Diccionario de la lengua española* (RAE, 2014), que pretenden ser representativas del léxico de una lengua se consideran una lista de exclusión suficiente para la detección de neologismos o, más específicamente, de candidatos a neologismos. No obstante, no son suficientes para la caracterización de estos ni para determinar la pervivencia de su carácter neológico. Por ende, en el presente estudio el criterio lexicográfico constituye una herramienta complementaria para el monitoreo del léxico utilizado por los hablantes.

## **2.5. Verbos: morfología, modelos de conjugación y neología verbal**

Los verbos constituyen una de las nueve clases de palabras en que se clasifican las unidades sintácticas. Entre estas clases, los verbos, debido a su capacidad flexiva, corresponden al tipo de palabras variables (RAE, 2009).

Por una parte, en cuanto a su estructura morfológica, se pueden identificar dos elementos o componentes:

la RAÍZ (también RADICAL o BASE LÉXICA), que expresa su significado, tal como lo describe el diccionario, y un conjunto de MORFEMAS FLEXIVOS a los que corresponden dos funciones: establecer la concordancia de número y persona con el sujeto gramatical y expresar las nociones de ‘modo’, ‘tiempo’ y ‘aspecto’ que corresponden al evento. (...) Así pues, los rasgos de número y persona aportan informaciones que afectan al sujeto, mientras que los demás precisan ciertos aspectos del suceso o de la situación a la que se hace referencia (RAE, 2009)

Tales morfemas flexivos son denominados *desinencias*. La desinencia de cada forma verbal conjugada se puede presentar de manera unitaria, como en *bail-abas* o *jug-abas*, donde se distingue únicamente raíz y desinencia, o bien, segmentada en tres elementos, donde el primero es la *vocal temática*, la cual es determinada por el infinitivo e indica a cuál de las tres conjugaciones del español pertenece (-ar, -er o -ir); el segundo indica *modo* y *tiempo* y el tercero corresponde al *morfema de concordancia* que indica *persona* y *número*. Siguiendo los ejemplos dados terminados en *-abas*, la segmentación sería de la siguiente manera, en

orden de colocación desde la raíz: *-a* es la vocal temática correspondiente a la primera conjugación, *-ba-* indica que corresponde al tiempo pretérito imperfecto del modo indicativo y *-s* señala que se trata de la 2.<sup>a</sup> persona singular. En ocasiones, no toda esta información se ve representada en la desinencia, lo que genera un debate respecto de la segmentación de la desinencia, como ocurre en formas como *bailé*, donde la desinencia corresponde únicamente al segmento *-é*. En tales casos no hay acuerdo en cuanto a si en tal segmento converge toda la información señalada o si se trata de una ausencia de segmentos fonológicos para su representación (RAE, 2010, pp. 49-50).

Además, a esta estructura puede adjuntarse un pronombre enclítico (*me, te, se, lo, los, la, las, le, les, os*) a modo de sufijo lo que puede implicar ciertos cambios:

Las formas personales del verbo solo se emplean hoy con pronombres enclíticos en la lengua literaria (*Diose por enterado*), con la excepción del imperativo (*Decímelo; Guárdatelas*), en el que constituye la única opción en la lengua estándar. En estos contextos el morfema *-mos* pierde la *-s* ante el pronombre enclítico *-nos*: *Alegrémonos por lo sucedido*, pero no ante los demás pronombres: *Digámosle la verdad*. En los mismos contextos, la *s-* del pronombre enclítico *se* se funde con la de *-mos*: *digámoselo, repitámoselo*. La desinencia de la 2.<sup>a</sup> persona del plural del imperativo en la variante *ama-d* pierde el segmento *-d* ante el pronombre enclítico *-os*: *amaos* (RAE, 2010, p. 54)

Por otra parte, las formas verbales se clasifican en formas personales y formas no personales, diferenciándose estas últimas por no admitir el segmento que indica persona y número. Así, las formas no personales son el infinitivo, el gerundio y el participio:

Con la raíz y la vocal temática se construye el tema de infinitivo (*ama-, teme-, parti-*), al que se agrega el morfema *-r*. La vocal temática no se altera en la formación de gerundios y participios de la primera conjugación (*am-a-ndo, am-a-do*), pero toma la forma del diptongo *-ie-* en los gerundios de la segunda y la tercera (*tem-ie-ndo, part-ie-ndo*). Los participios de la segunda se asimilan a los de la tercera en la vocal temática *-i-* (*tem-i-do, part-i-do*) (RAE, 2010, p. 50).

Tal como se señala en la *Nueva gramática de la lengua española*, NGLE (RAE, 2009), “las formas verbales, personales y no personales, se organizan en una serie de paradigmas cerrados, denominados TIEMPOS VERBALES, que se subordinan a los modos” (p. 51). Estos tiempos verbales se dividen en simples y compuestos de acuerdo a si en su estructura léxica se presenta uno o dos elementos respectivamente. Cabe destacar que para efectos de la presente investigación solo se consideran los tiempos verbales simples en tanto se trabaja con formas de un elemento. En función de estos tiempos verbales, los verbos del español se flexionan, lo que los define como un tipo de palabra variable. La Tabla 3 muestra esta clasificación con sus respectivos ejemplos.

<b>Modo</b>	<b>Forma del tiempo</b>	<b>Tiempo</b>	<b>Ejemplo</b>
Indicativo	Simples	Presente	<i>canto</i>
		Pretérito perfecto simple	<i>Canté</i>
		Pretérito imperfecto	<i>cantaba</i>
		Futuro simple	<i>cantaré</i>
		Condicional simple	<i>cantaría</i>
	Compuestos	Pretérito perfecto compuesto	<i>he cantado</i>
		Pretérito pluscuamperfecto	<i>había cantado</i>
		Pretérito anterior	<i>hube cantado</i>
		Futuro compuesto	<i>habré cantado</i>
		Condicional compuesto	<i>habría cantado</i>
Subjuntivo	Simples	Presente	<i>Cante</i>

		Pretérito imperfecto	<i>cantara o cantase</i>
		Futuro simple	<i>cantare</i>
	Compuestos	Pretérito perfecto compuesto	<i>haya cantado</i>
		Pretérito pluscuamperfecto	<i>hubiera o hubiese cantado</i>
		Futuro compuesto	<i>hubiere cantado</i>
Imperativo	-	-	<i>Canta</i>

**Tabla 3.** Modos y tiempos verbales del español (RAE, 2009).

De acuerdo con las regularidades que presentan sus formas personales o conjugadas, los verbos se clasifican en *verbos regulares* y *verbos irregulares*. Por una parte, la conjugación regular se caracteriza porque la raíz no se ve alterada o modificada en los tres paradigmas de esta conjugación (Tablas 4, 5 y 6), salvo excepciones dadas por la acentuación, donde se distinguen tres grupos: formas que llevan el acento en la vocal temática, formas que llevan el acento en el primer segmento de *tiempo* y *modo* y formas que llevan el acento en la última vocal de la raíz (RAE, 2009). Dentro de las principales características de este tipo de conjugación se encuentran:

La conjugación regular del español, a la que pertenecen la mayor parte de los verbos, está formada por una serie de paradigmas de formas flexivas para los distintos tiempos y modos que se adjuntan a los temas de presente, pretérito y futuro del verbo. Corresponden a tres modelos, que se identifican tradicionalmente por el timbre de la vocal temática del infinitivo: *-ar* (VT = *-a-*, *amar*), *-er* (VT = *-e-*, *temer*) e *-ir* (VT = *-i-*, *partir*) (RAE, 2010, p. 54).

# 1 AMAR

VERBO MODELO DE LA PRIMERA CONJUGACIÓN

## TIEMPOS SIMPLES

### FORMAS NO PERSONALES

Infinitivo	Participio	Gerundio
amar	amado	amando

### Indicativo

NÚMERO	PERSONAS DEL DISCURSO	PRONOMBRES PERSONALES	Presente	Pret. imperfecto / Copretérito	Pret. perfecto simple / Pretérito
Singular	1. <sup>a</sup>	yo	amo	amaba	amé
	2. <sup>a</sup>	tú/vos	amas / amás	amabas	amaste
		usted	ama	amaba	amó
Plural	3. <sup>a</sup>	él, ella	ama	amaba	amó
	1. <sup>a</sup>	nosotros, -tras	amamos	amábamos	amamos
	2. <sup>a</sup>	vosotros, -tras	amáis	amabais	amasteis
ustedes		aman	amaban	amaron	
Plural	3. <sup>a</sup>	ellos, ellas	aman	amaban	amaron

			Futuro simple / Futuro	Condicional simple / Pospretérito
Singular	1. <sup>a</sup>	yo	amaré	amaría
	2. <sup>a</sup>	tú/vos	amarás	amarías
		usted	amará	amaría
Plural	3. <sup>a</sup>	él, ella	amará	amaría
	1. <sup>a</sup>	nosotros, -tras	amaremos	amaríamos
	2. <sup>a</sup>	vosotros, -tras	amaréis	amaríais
ustedes		amarán	amarían	
Plural	3. <sup>a</sup>	ellos, ellas	amarán	amarían

### Subjuntivo

NÚMERO	PERSONAS DEL DISCURSO	PRONOMBRES PERSONALES	Presente	Pret. imperfecto / Pretérito	Futuro simple / Futuro
Singular	1. <sup>a</sup>	yo	ame	amara o amase	amare
	2. <sup>a</sup>	tú/vos	ames	amaras o amases	amares
		usted	ame	amara o amase	amare
Plural	3. <sup>a</sup>	él, ella	ame	amara o amase	amare
	1. <sup>a</sup>	nosotros, -tras	amemos	amáramos o amásemos	amáremos
	2. <sup>a</sup>	vosotros, -tras	améis	amarais o amaseis	amareis
ustedes		amen	amaran o amasen	amaren	
Plural	3. <sup>a</sup>	ellos, ellas	amen	amaran o amasen	amaren

### Imperativo

Singular	2. <sup>a</sup>	tú/vos	ama / amá	Plural	2. <sup>a</sup>	vosotros, -tras	amad
		usted	ame			ustedes	amen

Tabla 4. Modelo de conjugación *-ar* con verbo *amar* (RAE, 2009).

## 2 TEMER

VERBO MODELO DE LA SEGUNDA CONJUGACIÓN

### TIEMPOS SIMPLES

#### FORMAS NO PERSONALES

Infinitivo	Participio	Gerundio
temer	temido	temiendo

#### Indicativo

NÚMERO	PERSONAS DEL DISCURSO	PRONOMBRES PERSONALES	Presente	Pret. imperfecto / Copretérito	Pret. perfecto simple / Pretérito
Singular	1. <sup>a</sup>	yo	temo	temía	temí
	2. <sup>a</sup>	tú/vos	temes / temés	temías	temiste
		usted	teme	temía	temió
Plural	1. <sup>a</sup>	nosotros, -tras	tememos	temíamos	temimos
	2. <sup>a</sup>	vosotros, -tras	teméis	temíais	temisteis
		ustedes	temen	temían	temieron
			Futuro simple / Futuro	Condicional simple / Pospretérito	
Singular	1. <sup>a</sup>	yo	temeré	temería	
	2. <sup>a</sup>	tú/vos	temerás	temerías	
		usted	temerá	temería	
Plural	1. <sup>a</sup>	nosotros, -tras	temeremos	temeríamos	
	2. <sup>a</sup>	vosotros, -tras	temeréis	temeríais	
		ustedes	temerán	temerían	

#### Subjuntivo

NÚMERO	PERSONAS DEL DISCURSO	PRONOMBRES PERSONALES	Presente	Pret. imperfecto / Pretérito	Futuro simple / Futuro
Singular	1. <sup>a</sup>	yo	tema	temiera o temiese	temiere
	2. <sup>a</sup>	tú/vos	temas	temieras o temieses	temieres
		usted	tema	temiera o temiese	temiere
Plural	1. <sup>a</sup>	nosotros, -tras	temamos	temiéramos o temiésemos	temiéremos
	2. <sup>a</sup>	vosotros, -tras	temáis	temierais o temieseis	temiereis
		ustedes	teman	temieran o temiesen	temieren

#### Imperativo

Singular	2. <sup>a</sup>	tú/vos	teme / temé	Plural	2. <sup>a</sup>	vosotros, -tras	temed
		usted	tema			ustedes	teman

Tabla 5. Modelo de conjugación *-er* con verbo *temer* (RAE, 2009).

### 3 PARTIR

VERBO MODELO DE LA TERCERA CONJUGACIÓN

### TIEMPOS SIMPLES

#### FORMAS NO PERSONALES

Infinitivo	Participio	Gerundio
partir	partido	partiendo

#### Indicativo

NÚMERO	PERSONAS DEL DISCURSO	PRONOMBRES PERSONALES	Presente	Pret. imperfecto / Copretérito	Pret. perfecto simple / Pretérito
Singular	1. <sup>a</sup>	yo	parto	partía	partí
	2. <sup>a</sup>	tú/vos	partes / partís	partías	partiste
		usted	parte	partía	partió
3. <sup>a</sup>	él, ella				
Plural	1. <sup>a</sup>	nosotros, -tras	partimos	partíamos	partimos
	2. <sup>a</sup>	vosotros, -tras	partís	partíais	partisteis
		ustedes	parten	partían	partieron
3. <sup>a</sup>	ellos, ellas				
			Futuro simple / Futuro	Condicional simple / Pospretérito	
Singular	1. <sup>a</sup>	yo	partiré	partiría	
	2. <sup>a</sup>	tú/vos	partirás	partirías	
		usted	partirá	partiría	
3. <sup>a</sup>	él, ella				
Plural	1. <sup>a</sup>	nosotros, -tras	partiremos	partiríamos	
	2. <sup>a</sup>	vosotros, -tras	partiréis	partiríais	
		ustedes	partirán	partirían	
3. <sup>a</sup>	ellos, ellas				

#### Subjuntivo

NÚMERO	PERSONAS DEL DISCURSO	PRONOMBRES PERSONALES	Presente	Pret. imperfecto / Pretérito	Futuro simple / Futuro
Singular	1. <sup>a</sup>	yo	parta	partiera o partiese	partiere
	2. <sup>a</sup>	tú/vos	partas	partieras o partieses	partieres
		usted	parta	partiera o partiese	partiere
3. <sup>a</sup>	él, ella				
Plural	1. <sup>a</sup>	nosotros, -tras	partamos	partiéramos o partiésemos	partiéremos
	2. <sup>a</sup>	vosotros, -tras	partáis	partierais o partieseis	partiereis
		ustedes	partan	partieran o partiesen	partieren
3. <sup>a</sup>	ellos, ellas				

#### Imperativo

Singular	2. <sup>a</sup>	tú/vos	parte / partí	Plural	2. <sup>a</sup>	vosotros, -tras	partid
		usted	parta			ustedes	partan

Tabla 6. Modelo de conjugación *-ir* con verbo *partir* (RAE, 2009).

La conjugación irregular, por su parte, corresponde a aquellos verbos “cuya conjugación no se ajusta a los paradigmas de *amar, temer, partir*” (RAE, 2010, p. 57), es decir, aquellos en que los cambios generados por la acentuación no son regulares, o bien, aquellos en que la raíz sufre alteraciones en sus diferentes conjugaciones.

Debido a que, como se señala anteriormente, la mayor parte de los verbos del español pertenecen a la conjugación regular y a que en ella se encuentra el modelo más productivo en términos de los mecanismos de formación de palabras, esta investigación se concentra en los verbos regulares:

Aproximadamente el 90% de los verbos españoles pertenecen a la primera conjugación. Este es el paradigma que presenta una mayor proporción de verbos regulares, y el único modelo productivo, pues a él se ajustan casi todos los verbos que se crean mediante procesos de derivación (con los sufijos *-ar, -ear, -izar, -ificar*) y de parasíntesis (§ 8.2.3b). En cambio, es muy pequeño el conjunto de verbos regulares de la segunda y de la tercera conjugación (RAE, 2010, p. 55).

### 2.5.1. Participios

Particularmente, en cuanto al procedimiento de detección automática de verbos neológicos que se propone este estudio, los participios suponen cierta dificultad en tanto que comparten una serie de características morfológicas y sintácticas con los adjetivos: “Además de la flexión de género y número, muchos participios comparten con los adjetivos un buen número de estructuras sintácticas por el hecho de ser ambos elementos predicativos” (RAE, 2010, p. 524). Así, sus diferencias son observables, según la NGLÉ, a nivel del contexto en que son utilizados. Ejemplo de estas similitudes que dificultan la detección son *a)* y *b)*, extraídos de RAE (2010): en *a)* *elevado* es participio del verbo *elevantar*, mientras que en *b)* es adjetivo.

*a) La carga fue elevada al cuarto piso.*

*b) Pagó un precio muy elevado.*

Evidentemente, el contexto desambigua la categoría de *elevado* en uno y otro caso. Sin embargo, cuando se considera solo la unidad léxica aislada del contexto, la detección se torna

más compleja. Además, en ciertos contextos una misma unidad puede funcionar como participio o adjetivo indistintamente, como *situada* en c)

c) *Vivían en una cabaña situada en la playa*

### **3. Marco metodológico**

En este apartado se expondrá la metodología aplicada para la realización de la investigación. En primer lugar, se explicará el tipo de investigación (3.1); en segundo lugar, se presentarán las preguntas de investigación (3.2); en tercer lugar, se presentará el objetivo general (3.3.1) y los objetivos específicos (3.3.2); en cuarto lugar, en la sección de materiales y métodos, se detallarán los corpus utilizados (3.4.1) y las fases metodológicas del estudio (3.4.2); y, por último, los criterios de análisis (3.5).

#### **3.1. Tipo de investigación**

La estrategia general que estructura la presente investigación está basada en lo planteado por Yuni y Urbano (2014) en cuanto al tipo de estudio y enfoque de investigación. En relación con el tipo de estudio, este es descriptivo en tanto se propone monitorear y describir el estado actual del léxico, particularmente verbos. Evidentemente, tal descripción surge de un análisis de corpus, donde este último funciona como una muestra representativa, mas no exhaustiva, de los verbos que utilizan los hablantes en la actualidad. Al mismo tiempo, se pretende describir una metodología apropiada para la detección automática de estas unidades a partir de su estructura morfológica, independiente de los diccionarios integrados a otras herramientas que se proponen la misma tarea, lo que hace más efectivo el reconocimiento de unidades neológicas. En relación con el enfoque de la investigación, esta posee un enfoque o estrategia metodológica de tipo mixto puesto que, por un lado, los datos se obtienen y analizan en términos numéricos y estadísticos, y por el otro se analizan variables cualitativas como los verbos neológicos detectados.

#### **3.2. Preguntas de investigación**

Las preguntas que guían la presente investigación son las siguientes:

- a) ¿Qué verbos se utilizan actualmente en castellano, detectables en un corpus general sincrónico?
- b) ¿Cómo podrían detectarse automáticamente estos verbos en el corpus?
- c) ¿Qué verbos de los detectados son neologismos?

### **3.3. Objetivos de investigación**

#### **3.3.1. Objetivo general**

Identificar los verbos del castellano utilizados actualmente por los hablantes, que hayan quedado registrados en un corpus general sincrónico del español.

#### **3.3.2. Objetivos específicos**

- a) Reconocer elementos que conforman la estructura morfológica de los verbos del castellano: raíces y desinencias.
- b) Detectar verbos en un listado de formas del castellano, extraído de un corpus a través de un procedimiento automático.
- c) Identificar verbos y neologismos verbales vigentes en el castellano actual de dicho corpus.

### **3.4. Materiales y métodos**

#### **3.4.1. Materiales**

En esta investigación se trabajó con dos tipos de materiales: el corpus general sincrónico del español EsTenTen (Kilgarriff y Renau, 2013) y un listado de exclusión correspondiente a los lemas verbales del *Diccionario de la lengua española*, DLE (RAE, 2014). El corpus se empleó para la detección de los verbos en fuentes de uso real y actual, y el listado de exclusión para saber cuáles de ellos aún no han sido incorporados al diccionario y, por tanto, es probable que puedan considerarse neologismos.

Con respecto al EsTenTen, se utilizó en concreto la versión Spanish Web 2011 (esTenTen11, Eu + Am). Se trata de un corpus que contiene unos 10.000 millones de palabras correspondientes a textos recopilados de internet y representa una fuente de datos empíricos

respecto del léxico utilizado por los hablantes del español. Sin embargo, con el objetivo de acotar el estudio, en la presente investigación solo se trabajó con la parte peninsular del corpus, que consta de 2.500 millones de palabras. Asimismo, para tales efectos, se extrajeron de este corpus todas las formas cuya frecuencia era igual o superior a 5. El material concreto de trabajo fue un listado de formas o *types* provenientes de este corpus.

Por otra parte, el listado de exclusión procedente del DLE se extrajo a través de la plataforma EnclaveRAE, ofrecida por la Real Academia Española para facilitar a los usuarios el acceso a los recursos lingüísticos de la institución (<https://enclave.rae.es>). En concreto, se extrajeron todos los artículos del diccionario correspondientes a verbos, y dicha información se limpió para obtener solamente los lemas. Como resultado, se obtuvo un listado de 11.893 lemas verbales.

### **3.4.2. Fases metodológicas de la investigación**

La investigación se llevó a cabo en las fases que se detallan a continuación.

#### **3.4.2.1. Conformación de un listado de las desinencias verbales del español**

Se hizo una lista de todas las desinencias de las conjugaciones de verbos regulares del español con terminación en *-ar*, *-er* e *-ir*, junto con las de sus formas no personales. Para ello, se utilizaron las Tablas 4, 5 y 6 de los modelos de conjugación extraídas de la *Nueva gramática de la lengua española* (RAE, 2009) para identificar las raíces de cada verbo modelo (*amar*, *temer* y *partir*), que se reemplazaron por un espacio vacío, obteniendo así solo las desinencias de cada conjugación. De estas se seleccionaron solo aquellas que constaban de tres letras o más, ya que las de menor cantidad de letras podían interferir en el proceso de detección automática por su coincidencia con las terminaciones de palabras pertenecientes a otras clases gramaticales. Además, se incorporaron pronombres enclíticos para asegurar la detección de verbos que los contuvieran. En total, el listado estaba compuesto por 107 elementos, que muestra la Tabla 7.

<b>Desinencias formas personales y no personales de verbos regulares</b>					
ar	arais	aríamos	Eré	iendo	ieses
er	áramos	arían	eréis	iera	imos
ir	aran	arías	eremos	ierais	irá
aba	arán	aron	Ería	iéramos	irán
abais	aras	ase	eríais	ieran	irás
ábamos	arás	aseis	eríamos	ieras	iré
aban	are	ásemos	erían	iere	iréis
abas	aré	asen	erías	iereis	iremos
ada	areis	ases	Íais	iéremos	iría
adas	aréis	aste	íamos	ieren	iríais
ado	aremos	asteis	Ían	ieres	iríamos
ados	áremos	éis	Ías	ieron	irían
áis	aren	emos	Ida	iese	irías
amos	ares	erá	Idas	ieseis	iste
ara	aría	erán	Ido	iésemos	isteis
ará	aríais	erás	Idos	iesen	ndo
<b>Pronombres enclíticos</b>					
me	te	se	nos	le	
les	la	lo	las	los	

**Tabla 7.** Listado de desinencias verbales regulares y pronombres enclíticos.

### 3.4.2.2. Conformación de un listado de formas

A partir del corpus EsTenTen se extrajeron automáticamente todas las formas que aparecían 5 o más veces en este a través de Jaguar (Nazar, Vivaldi y Cabré, 2008), que permite trabajar con grandes volúmenes de datos. Se consideró una frecuencia mínima de 5 por considerarse esta una cantidad suficiente (aunque ciertamente arbitraria) para determinar que la ocurrencia de cada forma no era accidental. Cabe recordar que una forma corresponde a cualquier segmento de letras separado por un espacio en el corpus, de manera que constituye una posible palabra y, por ende, un posible verbo.

Asimismo, cada forma estaba asociada a un número, correspondiente a su frecuencia de aparición. El listado contaba con un total de 1.054.411 formas ordenadas alfabéticamente y cada una en una línea para su posterior procesamiento.

### 3.4.2.3. Procesamiento del listado de formas

Se creó un script en código Perl para la realización de los siguientes pasos:

- a) Recorrer el listado de formas línea por línea comprobando si cada una de ellas contiene una de las desinencias que se recogieron según se indicó en la fase 1. Cuando ocurre esta coincidencia entre parte de una forma y una desinencia almacena esa información por separado (raíz/desinencia); por ejemplo, la raíz *rape-* y la desinencia *-aba* correspondientes a la forma del corpus *rapeaba*. Si el script recorre la lista y encuentra más formas con la misma raíz, combinada con, al menos, 4 desinencias diferentes, deduce que es un verbo; por ejemplo, *rape-aba*, *rape-aban*, *rape-ada* y *rape-adas*. Esta cantidad de desinencias es tal porque se asume que para que un verbo pueda ser considerado vigente debe ser usado en varias de sus formas conjugadas. Dentro de estas condiciones, el número 4 como límite es por supuesto arbitrario. El resultado de este paso es una lista de raíces asociadas con sus respectivas desinencias (Tabla 8).

Raíz	Desinencias encontradas
rape	aba, aban, ada, adas, ado, ados
reten	emos, ida, ido, idos, iendo, éis, ían
suprim	amos, ida, idas, ido, idos, iendo, iera, ieran, ieron, iese, iesen, irán, iré, iría, ían

**Tabla 8.** Ejemplos de resultado de la detección de posibles verbos.

- b) Reconstruir el infinitivo de cada raíz encontrada, comprobando si cada una de ellas, al ser combinada con una terminación *-ar*, *-er* o *-ir*, calza con alguna de las formas del corpus. Esto se realiza con miras a obtener como resultado final una lista de lemas y no de formas conjugadas, lo que implica que las formas cuyo infinitivo no fue encontrado en el corpus son descartadas. Sin embargo, esto no significa que todas las formas cuyo infinitivo sí fue encontrado en el corpus son efectivamente verbos, ya que cabe la posibilidad de encontrar errores, sobre todo ortográficos.

En el caso del ejemplo anterior, se probaría la raíz *reten* con *-ar*: *\*retenar*; con *-er*: *retener*; y con *-ir*: *\*retenir*, de las cuales selecciona *retener* por corresponderse con una de las formas del corpus. Si en dado caso una misma raíz coincidiera con más de una terminación, entonces el script decide quedarse con la más frecuente; por ejemplo, si encontrara tanto *retener* como *\*retenir* por algún error ortográfico en el corpus, decide quedarse con la más frecuente que, probablemente, es la forma correcta: *retener*. El resultado de esta acción es una lista de infinitivos o lemas con su respectiva frecuencia, que se corresponden con las raíces encontradas (Tabla 9).

Lema	FrecLema	Raíz	Desinencias
rapear	206	rape	aba, aban, ada, adas, ado, ados
retener	26390	reten	emos, ida, ido, idos, iendo, éis, ían
suprimir	32436	suprim	amos, ida, idas, ido, idos, iendo, iera, ieran, ieron, iese, iesen, irán, iré, iría, ían

**Tabla 9.** Ejemplos de resultados de la reconstrucción del infinitivo de cada verbo.

- c) Identificar los verbos que solo son utilizados como participios a través de las desinencias características de esta forma no personal de los verbos (-*ido*, -*idos*, -*ida*, -*idas*, -*ado*, -*ados*, -*ada*, -*adas*). El script guarda la información referente a los verbos cuya raíz se encontró asociada solo a desinencias de participio. Es decir, si entre dichas desinencias se encontraba al menos una desinencia correspondiente a una forma personal del verbo, entonces no se consideraba propiamente un participio, como se ejemplifica en la Tabla 10, en la que el verbo *aceitar* tiene como desinencias de participio -*ada*, -*adas*, -*ado*, -*ados*, pero también -*amos*. Esta información fue recogida para trabajo futuro con el fin de monitorear, paralelamente, el proceso en que un verbo deja de ser utilizado en sus formas personales y pasa a ser utilizado únicamente cumpliendo la función de adjetivo. El resultado obtenido de esta acción es un registro binario en el que todo verbo que presenta solo desinencias de participio se encuentra etiquetado con un 1 y, de lo contrario, con un 0.

Lema	FrecLema	Raíz	Participio	Desinencias
abotonar	24	aboton	1	ada, adas, ado, ados
aceitar	8261	aceit	0	ada, adas, ado, ados, amos

**Tabla 10.** Ejemplos de resultados de la identificación de participios.

#### 3.4.2.4. Comparación con fuente lexicográfica

Una vez reconstruidos los lemas, se procede a verificar, por medio del mismo script, si los lemas encontrados se encuentran registrados en el DLE, con el fin de saber si son neológicos o no siguiendo el criterio lexicográfico como indicador de neologicidad (v. apartado 2.4 del marco teórico). El script cuenta tanto los verbos encontrados como los no encontrados en la fuente lexicográfica. Al final, se obtiene un registro binario en el que todo lema que fue hallado en el diccionario está etiquetado con un 1 o, de lo contrario, con un 0 (Tabla 11). Como se pueda observar en la Tabla 11, *retener* y *suprimir* están marcados con un 1 en la columna *DLE*, mientras que *rapear* con un 0 porque es un neologismo.

Lema	FrecLema	Raíz	DLE	Participio	Desinencias
rapear	206	rape	0	0	aba, aban, ada, adas, ado, ados

retener	26390	reten	1	0	emos, ida, ido, idos, iendo, éis, ían
suprimir	32436	suprim	1	0	amos, ida, idas, ido, idos, iendo, iera, ieran, ieron, iese, iesen, irán, iré, iría, ían

**Tabla 11.** Ejemplos de resultados de la comparación del leuario extraídos del corpus con el del DLE.

Al término de la ejecución del script, se obtienen los datos cuantitativos correspondientes a i) la cantidad de verbos encontrados en el corpus, ii) la cantidad de verbos hallados en el corpus que están registrados en el DLE, iii) la cantidad de verbos hallados en el corpus que no se encuentran registrados en el DLE y que, por tanto, son posibles neologismos y iv) la cantidad de verbos hallados en el corpus que solo son utilizados en participio.

#### **3.4.2.5. Realización de prueba de confiabilidad**

Una vez obtenidos los verbos por el procedimiento metodológico detallado anteriormente se realizó una prueba de confiabilidad del instrumento con el objetivo de evaluar su cobertura. Esta prueba consistió en comparar directamente ambas fuentes, específicamente el listado de formas del EsTenTen y el listado de lemas del DLE, para determinar qué cantidad de verbos del DLE se encontraban en el EsTenTen y, a partir de ello, medir la proporción detectada por el script.

#### **3.4.2.6. Análisis de verbos no encontrados en el DLE**

En una planilla Excel se filtraron los verbos hallados en el corpus que no se encontraban registrados en el DLE y que, por ende, se consideraban como posibles neologismos. Estos

fueron analizados manualmente para determinar si los verbos eran efectivamente neológicos o si correspondían a errores ortográficos o fallas del procedimiento (Tabla 12). Para ello se siguieron dos pasos:

- a) Se analizó uno por uno cada caso y se marcaron con un 1, en la columna de “Neologismo”, aquellas formas que efectivamente eran neologismos. Se consideraron neologismos aquellas palabras en las que se reconocía una base léxica y su respectivo procedimiento de derivación a partir de afijos. Por el contrario, se marcaron con un 0 las formas no neológicas.
- b) Se clasificaron todos los casos de acuerdo a sus respectivas tipologías: los neologismos (marcados con 1) se etiquetaron en la columna “Tipo de acuerdo” a la clasificación de neologismos de OBNEO (2004), que se muestra en la Tabla 13; mientras que las formas no neológicas (marcadas con 0) fueron etiquetadas como errores ortográficos, fallas del método o desconocido según correspondiera.

Lema	Frec Lema	Raíz	DLE	Participio	Desinencias	Neologismo	Tipo
ablar	1090	abl	0	0	aba, aban, ada, ado, amos, ara, aran, are, aron, emos	0	ort
suicidar	19657	suicid	0	0	aba, aban, ada, adas, ados, amos, ara, aran, aras, ará, arán, aré, aría, ase, asen, emos	0	falla

aflamencar	14	aflamenc	0	0	ada, adas, ado, ados, aron	1	FPRSU
------------	----	----------	---	---	----------------------------------	---	-------

**Tabla 12.** Ejemplo de análisis de posibles neologismos.

### 3.5. Criterios de análisis

Los datos obtenidos respecto de los verbos detectados automáticamente en el corpus EsTenTen fueron analizados en virtud de tres criterios. En primer lugar, se utilizó el criterio lexicográfico para determinar si un verbo podía o no ser considerado un neologismo. En este sentido, todo aquel verbo que se encontraba registrado como una entrada en el corpus de exclusión se consideraba parte del léxico estable del español; mientras que todo aquel verbo que no se encontraba registrado en el corpus de exclusión se consideraba un candidato a neologismo.

En segundo lugar, los datos que pasaban a ser considerados neologismos fueron analizados siguiendo los tipos de neologismos extraídos del protocolo de vaciado de textos escritos de OBNEO (2004), cuya codificación se muestra en la Tabla 13. Particularmente, y como se indica anteriormente (2.3), se esperaba encontrar, por una parte, neologismos de forma (F), específicamente por sufijación (FSUF), prefijación (FPRE), interferencias entre sufijación y prefijación (FPRSU) y, por otra parte, préstamos adaptados (AM).

<b>Neologismo de forma</b>	<b>F</b>
Sufijación	FSUF
Prefijación	FPRE
Interferencias entre sufijación y prefijación	FPRSU
Composición	FCOM
Composición culta	FCULT

Lexicalización	FLEX
Conversión sintáctica	FCONV
Sintagmación	FSINT
Siglación	FTSIG
Acronimia	FTACR
Abreviación	FTABR
Variación	FVAR
<b>Neologismo sintáctico</b>	<b>SINT</b>
<b>Neologismo semántico</b>	<b>S</b>
<b>Préstamo</b>	<b>M</b>
Préstamo adaptado	AM
<b>Otros</b>	<b>A</b>

**Tabla 13.** Codificación de neologismos (OBNEO, 2004).

Cabe destacar que los datos que no se consideraban neologismos y que, por ende, no se ceñían a esta clasificación, fueron etiquetados como errores ortográficos con el código *ort*, como en el caso de *\*contruir*, correspondiente al verbo *construir* o *\*ablar* correspondiente al verbo *hablar*; fallas del método con el código *falla*, como en el caso de *adormilar*, donde no pudo ser reconstruido el lema en su forma pronominal *adormilarse*; o casos desconocidos con el código *desc*, como *\*íar* o *\*agar*, donde no es posible reconocer un verbo, según correspondiera. En el primer caso, se consideraron verbos escritos incorrectamente ya sea por falta de alguna letra, omisión de tildes, errores de tipeo, etc. En el segundo caso, se consideraron verbos que erróneamente fueron etiquetados como posibles neologismos, es decir, como inexistentes en el corpus de exclusión por una falla en la reconstrucción del

infinitivo. En última instancia, se etiquetaron como desconocidos todos aquellos casos donde no era posible determinar si el error era producto de un error ortográfico o de una falla del método.

Por último, como se señala anteriormente, se recogió información respecto de la presencia de participios en el conjunto de verbos detectados por el script. El criterio para determinar si el verbo detectado era o no un participio fue la naturaleza de las desinencias en las que se encontraba conjugado. Es decir, todo aquel verbo que solo se conjugara con las desinencias *-ido, -idos, -ida, -idas, -ado, -ados, -ada, -adas* fue considerado un participio; mientras que la sola presencia de una desinencia diferente a estas permitía dejar de considerarlo un participio.

#### 4. Resultados y discusión

En este apartado se exponen los resultados obtenidos por medio de la aplicación del método propuesto y su respectiva interpretación. En primer lugar, se presentan los resultados generales (4.1). En segundo lugar, se exponen los resultados de la prueba de confiabilidad que valida el método propuesto (4.1.1). Luego, se presentan los resultados referentes tanto a los verbos no neológicos (4.2) como neológicos (4.3), donde se detallan los diferentes tipos de neologismos encontrados (4.3.1) para esbozar, por último, los resultados referentes a los participios encontrados en el corpus de estudio (4.4).

##### 4.1. Resultados generales

A partir de la aplicación del script, habiendo 1.054.411 formas del EsTenTen, se detectaron, en primer lugar, 8.547 (0.81%) formas que, por su estructura morfológica, se presentaban como candidatas a verbos, es decir, posibles raíces que se encontraban asociadas a 4 o más desinencias. En segundo lugar, de esas 8.547 formas que posiblemente eran verbos, se identificaron 6.025, es decir un 70.49% del total de posibles verbos y un 0.57% del total de formas del corpus, que se consideraron como verbos por haber sido encontrada su forma en infinitivo en el corpus. Por el contrario, 2.522 formas, un 29.50% del total de posibles verbos, lo que representa un 0.23% del total de formas del corpus, no clasificaron como verbos por no encontrarse su forma en infinitivo en el corpus. Estos resultados se muestran en la Tabla 14. Cabe recordar que la importancia de la reconstrucción del infinitivo radica en la necesidad de conformar un listado final de verbos utilizados en español, presentados como infinitivos o lemas, por lo que las 2.522 formas cuyo infinitivo no pudo ser reconstruido son descartadas para el listado final de verbos (Anexo 1).

Posibles verbos	Reconstrucción del infinitivo	%
8.547	6.025 con infinitivo	70.49
	2.522 sin infinitivo	29.50

**Tabla 14.** Verbos detectados en el corpus.

Luego, del total de 6.025 verbos detectados en el EsTenTen (cuyo infinitivo sí pudo ser reconstruido), 4.415, es decir, un 73.27% se encontraban registrados en el corpus de exclusión (DLE), lo que los clasifica como verbos no neológicos del español. Por el contrario, 1.610, un 26.72%, no se encontraron en el DLE, por lo que se postulaban como posibles neologismos verbales del español (Tabla 15).

<b>Verbos detectados</b>	<b>Comparación con DLE</b>	<b>%</b>
6.025	4.415 registrados en el DLE	73.27
	1.610 no registrados en el DLE	26.72

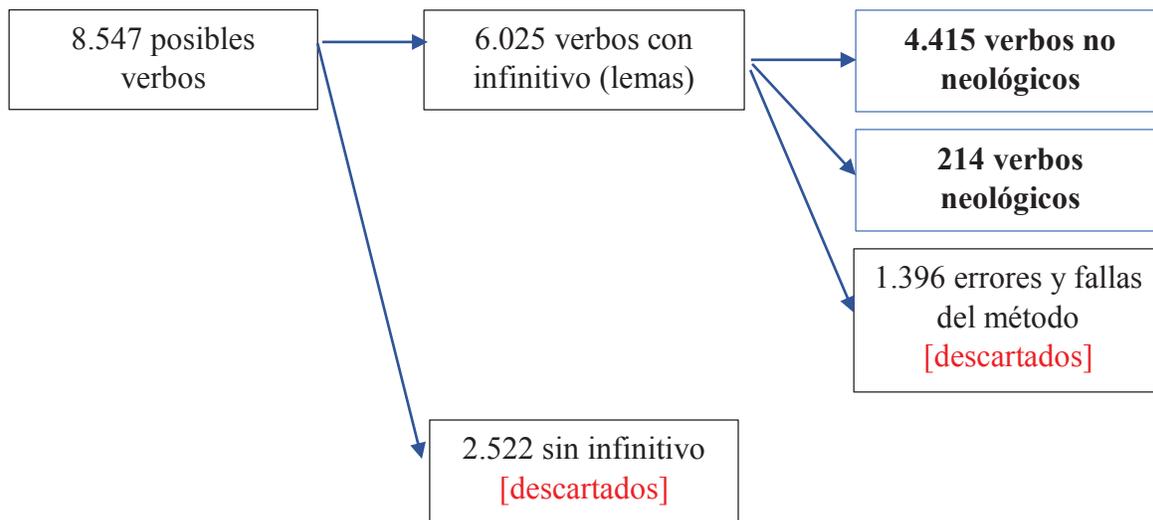
**Tabla 15.** Resultados de la comparación con DLE.

A su vez, de esos 1.610 posibles neologismos, por medio del análisis manual, se detectaron 214 verbos efectivamente neológicos, es decir, un 13.29% del total de posibles neologismos y un 3.55% del total de verbos detectados; mientras que los otros 1.396 posibles neologismos, es decir, un 86.27% del total de posibles neologismos y un 23.17% del total de verbos detectados, corresponde a diversos errores y fallas del método (Tabla 16), que, evidentemente, también fueron descartados para efectos del listado final de verbos pues no constituyen propiamente verbos del español.

<b>Candidatos a neologismos</b>	<b>Análisis manual</b>	<b>%</b>
1.610	214 neologismos	13.29
	1.396 errores ortográficos y fallas del método	86.70

**Tabla 16.** Resultados del análisis manual para la detección de neologismos.

En suma, el total de verbos detectados en el corpus de estudio, tanto neológicos como no neológicos, fue de 4.629 que constituyen el listado final de verbos (Anexo 5). La Figura 3 esquematiza los resultados generales.



**Figura 3.** Esquema de resultados generales.

Como se puede observar, del total de 4.629 verbos del español detectados a través del método propuesto, un 95.37% corresponde a verbos no neológicos, de acuerdo al criterio lexicográfico aplicado; mientras que el otro 4.62% corresponde a verbos neológicos.

Por último, del total de 8.547 posibles verbos detectados inicialmente, 1.339, es decir, un 15.66% eran participios, según los criterios de análisis aplicados (3.5). De ellos, 734, es decir, un 54.81% fue hallado en forma de infinitivo en el corpus; mientras que del otro 45.18% no fue posible reconstruir su infinitivo (Tabla 17). Como se señala anteriormente (3.4.2.3), la recolección de estos datos es paralela a la presente investigación y constituye un nicho de investigación para trabajo futuro.

Posibles verbos	Detección de participios	Reconstrucción de infinitivo	%
8.547	1.339	734 con infinitivo	54.81
		605 sin infinitivo	45.18

**Tabla 17.** Resultados detección de participios.

#### 4.1.1. Prueba de confiabilidad

Como se indica en la fase nº 4 de la investigación (v. apartado 3.4.2), se realizó una prueba de confiabilidad para evaluar la validez del método utilizado. A partir de esta prueba se pudo observar, por una parte, que, del total de 11.893 verbos registrados en el DLE, 6.219 se encuentran en el corpus EsTenTen, de los cuales el método propuesto fue capaz de detectar 4.415, es decir, un 70.99%. Esta cifra representa la cobertura del método propuesto en la detección de verbos no neológicos, la cual se considera alta y permite validarlo.

En otros términos, esta prueba de confiabilidad permitió observar la representatividad del corpus EsTenTen, particularmente de su variante peninsular, en el monitoreo de los verbos registrados en el DLE. Al respecto, se pudo determinar que un 47.70% de los verbos del DLE no se encuentran en el corpus de estudio utilizado, lo que representa una limitación en tanto cerca de la mitad de los verbos registrados en el diccionario no podrían ser detectados en el corpus, lo que no significa, necesariamente, que tal porcentaje de verbos no sean utilizados por los hablantes. Entre los verbos pertenecientes al DLE que no se encuentran registrados en el corpus EsTenTen están *murmujear*, *obtemperar* y *taucar*, cuya ausencia podría deberse, a simple vista, al escaso uso de estos verbos; no obstante, también se encuentran verbos como *hociconear* o *infraccionar* que, posiblemente, podrían hallarse en otro corpus más extenso.

#### 4.2. Verbos detectados en el corpus que se encuentran registrados en el DLE

Como se indica anteriormente, del total de verbos detectados en el corpus de estudio EsTenTen, 4.415 se encuentran registrados en el corpus de exclusión DLE, es decir, se trata de verbos no neológicos (Anexo 2). Esto significa que sobre el 95% de los verbos que, según los criterios aplicados en el presente estudio, se encuentran vigentes están integrados de manera estable en el caudal léxico del español. En este sentido, cobra relevancia la idea de

que el léxico, a pesar de ser objeto de cambios, goza de un alto grado de estabilidad que permite la pervivencia de las palabras a través del tiempo.

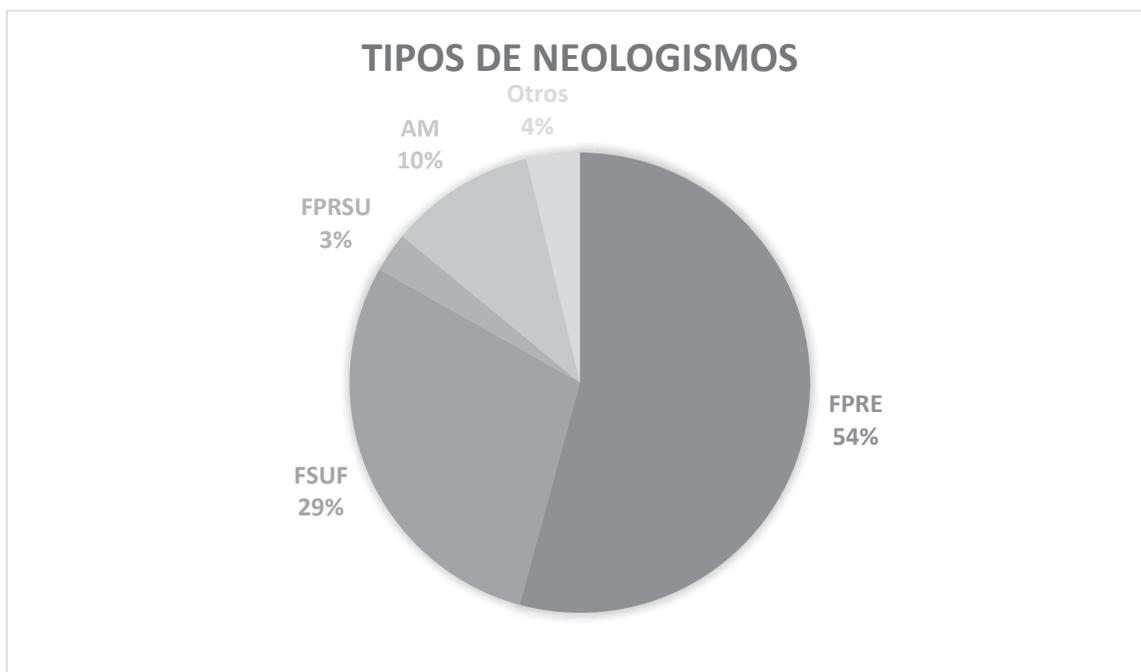
### 4.3. Verbos detectados en el corpus que no se encuentran registrados en el DLE

Como se menciona en la sección de resultados generales (4.1), del total de 1.610 posibles neologismos, 214 efectivamente lo eran, es decir, un 13,29% del total de candidatos; mientras que 1.396 correspondían a errores y fallas del método, es decir, el 86,70% (Tabla 16).

En términos generales, se puede determinar que la precisión del método en cuanto a la detección de verbos neológicos del español se encuentra mediada, como se expondrá posteriormente (apartado 4.3.2), por la alta proporción de errores ortográficos encontrados en el corpus de estudio.

#### 4.3.1. Neologismos

En total, se detectaron 214 verbos neológicos en el corpus de estudio, que representan un 4,62% del total de verbos que se consideran vigentes de acuerdo con los criterios aplicados y cuya clasificación se representa gráficamente en la Figura 4.



**Figura 4.** Tipos de neologismos encontrados.

En los siguientes subapartados se comentarán estos resultados.

#### 4.3.1.1. Derivación

##### 4.3.1.1.1. Prefijación (FPRE)

Entre los verbos neológicos detectados en el corpus de estudio, se detectaron 116 neologismos por prefijación, es decir, un 54% del total de verbos neológicos encontrados, lo que erige a este mecanismo como el más productivo. La Tabla 18 muestra estos neologismos. Entre ellos, se observan, por ejemplo, verbos como *desestructurar*, cuya base es *estructurar*, que significa “Articular, distribuir, ordenar las partes de un conjunto” (RAE, 2014, acep. 1) y que mediante la adhesión del prefijo *des-* que “Denota negación o inversión del significado de la palabra simple a la que va antepuesto” pasa a referir la acción de alterar el orden o la distribución de las partes de un conjunto.

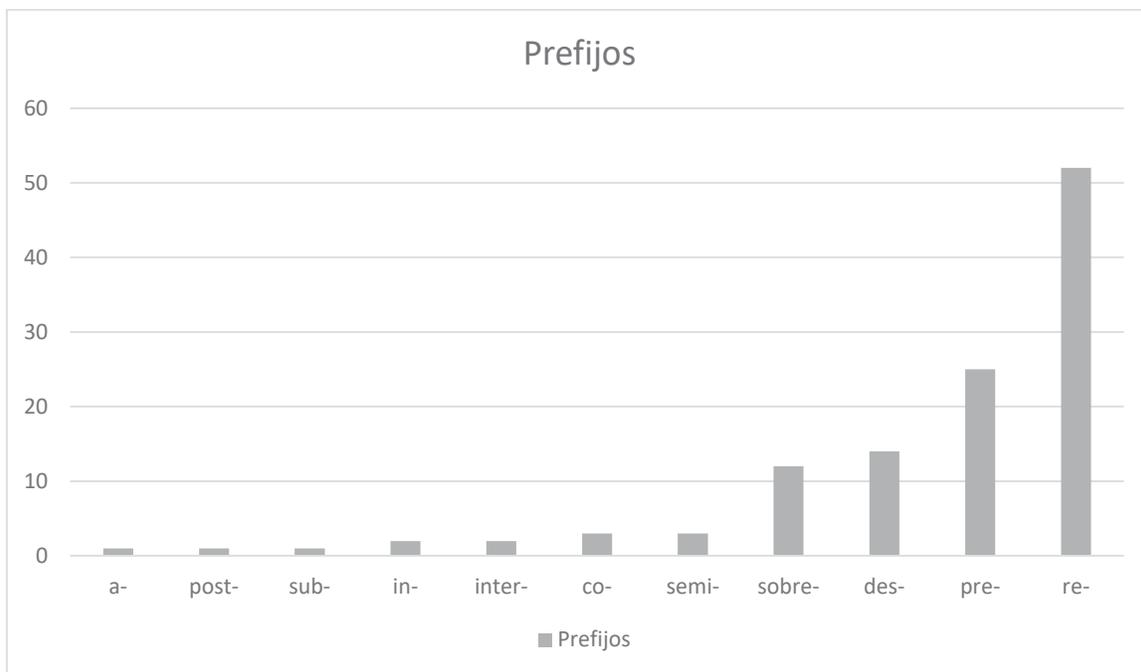
FPRE	Prefijo	Base	FPRE	Prefijo	Base
arreguntar	a-	reguntar	recodificar	re-	codificar
coorganizar	co-	organizar	recolocar	re-	colocar
coparticipar	co-	participar	recombinar	re-	combinar
copresidir	co-	presidir	recompilar	re-	compilar
desactualizar	des-	actualizar	reconfigurar	re-	configurar
desaturar	des-	saturar	reconfirmar	re-	confirmar
desconfigurar	des-	configurar	redenominar	re-	denominar
desensibilizar	des-	sensibilizar	redibujar	re-	dibujar
desespinar	des-	espinar	redimensionar	re-	dimensionar
desestacionalizar	des-	estacionalizar	redireccionar	re-	direccionar
desestructurar	des-	estructurar	redirigir	re-	dirigir
desfragmentar	des-	fragmentar	rediseñar	re-	diseñar
deshabilitar	des-	habilitar	reemitir	re-	emitir
desincronizar	des-	sincronizar	reequilibrar	re-	equilibrar

desmaterializar	des-	materializar	reescalar	re-	escalar
despigmentar	des-	pigmentar	reestablecer	re-	establecer
desprecintar	des-	precintar	reevaluar	re-	evaluar
desterritorializar	des-	territorializar	reexpresar	re-	expresar
inaplicar	in-	aplicar	refinanciar	re-	financiar
intercomunicar	in-	comunicar	reformular	re-	formular
interconexionar	inter-	conexionar	regrabar	re-	grabar
interrelacionar	inter-	relacionar	rehidratar	re-	hidratar
postproducir	post-	producir	reinaugurar	re-	inaugurar
preajustar	pre-	ajustar	reinstaurar	re-	instaurar
preanunciar	pre-	anunciar	reinterpretar	re-	interpretar
preasignar	pre-	asignar	reintroducir	re-	introducir
preavisar	pre-	avisar	reinventar	re-	inventar
precalcular	pre-	calcular	relocalizar	re-	localizar
precalentar	pre-	calentar	remezclar	re-	mezclar
precalificar	pre-	calificar	renegociar	re-	negociar
precargar	pre-	cargar	renumerar	re-	enumerar
preclasificar	pre-	clasificar	repavimentar	re-	pavimentar
precocer	pre-	cocer	reposicionar	re-	posicionar
precocinar	pre-	cocinar	reprocesar	re-	procesar
preconfigurar	pre-	configurar	reprogramar	re-	programar
prediseñar	pre-	diseñar	republicar	re-	publicar
preelaborar	pre-	elaborar	resignificar	re-	significar
preencolar	pre-	encolar	reubicar	re-	ubicar
preenvasar	pre-	envasar	reurbanizar	re-	urbanizar
preestablecer	pre-	establecer	reusar	re-	usar
preestrenar	pre-	estrenar	revegetar	re-	vegetar
prefabricar	pre-	fabricar	revisitar	re-	visitar
preinstalar	pre-	instalar	semienterrar	semi-	enterrar
premontar	pre-	montar	semimontar	semi-	montar

preordenar	pre-	ordenar	semiproteger	semi-	proteger
prepagar	pre-	pagar	sobredeterminar	sobre-	determinar
preprogramar	pre-	programar	sobreelevar	sobre-	elevar
previsualizar	pre-	visualizar	sobreescribir	sobre-	escribir
reactualizar	re-	actualizar	sobreestimar	sobre-	estimar
readaptar	re-	adaptar	sobreexpresar	sobre-	expresar
reaprovechar	re-	aprovechar	sobreimpresionar	sobre-	impresionar
reasfaltar	re-	asfaltar	sobremadurar	sobre-	madurar
reasignar	re-	asignar	sobreproteger	sobre-	proteger
recalcular	re-	calcular	sobrerrepresentar	sobre-	representar
recalibrar	re-	calibrar	sobreutilizar	sobre-	utilizar
recapturar	re-	capturar	sobrevaluar	sobre-	evaluar
recircular	re-	circular	sobreexplotar	sobre-	explotar
reclasificar	re-	clasificar	subvalorar	sub-	valorar

**Tabla 18.** Neologismos por prefijación encontrados en el corpus de estudio.

Como se puede observar, se encontraron verbos formados por 11 diferentes prefijos, entre los cuales se detectaron algunos más productivos, es decir, que permiten derivar mayor cantidad de neologismos a partir de palabras preexistentes (Figura 5). Entre ellos, los prefijos más productivos son *re-*, *pre-*, *des-* y *sobre-*, cuya frecuencia es de 52, 25, 14 y 12 respectivamente, del total de 116 neologismos por derivación. Evidentemente, el alto nivel de productividad de estos prefijos justifica que muchos de los verbos que los contienen, aunque no se consideren totalmente neológicos como en el caso de *precalentar* o *sobreproteger*, no se encuentren registrados en los diccionarios ya que esto supondría una tarea inabarcable.



**Figura 5.** Prefijos que forman parte de los neologismos encontrados, por orden ascendente de frecuencia.

Además, es posible advertir que la mayoría de estos verbos neológicos pertenecen a la primera conjugación (-ar), precisamente un 93.96%; mientras que solo un 3.44% y 2.58% corresponden a la segunda y tercera conjugación, respectivamente (-er e -ir), lo que demuestra el alto nivel de productividad de la primera conjugación que ya se anticipaba anteriormente (v. apartado 2.5).

#### 4.3.1.1.2. Sufijación (FSUF)

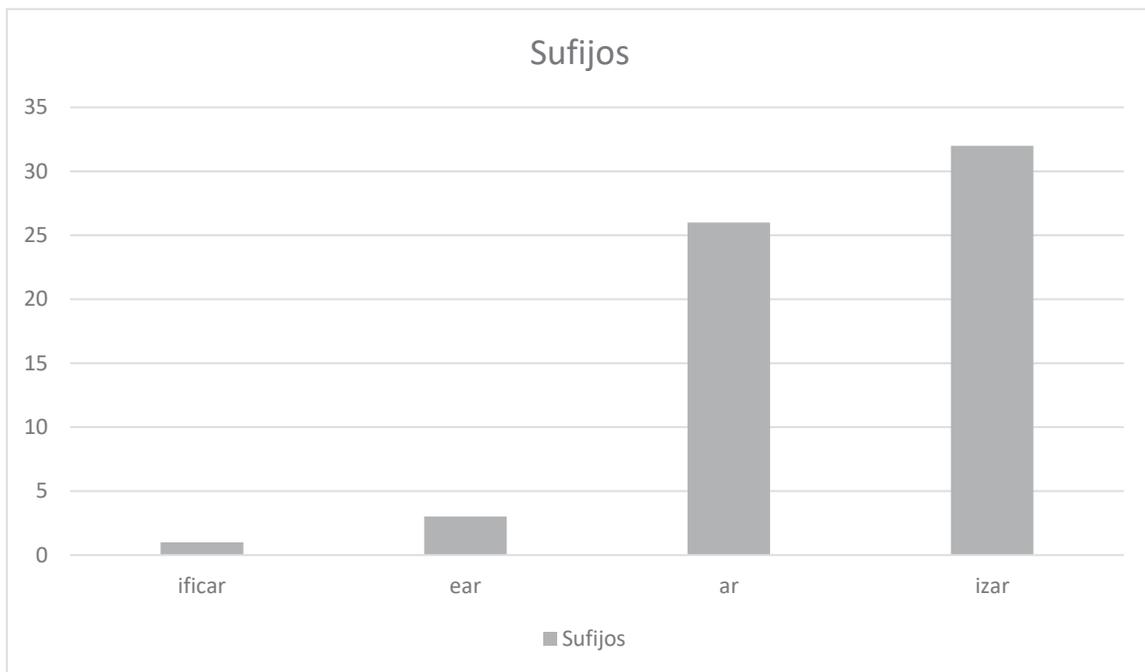
Entre los verbos neológicos detectados en el corpus de estudio, se detectaron 62 neologismos por sufijación, es decir, un 28.97% del total de verbos neológicos encontrados. La Tabla 19 muestra los neologismos por sufijación presentes en el corpus. Estos verbos provienen de un nombre como, por ejemplo, *agenda*, que mediante la adhesión del sufijo -ar refiere a la acción de establecer un evento en un horario en particular, o bien, de un adjetivo, como *virtual*, que mediante la adhesión del sufijo -izar refiere a la acción de dotar a algo de tal cualidad.

<b>FSUF</b>	<b>Base</b>	<b>Sufijo</b>	<b>FSUF</b>	<b>Base</b>	<b>Sufijo</b>
acetilar	acetil	-ar	monitorear	monitor	-ear
aditivar	aditivo	-ar	musealizar	museal	-izar
agendar	agenda	-ar	mutualizar	mutual	-izar
antologar	antología	-ar	operacionalizar	operacional	-izar
aperturar	apertura	-ar	operativizar	operativo	-izar
catastrar	catastro	-ar	paletizar	paleta	-izar
chocolatear	chocolate	-ear	paquetizar	paquete	-izar
conveniar	convenio	-ar	paralelizar	paralelo	-izar
cooperativizar	cooperativa	-izar	particionar	partición	-ar
discretizar	discreto	-izar	polimerizar	polímero	-izar
domotizar	domótica	-izar	provisionar	provisión	-ar
esencializar	esencia	-izar	punzonar	punzón	-ar
espacializar	espacial	-izar	referenciar	referencia	-ar
exaccionar	exacción	-ar	reificar	reificación	-ar
ficcionar	ficción	-ar	reversionar	reversión	-ar
ficcionalizar	ficcional	-izar	saborizar	sabor	-izar
fragilizar	frágil	-izar	sectorizar	sector	-izar
funcionalizar	funcional	-izar	securizar	seguro	-izar
gelatinizar	gelatina	-izar	serializar	serial	-izar
gelificar	gel	-ificar	siniestrar	siniestro	-ar
hegemonizar	hegemonía	-izar	subjetivar	subjetivo	-ar
hidrogenar	hidrógeno	-ar	suplementar	suplemento	-ar
infusionar	infusión	-ar	telonear	telón	-ear
intencionar	intención	-ar	tercerizar	tercero	-izar
internalizar	interno	-izar	termostatizar	termostato	-izar
invisibilizar	invisible	-izar	traccionar	tracción	-ar
linealizar	lineal	-izar	vectorizar	vector	-izar
mandatar	mandato	-ar	vehiculizar	vehículo	-izar

mantecar	manteca	-ar	virtualizar	virtual	-izar
marginalizar	marginal	-izar	vitaminar	vitamina	-ar
minorizar	menor	-izar	vivenciar	vivencia	-ar

**Tabla 19.** Neologismos por sufijación encontrados en el corpus de estudio.

Como se menciona en el apartado 2.3, los sufijos presentados corresponden a sufijos verbalizadores que, como su nombre lo indica, permiten la formación de verbos a partir de nombres o adjetivos. En este caso, como muestra la Figura 6, el sufijo más productivo es *izar-*, seguido de cerca por *ar-*. Además, como se señala anteriormente (2.5), en este tipo de verbos neológicos la conjugación más productiva es también la primera (*-ar*), no encontrándose ninguno correspondiente a la segunda o tercera conjugación (*-er* o *-ir*).



**Figura 6.** Frecuencia de sufijos.

#### 4.3.1.1.3. Parasíntesis (FPRSU)

Entre los verbos neológicos detectados en el corpus de estudio, se detectaron 6 neologismos por parasíntesis, es decir, un 3% del total de verbos neológicos encontrados, situándose como el mecanismo derivativo menos productivo. La Tabla 20 muestra los verbos parasintéticos encontrados en el corpus.

FPRSU	Prefijo	Base	Sufijo
adinerar	a-	dinero	-ar
aflamencar	a-	flamenco	-ar
asilvestrar	a-	silvestre	-ar
desvirtualizar	des-	virtual	-izar
enrevesar	en-	revés	-ar
enrutar	en-	ruta	-ar

**Tabla 20.** Neologismos por parasíntesis encontrados en el corpus de estudio.

Como se puede observar, se detectaron, por una parte, 4 verbos parasintéticos denominales: *adinerar*, *aflamencar*, *enrevesar* y *enrutar*; y, por otra parte, solo 2 verbos parasintéticos deadjetivales: *asilvestrar* y *desvirtualizar*. En los primeros, por ejemplo, *enrutar* proviene del sustantivo *ruta* y la adhesión del circunfijo *en...ar* lo convierte en un verbo que refiere a la acción de poner algo en una ruta o camino en particular. En el caso de los verbos parasintéticos deadjetivales, se presenta el verbo *asilvestrar* que proviene del adjetivo *silvestre* al que le fue adherido el circunfijo *a...ar* para denotar la acción de dotar a algo de la cualidad de silvestre.

Asimismo, en cuanto a los esquemas de formación, se puede advertir que los más frecuentes son aquellos que contienen el sufijo *-ar*: *a-N-ar* y *en-N-ar*. Por último, como en los casos anteriores, todos los casos corresponden a la primera conjugación (*-ar*).

#### 4.3.1.2. Préstamos adaptados (AM)

Entre los verbos neológicos detectados en el corpus de estudio, se detectaron 22 neologismos correspondientes a préstamos adaptados, es decir, un 10% del total de verbos neológicos encontrados. La Tabla 21 muestra los préstamos adaptados presentes en el corpus.

AM	Voz de origen	Tipo de adaptación
banear	Del inglés <i>ban</i> : ‘prohibir’	Sufijación a través de <i>-ear</i>
crackear	Del inglés <i>crack</i> : ‘quebrar’	Sufijación a través de <i>-ear</i>
customizar	Del inglés <i>customize</i> : ‘personalizar’	Sufijación a través de <i>-izar</i>
draftear	Del inglés <i>draft</i> : ‘preparar’	Sufijación a través de <i>-ear</i>
esponsorizar	Del inglés <i>sponsor</i> : ‘patrocinar’	Sufijación a través de <i>-izar</i>
flashear	Del inglés <i>flash</i> : -	Sufijación a través de <i>-ear</i>
foldear	Del inglés <i>fold</i> : -	Sufijación a través de <i>-ear</i>
grafitear	Del inglés <i>graffiti</i> : ‘pintar’	Sufijación a través de <i>-ear</i>
linkear	Del inglés <i>link</i> : ‘enlazar’	Sufijación a través de <i>-ear</i>
loguear	Del inglés <i>log</i> : ‘iniciar sesión’	Sufijación a través de <i>-ear</i>
masterizar	Del inglés <i>master</i> : -	Sufijación a través de <i>-izar</i>
overclockear	Del inglés <i>overclock</i> : ‘aumentar velocidad del reloj’	Sufijación a través de <i>-ear</i>
photoshopear	Del inglés <i>Photoshop</i> : ‘editar imagen’	Sufijación a través de <i>-ear</i>
pixelar	Del inglés <i>pixel</i> : -	Sufijación a través de <i>-ar</i>
rankear	Del inglés <i>ranking</i> : ‘clasificar’	Sufijación a través de <i>-ear</i>
rapear	Del inglés <i>rap</i> : ‘golpear/ criticar’	Sufijación a través de <i>-ear</i>
rasterizar	Del inglés <i>raster</i> : -	Sufijación a través de <i>-izar</i>
renderizar	Del inglés <i>render</i> : ‘generar’	Sufijación a través de <i>-izar</i>
resetear	Del inglés <i>reset</i> : ‘reiniciar’	Sufijación a través de <i>-ear</i>
ripear	Del inglés <i>rip</i> : ‘convertir’	Sufijación a través de <i>-ear</i>
spoilear	Del inglés <i>spoiler</i> : ‘anticipar’	Sufijación a través de <i>-ear</i>
twittear	Del inglés <i>tweet</i> : ‘piar’	Sufijación a través de <i>-ear</i>

**Tabla 21.** Préstamos adaptados neológicos encontrados en el corpus de estudio.

En cuanto a este tipo de neologismos, se puede observar que todos derivan de préstamos del inglés, los que han sido adaptados en la lengua receptora, el español, en diferentes grados. Por una parte, la predominancia del inglés ya es indicada por Álvarez de Miranda (2009), quien señala, respecto del contacto entre lenguas que posibilita los préstamos, que “la fluidez y la universalización de las comunicaciones han hecho que hoy el inglés pueda influir sobre lenguas muy alejadas de él, prácticamente sobre todas las del planeta” (p. 143). Por otra parte, en relación con el grado de adaptación de este tipo de neologismos, es posible observar que hay verbos que han experimentado notorias modificaciones morfológicas para adaptarse a las normas ortográficas del español, como, por ejemplo, en el caso de *loguear*, al que se le ha añadido la *u* requerida por las normas ortográficas, o *grafitear*, al que se le ha suprimido una *f*; mientras que otros verbos neológicos de este tipo han mantenido sus características morfológicas de origen, logrando un menor grado de adaptación a las normas ortográficas del español, como en el caso de *overclokear*, que únicamente pierde la combinación *ck* original o *photoshopear*.

Otro de los aspectos que caracteriza los préstamos adaptados encontrados en el corpus es que todos corresponden a la primera conjugación del español (-ar) y, además, todos ellos son adaptados a la estructura morfológica de los verbos del español a través de la sufijación. En este sentido, el sufijo verbalizador más productivo es *-ear* con una frecuencia de 16 casos, frente a los 5 y 1 casos con sufijo *-izar* y *-ar*, respectivamente. Con esto, es posible advertir que para la sufijación propia de las palabras del español y para aquella aplicada a los préstamos de otras lenguas, los sufijos de preferencia varían: en el primer caso, resultan más productivos los sufijos *-izar* y *-ar*; mientras que en el segundo caso lo hace el sufijo *-ear*.

**4.3.1.3. Otros mecanismos de formación de palabras**

Si bien el foco de esta investigación se encuentra en los neologismos formados por derivación, de igual manera se detectaron neologismos producto de otros mecanismos de formación de palabras, particularmente de composición. Esto debido a las características morfológicas que comparten los verbos en general. Es decir, no se pretendía cubrir la

totalidad de neologismos compuestos, mas las reglas del script no resultaban excluyentes para su hallazgo. En este sentido, se detectó 1 caso de composición (FCOM) y 7 de composición culta (FCULT) (Tabla 22).

<b>FCOM</b>	<b>Base 1</b>	<b>Base 2</b>	<b>-</b>
malpensar	mal	pensar	-
<b>FCULT</b>	<b>Prefijo</b>	<b>Base 2</b>	<b>Sufijo</b>
criogenizar	<i>crio-</i>	gen	-izar
criopreservar	<i>crio-</i>	preservar	-
geolocalizar	<i>geo-</i>	localizar	-
geoposicionar	<i>geo-</i>	posicionar	-
georreferenciar	<i>geo-</i>	referenciar	-
retroiluminar	<i>retro-</i>	iluminar	-
teletransportar	<i>tele-</i>	transportar	-

**Tabla 22.** Composiciones y composiciones cultas detectadas en el corpus de estudio.

Por una parte, en cuanto al caso de composición, se puede observar que se trata de un verbo que únicamente se utiliza como participio (Anexo 3) y como infinitivo, en tanto este fue detectado en el corpus. Esto resulta interesante si se considera que el adjetivo *malpensado* sí se encuentra registrado en el listado de exclusión, por lo que la condición neológica de *malpensar* es limítrofe. Específicamente, lo referente a los participios y su relación con los adjetivos se aborda en los apartados 2.5.1 y 4.4.

Por otra parte, los casos de composición culta se caracterizan por poseer prefijos cultos, no incluidas en la tabla de prefijos, tal como indica su definición (OBNEO, 2004). Así, se observan los prefijos *tele-*, *geo-* y *crio-* provenientes del griego y el prefijo *retro-* proveniente del latín. Particularmente, el neologismo *criogenizar* constituye un caso de parasíntesis (FPRSU); mientras que los demás son casos de prefijación, por lo que este mecanismo continúa siendo el más productivo.

### 4.3.2. Errores ortográficos y fallas del método

Como se menciona en el apartado 4.1, de los 1.610 posibles neologismos detectados en el corpus de estudio, 1.396, es decir, un 23,17%, corresponde a diversos errores ortográficos y fallas del método. Para el análisis de estos problemas, se filtraron estos 1.396 datos y de ellos se extrajo una muestra aleatoria de 100 formas, que fueron etiquetadas manualmente como *errores ortográficos*, *fallas del método* o *desconocido* (Anexo 4).

De la muestra aleatoria, el 54% fueron errores ortográficos, el 38% fallas del método y el 27% casos desconocidos. Con estos datos es posible medir la precisión del método propuesto. Por una parte, como se menciona, el porcentaje de fallas de este es de un 38%, lo que fija la precisión en 62%. No obstante, cabe recordar que el porcentaje de errores ortográficos es alto, por lo que a futuro sería pertinente proponer un modo de subsanar tal dificultad.

En lo que concierne a la naturaleza de esos datos, se pudo observar que, por un lado, los errores ortográficos más comunes fueron la ausencia de letras de la palabra o tilde y sustitución de letras. Por otro lado, en cuanto a las fallas del método, las más comunes fueron las dificultades en la reconstrucción del infinitivo, ya sea porque se trataba de un verbo irregular o porque el verbo solo se utiliza en su forma pronominal. Evidentemente, los datos etiquetados como *desconocidos* eran de diversa naturaleza. La Tabla 23 ejemplifica algunos de estos casos.

Tipo de error	Ejemplo	Detalle
Error ortográfico (ort)	<b>Lema:</b> <i>deshechar</i>	El lema encontrado corresponde al verbo <i>desechar</i> , pero presenta un error ortográfico (adición de <i>h</i> ) que genera que el script no encuentre el lema en el listado de exclusión y lo proponga como un candidato a neologismo.
Falla del método (falla)	<b>Lema:</b> <i>year</i> <b>Raíz:</b> <i>ye</i> <b>Desinencias:</b> <i>ndome, ndonos,</i>	El script reconoce el elemento estable de las conjugaciones encontradas, correspondiente a la raíz <i>ye</i> . Sin embargo, por tratarse de un verbo irregular (verbo <i>ir</i> ), no es capaz de reconstruir

	<i>ndose, ndote,</i> etc.	apropiadamente el infinitivo. En su lugar, asigna la terminación <i>-ar</i> , que existe en el corpus pero que no se corresponde con el verbo en cuestión.
Falla del método (falla)	<b>Lema:</b> <i>gangrenar</i>	El script detecta correctamente la existencia de un verbo a partir de su raíz y desinencias. Sin embargo, reconstruye el infinitivo del verbo en su forma no pronominal, tratándose de un verbo pronominal. Por ende, al no encontrar registrada la forma no pronominal en el corpus de exclusión infiere incorrectamente que se trata de un posible neologismo.
Desconocido (desc)	<b>Lema:</b> <i>qir</i> <b>Raíz:</b> <i>q</i> <b>Desinencias:</b> <i>aba, ado, ara,</i> <i>aran, ería, etc.</i>	El script reconoce cierta regularidad en la combinación de una raíz y sus desinencias, pero no es posible identificar algún verbo del español. Únicamente la desinencia <i>-ería</i> sugiere que se podría tratar del verbo <i>querer</i> .

**Tabla 23.** Ejemplos de errores y fallas del método.

Como se puede observar, un alto porcentaje de datos del corpus dificultan la búsqueda y detección de verbos. Frecuentemente las formas contienen elementos morfológicos propios de los verbos del español e incluso presentan ciertas regularidades, pero no en todos los casos esto implica que se trata efectivamente de un verbo, lo que disminuye la precisión del método. En consecuencia, deberían proponerse otros criterios que permitan al método subsanar tales dificultades, para potenciar la detección automática de verbos neológicos y no neológicos.

#### 4.4. Participios

En lo que concierne a los participios, como ya se mencionó anteriormente (3.4.2.3), los datos fueron recopilados con miras a un trabajo futuro. En principio, es posible observar que, debido a las características morfológicas de participios y adjetivos, su detección es un problema no resuelto.

En los resultados obtenidos, se detectaron 734 raíces conjugadas únicamente con desinencias de participio y en las que fue posible reconstruir el infinitivo; mientras que de otras 605 raíces conjugadas solo con desinencias de participio no fue posible reconstruirlo. Esto permite suponer que existe una gran cantidad de verbos que en el uso han perdido sus formas personales, para ser utilizados únicamente como adjetivos derivados de sus respectivos participios. La Tabla 24 muestra algunos ejemplos de los resultados obtenidos al respecto: en el primer caso, la raíz *afortun* se encontró asociada solo a desinencias de participio y, luego, fue posible reconstruir su infinitivo o lema, lo que significa que en el corpus este verbo era utilizado en todas las formas de participio y de infinitivo; mientras que en el segundo caso, la raíz *afric*, que también se encontró asociada solo a desinencia de participio, no fue encontrada en su forma de infinitivo, por lo que no pudo ser reconstruido su lema.

Lema	Raíz	Desinencias
afortunar	afortun	ada, adas, ado, ados
-	afric	ada, adas, ado, ados

**Tabla 24.** Ejemplos de participios detectados en el corpus de estudio.

## 5. Conclusiones

El léxico es uno de los componentes de la lengua que permite observar en toda su amplitud el carácter mutable e inmutable que caracteriza el signo lingüístico. En este sentido, el presente estudio ha permitido comprobar que el léxico goza de un alto grado de estabilidad producto de su naturaleza inmutable dada por el carácter convencional de las reglas que lo rigen. Esto se hace evidente al observar que el 95.37% de los 4.629 verbos que, según la presente investigación, forman parte del caudal léxico de manera estable no son neológicos. Prueba de ello es que se encuentran registrados en una obra lexicográfica de gran extensión, como lo es el DLE y que ha quedado evidencia empírica de su uso en el corpus de estudio. En contraste, el carácter inmutable se evidencia en la pequeña proporción de neologismos, que corresponde a un 4.62%. Sin embargo, a pesar de ser esta una cifra considerablemente

menor a la proporción de verbos estables, es fundamental en el desarrollo del léxico. Como se indica en este estudio, los procesos de lexicogénesis ocurren de manera inevitable y constante por las diferentes necesidades comunicativas y expresivas de los hablantes.

Particularmente, en cuanto a los neologismos y en el caso de la variante peninsular del español, se pudo observar que de los mecanismos de formación de palabras que constituían el foco de este estudio, el más productivo es la prefijación, seguido por la sufijación y la parasíntesis. Algunos de ellos son bastante frecuentes, como *monitorear* o *precalentar*, lo que evidencia que la frecuencia de uso no siempre es un criterio para su incorporación en los diccionarios, sobre todo si este registro no está basado en análisis de corpus.

Asimismo, se detectaron neologismos cuya ausencia en el listado de exclusión podría deberse a que su uso es reciente o poco generalizado y no ha alcanzado la estabilidad necesaria para incorporarse en la obra lexicográfica. En este contexto, resulta interesante el caso del neologismo *virtualizar*, derivado por sufijación del adjetivo *virtual* y perteneciente al campo conceptual de las nuevas tecnologías, que ha experimentado, a su vez, un proceso de prefijación, dando lugar al verbo *desvirtualizar*. Esto indica que la *novedad* de su uso es una cuestión limítrofe: por un lado, se desconoce cuán reciente es su uso, pero, por otro, se hace evidente que goza de cierto grado de estabilidad que le ha permitido derivar otro neologismo. Esto se opone a los principios del criterio lexicográfico en tanto se demuestra que *virtualizar* es un verbo estable en el uso, signo de ello es su productividad, lo que no se condice con su ausencia en el listado de exclusión. Con esto se comprueba que la delimitación de neologismos, especialmente en lo que concierne al factor tiempo, es un asunto particularmente complejo.

En lo que concierne a los préstamos, particularmente a los préstamos adaptados, se pudo observar y comprobar la supremacía del inglés como lengua dadora de unidades léxicas al español. En concreto, los préstamos detectados se caracterizan por tener diferentes grados de adaptación a las reglas ortográficas de la lengua receptora. En este sentido, es posible suponer que un préstamo escasamente adaptado y cuyo uso es reciente posiblemente pueda incorporarse posteriormente al diccionario.

Asimismo, en términos de los procesos derivativos de los verbos del español, y como se señala en el marco teórico (2.5), la primera conjugación (-ar) es la más productiva. Esto se hace evidente al comprobar que, del total de neologismos detectados, independiente del mecanismo de derivación aplicado, el 96% pertenecía a esta conjugación.

Desde otra perspectiva, los resultados de este estudio permitieron medir los aciertos y limitaciones del método propuesto. En primer lugar, con respecto a los aciertos, cabe destacar la alta cobertura del método en cuanto a la detección de verbos no neológicos. Esto reafirma la idea de que este tipo de métodos, como el aquí propuesto o el de otras herramientas como TreeTagger, pueden prescindir de un diccionario integrado y, en su lugar, pueden basarse en la estructura morfológica que es observable, sistemática y regular en la mayoría de los verbos del español. Todo esto se suma a las ventajas evidentes de todo proceso automatizado que disminuye la intervención humana y permite el manejo de un gran volumen de datos. En segundo lugar, en cuanto a las limitaciones del método, se puede indicar, por una parte, las dificultades del método para lidiar con las faltas ortográficas presentes en el corpus. Al respecto, la gran cantidad de datos que presentaban errores ortográficos en el corpus implicó un gran porcentaje de falsos neologismos. En este sentido, se tuvo que recurrir al análisis manual para separar este tipo de verbos. Por ello, resultaría pertinente integrar nuevas reglas y criterios al procesamiento automático que apunten a subsanar esta limitación. Por otra parte, se pueden identificar las limitaciones del método, como la imposibilidad de reconstruir el lema de las formas verbales pronominales, ya explicadas en el apartado 4.3.2. Por último, se advierten las limitaciones propias de un corpus cuya representatividad es variable dependiendo de la naturaleza de los textos que lo componen y de las variantes del idioma.

En consideración de lo anterior, se propone como trabajo futuro una mejora del script, orientada, entre otras cosas, al control de los errores ortográficos y de las formas pronominales para aumentar su nivel de precisión en la detección de verbos neológicos. Asimismo, se propone ampliar el corpus de estudio con el fin de realizar una investigación más representativa del español. Con ello, se podría replicar el método propuesto en otros corpus, como el de Google Books o el EsTenTen en su variante americana.

## Bibliografía

- Álvarez de Miranda, P. (2009). Neología y pérdida léxica. En E. De Miguel (ed.), *Panorama de la lexicología* (pp. 133-157). Barcelona: Ariel.
- Batiukova, O. (2009). Neología y pérdida léxica. En E. De Miguel (ed.), *Panorama de la lexicología* (pp. 487-519). Barcelona: Ariel.
- Cabré, T. (1993). *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Antártida.
- Cabré, T., y Bayà, M. R. Bernal, E. Freixa, J., Solé, E., Vallès, T. (2002). Evaluación de la vitalidad de una lengua a través de la neología: a propósito de la neología espontánea y de la neología planificada. *Lèxic i Neologia. Barcelona: Observatori de Neologia. Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra*, 159-201.
- Dapena, J. Á. P. (2002). *Manual de técnica lexicográfica*. Arco/libros.
- De Miguel, E. (2015). "Lexicología". En Gutiérrez-Rexach, J. (dir.), *Enciclopedia de lingüística hispánica*. Nueva York: Routledge.
- Feliu, E. (2009). Palabras con estructura interna. En E. De Miguel (ed.), *Panorama de la lexicología* (pp. 51-81). Barcelona: Ariel.
- Hanks, P. (2013). *Lexical analysis: norms and exploitations*. Cambridge: MIT Press.
- Janssen, M. (2009). Detección de neologismos: una perspectiva computacional. *Debate Terminológico*, 5, 68-75.
- Kilgarriff, A., & Renau, I. (2013). esTenTen, a vast web corpus of Peninsular and American Spanish. *Procedia-Social and Behavioral Sciences*, 95, 12-19.
- Observatori de Neologia (2004). *Metodología del trabajo en neología: criterios, materiales y procesos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Nazar, R.; Vivaldi, J.; Cabré, MT. (2008). A Suite to Compile and Analyze an LSP Corpus. Proceedings of LREC 2008 (The 6th edition of the Language Resources and Evaluation Conference) Marrakech (Morocco), May 28-30, 2008.
- Real Academia Española. Diccionario de la lengua española. 23.<sup>a</sup> ed. actualizada en línea. Consulta a través de EnclaveRAE. URL: <https://enclave.rae.es>. [Última consulta: 19/12/2018].

- Real Academia Española. (2009). Nueva gramática de la lengua española. Madrid: Espasa Calpe.
- Real Academia Española. (2010). Nueva gramática de la lengua española. Manual. Madrid: Espasa Calpe.
- Real Academia Española. (2014). *Diccionario de la lengua española*. 23.<sup>a</sup> ed. Madrid: Espasa.
- Renau, I., Nazar, R., Lecaros, V. (en prensa). La evolución de las marcas ortográficas y tipográficas en los procesos de lexicalización de neologismos: un estudio en el vocabulario de la crisis económica en prensa española. *Revista Española de Lingüística Aplicada*.
- Saussure, F. (1916[2007]). *Curso de lingüística general*. Buenos Aires: Losada.
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK. URL: <http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/treetagger.html>.
- Torner, S. (redacción); Acebo, S. (coord.); Lucena, N. (dir.). (2006). *Gramática de la Lengua Española*. 1 ed. Barcelona: Larousse.
- Varo, C. (2013). Aproximación teórico-práctica al procesamiento lingüístico de neologismos léxicos: A theoretical and practical approach to the linguistic processing of neologisms. *Revista signos*, 46(81), 132-152. <https://dx.doi.org/10.4067/S0718-09342013000100006>
- Yuni, J. A., y Urbano, C. A. (2014). *Técnicas para investigar I*. Editorial Brujas.