



UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Departamento de Matemáticas



TESIS DOCTORAL

RECONOCIMIENTO Y GENERACIÓN DE PALABRAS COMPUESTAS EN ESPAÑOL

Virginia Gutiérrez Rodríguez

Licenciada en Matemáticas, especialidad Estadística e Investigación Operativa

LAS PALMAS DE GRAN CANARIA, 2015



Departamento de Matemáticas

Programa de doctorado: Simulación Numérica en Ciencia y Tecnología
(Bienio 1995-1997)

TESIS DOCTORAL

***RECONOCIMIENTO Y GENERACIÓN
DE PALABRAS COMPUESTAS
EN ESPAÑOL***

Autora: D^a. Virginia Gutiérrez Rodríguez

Tesis Doctoral dirigida por

Dr. D. Octavio Santana Suárez

Dr. D. José R. Pérez Aguiar

Las Palmas de Gran Canaria, Noviembre 2015

A Gigi y Adri

Agradecimientos

Quiero expresar mi eterno agradecimiento a mi director Dr. D. Octavio Santana...
porque la vida me ha dado un regalo increíble cuando me puso en su camino...
gracias por ser una persona excepcional en el ámbito profesional...
como en el personal...

A mi director Dr. D. José Pérez Aguiar mi especial gratitud...
por su inestimable ayuda con la que siempre he contado...

A mi pequeña gran familia, hermanos, sobrinos, cuñados y grandes allegados...
por el inmenso cariño y respeto que les tengo...
por la paciencia y discreción tenida esperando...
que este trabajo obtuviera su fruto...
Ya llegó...

A mis grandes y queridos maestros Isa, Rafa, Bea, Robert...
y a todos mis compañeros Félix, Fernando, Conchi, Stephy, Roda...
que me han enseñado el verdadero significado...
de la palabra AMISTAD...

A todos los presentes, ausentes, a los cercanos, a los lejanos...
los que forman parte de mi vida, los de toda mi vida...
los que se han incorporado apenas hace nada...
porque si hoy he llegado aquí es gracias...
a que han sido parte de mi...

A todos los que pensaron que este momento llegaría...
Solo me faltaba un motivo...
Y mi motivo son todos USTEDES...

«*La mayor fortaleza del ser humano, otros seres humanos*» (Ellen Bercheid)

¡¡GRACIAS!!

Resumen

Uno de los medios con los que cuenta la lengua española para ampliar el conjunto de voces del idioma consiste en la utilización de mecanismos de tipo morfológico para formar nuevas palabras: **la composición** es, sin duda, uno de los procesos de formación de palabras de mayor importancia —aunque haya recibido escaso tratamiento por parte de la bibliografía¹— de los que dispone la lengua para la renovación y enriquecimiento del léxico, también recibe el nombre neologismo morfosintáctico, se basa en coordinar o interaccionar específicamente dos o más lexemas —o raíces cultas— para formar una nueva palabra que compone una unidad semántica y sintáctica. En esta Tesis se estudian únicamente aquellos casos en los que la palabra compuesta se ha consolidado como la unión gráfica de los elementos que la producen de forma regular o irregular. Se extraen de diferentes fuentes léxicas —Internet y el lenguaje periodístico recogen, con frecuencia, neologismos compositivos debido a la rápida evolución de los acontecimientos y a su inmediata transcripción al mundo de las Tecnologías de la Información— los distintos integrantes y se examinan los mecanismos aplicados para establecer su comportamiento morfosemántico y su generación. Se hacen imprescindibles unos procesos automáticos que sean capaces de identificar las palabras interesadas y situarlas en un contexto lingüístico adecuado.

El reconocimiento de palabras compuestas en español es útil en aplicaciones para el procesamiento automático del lenguaje natural, debido a que lleva implícito vínculos semánticos; asimismo, potencia las búsquedas en Internet, al ampliar el abanico de relaciones morfológicas deducidas de los compuestos bajo estudio, sus derivaciones y flexiones. Además del desarrollo de una herramienta lingüística que permite reconocer palabras compuestas, parasintéticas por composición o neologismos compositivos, entre otros procesos de lexicalización, se incluye la generación de las informadas alteraciones morfológicas — se tienen en cuenta las reglas de formación, fonéticas y excepciones e irregularidades detectadas.

¹ Conclusión que comparto con Buenafuentes de la Mata (2007) en su tesis *Procesos de gramaticalización y lexicalización en la formación de compuestos en español*.

Índice

Agradecimientos	vii
Resumen	ix
Índice de Figuras	xv
Índice de Tablas	xvii
CAPÍTULO 1.- LA COMPOSICIÓN EN ESPAÑOL	1
Introducción	1
La composición en español.....	4
Elementos constituyentes del compuesto	7
Elementos compositivos prefijales y sufijales.....	9
Número de elementos constituyentes.....	10
Formaciones con elementos patrimoniales.....	11
Clasificación de los compuestos según clase funcional del conjunto y de los elementos constituyentes.....	11
Clasificación de los compuestos atendiendo a su estructura formal	20
Otros procesos de formación de palabras	23
Los adjetivos acabados en –mente y las preposiciones	23
Parasíntesis por composición.....	24
¿Cómo estudiar los compuestos?	26
CAPÍTULO 2.- NEOLOGISMOS COMPOSITIVOS EN ESPAÑOL	31
Introducción	31
Los neologismos.....	33
Clasificación de los neologismos	34
La vitalidad de los neologismos compositivos.....	36

CAPÍTULO 3.- ESTUDIO DE LA COMPOSICIÓN YUXTAPUESTA EN ESPAÑOL. REGLAS DE FORMACIÓN Y DE CAMBIO GRÁFICO⁴⁵

Introducción	45
Corpus seleccionado	46
Las reglas de formación de compuestos	52
Reglas de formación	54
Regla 1 (unión)	54
Regla 2 (-a2, -e2,-o2, -ha, -ho).....	55
Regla 3 (-a+i, -o+i,-e+i, -eo+i, -ea+i).....	57
Regla 4 (+i)	58
Regla 5 (-a, -o, -e).....	59
Regla 6 (elemento compositivo)	60
Regla 7 (@preposición@)	61
Regla 8 (-vocal@preposición@).....	62
Tratamiento de las excepciones	62
Regla 9 (excepción: variantes)	63
Regla 10 (excepción: género)	63
Regla 11 (excepción: plural)	64
Regla 12 (excepción: paranomasia)	64
Regla 13 (excepción: 2º palabra).....	65
Regla 13.1 (2ªpalabra_plural).....	65
Regla 13.2 (-sílabas_duplicada).....	65
Regla 13.3 (-sílabas).....	65
Regla 13.4 (acronimia).....	65
Regla 14 (abreviamiento)	66
Regla 15 (parasíntesis)	66
Reglas fonéticas o de cambio gráfico.....	67

Regla 16(+rr)	67
Regla 17 (-ca+qui, -co+qui)	67
Regla 18 (-za+ci, -zo+ci, -zón+ci)	67
Regla 19 (-n+m)	68
Regla 20 (copulativa)	68
Regla 21 (-h+j)	69
Regla 22 (metafonía).....	69
Regla 23 (tilde).....	69
Regla 24 (diptongo)	70
Modelo de datos	70
CAPÍTULO 4.- RECONOCIMIENTO DE COMPUESTOS	73
Introducción	73
Automatización del reconocimiento de compuestos	74
Fases en el reconocedor de palabras compuestas	78
Criterios en el reconocimiento de nuevos compuestos.....	80
Obtención de las reglas de formación para el reconocimiento	82
Grafo de dependencias de las reglas de formación y de cambio gráfico.....	84
Base de datos de las soluciones.....	87
Ordenamiento de las soluciones según categorías	89
Obtención de la solución según significado	94
CAPÍTULO 5.- GENERACIÓN DE COMPUESTOS	99
Introducción	99
Automatización de la generación de compuestos.....	100
Fases en el generador de palabras compuestas.....	102
Grafo de dependencias de las reglas de generación	104
Ordenación de la solución según significado	106

CONCLUSIONES	117
Aportaciones originales	119
Futuras Líneas de Investigación	120
BIBLIOGRAFÍA	121
Bibliografía lingüística	121
Obras utilizadas para la elaboración del corpus. Diccionarios	126
Sitios web	126
APENDICES	129
APÉNDICE A: DIAGRAMAS DE FLUJO	129
APÉNDICE B: GRÁFICOS PARA LOS PROCESOS DE FORMACIÓN FORMALES DE NEOLOGISMOS.	131
LISTA DE ACRÓNIMOS Y ABREVIATURAS.....	135
GLOSARIO DE TÉRMINOS	137
LISTA DE VOCES CITADAS	139

Índice de Figuras

Figura 1.- Distribución de los tipos de composición atendiendo a la clase funcional del conjunto.....	12
Figura 2.- Distribución de los compuestos nominales recogidos en la base documental atendiendo a la categoría gramatical de los elementos componentes.....	14
Figura 3.- Distribución de los compuestos adjetivales recogidos en la base documental atendiendo a la categoría gramatical de los elementos componentes.....	17
Figura 4.- Distribución de los compuestos verbales recogidos en la base documental atendiendo a la categoría gramatical de los elementos componentes.....	18
Figura 5.- Distribución de la composición nominal, adjetiva y verbal en la base documental atendiendo a las categorías gramaticales de los elementos componentes.....	23
Figura 6.- Formaciones neologistas del banco de datos de BOBNEO durante el periodo 2004 al 2010 en valores absolutos.	37
Figura 7.- Formaciones neologistas del banco de datos de BOBNEO durante el periodo 2004 al 2010 en valores relativos.....	38
Figura 8.- Gráficos circulares y radiales para los procesos de formación formales de neologismos en el periodo 2004-2010.	40
Figura 9.- Gráficos radiales del proceso de formación de neologismos compositivos — composición y derivación— en el periodo 2004-2010.	41
Figura 10.- Grafo relacional de los compuestos recopilados de diversas fuentes lexicográficas.	47
Figura 11.- Modelo lógico de la Base de Datos BARGCoS.	51
Figura 12.- Producción de la regla 1 (unión) en base a la categoría gramatical de los constituyentes.	55
Figura 13.- Producción de la regla 2 (-a2, -e2, -o2, -ha, -ho) en base a la categoría gramatical de los constituyentes.....	56
Figura 14.- Producción de la regla 3 (-a+i, -o+i, -e+i, -eo+i, -ea+i) en base a la categoría gramatical de los constituyentes.....	58
Figura 15.- Producción de la regla 5 (-a, -o, -e) en base a la categoría gramatical de los constituyentes.	60

Figura 16.- Producción de la regla 7 (@prep@) en base a la categoría gramatical de los constituyentes.	62
Figura 17.- Producción de la regla 9 a la regla 13 de tratamiento de las excepciones, en base a la categoría gramatical de los constituyentes.	66
Figura 18.- Producción de la regla 4 (+i) y regla 20 (copulativa) en base a la categoría gramatical de los constituyentes.	68
Figura 19.- Modelo de lógico global de la BBDD ARGCoS.	70
Figura 20.- Diagrama de flujo del proceso de reconocimiento de compuestos.	76
Figura 21.- Portal web del Flexionador y Lematizador Morfológico y de Relaciones Morfológicas del Grupo de Investigación GEDLC de la ULPGC.	80
Figura 22.- Interfaz de la aplicación ARGCoS para el reconocimiento de compuestos.	81
Figura 23.- Extracción del grafo de dependencias de las reglas de formación y cambio gráfico para el reconocimiento.	86
Figura 24.- Modelo lógico global de la Base de Datos BARGCoS.	88
Figura 25.- Distribución del total de cada tipo composición según regla de formación en valores absolutos —imagen superior— y relativos —imagen inferior.	90
Figura 26.- Gráficos radiales para la distribución del total de cada tipo compositivo según la regla de formación en valores absolutos —imagen superior— y relativos —inferior.	91
Figura 27.- Distribución del total para cada regla de formación según tipo composición en valores absolutos —imagen superior— y relativos —imagen inferior.	92
Figura 28.- Gráficos radiales para cada regla de formación según tipo composición en valores absolutos y relativos, respectivamente.	93
Figura 29.- Diagrama de flujo del proceso de generación de compuestos.	102
Figura 30.- Interfaz de la aplicación ARGCoS para la generación de compuestos.	104
Figura 31.- Clasificación de la composición en español atendiendo a las categorías gramaticales del compuesto y de los elementos componentes.	105
Figura 32.- Extracción del grafo de dependencias de las reglas de formación y cambio gráfico para la generación.	106
Figura 33.- Distribución de las reglas de formación para compuestos sustantivos según la categoría gramatical de sus elementos constituyentes.	109
Figura 34.- Distribución de las reglas de formación para compuestos adjetivos según la categoría gramatical de sus elementos constituyentes.	110

Índice de Tablas

Tabla 1.- Combinaciones de distintas unidades léxicas para formar palabras compuestas.....	8
Tabla 2.- Porcentajes en las combinaciones de distintas unidades léxicas para formar palabras compuestas.....	8
Tabla 3.- Elementos prefijales y elementos sufijales.....	10
Tabla 4.- Clasificación de la composición según Bustos Gisbert.....	13
Tabla 5.- Compuestos nominales según la categoría gramatical de sus elementos componentes.....	15
Tabla 6.- Compuestos adjetivales según la categoría gramatical de sus elementos componentes.....	16
Tabla 7.- Clasificación de compuestos atendiendo a la categoría gramatical composicional y las categorías de sus elementos componentes.....	19
Tabla 8.- Porcentajes de diversos tipos de formaciones de neologismos formales —derivativos, compositivos, entre otros— del banco de datos de BOBNEO del 2004 al 2010.....	36
Tabla 9.- Tabla de contingencia para comprobar la homogeneidad de los diferentes tipos de composición —propia y culta— en las bases de datos BOBNEO vs BARGCoS.....	42
Tabla 10.- Número de soluciones para un conjunto de compuestos.....	82
Tabla 11.- Tabla de correspondencia entre reglas de generación y reglas de reconocimiento.....	83

«...Las palabras atraen, repelen, evocan, sugieren. Se dejan descomponer en letras, que a su vez, pueden combinarse formando nuevas palabras. Sus sonidos pueden resonar en nuestros oídos con matices poéticos inauditos. La maleabilidad de las palabras permite una infinidad de juegos que nos conducen, con mano segura y divertida a las entrañas mismas de la lengua.... Jugando con las palabras conseguimos: crear neologismos y observar el entorno de forma nueva e inesperada...»

(Rafael Hidalgo de la Torre, *Sugerencias sugestivas con las palabras*, 2010)

CAPÍTULO 1.- LA COMPOSICIÓN EN ESPAÑOL

RESUMEN: Este capítulo presenta una breve introducción al tema, el marco teórico aplicado al estudio de las palabras compuestas en español. El lector podrá hacerse una idea de qué es la composición y el objetivo de la Tesis.

Palabras usuales: composición, neologismo, reglas de formación, procesamiento del lenguaje natural, lingüística computacional.

Introducción

La etimología de la composición se encuentra en la raíz latina *compositio* que se define como la acción y efecto de componer —juntar varias cosas y colocarlas en orden para formar una o constituir algo. Uno de los recursos con los que cuentan muchas lenguas para aumentar su caudal léxico consiste en utilizar mecanismos de tipo morfológico—también llamados *neologismos morfosintácticos*²— para formar nuevas palabras como son la composición, la derivación o la parasíntesis, entre otros. En el mundo clásico, estos procedimientos formativos fueron unos de los que mejor supo explotar la lengua hasta convertirse en una de las de mayor riqueza y flexibilidad léxicas que más han influido en nuestro vocabulario actual. Sin duda, la composición es uno de los procesos de formación de palabras con mayor importancia de los que dispone la lengua para la renovación y enriquecimiento de su léxico, a partir de unidades léxicas, vocablos o giros nuevos. Los avances tecnológicos, los desarrollos científicos, el auge del uso de los medios sociales, son ejemplos que han hecho de catalizadores en estos procesos de cambios. Se observa la presencia cada vez mayor de este mecanismo para crear neologismos morfosintácticos, en particular, neologismos

² Los neologismos morfosintácticos son aquellos «productos léxicos conseguidos por derivación, composición o parasíntesis con fines expresivos, surgidos en un contexto poético y creados por un autor que en un momento dado siente la necesidad de crear como una afirmación de su libertad de expresión, como una muestra de originalidad frente a la lengua común a la que en consecuencia considera insuficiente, pobre o poco precisa» (M^a VICTORIA ROMERO GUALDA, *Hacia una tipología del neologismo literario*, Anales de la Universidad de Murcia, Murcia, 1978, pág.149).

compositivos, gracias al influjo provocado por los lenguajes técnicos, publicitarios y sociales, principalmente.

Mediante la composición, la lengua obtiene nuevos vocablos a partir de elementos ya presentes en el lenguaje o de otros prestados de fuera: reúne dos o más de estos elementos en uno solo, *con un orden determinado y una unidad íntima de sentido*. Por contraposición, en la derivación existe un elemento gramatical que no está libre, es decir, consiste en la creación de elementos léxicos nuevos mediante la adición o supresión a palabras ya existentes de elementos inseparables —afijos³. En la parasíntesis se combinan los mecanismos anteriores, bien por afijación que simultanea dos procesos derivativos —sufijación y prefijación— o bien por combinación de elementos de la composición y de la derivación como trata Serrano Dolader⁴. Entre la derivación y la composición se encuentra la utilización de raíces cultas greco-latinas en los procesos de formación de nuevas palabras, especialmente en los campos científicos y técnicos; las voces en cuya formación intervienen estos elementos podrían no considerarse propiamente compuestas, pues la mayoría de estas raíces no pueden aparecer de forma aislada, pero tampoco pueden considerarse derivadas puesto que tienen un comportamiento peculiar que los aleja de los auténticos afijos, es más, tienen un significado léxico más próximo al de las raíces que al de los afijos. A este tipo de raíces se le da el nombre de elementos prefijales o *pseudoprefijoides* o elementos sufijales o *pseudosufijoides* —*elementos compositivos*—, en función de si se anteponen o se posponen a otra raíz, respectivamente. En suma, en la presente Tesis se estudian aquellos casos de compuestos que se han consolidado como la unión gráfica de los elementos que intervienen⁵; además, se trata los pseudoprefijoides o pseudosufijoides como elementos compositivos, no como morfemas

³ Se entienden por afijos las partículas o secuencias lingüísticas que se adjuntan a una palabra o lexema, modificando su sentido o valor gramatical.

⁴ Véase, DAVID SERRANO DOLADER, *Las formaciones parasintéticas en español*, Arco/Libros, Madrid, 1995; DAVID SERRANO DOLADER, “Sobre los compuestos (para)sintéticos ¿en español?”, *Los límites de la morfología: estudios ofrecidos a Soledad Varela Ortega*, Universidad Autónoma de Madrid. págs. 427-442, 2012.

⁵ Hay que considerar que la composición puede dividirse en composición de tipo ortográficamente unidos y composición sintagmática, donde los componentes han alcanzado una coherencia semántica sin fusión ortográfica. La presente Tesis sólo se centrará en el primer tipo de composición.

No se estudian los compuestos sintagmáticos por la dificultad en su delimitación: ¿cuáles son los límites entre la composición sintagmática, los sintagmas libres de la sintaxis y, a su vez, se puede distinguir de otros fenómenos como las colocaciones y las locuciones?.

derivativos, y se incluye la parasíntesis por composición, desde un punto de vista principalmente morfológico, no sintáctico ni semántico, lo que no impide que en ocasiones sea necesario hacer referencias a tales aspectos. Son procedimientos que pueden crear neologismos, lo que constituye un fructífero manantial para el enriquecimiento de la lengua.

Las más recientes aportaciones realizadas al estudio sobre la composición en español, principalmente nominal, deja entrever el vacío tan grande que existe, pese a la presumible trascendencia de este proceso de formación de palabras, pues han resultado infructuosas las búsquedas sobre el procesamiento automático de la composición en español. En este sentido, esta Tesis trata de ampliar el conocimiento sobre la composición con la intención de cubrir algunos huecos existentes en la literatura y llenar el vacío informático en el tratamiento automatizado de la composición. Se pretende hacer una contribución a la identificación y generación automatizada de compuestos a través de las reglas de formación de carácter léxico y fonéticas o de cambio gráfico, y los criterios de aplicación que se deben llevar a cabo en cada situación. La metodología empleada consiste en extraer de diferentes diccionarios los distintos compuestos con el propósito de garantizar un conocimiento suficiente de los diferentes casos que se pueden dar y se estudian los mecanismos de unión aplicados a partir de la realidad explorada — se atiende a la clase funcional del compuesto, a su estructura formal y a la clase funcional de los elementos componentes—, que permitan solventar barreras lingüísticas a través del dominio informático y poner de manifiesto las irregularidades encontradas en su tratamiento. Con este objetivo se acometen diversos trabajos⁶ a favor de una apuesta empírica que se sustenta en un corpus de alrededor 11 000 compuestos recopilados de diversos repertorios lexicográficos de la lengua española.

⁶ OCTAVIO SANTANA SUÁREZ, VIRGINIA GUTIÉRREZ RODRÍGUEZ, Y JOSÉ PÉREZ AGUIAR, “Pragmatización en la automatización del reconocimiento de palabras compuestas en español”. *LEA Lingüística Española Aplicada*, XXXV/2, págs. 181-201, 2013; SANTANA SUÁREZ, OCTAVIO, GUTIÉRREZ RODRÍGUEZ, VIRGINIA, PÉREZ AGUIAR, JOSÉ Y SÁNCHEZ BERRIEL, ISABEL, “Reglas de formación de palabras compuestas en español para la automatización de su reconocimiento”, *Procesamiento del Lenguaje Natural*, vol.51, págs. 75-82, Jaén, 2012; SANTANA SUÁREZ, OCTAVIO, CARRERAS RIUDAVETS, FRANCISCO, PÉREZ AGUIAR, JOSÉ Y GUTIÉRREZ RODRÍGUEZ, VIRGINIA, “El Reconocimiento Automático de la Composición en Español”, *Conference Abstracts of the First International Conference of the Alliance of Digital Humanities Organizations*, Paris, 2006.

La composición en español

La composición es, sin duda, uno de los procesos de formación de palabras más relevantes con que cuenta la lengua para la renovación y enriquecimiento de su léxico. Se observa la presencia cada vez mayor de este mecanismo para crear neologismos morfosintácticos, en particular, neologismos compositivos, gracias al influjo provocado por los lenguajes técnicos, publicitarios y sociales, entre otros.

La perspectiva adoptada para el estudio de compuestos se basa en un análisis sincrónico debido a la dificultad que entraña la visión diacrónica ya que es prácticamente inexistente⁷ y difícil su automatización. En los estudios realizados sobre composición en español, se pone de manifiesto una falta de homogeneidad tanto en el establecimiento de los diferentes tipos de compuestos como en el esclarecimiento de sus características —se remite al lector a las obras de Alvar Ezquerra, M. Lang, Alemany Bolufer, Almela Pérez, Bustos Gisbert, Val Avaro, Varela Ortega⁸, entre otros —, sobre todo a la hora de fijar una frontera entre la composición y la derivación o bien entre los compuestos sintagmáticos y las

⁷ En la tesis doctoral de Buenafuentes de la Mata (*Procesos de gramaticalización y lexicalización en la formación de compuestos en español*, Tesis Doctoral, Universidad Autónoma de Barcelona, 2007) se demuestra que la diacronía juega un papel muy importante en el estudio de la formación de palabras y, en concreto, de la composición; además las teorías de la gramaticalización y la lexicalización son el mejor marco teórico para alcanzar dicho objetivo. En este sentido, contribuye a llenar en parte el vacío bibliográfico existente en el tratamiento histórico de la composición.

⁸ MANUEL ALVAR EZQUERRA, *La formación de las palabras en español*, Cuadernos de lengua española, Arco/Libros, Madrid, 2002; LANG MERVING, *Formación de palabras en español. Morfología derivativa productiva en léxico moderno*, Cátedra, Madrid, 1992; JOSÉ ALEMANY BOLUFER, *Tratado de formación de palabras en la lengua castellana. La derivación y la composición. Estudios de los sufijos y prefijos empleados en una y otra*, Victoriano Suárez, Madrid, 1920; RAMÓN ALMELA PÉREZ, *Procedimientos de formación de palabras en español*, Ariel, Barcelona, 1999; EUGENIO BUSTOS GISBERT, *La composición nominal en español*, Pub. Universidad, Salamanca, 1986; JESÚS VAL ÁLVARO, “La composición”. En IGNACIO BOSQUE Y VIOLETA DEMONTE (eds.), *Gramática descriptiva de la Lengua Española*, Espasa Calpe vol. 3, Madrid, 1999, págs. 4757-4841; SOLEDAD VARELA ORTEGA, *Fundamentos de Morfología*, Síntesis, Madrid, 1990.

colocaciones⁹ y locuciones¹⁰. Para Lang (1992) la composición consiste en la unión de palabras ya se trate de formas libres o morfemas léxicos; en la misma línea puede resultar la definición de Alvar (2002) donde en la composición participan dos o más unidades léxicas que pueden aparecer libres en la lengua; algo más precisa es la que nos da Varela (1990) considerando que en la composición se unen o se combinan dos o más formas libres para construir una forma compleja, la cual, desde el punto de vista significativo, fónico y funcional, representa una unidad léxica con un sentido único y constante¹¹. La definición más próxima a este estudio aparece en el Diccionario de la Real Academia Española¹² (DRAE):

“procedimiento por el cual se forman vocablos agregando a uno simple una o más preposiciones o partículas u otro vocablo íntegro o modificado por eufonía—coyotomate, quitaipón, cagalaolla, paraguas”

donde además se añade que:

«consiste en hacer de dos o más elementos —llamados elementos componentes— una nueva palabra cuyo significado —muchas veces en sentido figurado— suele ser distinto y más extenso que la suma de los significados de sus componentes. Los elementos componentes tiene un orden determinado y una unidad íntima de sentido»

⁹ Se consideran colocaciones a las combinaciones frecuentes de unidades léxicas fijadas en la norma o combinaciones de palabras que se distinguen por su alta frecuencia de uso, respondiendo a pautas de formación gramaticales y significado composicional.

¹⁰ Para JUAN MANUEL PÉREZ VIGARAY en “Locuciones y compuestos nominales. Aportaciones de Julio Casares al estudio de la formación de palabras” (*Philologia Canariensis*, 2-3, págs. 295-310, 1997) distingue «entre las creaciones léxicas construidas sobre las reglas de la sintaxis libre de nuestra lengua, de aquellas otras que se forman sobre la base de reglas propias y exclusivas, distintas de las de la sintaxis libre».

¹¹ Es en esto último, donde se puede establecer una pequeña distinción o frontera con las colocaciones ya que en la composición sintagmática el orden importa y no se puede alterar añadiendo elementos en medio —*mesa camilla* no puede aparecer con el adverbio grande de la forma “*mesa grande camilla*” sino, en todo caso, como “*mesa camilla grande*”. No obstante, no es relevante para nuestro estudio debido a que no se trata la composición sintagmática.

¹² Real Academia Española, *Diccionario de la Lengua Española*, Espasa-Calpe, Madrid, edición electrónica. 22ª, 2001.

Aunque existe un tipo de composición denominada *sintagmática*¹³, se deja al margen la mayoría de estos compuestos en la automatización de su reconocimiento, ya que los componentes han alcanzado una coherencia semántica sin fusión gráfica — *bases de datos, caja negra, mesa camilla, abeja reina*— y entrarían en conflicto con las colocaciones léxicas¹⁴ —*lanzarse al ataque, circular rumor, medida drástica, abanderar lucha, ronda de negociaciones*. Sin embargo, gracias al comportamiento motivado, regular e isomórfico¹⁵ de los elementos constituyentes¹⁶, el compuesto podría diferenciarse de las colocaciones y las locuciones. Según se recoge en el trabajo de Alvar¹⁷, la composición se sirve de procedimientos para la creación de palabras nuevas como la *yuxtaposición o léxicas compuestas*, donde la fusión gráfica de los elementos participantes en el compuesto es total, así como su lexicalización y su gramaticalización —*malsano, agridulce, quitapión, boquicerrado*. El presente trabajo se restringe a este tipo de composición, como se explica en el epígrafe correspondiente a la *Clasificación de los compuestos atendiendo a su estructura formal*.

¹³ Se entiende por composición sintagmática aquel tipo de composición que tiene como resultado unos compuestos que se aproximan bastante a los sintagmas nominales correspondientes, pero que no pueden ser considerados como sintagmas nominales, sino como unidades léxicas, dado que el compuesto se utiliza como una única palabra en la que han quedado cohesionados todos sus componentes.

¹⁴ No pueden adquirir la categoría de compuesto aunque semánticamente estén unidos como *mesa de jardín o medida drástica* frente a palabras que sí lo son como *libro de cocina o abeja reina*.

¹⁵ Según Buenafuentes de la Mata (2007): “el hecho de que una formación sea isomórfica, implica que pueda ser analizada o descompuesta en sus componentes. Por otra parte, la motivación se relaciona con la transparencia semántica que presenta la unidad: si es posible interpretar su significado sólo a partir de la formación será motivada”.

¹⁶ Véase, LEONOR RUIZ GURILLO, “Compuestos, colocaciones, locuciones: intento de delimitación”. En VEIGA, A., GONZÁLEZ PERERIRA, M., SOUTO GÓMEZ, M. (eds.), *Léxico y Gramática*. Tris Tram, Lugo, 2002, págs. 327-339.

¹⁷ MANUEL ALVAR EZQUERRA, *La formación de palabras en español*, op. cit.

Elementos constituyentes del compuesto

En esta Tesis se considera que los elementos componentes que forman una palabra compuesta yuxtapuesta¹⁸ pueden ser:

- palabras patrimoniales—aquellas que tiene un solo lexema o un lexema unido a morfemas flexivos o la que está formada por un morfema libre o unido a morfemas flexivos.
- o temas cultos de origen grecolatino denominados *elementos prefijales* o *sufijales* — *elementos compositivos*— según se antepongan o se pospongan.

Al fijar que los elementos compositivos son “temas cultos” se establece la diferencia con lo que se entiende por verdaderos afijos, tanto sean prefijos como sufijos¹⁹. La Tabla 1 muestra ejemplos de palabras compuestas por las combinaciones posibles de estos dos tipos de unidades léxicas.

COMPOSICIÓN	PALABRA	ELEMENTO COMPOSITIVO
PALABRA	<i>hojalata albasol malgastar uñalbo alborotapueblos aguavina maestrescuela malacara pararrayos puntapié pintamonas telarañas pinchaúvas franjirrojo alapivot juntarrimas abrazajarolas apagafuegos azulcrema sillónball pinchaúva vendepatrias</i>	<i>germanófilo timbrología penología nectarívoro ondímetro urínifero oxigenoterapia patinódromo laserterapia sambódromo</i>

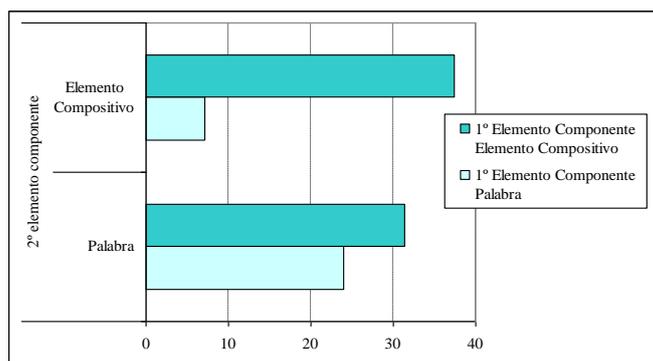
¹⁸ Existen autores que distinguen palabras yuxtapuestas a compuestas, es decir, no incluyen la yuxtaposición como una clasificación de las palabras compuestas, son términos distintos. Mientras que las palabras yuxtapuestas están formadas por palabras patrimoniales, las compuestas lo estarían por raíces, por lo general, griegas o latinas, aunque también podrían proceder las yuxtapuestas, pero estas últimas son mucho más moderna su creación y sus elementos componentes tienen sentido como palabras sueltas, que es lo que diferencia de los compuestos como *agricultura* (existe *cultura* pero no *agri*-).

¹⁹ Véase MANUEL FERNANDO PÉREZ LAGOS, "¿Entre composición y afijación? Naturaleza de los elementos de formación culta", Servicio de Publicaciones de la Universidad de Granada vol. 3, págs. 361-369, 1997.

ELEMENTO COMPOSITIVO	<i>ecosistema cardioprotector cornidelantero denticonejuno euroasiático indogermánico necrodactilar vitrocerámica zooplancton cibercampaña anorexígeno arhiconocido</i>	<i>logopeda filántropo acuametría adipogénesis aerobio agorafobia aluminografía andrófobo fibromialgia</i>
-----------------------------	--	---

Tabla 1.- Combinaciones de distintas unidades léxicas para formar palabras compuestas.

Para Buenafuentes de la Mata, existen tres tipos de estudios compositivos en español: la composición sintagmática —no será tratada en la presente Tesis—, la composición culta —participan elementos compositivos prefijales y sufijales— y la composición léxica —o composición propia, según Bustos. La Tabla 2 muestra el porcentaje en la formación de palabras compuestas según las combinaciones posibles de estos dos tipos de unidades léxicas, recogidas en la base documental²⁰ trabajada en esta Tesis.



		2º Elemento Componente	
		Palabra	Elemento Compositivo
1º Elemento Componente	Palabra	23,97%	7,15%
	Elemento Compositivo	31,42%	37,44%

Tabla 2.- Porcentajes en las combinaciones de distintas unidades léxicas para formar palabras compuestas.

²⁰ La investigación está fundamentada en un corpus de alrededor de 11 000 compuestos recopilados de diversos repertorios lexicográficos de la lengua española, la cual se detalla en el siguiente capítulo.

Elementos compositivos prefijales y sufijales

En el lenguaje científico y tecnológico es muy habitual la formación de nuevas palabras o neologismos utilizando raíces cultas grecolatinas. Se consideran dentro de la composición los cultismos o composición culta —palabras cuya morfología sigue muy estrechamente su origen etimológico griego o latino. Las voces en cuya formación intervienen estos elementos podrían no considerarse propiamente compuestas, ya que la mayoría de estas raíces no pueden aparecer de forma aislada, aunque tampoco pueden considerarse derivadas²¹ pues tienen un comportamiento peculiar que los aleja de los auténticos afijos; es más, tienen un significado léxico más próximo al de las raíces que al de los afijos²²; incluso estos elementos compositivos pueden ocupar tanto la posición inicial como la posición final en el compuesto frente a la posición predeterminada de los afijos, como se observa en:

cefalópodo o *cefalotórax* → *cefalo-* elemento compositivo prefijal

dolicocéfalo o *mesocéfalo* → *-céfalo* elemento compositivo sufijal

Gracias a la naturaleza léxica de los elementos compositivos, también llamados *pseudoprefijoideas* o *pseudosufijoideas*, no entrarían a formar parte del proceso derivativo sino compositivo, pues no son considerados como verdaderos morfemas derivativos que forman palabras nuevas al añadirlos a los materiales con los que cuenta la lengua, como sucede con los prefijos y sufijos —*re-tocar*, *joy-ero*.

Se llevó a cabo una búsqueda sobre un corpus de alrededor 53 000 palabras, donde para cada elemento prefijal se estudiaron en media 30 palabras asociadas y para cada elemento sufijal 18 palabras relacionadas. La Tabla 3 muestra información de los elementos compositivos estudiados, el número de elementos prefijales y sufijales considerados, el

²¹ Hay que tener en cuenta que los elementos compositivos prefijales no equivalen a una palabra derivada por prefijación (*bienvenir* es una palabra compuesta, no derivada del verbo *venir* con los prefijos *bi-* *en-* sino proviene del elemento compositivo *bien-*).

²² Un reconocimiento automatizado de compuestos formados por elementos compositivos tiene que considerar la distinción entre elementos compositivos y afijos: la combinación entre verdaderos afijos no produce palabras en la lengua (*in-dad, *pre-ción o cualesquiera otras combinaciones de prefijo y sufijo dan siempre malas formaciones léxicas, según SOLEDAD VARELA ORTEGA, *Morfología Léxica: La formación de palabras*, Gredos, Madrid, 2005).

número de compuestos encontrados y el total de palabras tratadas —contienen los elementos compositivos fijados.

Elementos compositivos	
Elementos prefijales	<p><i>estrati- porta- galo- urtic- talaso-trombo- pla- uro- bronco- men- mal- somato- pseudo- psico- hip- conco- urtic-ξoo- tel- vagin- ton- tono- temporo- mal- pict- geo- adipo- acu- vitro- hidr- xero- xilo- torξ- fil- hidr- cinam- cine- cinemat- cinema- cinemato- cinesi- cinesio- cineso- quin- mari- teno- peni- gen- dem- electro- hem- psitac- foto- ren- opo- neuro- cali- neo- sono- teo- term- mamó- tiro- faco- colo- fon- bio- querato- aer- gaso- luc- rodo- flori- calci- alo- cuadri- sico- viti- pluv- leg- bue- mis- icter- moto- acido- aceto- aceti- casei- acaro- alectoro- plesio- hepato- plagio-</i></p> <p>TOTAL: 1 647 Compuestos encontrados: 6 952 Palabras tratadas: 48 493</p>
Elementos sufijales	<p><i>-fero -oide -cito -osis -uro -forme -oma -anto -metro -logía -mero -fico -bio -lito -omo -mano -logo -grafo -oide -oideo -grado -terio -cola -arca -grafía -andro -geno -cario -cida -filo -uria -emia -baro -terapia -metría -megalia -cinesis -pirexia -branquia -ónfalo -ptosís -sepsia -epático -odoncia -frenia</i></p> <p>TOTAL: 274 Compuestos encontrados: 1 650 Palabras tratadas: 5 005</p>

Tabla 3.- Elementos prefijales y elementos sufijales.

Número de elementos constituyentes

Se hace necesario definir procesos automáticos que sean capaces de identificar y crear palabras compuestas y situarlas en un contexto lingüístico idóneo. Por ello, en un proceso automático de reconocimiento de compuestos hay que tener en cuenta el número de constituyentes que lo forman.

Por lo general, consta de dos elementos (97,7%), en menor medida existen casos de tres —*almempena, maldejo, quetetróncho, trampantojo*—, o incluso cuatro componentes —*correquetecagas, culodemalasiento, correveidile*—, aunque, en la mayoría de las ocasiones suelen insertarse elementos monosilábicos átonos como preposiciones, conjunciones copulativas, pronombres, artículos, entre otros, debido a que han sufrido un proceso de aglutinación sin pérdida de dicho elemento del sintagma nominal u oración original del que provienen.

dedodadama, porsiacaso, tentenelaire, azuliblanco, hijodalgo, ahimelollevas, besalamano, bienmesabe, bienteveo, cagalaolla, cenaaooscuros, diostedé, esperalaultima, guardalagua, juanencueros, lavatiquevoy, maldejojo, metomentodo, montambanco, pandesapo, pontentodo, quetetracho, quitaipón, rabodejunco, salsípuedes, tenconten, tentempié, tiraiafloja, vaivén

Formaciones con elementos patrimoniales

En cuanto a las combinaciones de palabras sin elementos compositivos cultos, pueden crearse a partir de conceptos que estén relacionados, o mejor dicho, de categorías gramaticales como sustantivos, adjetivos y verbos; por tanto, la estructura de los compuestos puede presentar múltiples formas²³, según la categoría gramatical de los componentes y la categoría del resultado final.

Con la variedad de formantes y de resultados que presentan, las relaciones entre los elementos participantes son de muy diversa índole —destaca la importancia que tiene el orden de los constituyentes, especialmente desde un punto de vista semántico. Según la frecuencia de aparición de la clase funcional del conjunto, uno de los compuestos más productivos son los que dan como resultado un *sustantivo* —composición más común y sirve para denominar objetos—, le sigue la composición adjetiva y, en menor medida, otras categorías gramaticales adicionales, tales como verbos, adverbios, pronombres, conjunciones, Figura 1.

Clasificación de los compuestos según clase funcional del conjunto y de los elementos constituyentes.

En función de la clase funcional de los elementos componentes se obtienen los siguientes esquemas de combinación más típicos:

Verbo+Complemento

Sustantivo+Sustantivo

Sustantivo+Adjetivo

²³ Hay que tener en cuenta la dificultad que supone establecer límites a la hora de designar una clasificación de compuestos en un terreno tan inestable como el de la composición.

Adjetivo+Sustantivo
Adjetivo+Adjetivo
Verbo+Verbo
Sustantivo+Preposición+Sustantivo

donde Bustos Gisbert plantea una clasificación atendiendo a las categorías gramaticales de los elementos componentes, así como la categoría del compuesto como se indica en la Tabla 4.

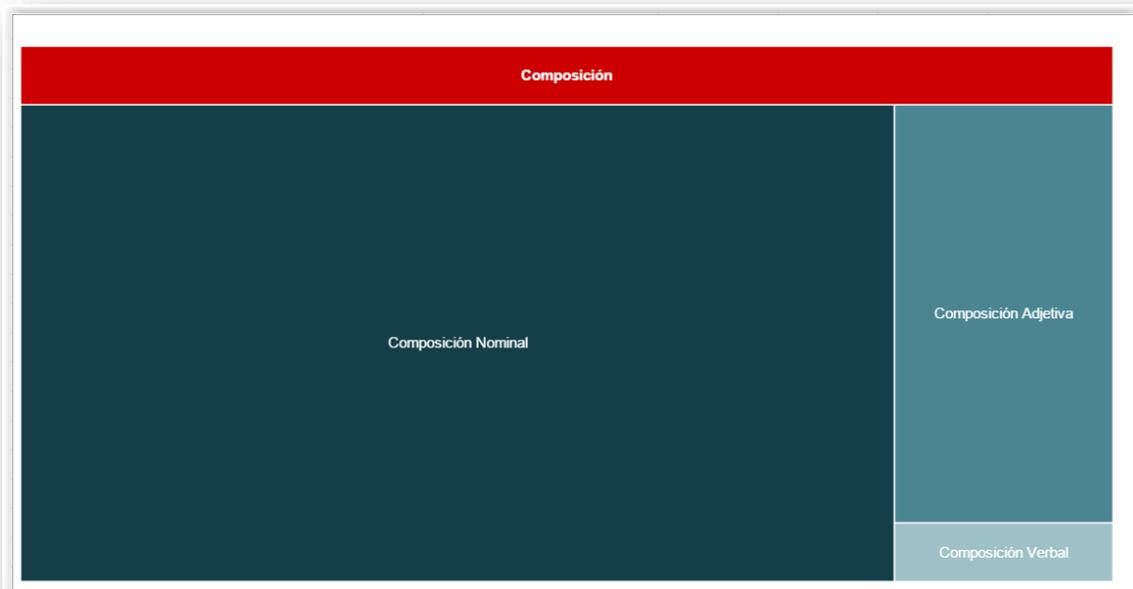
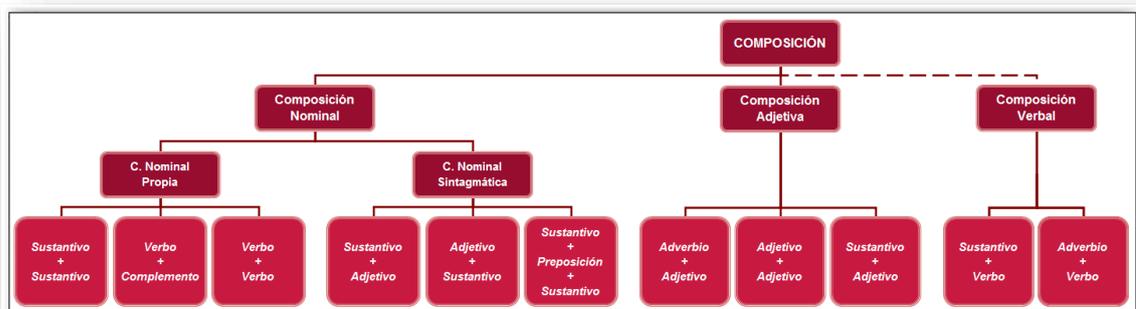


Figura 1.- Distribución de los tipos de composición atendiendo a la clase funcional del conjunto.



Composición Adjetiva		<i>Adverbio+Adjetivo</i> <i>Adjetivo+Adjetivo</i> <i>Sustantivo+Adjetivo -</i>
Composición Nominal	Composición Propia	<i>Sustantivo+Sustantivo</i> <i>Verbo+Verbo</i> <i>Verbo+Complemento</i>
	Composición Sintagmática	<i>Sustantivo+Adjetivo o</i> <i>Adjetivo+Sustantivo</i> <i>Sustantivo+Preposición+Sustantivo</i>

Tabla 4.- Clasificación de la composición según Bustos Gisbert.

Aunque existe un tipo de composición denominada *sintagmática*²⁴ —aquella que tiene como resultado unos compuestos que se aproximan bastantes a los sintagmas nominales correspondientes, pero no pueden ser considerados sintagmas nominales sino unidades léxicas—, se deja al margen la mayoría de estos compuestos en la automatización de su reconocimiento, ya que los componentes han alcanzado una coherencia semántica sin fusión gráfica. No obstante, en la base documental se encuentran, en menor cuantía, compuestos sintagmáticos que han sufrido un proceso de aglutinación de los elementos que lo forman Figura 5.

Por otro lado, la composición adjetiva y la composición propia nominal, también conocida como composición léxica, está formada por dos o más palabras o bases con alguna modificación fónica, generalmente con unión gráfica de los elementos que la componen. Desde un punto de vista histórico no existe una separación tajante entre los compuestos

²⁴ La composición sintagmática es un proceso innovador pues la mayoría de los estudios sobre formación de palabras no hacen mención a estas clases de compuestos, a excepción de Bustos Gisbert (1986), Val Alvaro (1999) y Buenafuentes de la Mata (2007), entre otros, siendo uno de los mecanismos empleados en la lengua latina y teniendo una productividad enorme en español. Con lo cual se puede llegar a afirmar, según Buenafuentes de la Mata, que la composición sintagmática es realmente el mecanismo que confiere vitalidad a la composición en español. Sin embargo, en la presenta Tesis no se estudia este tipo de composición.

sintagmáticos y determinados ejemplos de compuestos propios, por cuanto que estos, en ciertos casos, pueden ser el resultado de un proceso de aglutinación.

hoja de lata → *bojalata*

hidalgo → *hijo de algo*²⁵

Composición Nominal		
Composición Propia		Composición Sintagmática
Verbo + Complemento	Sustantivo + Sustantivo	Sustantivo + Adjetivo
	Verbo + Verbo	Adjetivo + Sustantivo
		Sustantivo + Prep. + Sustantivo

Figura 2.- Distribución de los compuestos nominales recogidos en la base documental atendiendo a la categoría gramatical de los elementos componentes.

En la Figura 2 se aprecia la distribución de la composición nominal teniendo en cuenta las categorías gramaticales de los elementos que la componen. Se destaca, dentro de

²⁵ Desde un punto de vista sincrónico, el no reconocimiento de grupos sintácticos en alguno de estos tipos de compuestos es un problema de tradición lingüística o de realidad de los hechos del lenguaje. Es por ello que se dejan al margen la mayoría de los compuestos sintagmáticos y por lo tanto no se van a considerar en el presente estudio.

la composición nominal, la combinación de *Verbo+Complemento*, como una de las más productivas²⁶, debido a la constitución interna que presentan así como a la comodidad que le produce a un hablante esta estructura por la semántica que sus compuestos implican, ya que caracterizan al referente a través de su actividad, como indica Bustos (1986). Le sigue la construcción *Sustantivo+Sustantivo*. En la Tabla 5 se aprecian diferentes formaciones de compuestos nominales.

	Sustantivo	Adjetivo
Sustantivo	<i>bocacalles casatienda carricoche telaraña</i>	<i>aguardiente hierbabuena artimaña pelirrojo</i>
Adjetivo	<i>ciempiés malasangre mediodía malpaís</i>	
Verbo	<i>bincapié trotamundos quemasangres cubrecamas saltarrostro quitapenas</i>	
Adverbio	<i>bienandanza menoscuenta bienquerencia</i>	<i>bienpensante maleducado</i>
	Verbo	Adverbio
Verbo	<i>compraventa duermevela ganapierte vaivén</i>	<i>bogavante atalejo mandamás</i>
Adverbio	<i>bienmesabe</i>	

Tabla 5.- Compuestos nominales según la categoría gramatical de sus elementos componentes.

Con respecto a la composición sintagmática, se aprecia en la Figura 2 que la combinación *Sustantivo+Prep+Sustantivo* es más numerosa que la *Sustantivo+Adjetivo*; sin embargo, no se considera en este estudio aquellos compuestos donde los elementos que intervienen no se han consolidado gráficamente y aparecen separados por un guion o nexos (espacio o preposición), como se aprecia en la Figura 5.

Existen compuestos nominales que representan sistemas marginales o bien resulta escaso el número de casos que producen o son casos particulares de los anteriores:

PREPOSICIÓN+CONJUNCIÓN CONDICIONAL +ADVERBIO DE MODO: *por si acaso*.

NUMERAL+PREPOSICIÓN+SUSTANTIVO: *cientoemboca, milenrama*.

²⁶ Es el más importante cuantitativamente hablando porque es aquel del que se recogen más ejemplos en todos los repertorios lexicográficos consultados.

VERBO+CONJUNCIÓN: *creique, penseque*.

Los compuestos que constan de más de dos piezas léxicas dan como resultado un sustantivo formado por las combinaciones, entre otras:

VERBO+Y+VERBO: *tiraiafloja, quitaipón*.

VERBO+VERBO+Y+VERBO: *correveidile*.

ADVERBIO+PRONOMBRE PERSONAL+VERBO: *bienmesabe*.

VERBO+PRONOMBRE PERSONAL+ PRON. IMPERSONAL: *sabelotodo*.

VERBO+PRONOMBRE PERSONAL+VERBO: *hazmerreír*.

VERBO+PRONOMBRE PERSONAL+PREPOSICIÓN+ PRON. IMPERSONAL: *metomentodo*.

El segundo caso en productividad son los compuestos *adjetivos*. La estructura más profusa es *Sustantivo+Adjetivo* donde existe una marcada relación entre los elementos del compuesto ya que el segundo miembro predica una cualidad del primero, sustantivo²⁷ que generalmente designa partes exteriores del cuerpo humano o animal. Existen componentes que parecen admitir dos interpretaciones diferentes —*adjetivos y sustantivos*—, y no se posee información suficiente para decidir cual resulta más correcta —*cardocuco “cardo silvestre”* no está claro si *cuco* es adjetivo o sustantivo. En la Figura 3 se aprecia la distribución de la composición adjetiva, atendiendo a las clases funcionales de los elementos que la componen. En la Tabla 6 se aprecian diferentes formaciones de compuestos adjetivales.

	Adjetivo (o Participio)
Sustantivo	<i>alicaído cejijunto cariparejo pelirrojo teticiega patidifuso cuellilargo faldicorto ojitruco lengüilargo</i>
Adjetivo	<i>tonticiego grandilocuente agridulce hispanohablante rectinervio cultipicaño</i>
Adverbio	<i>bienintencionado, malaconsejado bienhablado</i>

Tabla 6.- Compuestos adjetivales según la categoría gramatical de sus elementos componentes.

²⁷ Existe una marcada tendencia a colocar en el primer miembro sustantivos bislabos.

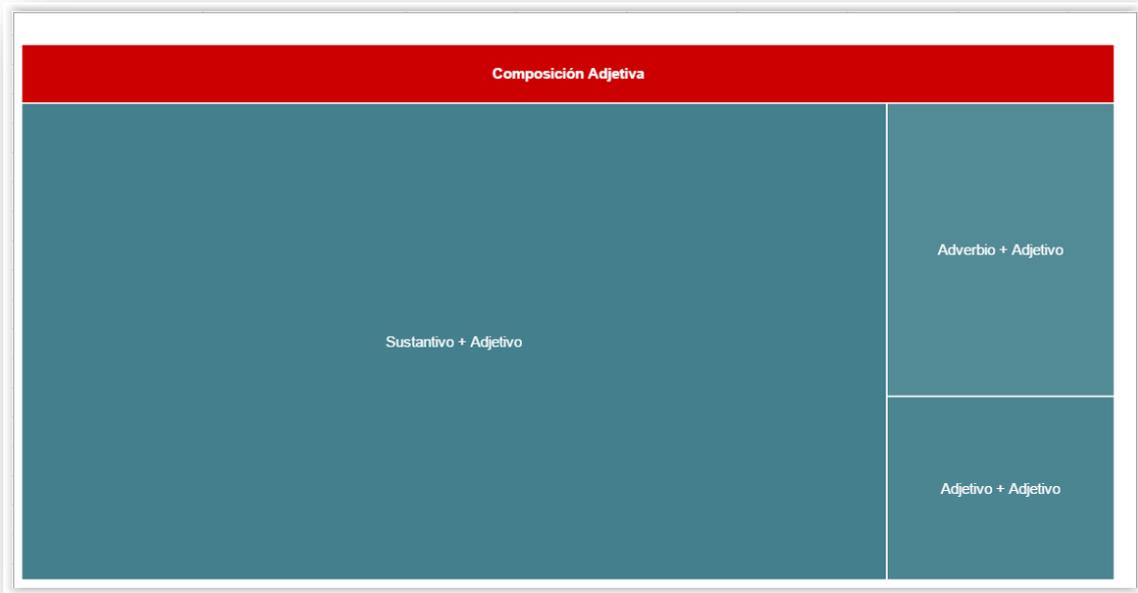


Figura 3.- Distribución de los compuestos adjetivales recogidos en la base documental atendiendo a la categoría gramatical de los elementos componentes.

Por regla general, el resultado de fusionar dos palabras para formar un compuesto es un sustantivo o un adjetivo. Pero existen, aunque en menor medida, diversas categorías gramaticales adicionales:

- ❖ Cuando dan lugar a *verbo*. No se ha detectado que esta categoría gramatical combine dos constituyentes de su misma categoría gramatical, es decir, *Verbo+Verbo*, constatándose esta misma restricción en otras lenguas. En la Figura 4 se aprecia la distribución de la composición verbal, atendiendo a las clases funcionales de los elementos que la componen.

SUSTANTIVO+VERBO: *maniatar, aliquebrar, manuscibir, heroificar*.

ADVERBIO+VERBO: *maldecir, malcomer, bienpensar, menospreciar, malcasar*.

- ❖ Que dan lugar a *adverbio*

CONJUNCIÓN+VERBO: *siquiera* (adverbio o conjunción).

ADVERBIO DE CANTIDAD+ADJETIVO O ADVERBIO: *tampoco*.

ELEMENTO COMPOSITIVO+ADVERBIO DE TIEMPO: *anteayer*.

ADJETIVO+ADVERBIO DE MODO: *otrosí*.

Raros:

ADVERBIO+VERBO: *dondequiera*.

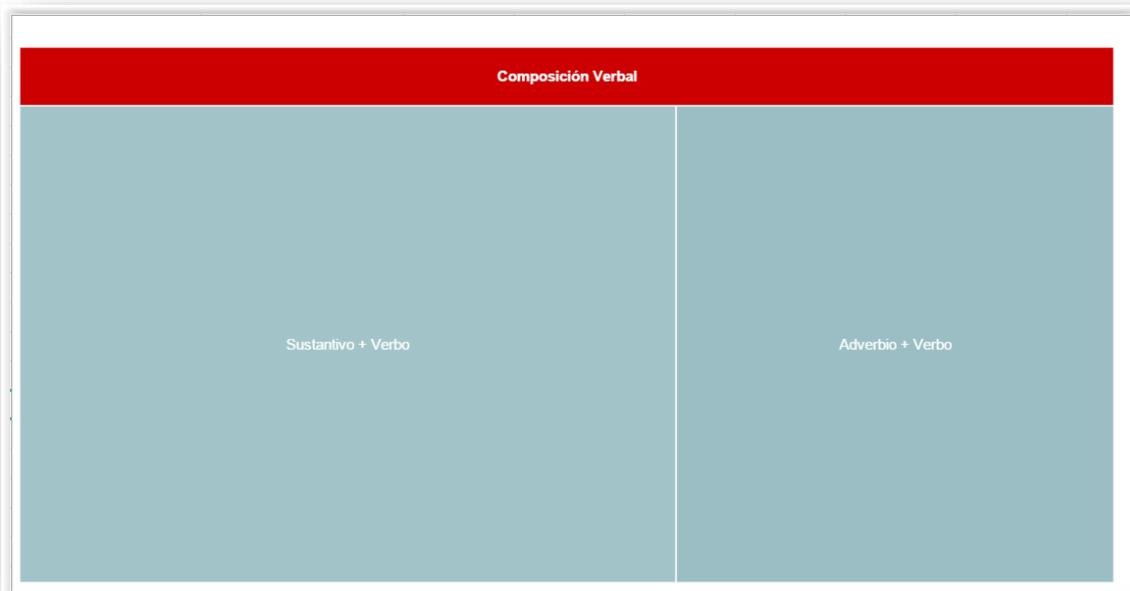


Figura 4.- Distribución de los compuestos verbales recogidos en la base documental atendiendo a la categoría gramatical de los elementos componentes.

❖ Que dan lugar a *pronombres*:

PRONOMBRE RELATIVO+VERBO: *cualquiera, quienquiera*.

PRON.PERSONAL+ADJETIVO: *nosotros*.

❖ Que dan lugar a *conjunciones*:

ADVERBIO DE TIEMPO O DE MODO + PRONOMBRE RELATIVO: *aunque*.

PREFIJO+SUSTANTIVO U.T.C. CONJUNCIÓN ADVERSATIVA: *sin embargo*.

CONJUNCIÓN+ADVERBIO DE MODO: *sino*.

❖ Que dan lugar a *numerales*:

veinticinco → *veinte+cinco*, *dieciséis* → *diez+seis*

En definitiva, según la frecuencia de aparición de la categoría gramatical del compuesto, así como de sus constituyentes, se pueden detallar las diferentes formaciones de compuestos como muestra la Tabla 7 —la intensidad del color marca la mayor o menor frecuencia de la combinación.

CATEGORÍA GRAMATICAL COMPUESTO	CATEGORÍA GRAMATICAL ELEMENTOS COMPONENTES				
	SUSTANTIVO	ADJETIVO	VERBO	ADVERBIO	
SUSTANTIVO	SUSTANTIVO	<i>casatienda bocamanga carricoche telaraña</i>	<i>aguardiente hierbabuena artimaña pelirrojo</i>		
	ADJETIVO	<i>ciempiés malasangre mediodía malpaís</i>			
	VERBO	<i>hincapié matamoscas quemasangres salirrostro quitapenas trotamundos</i>		<i>compraventa duermeverla ganapierde tirajalloja quitapón correvidile hazmerreír</i>	<i>bogavante atalejo mandamás creique penseque sabelotodo metomentodo</i>
	ADVERBIO	<i>bienandanza menoscuenta bienquerencia cientoemboca milenrama</i>		<i>bienmesabe</i>	<i>porsiacaso</i>
ADJETIVO	SUSTANTIVO		<i>alicaído cejiunto pelirrojo teticiega patidifuso</i>		
	ADJETIVO		<i>tonticiego grandilocuente agridulce hispanohablante</i>		
	VERBO				
	ADVERBIO		<i>bienintencionado</i>		
VERBO	SUSTANTIVO		<i>maniatar aliquebrar manuscibir</i>		
	ADJETIVO				
	VERBO				
	ADVERBIO		<i>maldecir malcomer bienpensar menospreciar</i>		
ADVERBIO, PRONOMBRE, CONJUNCIÓN,...	SUSTANTIVO				
	ADJETIVO			<i>otrosi</i>	
	VERBO				
	ADVERBIO	<i>tampoco nosotros</i>	<i>siquiera dondequiera cualquiera quienquiera</i>	<i>anteayer aunque sino</i>	

Tabla 7.- Clasificación de compuestos atendiendo a la categoría gramatical composicional y las categorías de sus elementos componentes.

Clasificación de los compuestos atendiendo a su estructura formal

Por otro lado, según se recoge en el trabajo de Alvar Ezquerro (2003), la composición se sirve de procedimientos para la creación de palabras nuevas como: la sinapsia, disyunción, contraposición, yuxtaposición, prefijos vulgares y acortamiento —abreviatura, abreviatura simple, acronimia, abreviatura compuesta, abreviatura compleja. El más caudaloso de los procesos de composición es la *yuxtaposición o lexías compuestas*, donde la fusión gráfica de los elementos participantes en el compuesto es total, así como su lexicalización y su gramaticalización, como se aprecia en:

malsano, agridulce, quitaipón, boquicerrado, carnicol, malqueda, cochitril, hincapié

Sin embargo, la unión de los miembros de la sinapsia es de naturaleza sintáctica, no morfológica, por lo que es difícil determinar si se ha producido lexicalización o no. Suele existir un nexo de unión entre las dos palabras que dan lugar al nuevo término, generalmente con las preposiciones *de* y *a*:

pan de azúcar, paso a nivel, cuerda sin fin, flor de la abeja

estos compuestos se corresponde con la clasificación sintagmática que hace Bustos Gisbert de *Sustantivo+Preposición+Sustantivo*.

Por más que la lexicalización sea un hecho, la *disyunción* da origen a un tipo de lexías en la que los dos elementos participantes no se han soldado gráficamente, por más que la lexicalización sea un hecho:

alta mar, peso pluma, pájaro mosca

se corresponde con diversos compuestos *Sustantivo+Sustantivo* o *Sustantivo+Adjetivo* o *Adjetivo+Sustantivo*. En tales composiciones se llega incluso a la unión gráfica de los elementos:

tela de araña→telaraña

agua nieve→aguanieve

ave fría→avefría

En un grado más alto de unión gráfica figura la *contraposición*, donde los elementos que participan se escriben unidos por un guion, aunque generalmente el resultado aparecerá sin el debido a las restricciones del uso del guion:

coche bomba → *coche-bomba*
falda pantalón → *falda-pantalón*

Muchos autores consideran el *acortamiento* como un procedimiento de formación de nuevas palabras o neologismos que, por su naturaleza, escaparían en principio, a una teoría morfológica —no la asumen como parte de la composición. La frontera entre derivación y composición no resulta clara, sobre todo en el caso de abreviamiento:

coyotomate → *coyote+tomate*

o en la acronimia²⁸ —consiste en unir el inicio de una palabra con el final de otra—, como:

informática → *información automática*
tergal → *poliéster galo*
secrefata → *secretaria azafata*
itañol → *italiano español*
autobús → *automóvil ómnibus*
cantautor → *cantante autor*
transistor → *transfer resistor*
eurovision → *europa televisión*
tecnotrónica → *tecnología electrónica*

o en el caso del abreviamiento, la cual tiene diversas variantes, se considera el acortamiento léxico, como:

cine → *cinematógrafo*
foto → *fotografía*
bus → *autobús*

²⁸ JOSE ALBERTO MIRANDA, *Formación de palabras en español*, Ediciones Colegio de España, 1994. pág. 161.

profe → *profesor*

mate → *matemáticas*

aunque algunos se consideran en la base de elementos compositivos prefijales o sufijales según proceda:

euro- → *europa*

zoo- → *zoológico*

tele- → *televisor*

En la definición de composición considerada, se debe tener en cuenta que un neologismo compositivo ha de funcionar como una unidad léxica, que signifique un concepto unitario y designe una realidad específica. Por ello, algunos autores han clasificado la composición como:

- ❖ compuestos imperfectos, aquellos que se escriben separados por guion —*físico-químico*—, u otros, aunque tengan unidad de significado, se escriben con signos diferentes o con más elementos —*guerra civil, golpe de estado*.
- ❖ compuestos perfectos, forman una unidad no sólo en su significado sino también en la escritura y en su funcionamiento gramatical.

Los compuestos perfectos son los más abundantes y en los que se centra el presente estudio.

Según la categoría gramatical de los elementos que integran los compuestos perfectos o yuxtapuestos, se aprecia en la Figura 5 los tipos nominales, adjetivales y verbales que se recogen en la base de referencia. Se contrastan con los compuestos imperfectos que aparecen separados por un guion o nexo —espacio en blanco o preposición, principalmente *de*. Se observa una proliferación de casos de compuestos con nexos en la categoría composicional nominal de *Sustantivo+Adjetivo* frente a las categorías nominal y adjetiva de los compuestos yuxtapuestos.

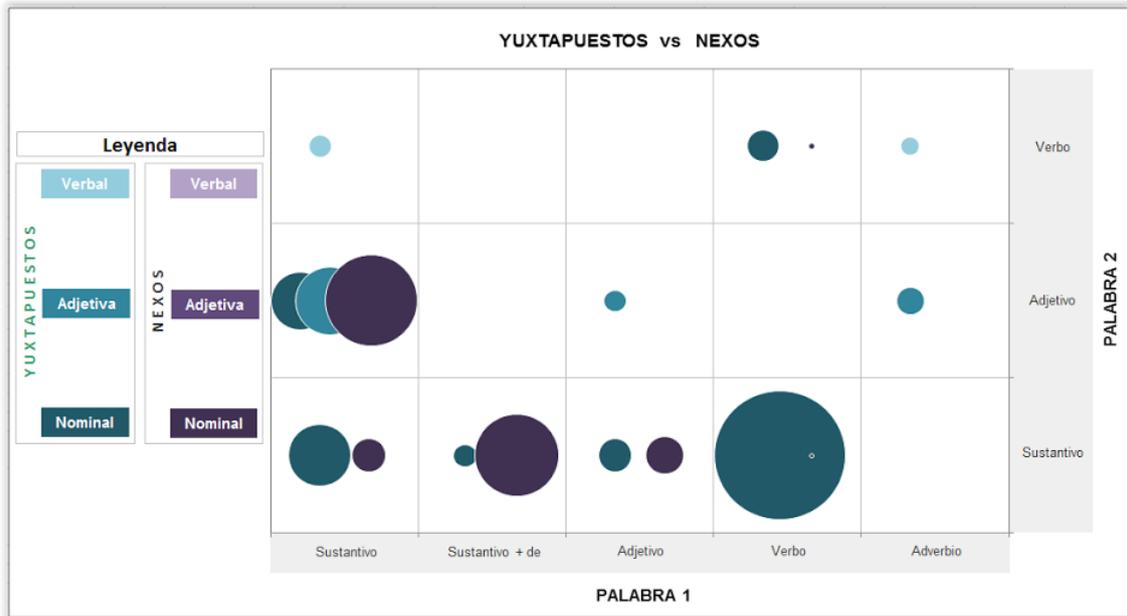


Figura 5.- Distribución de la composición nominal, adjetiva y verbal en la base documental atendiendo a las categorías gramaticales de los elementos componentes.

Otros procesos de formación de palabras

Los adjetivos acabados en *-mente* y las preposiciones

Se hace una mención especial a la combinación de una preposición con otra categoría gramatical, como por ejemplo:

PREPOSICIÓN + SUSTANTIVO: *sobredosis, entreacto, contradanza*

PREPOSICIÓN + VERBO: *sobredimensionar, contradecir, entresacar*

PREPOSICIÓN + ADJETIVO: *sobreabundante, contrachapado, entremedio*

pues generalmente se suelen confundir estos casos con composición cuando en realidad se trata de un proceso derivativo como indica Varela (1990). Alvar (2002) considera la formación de palabras mediante prefijos vulgares como parte de la composición cuando estos prefijos coinciden con las preposiciones, esto es, se unen dos elementos independientes de la lengua.

No se trata la formación del tipo:

ADJETIVO + *-mente* → ADVERBIO: *intrínsecamente, fatídicamente*

aunque algunos autores²⁹ la reconocen como un proceso de composición, figura consolidada como una formación derivativa y así lo recoge el Diccionario General de la Lengua Española³⁰ (VOX) pues considera *-mente* como un sufijo, no un elemento sufijal —entra en la formación de adverbios de modo pospuesto a los adjetivos en su forma femenina³¹ —*malamente*.

Parasíntesis por composición

En los procesos parasintéticos se aprovecha simultáneamente bien la derivación y composición³² —*ropavejero, doceañista, picapedrero, quinceañera, sietemesino*—, o bien se tienen simultáneamente prefijación y sufijación —*antibacteriano, contrarreformista*— denominando a estos procesos *parasintéticos por derivación*, cuyo estudio no entra en la presente Tesis.

Este trabajo se centra en la denominada *parasíntesis por composición* donde existen palabras que se forman a partir de una combinación de otras dos más una sufijación que siguen la estructura *A+B+Sufijación*, según Serrano Dolader (1995):

- no existe la combinación *A+B* —se consideraría derivación de palabra compuesta.
- tampoco existe la palabra *B+Sufijación* —supondría derivado por sufijación; en caso de aparecer, no serían el origen de la palabra final.

Este es el caso:

burgamandera → no existe *burgamanda** ni *mandera**

anquimuleño → no existe *muleño**

²⁹ Véase WALDO PÉREZ CINO, *Manual Práctico de formación de palabras en español I*, Verbum, Madrid, 2002.

³⁰ *Diccionario General de la Lengua Española VOX*, Barcelona, 2003.

³¹ El adjetivo adopta siempre la forma femenina, si la tiene, pues *-mente* es femenino en latín y conserva el acento.

³² Se tiene que señalar claramente la distinción con derivación de palabras compuestas —como *librecambista* cuya relación morfológica es la forma primitiva *librecambio*—; al igual que no se ha de considerar que la palabra compuesta pueda soportar que alguno de sus elementos constituyentes se encuentre derivado.

misacantano → no existe *misacanta** ni *cantano**

ropavejero → no existe *vejero**

doceañista → no existe *añista**, ni es derivado de *doceño*

aguamanil → no existe *mano+il*, ni es derivado de *aguamano*

La parasíntesis es un proceso bastante complejo en el que se añaden dos o más morfemas, sin que exista claramente una forma más simple intermedia. André Martinet³³ propuso analizar la parasíntesis, como un esquema de formación de palabras en el que la forma léxica resultante tiene la forma:

$$[M_1^L + M_2^L + M^{nL}]M_3^L$$

*Morfema liberable*³⁴ + *Morfema liberable* + *Morfema no liberable*

Por tanto, una condición importante para considerar un proceso lexicogénico de formación de palabras, como la parasíntesis, debe involucrar simultáneamente la composición y la derivación por sufijación, ocurriendo esta triple conjunción de una sola vez. Esto implica que ni el compuesto ni el derivado deben existir anteriormente en la lengua³⁵. Un ejemplo de este proceso: *aguamanil* y *quinceaño*

$$[agua + man(o) + (-il)] = aguamanil$$

³³ ANDRÉ MARTINET, Gramática funcional del francés. Barcelona, Ariel, 1984 [1979].

³⁴ Un morfema liberable normalmente lo constituye un lexema que puede ocurrir en forma libre o en otras combinaciones de palabras y un morfema no liberable forma habitualmente un afijo derivativo.

³⁵ Es importante distinguir la parasíntesis que es un único proceso de la aplicación doble de derivación, donde se reconoce la ocurrencia de un morfema liberable intermedio. Un ejemplo de esto es la palabra *mariposita* y *paraguazo* cuya formación puede analizarse así:

$$[[M_1^L + M_2^L] M_3^L + M^{nL}] M_4^L - M_nL$$

$$[[\text{mari-} + \text{pos(ar)}] \text{-ita}]$$

$$[[\text{para} + \text{agua(s)}] \text{-azo}]$$

donde se reconoce claramente una forma intermedia *mariposa* y *paraguas*, respectivamente, de las cuales a su vez se derivan mediante la adición de un sufijo *mariposita* y *paraguazo*, precisamente el que sea reconocible una forma intermedia es lo que impide considerar en el proceso la parasíntesis.

[*quince* + añ(o) + (-ero)] = *quinceaño*

donde se reconocen dos morfemas liberables: *agua* + *mano* y un morfema no liberable o ligado como es *-il*. Bajo esta propuesta de Martinet se esclarece el proceso de parasíntesis, que considera que responde a sus propias reglas.

¿Cómo estudiar los compuestos?

Existe un polimorfismo subyacente bajo las formaciones compuestas pues los compuestos son unidades lingüísticas muy especiales que presentan características de tipo sintáctica — debido a su estructura casi oracional—, morfológicas —se trata de un proceso de formación de palabras—, y léxicas —al fin y al cabo, un compuesto es una unidad léxica. A través de un análisis histórico, los compuestos reflejan claramente las conexiones que se producen entre sintaxis, morfología y léxico, como se aprecia en el siguiente ejemplo:

tela de araña (compuesto imperfecto³⁶ o impropio) → *telaraña* (compuesto perfecto)

En el tratamiento automatizado de la composición, en el que se pretende hacer una contribución a la identificación automatizada de compuestos, se deben proyectar los distintos tipos de compuestos —nominales, adjetivales y verbales— en cuatro niveles de estudio:

- ❖ Fónico: en el que se tratan como unidades monoacentuadas y se integran las dos unidades léxicas —se obvia aquellas que contengan guion (-) o nexos (espacio o preposición).
- ❖ Morfológico: se considera el número de unidades que formen el compuesto, generalmente dos, así como su naturaleza; se tiene en cuenta el comportamiento de los elementos componentes con respecto a la variación del género y del número. Además se busca la existencia de marcas de composición —inclusión del interfijo

³⁶ Algunos autores también la reconocen como composición sintagmática.

“¿”, es tratado en el epígrafe *Las reglas de formación de compuestos*, sin dejar de lado la frontera entre derivación, composición y parasíntesis.

- ❖ Sintáctico: se comprueba si revela un comportamiento como unidades de funcionamiento tales que:
 - Exista un orden fijo de constitución de los compuestos específicamente.
- ❖ Semántico³⁷: en el proceso de creación de nuevas palabras se tratan como una unidad y se estudian las reglas de naturaleza semántica de compuestos conforme a:
 - las relaciones de endocentrismo³⁸ —aquel que representa una especialización con respecto a su núcleo referencial, como por ejemplo: *agua de nieve*, *lavavajillas*,...— y exocentrismo —aquel que designa una realidad no referida por ninguno de los elementos componentes, como por ejemplo: *agua sal*, *aguardiente*...
 - la función del compuesto como elemento definidor o como elemento caracterizador del referente, y los valores connotativos que puede tener el compuesto en virtud de esta función.

³⁷ Para EUGENIO COSERIU (“Los procedimientos semánticos en la formación de palabras”, *Odisea* n°3, págs. 179-189, 2003): “las diferentes teorías sobre la formación de palabras —tanto las teorías tradicionales como la mayoría de las teorías modernas— no pueden responder adecuadamente a este fenómeno, en el mismo sentido que intuitivamente reconocen los hablantes y que se manifiesta en la actividad lingüística de estos... Se confunden designación y significado de lengua... La no distinción entre designación y significado de lengua, el caso más sintomático es, sin duda, el de los compuestos llamados endocéntricos —designan algo que constituyen al mismo tiempo su *determinatum*— y exocéntricos —lo que designan no es algo que no expresan ellos mismos.

³⁸ Véase JUAN MANUEL PÉREZ VIGARAY, *La composición nominal en español*, Tesis Doctoral, Universidad Las Palmas de Gran Canaria, 1994.

Sin embargo, se deja para estudios posteriores el tratamiento semántico de los compuestos endocéntricos³⁹ —su automatización resulta artificiosa por su construcción—, frente a los compuestos exocéntricos que escapan actualmente al dominio informático —su significado no se desprende de la suma de los de sus integrantes, sino que ostenta un significado añadido, imprevisible y de forma figurada generalmente.

³⁹ La construcción endocéntrica es un constituyente compuesto que consta de un núcleo, que es elemento indispensable, y de otros elementos prescindibles que complementan al núcleo.(IGNACIO BOSQUE, JAVIER GUTIÉRREZ-REXACH, *Fundamentos de Sintaxis Formal*, Akal (1ª edición), Madrid, 2009).

Conclusiones del capítulo → En la presente Tesis se procede a estudiar, desde un punto de vista morfológico, los compuestos yuxtapuestos o lexías compuestas nominal y adjetiva principalmente, al igual que algunos casos especiales de acortamiento, elementos compositivos prefijales/sufijales y parasíntesis por composición. Los restantes tipos no se consideran debido a la dificultad para justificar que constituyen un verdadero compuesto en español, ya que habría que tener en cuenta factores sintácticos y semánticos que escapan al dominio informático. Se dejan al margen la mayoría de los compuestos de tipo sintagmático, ya que los componentes han alcanzado una coherencia semántica pero sin fusión gráfica debido a que pueden entrar en conflicto con las colocaciones léxicas en la automatización de su reconocimiento. Es importante tener presente el *orden fijo* de los elementos constituyentes con respecto a la categoría gramatical composicional —no se han encontrado compuestos verbales o adjetivales formados por un *Verbo+Sustantivo*. Se deja para estudios posteriores el tratamiento semántico de los compuestos endocéntricos —su automatización resulta artificiosa por su construcción—, frente a los compuestos exocéntricos que escapan actualmente al dominio informático —su significado no se desprende de la suma de los de sus integrantes, sino que ostenta un significado añadido, imprevisible y de forma figurada generalmente.

CAPÍTULO 2.- NEOLOGISMOS COMPOSITIVOS EN ESPAÑOL

RESUMEN: Este capítulo presenta un estudio sobre los neologismos compositivos, probablemente sea el proceso más universal de formación de palabras que tienen las lenguas para su enriquecimiento.

Introducción

Aunque tradicionalmente la composición ha jugado un papel secundario frente a la derivación en la formación de palabras en español, tanto desde la perspectiva de su productividad como por la escasez de bibliografía, se observa la presencia cada vez mayor de este mecanismo para crear neologismos compositivos, llamados *neologismos morfosintánticos*, donde Romero Gualda (1978: 176) los define como aquellos:

«productos léxicos conseguidos por derivación, composición o parasíntesis con fines expresivos, surgidos en un contexto poético y creados por un autor que en un momento dado siente la necesidad de crear como una afirmación de su libertad de expresión, como una muestra de originalidad frente a la lengua común a la que en consecuencia considera insuficiente, pobre o poco precisa»

La creación de neologismos se produce por modas o bien por necesidades de nuevas denominaciones o significados, que acaban por incorporarse al vocabulario de una determinada lengua, ya que la sociedad —principal fuente de creación léxica— actualmente favorece la creación de términos que corresponden a nuevos conceptos tanto materiales como intelectuales⁴⁰: ya se trate “de un progreso industrial, de una modificación de la vida social, de un movimiento ideológico, de una nueva manera de sentir o de comprender, de un enriquecimiento del dominio moral, el neologismo constituye una necesidad imperiosa, es

⁴⁰ Véase GLORIA GUERRERO RAMOS, *Neologismos en el español actual*, Arco/Libros, Madrid, 2010, [1995].

una de las manifestaciones principales de la vitalidad de una lengua”. Casado Velarde⁴¹ (2015) utiliza el concepto de innovación léxica por medios morfológicos en la creación instantánea de neologismos y plantea que “los diccionarios y gramáticas, si quieren reflejar el estado de la lengua de cada momento, tendrán que ir cambiando la foto del idioma; los grandes innovadores lingüísticos son los poetas, pero también los grandes pensadores, inventores, personas dedicadas a la ciencia y a sus aplicaciones, periodistas y otros profesionales de la comunicación”.

Los procedimientos neológicos que operan sobre las unidades de los sistemas lingüísticos, la neología de forma o morfosintácticos —fabricación de nuevas unidades léxicas a partir de elementos que pertenecen a un sistema morfológico de la lengua en cuestión o a sistemas extranjeros antiguos o actuales— frente a la neología de significado —utilización de un término ya existente en la lengua con un contenido semántico nuevo— constituye uno de los procedimientos de construcción de neologismos más productivos en español, y en particular la composición culta, productora continua de numerosas palabras técnicas y científicas.

El propósito de este capítulo es presentar un estudio analítico cuantitativo y cualitativo a partir de información obtenida en el banco de neologismos BOBNEO⁴² vs la base documental BARGCoS⁴³, con el fin de establecer cuáles son los procedimientos de creación léxica más fructíferos en la fabricación de neologismos, con énfasis en el mecanismo de la composición —probablemente el proceso más universal de formación de palabras que tienen las lenguas.

⁴¹ MANUEL CASADO VELARDE, *La innovación léxica en el español actual*, Síntesis S. A., Madrid, 2015.

⁴² La plataforma BOBNEO se puede consultar en la siguiente dirección: <<http://obneo.iula.upf.edu/bobneo/index.php>> [01/10/2015]. Se trata de una base de datos del Observatorio de Neología (OBNEO) de la Universidad Pompeu Fabra (UPF) que ofrece los neologismos léxicos procedentes de los medios de comunicación, escritos y orales, en catalán y español, recogidos desde 1992. La herramienta ofrece una gran variedad de combinaciones para la selección de las posibles búsquedas por fecha, región, fuente y, por supuesto, tipo de neologismo.

⁴³ Base de datos de la herramienta ARGCoS “Reconocimiento y Generación Automática de Compuestos en español”, que recoge alrededor de 11 000 compuestos recopilados de diversas fuentes lexicográficas (cf. Capítulo 3, *Corpus seleccionado*, pg.48 y Capítulo 4, *Base de datos de las soluciones*, pg.88).

Los neologismos

Citando a Rafael Hidalgo de la Torre (2010) en su libro “Sugerencias sugestivas con las palabras”:

«...Las palabras atraen, repelen, evocan, sugieren. Se dejan descomponer en letras, que a su vez, pueden combinarse formando nuevas palabras. Sus sonidos pueden resonar en nuestros oídos con matices poéticos inauditos. La maleabilidad de las palabras permite una infinidad de juegos que nos conducen, con mano segura y divertida a las entrañas mismas de la lengua.... Jugando con las palabras conseguimos: crear neologismos y observar el entorno de forma nueva e inesperada...»

Según Álvarez de Miranda⁴⁴ (2009) todos los vocablos de una lengua han sido alguna vez un neologismo, por ello el concepto de neología es fácilmente formulable como la incorporación al léxico de un nuevo elemento. El Diccionario General de la Lengua Española⁴⁵ define neologismos como:

*“vocablo acepción o giro nuevo en una lengua;
proviene de neo- (palabra) y del gr. logismós (razonamiento)”*

La incorporación de un neologismo a los diccionarios se ha considerado a menudo como prueba de su afianzamiento. Las causas para la aparición de neologismos van, desde su establecimiento en la gramática por un uso cada vez mayor y más extendido entre la población, hasta la necesidad emergente de designar nuevos conceptos ligada al desarrollo científico-tecnológico. Las características de una sociedad moderna altamente tecnificada hace que se presenten situaciones de necesidad de un término nuevo. Esta necesidad de creación de léxico especializado presenta dos fases diferentes⁴⁶:

- En un primer momento se produce la acuñación de un término por los propios especialistas.

⁴⁴ PEDRO ÁLVAREZ DE MIRANDA, “Neología y pérdida Léxica”, en ELENA DE MIGUEL (eda.), *Panorama de la lexicología*, Ariel, Barcelona, 2009, págs. 133-158.

⁴⁵ *Diccionario General de la Lengua Española VOX*, op. cit.

⁴⁶ JOSEFA GÓMEZ DE ENTERRÍA, “Últimas tendencias neológicas en la prensa económica” en M.T. CABRÉ, J. FREIXA y E. SOLÉ (eds.), *La neología en el tombant de segle*, Observatorio de Neologia IULA Barcelona, págs. 75-84, 2000.

- En una segunda fase, se lleva a cabo la difusión de los neologismos entre los miembros que forman la comunidad lingüística, con la consiguiente aceptación o rechazo de los mismos hasta lograr su instalación en el sistema de la lengua. Para Alarcos⁴⁷:

«un vocablo se despoja de su carácter neológico cuando pasa inadvertido entre todos los demás tradicionales»

En las lenguas mejor estudiadas y mejor servidas lexicográficamente, según comenta Álvarez de Miranda existen editoriales que renuevan anualmente sus diccionarios para incorporar los nuevos neologismos, han proliferado los observatorios de neología, las antenas neológicas o los equipos de vigilancia neológica.

Clasificación de los neologismos

Según B. Arrieta⁴⁸ el dinamismo que caracteriza al lenguaje permite la generación de nuevos términos que incrementan los repertorios lexicales de los distintos registros discursivos. Tal dinamismo enriquecedor del lenguaje es más notorio, aunque no exclusivo, en el registro científico debido al constante desarrollo de nuevas tecnologías que necesitan una rápida difusión a través de los diferentes medios de comunicación. Esta divulgación trae consigo la necesidad inmediata de asignarle una palabra unívoca a la cosa inventada que sirva para su identificación y es así como se producen los neologismos. El hecho de detectarlos e inventariarlos permite realizar un seguimiento eficaz de su futuro e implementación⁴⁹. Corresponde entonces, tras haber recolectado un corpus considerable de estos términos,

⁴⁷ EMILIO. ALARCOS LLORACH, “Consideraciones sobre el neologismo”, *El neologismo necesario*, Fundación EFE, Madrid, págs. 19-29, 1992.

⁴⁸ BEATRIZ ARRIETA, RAFAEL MEZA, JUDITH BATISTA, “Procedimientos morfológicos para la creación de neologismos en el discurso académico”, *Revista de Investigación Lingüística*, nº 13, págs. 219-240, 2010.

⁴⁹ Véanse también los estudios realizados por M. CABRÉ (“La clasificación de los neologismos: Una tarea compleja”. *Alfa*. São Paulo, 50 (2), págs. 229-250, 2006), J. MARTÍN CAMACHO (“Los procesos neológicos del léxico científico. Un esbozo de clasificación”, *Anuario de Estudios Filológicos*, vol. XXVIII, págs.157-174, 2004), I. CHUMACEIRO (“Morfolología”. *Español Actual* vol. 69, págs.51-66, Madrid, 1998), donde ha quedado demostrado el gran caudal de términos aportados al idioma español por los académicos universitarios a través del lenguaje científico-técnico.

agruparlos y clasificarlos. Para ello Cabré Castellvi⁵⁰ aporta una clasificación o tipología neológica que distingue los siguientes tipos excluyentes entre sí:

- ✓ Neologismos de forma o formales: nuevas unidades formales, antes inexistentes en el léxico de la lengua, que incluyen los procedimientos de formación de palabras como la derivación —prefijación y/o sufijación—, composición y composición culta, acronimia, sintagmación, entre otros procesos formativos.
- ✓ Neologismos sintácticos: implica cambio de subcategoría gramatical en una base léxica.
- ✓ Neologismos semánticos: formados por la modificación de una base léxica.
- ✓ Préstamos: unidades importadas de otras lenguas.
- ✓ Otros: para los casos difíciles de etiquetar.

Dentro de las diferentes posibilidades existentes para la aparición de neologismos, se considera la composición como uno de los más importantes recursos neológicos, aunque no se descarta el peso cuantitativo que tiene en español el léxico multiplicativo, el que resulta de mecanismos derivativos y en particular la prefijación. Sin embargo existen diversos autores que sitúan la creación de palabras por medio de prefijos junto a la composición. Es frecuente observar en la formación de nuevas palabras fenómenos que son, a causa de sus formantes, creaciones muy cultas, y es evidente la continua presencia del léxico denominado por García Platero⁵¹, ‘jerga de portavoz’ y la posterior estandarización —adaptación a la norma común— de las diferentes unidades en virtud de la demanda del público receptor, cada vez más familiarizado con este vocabulario.

⁵⁰ MARÍA TERESA CABRÉ CASTELLVI, “La clasificación de los neologismos: Una tarea compleja”, *op.cit.*

⁵¹JUAN MANUEL GARCÍA PLATERO, "Procedimientos lexicogénicos en el discurso periodístico" en J. A. MOLINA REDONDO. y J DE D. LUQUE DURÁN, *Estudios de lingüística general. Actas del II Congreso Nacional de Lingüística General*, Granada, 1996. Servicio de Publicaciones de la Universidad de Granada, vol. 2, págs. 139-145, 1997.

La vitalidad de los neologismos compositivos

Para el Centro Virtual Cervantes (CVC) “la lengua, es dinámica por definición y tanto el incremento de la compartimentación del lenguaje técnico como la rapidez necesaria de los medios de comunicación, produce neologismos”. Se hace necesario detectarlos e inventariarlos para permitir realizar un seguimiento e implantación. Por ello, CVC dispone de un portal web⁵² que tiene como objetivo contribuir a la difusión de neologismos⁵³, detectados en diversos medios de información en un periodo cronológico, que surge de un convenio de colaboración entre la Universidad Pompeu Fabra y el Instituto Cervantes; presenta su banco de Neologismos, producto de la extracción de las bases de datos de BOBNEO⁵⁴ del Observatorio de Neología del Instituto Universitario de Lingüística Aplicada (IULA), que recoge desde 1988 los neologismos léxicos procedentes de los medios de comunicación, escritos y orales, en catalán y español.

		Valor absoluto	Porcentaje
COMPOSICIÓN	COMPOSITIVOS	4 513	6,36%
	CULTOS	8 205	11,56%
DERIVACIÓN	PREFIJACIÓN	11 487	16,2%
	PREFIJACIÓN-SUFIJACIÓN	280	~0,4%
	SUFIJACIÓN	14 197	20%
OTRAS FORMACIONES	LEXICALIZACIÓN	222	0,3%
	SINTAGMACIÓN	4 019	5,7%
	NEOLOGISMO SEMANTICO	4 865	6,9%
	NEOLOGISMO SINTÁCTICO	470	0,7%
	RESTO	22 725	32,01%
TOTAL		70 983	100%

Tabla 8.- Porcentajes de diversos tipos de formaciones de neologismos formales — derivativos, compositivos, entre otros— del banco de datos de BOBNEO del 2004 al 2010.

⁵² <http://cvc.cervantes.es/lengua/banco_neologismos/busqueda.asp> [1/10/2015]

⁵³ CVC manifiesta que la lista de neologismos publicados en dicho portal no posee un carácter valorativo, tiene un propósito de inventario etiquetado a partir del cual los investigadores pueden establecer diagnósticos y realizar trabajos analíticos sobre el uso y la implantación de los neologismos en español y catalán.

⁵⁴ <<http://obneo.iula.upf.edu/bobneo/index.php>>[1/10/2015]

Los resultados analizados sobre neologismos compositivos son de interés en el estudio de la presente Tesis y refuerzan que es un proceso de creatividad léxica abundante, pese a que la composición no es el proceso de formación de palabras más empleado en español, se ha podido ver que el porcentaje de creaciones léxicas que se basan en la formación por composición propia, sintagmática o culta (~ 24%) —*albiazul, aerogel, politoxicomanía, sanbenito, sociobiología*— no es relativamente tan bajo con respecto a las creaciones formales constituidos por derivación (aproximadamente 36%), considerado hasta el momento como uno de los procesos más fecundos con los que cuenta nuestra lengua para su enriquecimiento; los ejemplos restantes, los préstamos y las formaciones sintagmáticas, lexicalizaciones y los neologismos semántico-sintácticos, configuran el restante porcentaje, como se muestra en la Tabla 8.

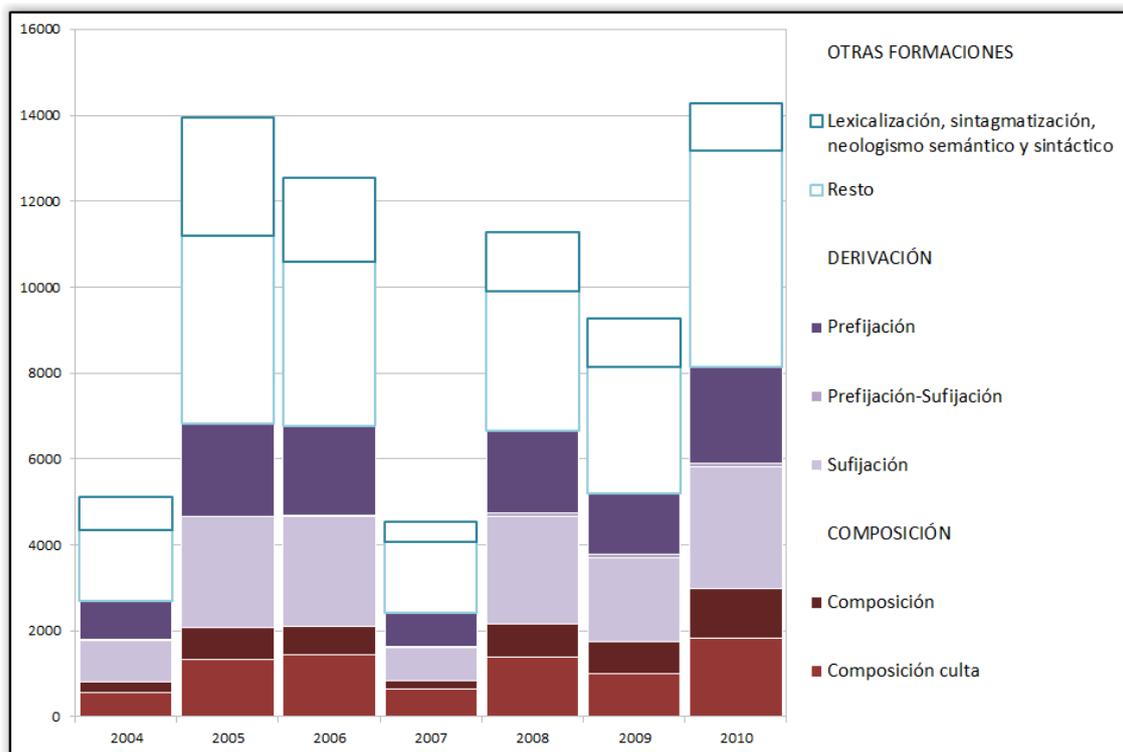


Figura 6.- Formaciones neológicas del banco de datos de BOBNEO durante el periodo 2004 al 2010 en valores absolutos.

En las Figura 6 y Figura 7 se aprecia la distribución de las diversas formaciones neológicas recolectadas en el banco de datos de BOBNEO a lo largo del periodo 2004-

2010, representadas en valores absolutos y relativos, respectivamente. Se observa la baja productividad en determinados años —2004 y 2007— y la proporción de, al menos un 50 %, las formaciones por composición y derivación frente al resto que se mantiene con igual distribución a lo largo del periodo.

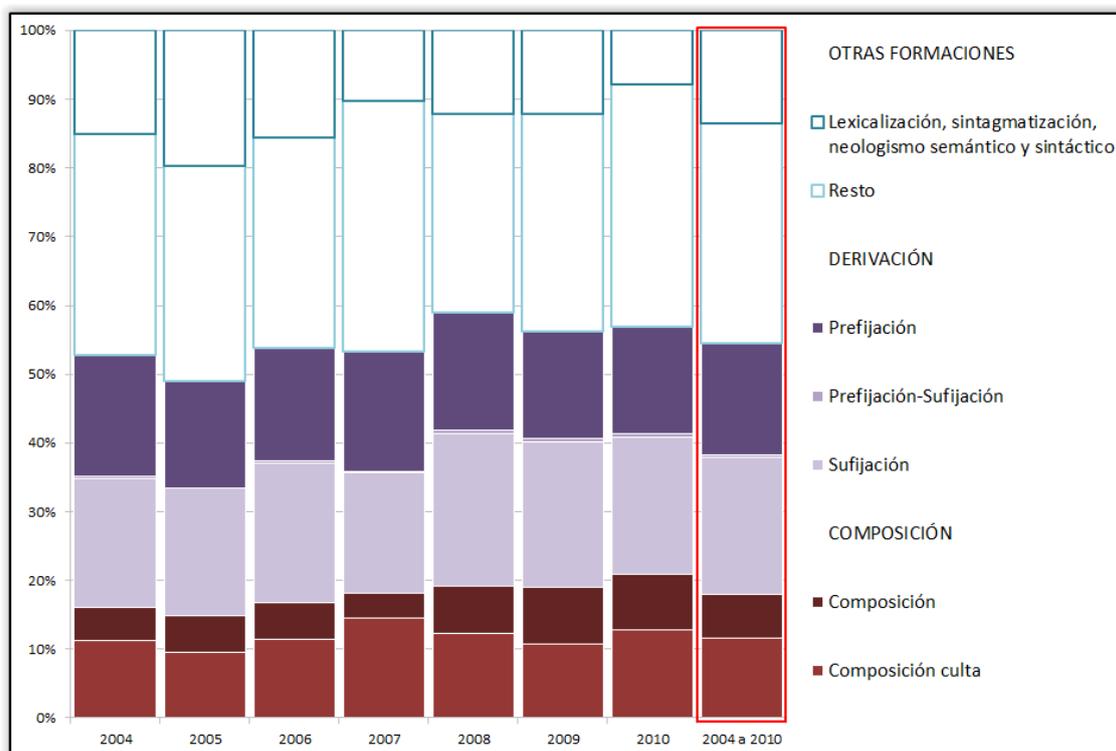
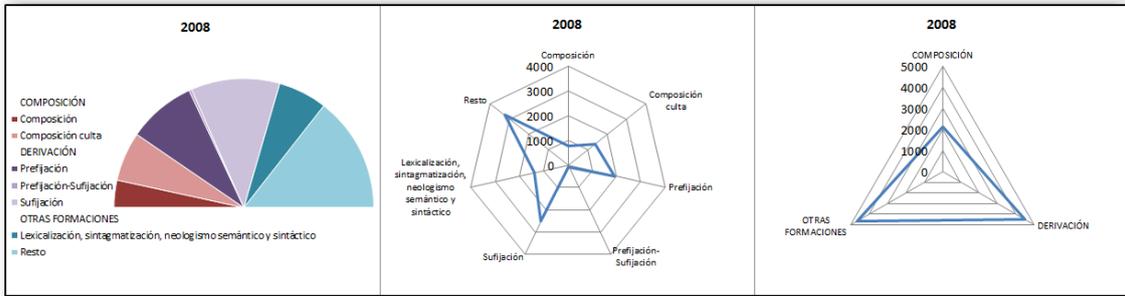
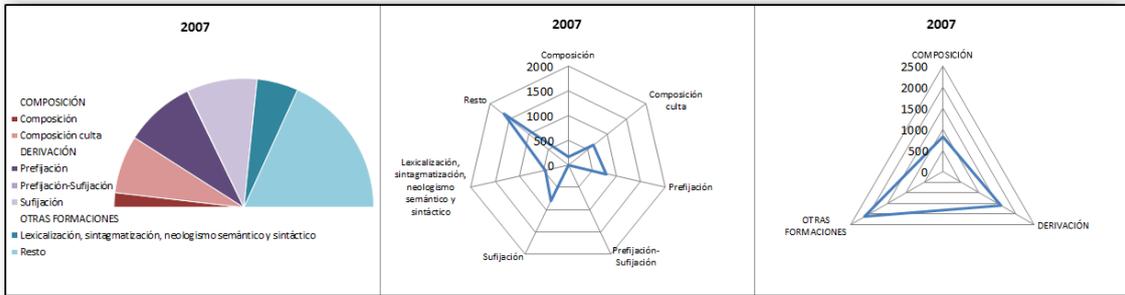
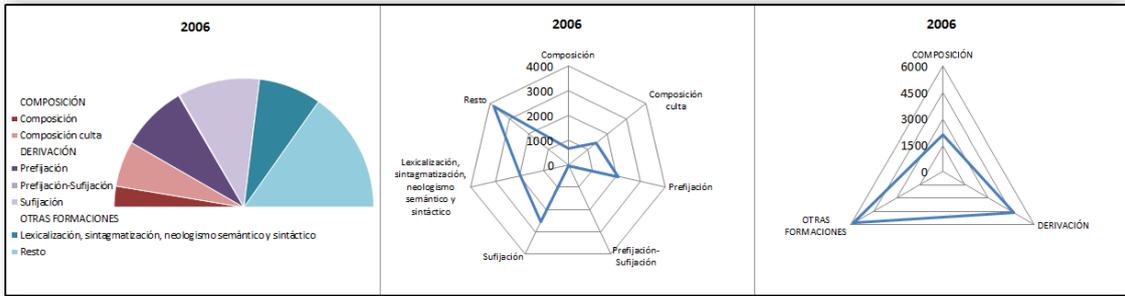
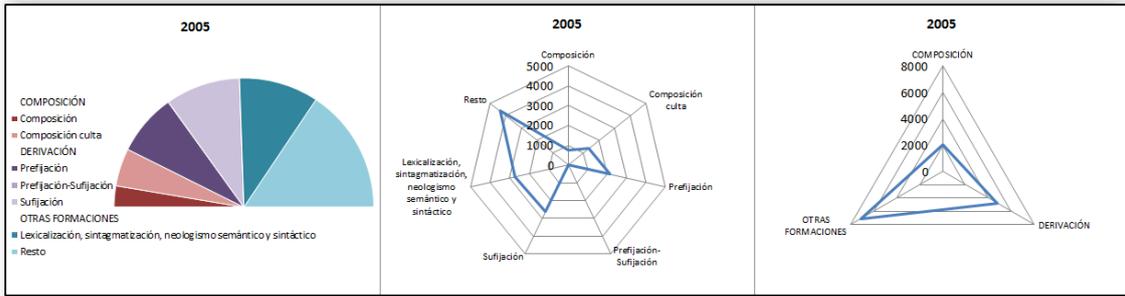
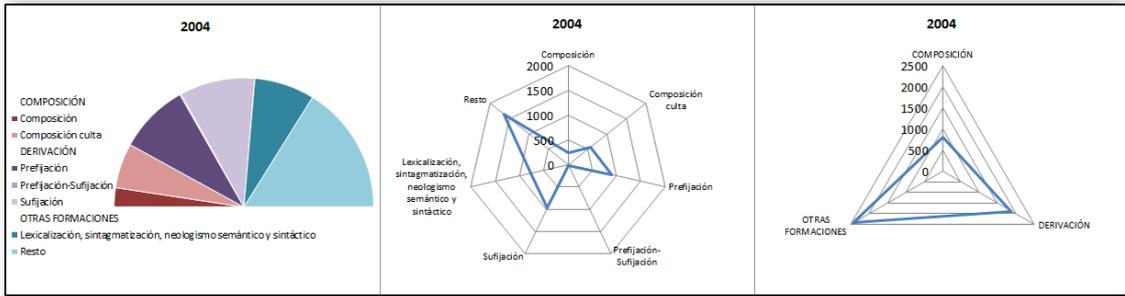


Figura 7.- Formaciones neológicas del banco de datos de BOBNEO durante el periodo 2004 al 2010 en valores relativos

La Figura 8 muestra una serie de gráficos circulares y radiales que representan la distribución de los diferentes procesos de formación neológica —compositiva, derivativa y otras formaciones— correspondiente a cada año durante el periodo 2004 al 2010. Se observa un patrón de comportamiento similar a lo largo de cada año en la productividad de tales formaciones, como se aprecia en el último gráfico. Los cambios que se aprecian corresponden a procesos neológicos heterogéneos que se han aglutinado en otras formaciones —lexicalización, neologismos semántico, sintáctico, entre otros.



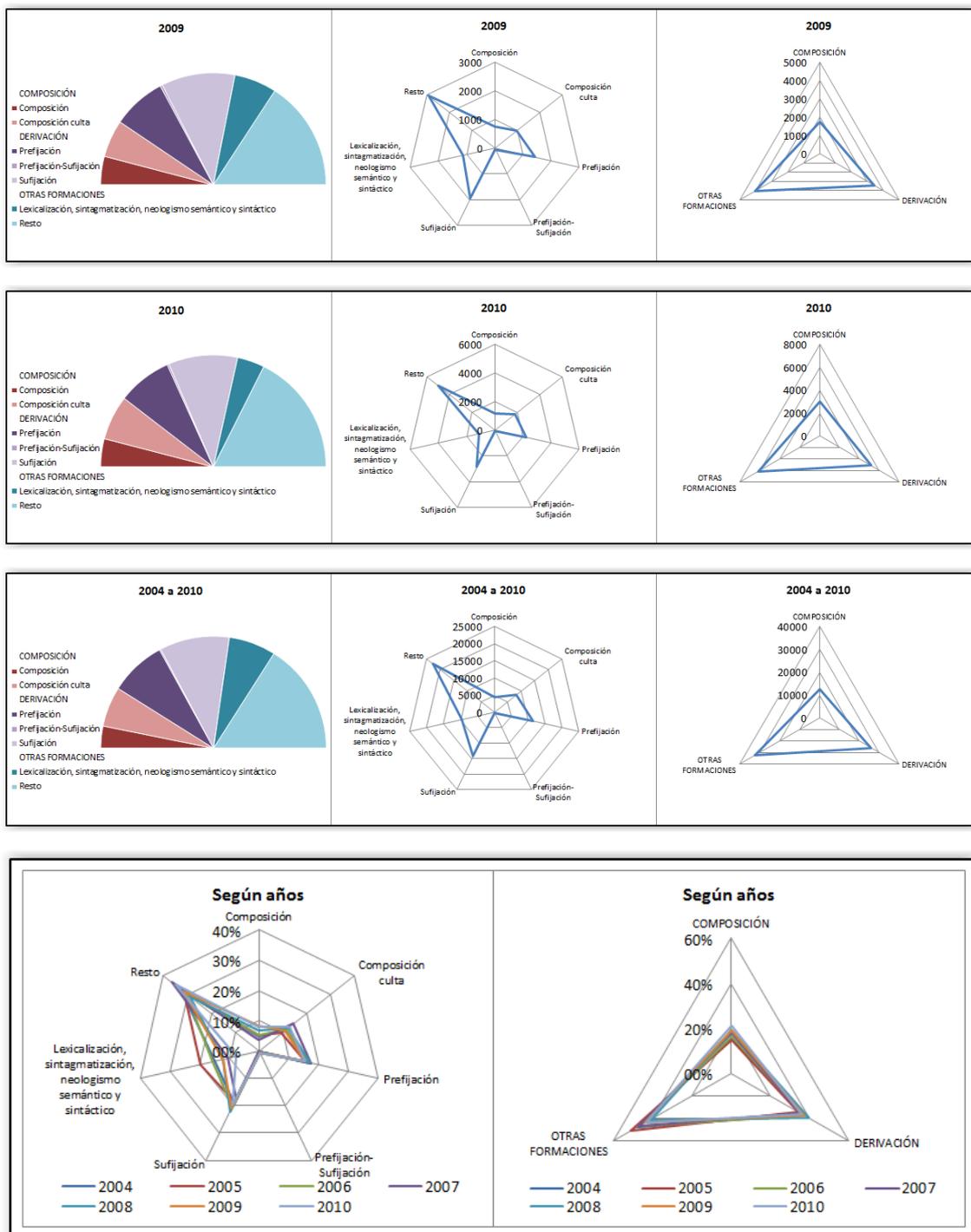


Figura 8.- Gráficos circulares y radiales⁵⁵ para los procesos de formación formales de neologismos en el periodo 2004-2010.

⁵⁵ Ver APÉNDICE B: GRÁFICOS PARA LOS PROCESOS DE FORMACIÓN FORMALES DE NEOLOGISMOS., donde se pueden apreciar los valores en imágenes con mayor resolución.

A través de la Figura 9 se muestra el porcentaje por filas de la producción de neologismos compositivos y derivados; se aprecia que mantiene el mismo patrón de comportamiento según la productividad neologistas a lo largo del periodo 2004 al 2010. Sin embargo, no se puede decir lo mismo para otras formaciones debido a la heterogeneidad de las formaciones aglutinadas.

Mediante una prueba de la Ji-cuadrado de bondad de ajuste se comprueba que los datos disponibles en la muestra tomada de la base de datos de BOBNEO, siguiendo la clasificación de neologismos propuesto por M. Cabré —derivación y composición— se corresponde con una distribución uniforme ($p\text{-valor}=0,0524$)⁵⁶.

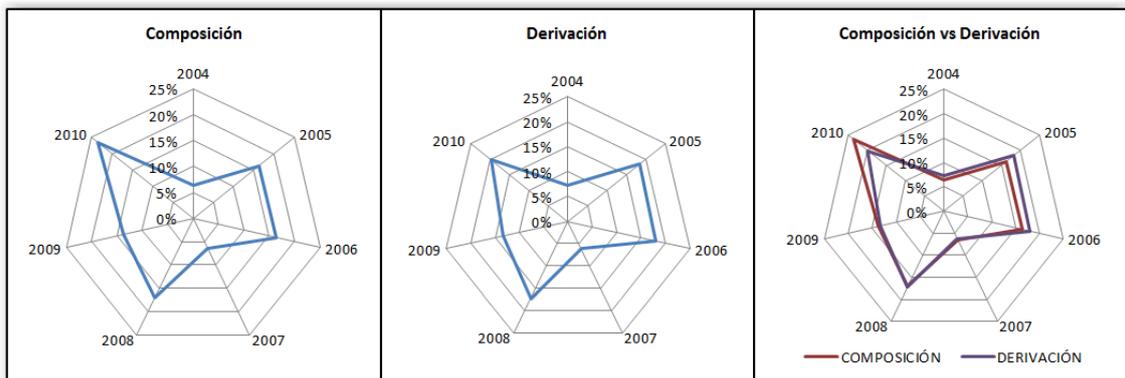


Figura 9.- Gráficos radiales del proceso de formación de neologismos compositivos —composición y derivación— en el periodo 2004-2010.

⁵⁶ Como el p-valor es mayor al criterio α menos exigente (5%) se dice que el contraste no es significativo.

Se realiza una prueba de la Ji-cuadrado de homogeneidad de muestras para comprobar si las proporciones de formación de compuestos propios y cultos en las bases de datos BOBNEO y BARGCoS son iguales. Para ello se construye la tabla de contingencia (ver Tabla 9) y se realiza el tratamiento estadístico basado en la distribución Ji-cuadrado con un grado de libertad. Se concluye que no existe diferencia significativa entre las dos bases de datos en cuanto a la productividad de los diferentes tipos de compuestos: propios y cultos (p-valor=0,05107)

	Composición	Composición Culta
Banco de datos OBNEO	4 513	8 205
Base de Datos ARGCoS ⁵⁷	4 147 ⁵⁸	8 363

Tabla 9.- Tabla de contingencia para comprobar la homogeneidad de los diferentes tipos de composición —propia y culta— en las bases de datos BOBNEO vs BARGCoS.

El comportamiento observado en ambas bases de datos (BBDD) deja ver que la productividad en la formación de compuestos cultos es mayor que la composición propia, hecho que se contrasta con la base de datos de referencia.

⁵⁷ Los datos correspondientes a composición propia y culta recogidas en la base documental BARGCoS, se toman de la Tabla 2 que muestra los porcentajes en la formación de palabras compuestas según las combinaciones posibles de los tipos de unidades léxicas en que se ha dividido nuestro estudio: palabras patrimoniales y elementos compositivos.

⁵⁸ Se considera añadir a BARGCoS 1 059 compuestos no yuxtapuestos de Bustos Gisbert (*La composición nominal en español*, op. cit. págs. 367-463).

Conclusiones del capítulo → Gracias a los avances tecnológicos y científicos, al auge en el uso de los nuevos dispositivos de comunicación social, donde los contenidos son creados por los usuarios —principal fuente de creación léxica— o al lenguaje periodístico, se recogen con frecuencia neologismos compositivos. La composición propia y culta es uno de los procesos más universal de formación de palabras que tienen la lengua para el enriquecimiento de su léxico. La productividad de formaciones compositivas culta es bastante superior con respecto a las creaciones compositivas propias, contrastado este comportamiento con los compuestos de la base de datos BARGCoS. Por todo ello se hacen imprescindibles procesos automáticos que sean capaces de situarlas en un contexto lingüístico adecuado desde un punto de vista morfológico y semántico.

CAPÍTULO 3.- ESTUDIO DE LA COMPOSICIÓN YUXTAPUESTA EN ESPAÑOL. REGLAS DE FORMACIÓN Y DE CAMBIO GRÁFICO

RESUMEN: En este capítulo se procede a estudiar los compuestos yuxtapuestos, se presenta un conjunto de reglas de carácter léxico y de cambio gráfico construidas a partir del comportamiento mayoritario observado en las distintas bases documentales tratadas.

Introducción

En el Capítulo 2 se estudia la composición como uno de los procesos de formación de palabras con mayor importancia de los que dispone muchas lenguas para aumentar su caudal léxico; se echa mano de tal mecanismo de tipo morfológico para formar nuevas palabras — *neologismos morfosintánticos*— a partir de unidades léxicas, vocablos o giros gracias al tirón provocado por los lenguajes técnicos, publicitarios, sociales y demás⁵⁹. Qué duda cabe que se hacen imprescindibles procesos automáticos que sean capaces de identificar las palabras en cuestión y situarlas en un contexto lingüístico adecuado.

En el presente capítulo se tienen en cuenta únicamente aquellos casos en los que la palabra compuesta se ha consolidado como la unión gráfica de los elementos que la componen de forma regular o irregular, con el propósito de garantizar un conocimiento suficiente de la casuística que concierne a este proceso de formación. Se consideran solo aquellos compuestos consolidados o perfectos, donde el grado de fusión de los elementos

⁵⁹ Internet y el lenguaje periodístico recogen, con frecuencia, neologismos compositivos debido a la rápida evolución de los acontecimientos y a su inmediata transcripción al mundo de las Tecnologías de la Información.

del compuesto es total⁶⁰; se tratan los pseudoprefijoides y pseudosufijoides como elementos compositivos no como morfemas derivativos; desde un punto de vista principalmente morfológico se incluye la parasíntesis por composición y algunos casos de acortamiento, aunque no se cuenta con perspectivas sintácticas, ni semánticas, en ocasiones procede hacer referencia a tales aspectos.

Se extraen de diferentes fuentes léxicas los distintos compuestos y se examina los mecanismos aplicados para establecer su comportamiento morfosemántico y su generación. Se obtiene así un conjunto de reglas —*formación y cambio gráfico*— y sus excepciones que permiten el reconocimiento y generación automática de palabras compuestas, construidas a partir del comportamiento mayoritario observado. Se exponen las normas aplicadas que se inducen a partir de la realidad en exploración que permitan solventar barreras lingüísticas a través del dominio informático y que pongan de manifiesto las irregularidades que afloran en el tratamiento.

Corpus seleccionado

Se procede a recolectar un corpus lo suficientemente amplio como para analizar pormenorizadamente la composición en español y poder extraer conclusiones generalizadoras. Se estudian los compuestos yuxtapuestos o lexías compuestas en composición nominal y adjetiva, al igual que algunos casos especiales de acortamiento, elementos compositivos y parasintéticos por composición. La investigación se fundamenta en un corpus con una cantidad significativa de compuestos, alrededor de 11 000, que provienen de diversos repertorios lexicográficos de la lengua española.

Se parte de los compuestos recogidos del glosario de compuestos de Bustos Gisbert⁶¹, que para su elaboración se utilizaron obras de carácter general —DRAE— y otras de carácter regional o dialectal, como hablas leonesas, aragonesas, meridionales y español de América. Se procede a clasificar los diferentes compuestos en grupos según la categoría gramatical de sus constituyentes y del compuesto (véase clasificación según Bustos Gisbert en Tabla 2). La

⁶⁰ No se estudian aquellos compuestos donde sus elementos componentes han alcanzado una coherencia sintáctica y/o semántica pero sin fusión ortográfica.

⁶¹ Véase EUGENIO BUSTOS, *La composición nominal en español*, op. cit. págs. 367-463.

base documental de partida —BARGCoS⁶²— se muestra en la Figura 10 donde, en una fase inicial, se recogen en la Base de Datos los compuestos recopilados con su información asociada: definición, categoría, tamaño, especialmente —se aprecia la relación existente entre los compuestos, la fuente de procedencia y clasificación a la que pertenece atendiendo a sus categorías gramaticales.

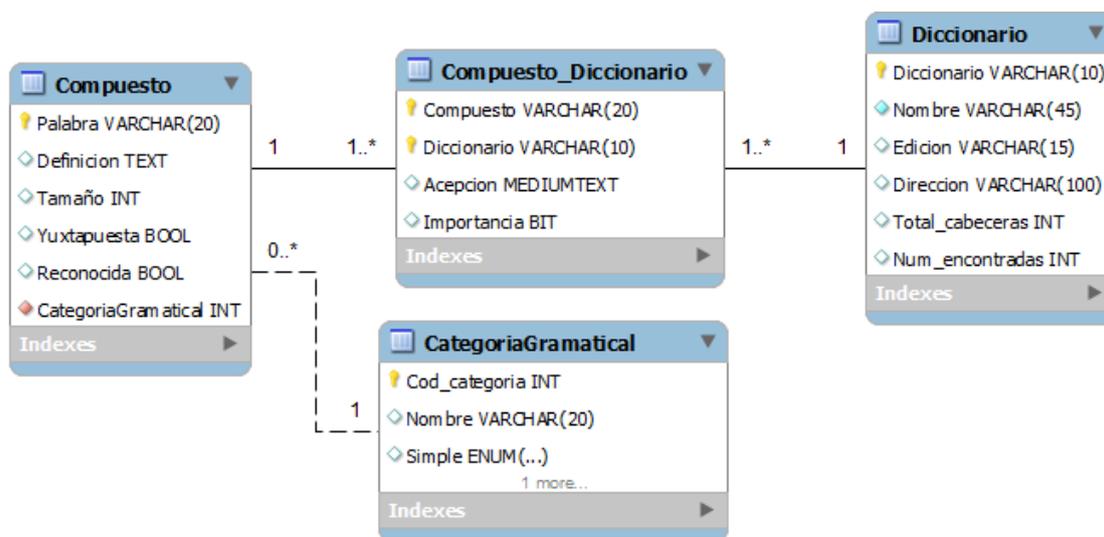


Figura 10.- Grafo relacional de los compuestos recopilados de diversas fuentes lexicográficas.

A continuación se describen cada una de las entidades y sus propiedades.

Entidad
Compuesto
Descripción
Entidad que contiene la información de las palabras compuestas estudiadas.
Atributos
Palabra: Palabra compuesta. Definición: Definición de la palabra compuesta. Tamaño: Tamaño de la palabra. Yuxtapuesta: Representa si la palabra es yuxtapuesta o no. Nexo: Determina si existe nexo de unión (guion, espacio o preposición).

⁶² Base de datos BARGCoS “Reconocimiento y Generación Automática de Compuestos en español”

Reconocida: Determina si la palabra es reconocida por el lematizador del GEDLC⁶³.

Categoría: Categoría gramatical de la palabra compuesta.

Entidad

Diccionario

Descripción

Entidad que contiene la información de los diversos repertorios lexicográficos con los que se trabajan.

Atributos

Diccionario: Siglas del diccionario.

Nombre: Nombre completo del diccionario.

Edición: Formato y fecha de edición del diccionario.

Dirección: Fichero (.XML,..) o URL del diccionario.

Num_Cabeceras: Cantidad total de palabras en el diccionario.

Num_encontradas: Cantidad de palabras compuestas encontradas.

Entidad

CategoríaGramatical

Descripción

Entidad que contiene la información de categorías gramaticales.

Atributos

Cod_categoria: Código de la categoría gramatical.

Nombre: Nombre categoría gramatical.

Simple: indica si la categoría es simple o compuesta.

⁶³ GEDLC Grupo de Estructuras de datos y Lingüística Computacional de la Universidad de Las Palmas de Gran Canaria <<http://www.gedlc.ulpgc.es/investigacion/scogeme02/lematiza.htm>> [1/10/2015]

Se amplía la base documental según, entre otros criterios⁶⁴, la información etimológica que suministran los principales repertorios lexicográficos de la lengua española (Bibliograf, 2003; RAE, 2001; Clave, 1997; Larousse, 1996; Moliner, 1996)⁶⁵. Los patrones que se observan son los siguientes:

- ❖ DRAE. → De...palabra... y... palabra....

mariposa.

De *Mari*, apóc. de *María*, y *posa*, 2.^a pers. sing. del imperat. de *posar*.

zoología.

De *zoo-* y *-logía*.

- ❖ VOX → (...palabra...+... palabra...)

mariposa

(*mari-* + *pósate*, de dichos infantiles)

zoología

(*zoo-* + *-logía*)

⁶⁴ En una segunda fase se consideran solamente los compuestos consolidados o perfectos, donde las dos (o más) unidades léxicas están totalmente integradas, se obvian aquellas que contengan guion o nexos (espacio o preposición).

⁶⁵ *Diccionario General de la Lengua Española VOX*, op.cit.; REAL ACADEMIA ESPAÑOLA, *Diccionario de la Lengua Española*, op.cit.; *Diccionario de Uso del Español Actual*, Clave S.M., Madrid, 1997; *Gran Diccionario de la Lengua Española*, Larousse Planeta, S.A., Barcelona, 1996; MARÍA MOLINER, *Diccionario de Uso del Español*, Gredos, Madrid, 1996.

❖ CLAVE → De...palabra... y...palabra...

marisabidillo, lla

sustantivo

coloquial. Persona que presume de lista o de bien informada:

Aunque no sepa lo que le preguntes te contestará, porque es una marisabidilla.

Notas

Etimología: De *María* y *sabidillo*.

zoología

sustantivo femenino

Ciencia que estudia los organismos animales:

Para este examen de zoología tengo que aprenderme las características de los mamíferos.

Notas

Etimología: De *zoo-* (animal) y *-logía* (estudio, ciencia).

❖ LAROUSSE → No sigue un único patrón.

zoología (Del gr. *zoon*, animal + *logos*, ciencia.) Ciencia que estudia los animales.

mariposa (De *María*, *pósate*, expresión procedente de una canción infantil.)

correveidile (Derivado de *corre* + *ve* + *dile*.)

Finalmente, se construye una base de estudio de unos 11 000 compuestos recopilados de diversas fuentes lexicográficas y se buscan inclusiones de compuestos procedentes de nuevas fuentes neológicas o nuevos diccionarios⁶⁶. En Figura 11 se muestra parte del grafo relacional de la Base de Datos BARGCoS que refleja cómo está organizada y clasificada la información.

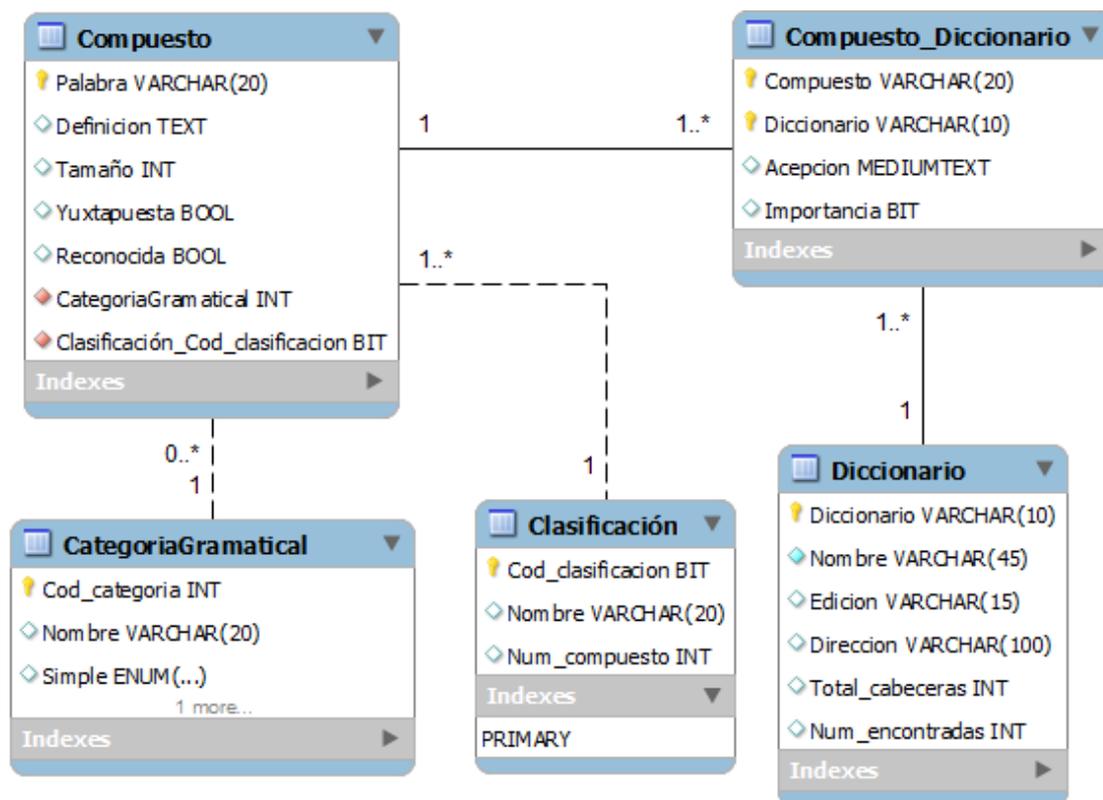


Figura 11.- Modelo lógico de la Base de Datos BARGCoS.

Se describe a continuación las entidades que se añadieron con sus propiedades correspondientes:

⁶⁶ MANUEL ALVAR EZQUERRA, *Nuevo Diccionario de voces de uso actual*, Arco/Libros, Madrid, 2003; *Diccionario de neologismos de la lengua española*, Larousse, S.A., Barcelona, 1998.

Entidad
Clasificación
Descripción
Entidad que contiene las clasificaciones según la categoría gramatical.
Atributos
Diccionario: Código de clasificaciones. Nombre: Nombre categoría gramatical. Num_compuestos: Cantidad de palabras compuestas encontradas.

Una vez recogido el corpus, se estudia los elementos que lo forman y las alteraciones que han sufrido los mismos para poder deducir las reglas de formación y los cambios gráficos que permita la identificación automatizada de palabras compuestas yuxtapuestas. Se obtiene así un conjunto de reglas y sus excepciones que facultan el reconocimiento y generación automática de palabras compuestas.

Las reglas de formación de compuestos

Las reglas de formación de compuestos que se estudian son de naturaleza léxica y no sintáctica, semántica o fonológica, lo que no excluye que, en ocasiones, se precise hacer referencia a tales aspectos dado que representan características propias de tales procesos de construcción de palabras. Ya que las formas libres que constituyen los compuestos guardan relaciones gramaticales inequívocamente sintagmáticas, resulta difícil fijar la frontera entre la sintaxis y la morfología.

Las reglas de formación de compuestos que se buscan, se aplican a primitivas que deben ser palabras consolidadas o neologismos, pero nunca incorrectas —ortografía irregular o términos bloqueados como **grabamiento* por *grabación*. La existencia en el lexicon⁶⁷ de

⁶⁷ Debido a que generalmente los compuestos no se han consolidado en los diccionarios de uso de la lengua, se percibe una escasez de formas compuestas de actualidad en tales repertorios lexicográficos —en ocasiones, los ejemplos utilizados echan mano de formaciones neológicas, no necesariamente documentadas en la base de referencia.

innumerables particularidades no debería imponer una barrera para la automatización del proceso, tanto en el reconocimiento como en la generación, dado que algunas de las palabras que presentan irregularidades admiten un reglado y el resto pueden llevar un tratamiento especial.

Para la deducción de las reglas de formación que permiten la identificación automatizada de palabras compuestas yuxtapuestas, se parte de la base documental de compuestos extraídos de diferentes repertorios lexicográficos y se estudian los cambios gráficos que han sufrido las diferentes palabras que los forman así como sus categorías gramaticales. Hay que tener en cuenta que un estudio cuyo objetivo sea la automatización de la morfología por medios informáticos, los aspectos teóricos o formales no tienen por qué coincidir con los estrictamente lingüísticos. Así, la falsa composición **verdenegro* —debería ser *verdinegro*— no tendría por qué tratarse de una formación errónea ya que no contraviene ninguna regla compositiva del español. Se obtiene así un conjunto de reglas y sus excepciones que hagan factible el reconocimiento y posterior generación automática de palabras compuestas.

A partir del comportamiento de los vocablos constituyentes del compuesto, se deducen sus reglas de formación; algunas coinciden con las tratadas por lingüistas con determinadas adaptaciones informáticas que se justifican por el comportamiento mayoritariamente observado. Según el Diccionario General de la Lengua Española VOX señala que la composición *aeriforme* proviene de:

aeriforme
(*aeri-* + *-forme*)
1 **Adjetivo**. FÍSICA. Parecido al aire: *fluidos aeriformes*.

y sin embargo, el comportamiento mayoritario observado es el elemento prefijal *aero-* no *aeri-*.

Con lo cual, se define como *regla de formación*:

«a toda pauta de máxima frecuencia que posibilite establecer un mecanismo que relacione los elementos que constituyen el compuesto para su reconocimiento por medios informáticos»

Para un correcto tratamiento informático, importa tener en cuenta que cuando se aplican las reglas para el reconocimiento y generación de compuestos se han de considerar

los cambios gráficos que se pueden producir como consecuencia de aplicar la regla. Estos cambios gráficos reciben el nombre de *reglas fonéticas*.

Reglas de formación

Los procedimientos mediante los que se forman palabras compuestas pueden dividirse en dos grandes grupos atendiendo al grado de modificación que sufran los elementos originales:

- bien por la mera adición de dos o más términos sin que ninguno de ellos se modifique

rompeolas, abrelatas, mediodía, cenaoscuras, bullebulle, mariposa, ajoqueso

- o bien por la unión que conlleve algún tipo de modificación gráfica en alguno de los elementos que intervienen en la composición —generalmente ocurre en el primero de los componentes— o por adición al resultado final

agridulce, rojiblanco, coliflor, balompié, liquidámbar, anquirredondo, claroscuro, cochitritl.

Se presentan las reglas de formación de palabras compuestas que se ponen de manifiesto en la base de estudio. Se usan tanto para el reconocimiento como para la generación de compuestos en español: recorrerlas en sentido contrario implica que funcionen en una u otra modalidad. Se muestran desde el punto de vista de la generación de los compuestos.

Regla 1 (unión)

Unión sin pérdida, adición, ni cambio —simplemente se unen dos o más términos sin que ninguno sufra modificación alguna. Esta regla se aplica de forma general a todas las categorías gramaticales. Usualmente, en la composición adjetiva de *Adjetivo+Adjetivo* se utiliza con adjetivos de nacionalidad, y en la composición nominal de *Verbo+Verbo* se duplica el lexema verbal. Es la regla que más se utiliza en el compuesto nominal de *Verbo+Complemento* —este tipo de formación representa el mecanismo más productivo de la lengua española: se une el lexema verbal en forma imperativa con el complemento. En la Figura 12 se muestra la producción de compuestos en BARGCoS que aplican la regla de unión, se presenta según la categoría gramatical de los elementos componentes y la categoría del compuesto.

utilizados en esta regla. En la Figura 13 se muestra la producción de compuestos de BARGCoS, en base a la categoría gramatical de los constituyentes.

[ADJETIVO+ADJETIVO]_{ADJETIVO}: *claroscuro*.

[SUSTANTIVO+ADJETIVO]_{ADJETIVO}: *uñalbo, cariacontencido, gargantiazul*.

SUSTANTIVO+SUSTANTIVO: *maestrescuela, telaraña, pesamedello, manobra, ajolio, rosadelfa*.

VERBO+COMPLEMENTO O VICEVERSA: *guardaguas, paraguas, rompesquinas, tragaños, saltatrás, botaguas, picabejero, bogavante, matalbahaca, quebrarado*.

SUSTANTIVO+PREPOSICIÓN+SUSTANTIVO: *trampantojo*

SUSTANTIVO+ADJETIVO O VICEVERSA: *bocabajo, telaraña, doblescudo, cuatrojos, mediagna*.

Caso adicional de *ha* u *ho*: Se elimina la *h* del punto de unión si se encuentra entre dos vocales iguales, como resultado de la unión de los constituyentes.

quebracho ← *quebra* + *acho* ← *quebra* + *hacha*

sabiondo ← *sabio* + *ondo* ← *sabio* + *hondo*

matambre ← *mata* + *ambre* ← *mata* + *hambre*



Figura 13.- Producción de la regla 2 (-a2, -e2, -o2, -ha, -ho) en base a la categoría gramatical de los constituyentes.

Regla 3 (-a+i, -o+i, -e+i, -eo+i, -ea+i)

Sustitución de las vocales átonas finales *a*, *e* u *o* correspondientes al primer constituyente por el infijo compositivo *i*, o en el caso de *Verbo+Complemento* se reemplaza la terminación verbal, acabada en *a* para los verbos de la 1ª conjugación y en *e* para los de la 2ª y 3ª, por el infijo compositivo *i*, y a continuación llevar a cabo la unión sin pérdida. Generalmente, en la composición *Adjetivo+Adjetivo* se utiliza con adjetivos de color. En la mayoría de los casos, la forma del tiempo verbal que se usa en los compuestos *Verbo+Verbo*, presentan la vocal átona *e*. En los compuestos de tipo *Sustantivo+Adjetivo* o *Adjetivo+Sustantivo*, son más las ocasiones en las que hay que sustituir el morfema *o*. Debido a la estructura peculiar del compuesto [*Sustantivo+Adjetivo*]_{Adjetivo} —la más productiva— hay que tener en cuenta que existen palabras que aplicarían la regla anterior (regla: -a2, -o2, -e2) de no ser porque previamente han incorporado el infijo compositivo *i*. En la Figura 14 se muestra la producción de compuestos en BARGCoS.

[ADJETIVO+ADJETIVO]_{ADJETIVO}: *altibajo, rojinegro, galicursi, cojitranco, anchicorto, cortiancho, verdinegro, dulciagrío, galicursi*.

[SUSTANTIVO+ADJETIVO]_{ADJETIVO}: *paticojo, alicaído, rabicorto, alirrojo, barbicañaño, clarividente, culinegro, sangrígordo, cariancho, ubriciega, alicaído, manivacio*.

VERBO+VERBO: *subibaja, comicalla, corriverás, bativoleo, botivoleo, chiticalla, lavatiquevoy, rompirraja*.

SUSTANTIVO+SUSTANTIVO: *ajiaceite, carininfo, carnicol, carricoche, catricofre, capisayo, ajipuerro*.

VERBO+COMPLEMENTO O VICEVERSA: *abriboca, frunciboca, baticola, tentibien, saltigalgo, batiaguas, botifuera, ardiviejas*.

SUSTANTIVO+PREPOSICIÓN+SUSTANTIVO: *patiporsuelo*

SUSTANTIVO+ADJETIVO O VICEVERSA: *burriciego, capipardo, clarividencia, cochifrito, coliblanca, caridura*.



Figura 14.- Producción de la regla 3 (-a+i, -o+i, -e+i, -eo+i, -ea+i) en base a la categoría gramatical de los constituyentes.

Regla 4 (+i)

Si el primer elemento acaba en consonante y el segundo empieza en consonante, se introduce la *i* entre ambos y se lleva a cabo la unión. No se aprecia que se produzca en todos los tipos de composición, sino en aquellos casos en los que ambos componentes tienen igual categoría, a excepción de la combinación *Verbo+Verbo*⁶⁸.

ADJETIVO+ADJETIVO: *azuliblanco*.

SUSTANTIVO+SUSTANTIVO: *calicanto, coliflor, colinabo, paniqueso, salipez*.

El papel que parece desempeñar el morfema compositivo *i* es de función conjuntiva que, aunque parece probable desde la perspectiva semántica —*carricoche, coliflor*—, no se justifica en otros casos, especialmente, en algunos compuestos adjetivales, porque en tales no se atisba claramente ninguna razón de tipo semántico o fonológico.

⁶⁸ En este tipo de composición se utiliza la *Regla 20 (copulativa)*: se sustituye la *y* copulativa que funciona como elemento conjuntivo por la vocal *i* y se lleva a cabo la unión. Esto es debido a la flexión del verbo.

VERBO+VERBO: *correiverás, quitaipón, tiraiafloja, correvidile*

Regla 5 (-a, -o, -e)

Si el primer constituyente acaba en vocal fuerte, la pierde. Cuando el final vocálico del primer componente y el principio vocálico del segundo tienen la posibilidad de formación de un diptongo o de contracción, se procede a la unión con pérdida de la vocal del primer elemento. En la Figura 15 se muestra la producción de compuestos en BARGCoS según la categoría gramatical de los constituyentes y del compuesto. Se aprecia que es una regla que se aplica en todos los tipos de compuestos.

ADJETIVO+ADJETIVO: *eurasiático, agridulce*.

[SUSTANTIVO+ADJETIVO]_{ADJETIVO}: *papalba (papo o papu+alba), manvacío, cabezaído*.

SUSTANTIVO+SUSTANTIVO: *gatuña, ajaceite, mancuadra, rabiguana, liquidámbar*.

VERBO+COMPLEMENTO O VICEVERSA: *chambeculos, abrojo, desmochollas*.

SUSTANTIVO+ADJETIVO O VICEVERSA: *brazalbo, duraluminio, mandoble, manzorda, disanto, sabibondo*.

Existen casos en los que parecería que más que aplicar esta regla se podría haber utilizado la de sustitución del morfema por el infijo compositivo *i*, como puede verse en:

labibendido ← *labio+hendido* (no es solución *labe+hendido*)

disanto ← *dia+santo* (no es solución *do+santo*)

sabibondo ← *sabio+hondo*

agridulce ← *agrio+dulce* (no es solución: *agro+dulce, agre+dulce, agra+dulce*)

En otros casos, se dice que ha sufrido un proceso de pérdida del infijo compositivo *i*:

cabezaído ← *cabecicaído* ← *cabeza+caído*

manvacío ← *manivacío* ← *mano+vacío*

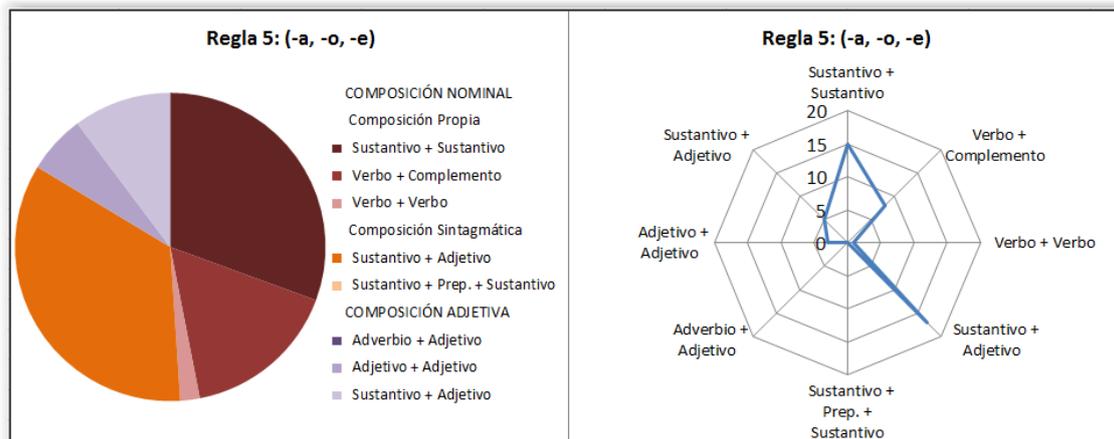


Figura 15.- Producción de la regla 5 (-a, -o, -e) en base a la categoría gramatical de los constituyentes.

Regla 6 (elemento compositivo)

Los elementos compositivos pueden estar formados por la agregación a una palabra española de raíces cultas greco-latinas, antepuestas o pospuestas —también llamada *composición híbrida*⁶⁹—, como:

acutifolio ← *acuti-* (del latín *acutu*, agudo) + *folio* (hoja)

nitroglicerina ← *nitro-* (del griego *nitrón*, nitro) + *glicerina*

abulomanía ← *abulo-* (del griego *aboulía*, sin voluntad) + *manía*

petrolífero ← *petróleo* + *-fero* (del latín *ferre*, conducir)

ecosonda ← *eco-* (del latín y griego *eco*, resonancia) + *sonda*

o bien por la combinación de raíces cultas:

⁶⁹ La composición híbrida es aquella que presenta entre sus formantes un elemento culto y otro de origen patrimonial (véase, CRISTINA BUENAFUENTES, *Procesos de gramaticalización y lexicalización en la formación de compuestos en español*, op.cit., pág. 343). Pueden existir problemas al considerar uno de los formantes no como un elemento culto, sino ya como una palabra evolucionada del español, como por ejemplo: el elemento compositivo *-fobia*, como en *xenofobia* ← *xeno-* + *fobia*, pues se puede caracterizar como compuesto culto o bien clasificarlo como compuesto híbrido al considerar que viene de la palabra patrimonial *fobia*.

- raíz prefija y sufija griegas o latinas, como:

teléfono ← *tele-* (del griego *tele*, lejos) + *-fono* (del griego *phoné*, voz o sonido)

filicidio ← *fili-* (del latín *filius*, hijo) + *-cidio* (del latín *cidium*, matar)

taquicardia ← *taqui-* (del griego *tachys*, rápido) + *-cardia* (del griego *kardía*, corazón)

- raíz prefija griega y sufija latina:

automóvil ← *auto-* (del griego *autos*, mismo o propio) + *-móvil* (del latín *mobilis*, mover)

- raíz prefija latina y sufija griega:

lucífilo ← *luci-* (del latín *lux*, *lucis*, luz) + *-fílo* (del griego *philos*, amante)

o por la incorporación de otro tipo de raíces:

arisblanco ← *aris-* (de arista) + *blanco*

eurocheque ← *euro-* (de europeo) + *cheque*

galíparla ← *gali-* (de galo, francés) + *parla*

Regla 7 (@preposición@)

El sintagma nominal ha sufrido un proceso de aglutinación sin pérdida de la preposición — *de*, *a*, *ante*, *con*, *en*, *por*—, del artículo —*la*, *el*—, de la conjunción (copulativa, condicional: “*sí*”, “*que*”...) o del pronombre personal clítico o enclítico —*me*, *te*, *lo*, *la*— presentes. La Figura 16 presenta la productividad de compuestos clasificados según la categoría gramatical de los componentes. Es la regla por excelencia para el tipo de compuesto *Sustantivo + Preposición + Sustantivo*.

VERBO+VERBO: *lavatiquevoy*, *correquetecagas*, *tenconten*, *salsipuedes*.

VERBO+COMPLEMENTO: *saltaembarca*, *besalamano*, *bienmesabe*, *abimelollevas*, *bienteveo*, *cenaaoscuras*, *esperalaultima*, *parlaembalde*, *diostedé*.

SUSTANTIVO+PREPOSICIÓN+SUSTANTIVO: *rabodejunco*, *juanencueros*, *patíporsuelo*, *almaempena*.

Regla 9 (excepción: variantes)

Existen variantes lexemáticas en las que se ha producido la síncope de uno o más elementos fonéticos del lexema: primero se procede a su modificación y luego se lleva a cabo la unión sin pérdida.

botesela: sela ← silla

capaniu: niu ← nido

pejemuller: peje ← pez,

cagarrache: rache ← erraj

picatrueu: trueu ← tuero

ajilimójili: ajili ← ajo mójili ← moje

guadarnés: guad ← guarda

cuerdacarru: carru ← carro

rabucocandil: rabuco ← rabo

Regla 10 (excepción: género)

- Si el primer elemento acaba en *a*, se sustituye por la vocal *o* y se lleva a cabo la unión

*ceromiel*⁷⁰ ← *cera + miel*

*perrogorda*⁷¹ ← *perra + gorda*

- Si el primer elemento acaba en *o*, se sustituye por la vocal *a* y se procede a la unión

*cerrapan*⁷² ← *cerro + pan*

- De forma similar ocurre con la segunda palabra

aguagriero ← *agua + agriera*

batiporte ← *bate + porta*.

⁷⁰ Definición de *ceromiel*: “Mezcla de una parte de cera y dos de miel, que antiguamente empleaban en la cura de las úlceras y heridas” (DRAE).

⁷¹ Definición de *perrogorda*: “Moneda de dos céntimos” (Véase, ANTONIO ALCALÁ VENCESLADA, *El Vocabulario Andaluz*, Madrid, 1951).

⁷² Definición de *cerrapan*: “Rebanada de pan” (cf. cast. cerro 'lomo, espinazo') (Véase, MANUEL ALVAR, *Palabras y cosas en la Aezcoa*, Zaragoza, 1947).

Regla 11 (excepción: plural)⁷³

Se conserva el plural en los casos en que se produzca.

*ajaspajas*⁷⁴ ← *ajos* + *pajas*

malaslenguas ← *malas* + *lenguas* ← *mal* + *lengua* [no es el plural de **mallengua*]

malastripas ← *malas* + *tripas* ← *mal* + *tripa* [no es el plural de **maltripa*]

cualesquiera ← *cuales* + *quiera*

quienesquiera ← *quienes* + *quiera*

Regla 12 (excepción: paranomasia)

Los elementos constituyentes del compuesto pueden presentar una construcción paranomásica —creados por semejanza de sonidos con otros vocablos.

sabidondo, sabilongo ← *sabihondo*

*pedabobo*⁷⁵ ← *pedagogo* ← *paído-* + *-ago*

balsopeto ← *bolso* + *peto*

⁷³ La marca de plural en los compuestos ortográficamente unidos se forma, por regla general, añadiendo los morfemas *-s* o *-es* al final del compuesto, es decir, la flexión es externa al compuesto (singular *mariposa*, plural *mariposas*; singular *rascacielos*, plural *rascacielos*). Pero algunas veces la flexión de número se aplica a ambos componentes, como se plasma en esta regla, o bien en expresiones anticuadas la flexión de número se aplica en ambos componentes, se marca el plural tanto interna como externamente (singular *ricabembra*, plural *ricasbembras*; singular *gentilhombre*, plural *gentileshombres*; singular *ricadueña*, plural *ricasdueñas*).

⁷⁴ Definición de *ajaspajas*: “Paja que queda en la ristra de ajos después de quitar la cabeza de éstos” (DRAE). Por ello, se aplica primero a este compuesto la regla 10 (excepción: género) donde se cambia *ajo* por *aja* y se mantiene el plural.

⁷⁵ Véase, MARÍA DEL PILAR ORTEGA (“Neología y Prensa: Un binomio eficaz”, *Espéculo* n° 18, (Universidad Complutense de Madrid, 2001, <<https://pendientedemigracion.ucm.es/info/especulo/numero18/neologism.html>> [1/10/2015]): “Los pedabobos ya empezaban a causar estragos en las filas inocentes del idioma...” (ABC 24- 12- 98). La formación se crea por semejanza con pedagogo.

Regla 13 (excepción: 2ª palabra)

Generalmente, las reglas se aplican al primer constituyente del compuesto, salvo en excepciones que actúan sobre el segundo:

Regla 13.1 (2ª palabra plural)

- En ocasiones, se forma un compuesto con marca de plural, aunque el segundo componente sólo exista en singular —para su reconocimiento habría que hacer desaparecer la terminación de plural del compuesto.

vuelvaluegos ← *vuelva* + *luegos* ← *vuelva* + *luego*

tragafées ← *traga* + *fees* ← *traga* + *fe*

Regla 13.2 (-sílabas duplicadas)

- Reducción de la estructura silábica duplicada que aparece al final del primer elemento y al inicio del segundo.

cascorvo ← *casco* + *corvo*

Regla 13.3 (-sílabas)

- Unión con pérdida de la sílaba final del primer elemento cuando existe idéntica consonante en la sílaba final del primer lexema y en la inicial del segundo:

coyotomate ← *coyote* + *tomate*

cejunto ← *ceja* + *junto*

Regla 13.4 (acronimia)

- Eliminación de la sílaba final del primer lexema, la inicial del segundo o incluso ambas:

cascalbo ← *cáscara* + *albo*

cardimuelle ← *cardinche* + *armuelle*

cartapel ← *carta* + *papel*

cochitril ← *cocho* + *cortil*

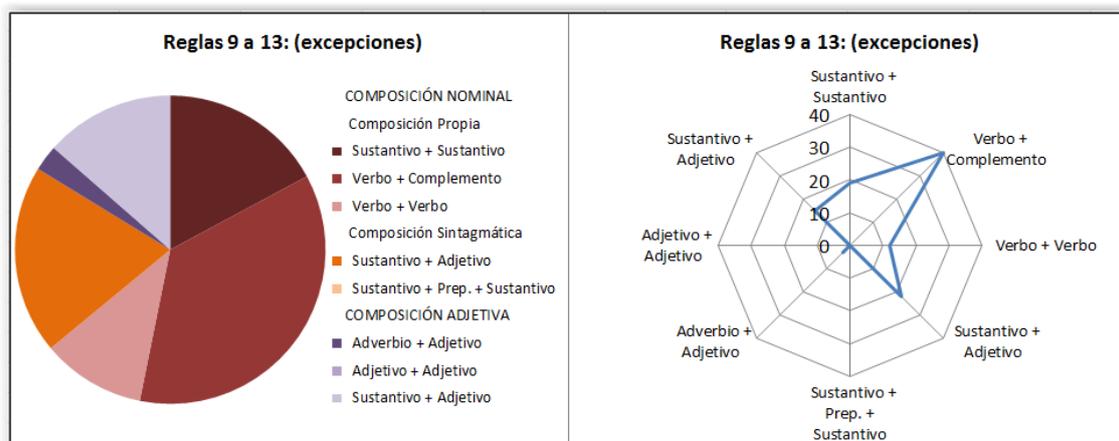


Figura 17.- Producción de la regla 9 a la regla 13 de tratamiento de las excepciones, en base a la categoría gramatical de los constituyentes.

Regla 14 (abreviamiento)

Se considera el abreviamiento o acortamiento léxico, la cual tiene diversas variantes: primero se procede a su cambio —puede darse que cualquiera de los constituyentes del compuesto— y luego se lleva a cabo la unión sin pérdida.

cine ← *cinematógrafo*

foto ← *fotografía*

bus ← *autobús*

profe ← *profesor*

mate ← *matemáticas*

Regla 15 (parasíntesis)

Se unen dos o más términos donde para el segundo elemento (B) se obtiene la forma sufijada propuesta, siempre que no exista la combinación de las constituyentes (A+B) ni exista la palabra *B+Sufijación* —en caso de aparecer, no serían el origen de la palabra final.

anquimuleño → *anca* + *mulo* (-eño) [no existe **muleño*, ni **anquimulo*]

misacantano → *misa* + *canta* (-ano) [no existe **misacanta*, ni **cantano*]

ropavejero → *ropa* + *viejo* (-ero) [no existe **vejero*, ni *ropavejero* es derivado de *ropavieja*]

aguamanil → *agua* + *manos* (-il) [no existe *mano+il*, ni es derivado de *aguamano*s]

Reglas fonéticas o de cambio gráfico

Como pueden producirse cambios gráficos en el compuesto final cuando se aplican las reglas anteriores; se han de tener en cuenta las siguientes reglas fonéticas o de cambio gráfico necesarias para lograr un correcto tratamiento informático en el reconocimiento y en la generación de tales palabras. Se muestran desde el punto de vista de la generación de los compuestos.

Regla 16(+rr)

Si el segundo elemento empieza por r, se debe sustituir por rr:

aguarrás, alirrojo, andarríos, alzarraños, arrancarrabia, buscarruidos, barbirrosto, boquirroto, carirraído, greccorromano, matarratas, pararrayos, pejerrey, yerbarreuma, colirrojo

salvo cuando va precedida de “l” o “n”:

milenrama ← mil+en+rama.

Regla 17 (-ca+qui, -co+qui)

Si al sustituir las vocales átonas finales *a* y *o* correspondientes al primer constituyente por el infijo compositivo *i* acaba en la sílaba *ci*, se sustituye por *qui* y se lleva a cabo la unión.

anquialmendrado, arquibanco, arquimesa, blanquinegro, blanquiazul, boquiancho, boquicerrado, coquiduro, moquicaído, piquijuye, piquituerto/a

Regla 18 (-za+ci, -zo+ci, -zón+ci)

Si al sustituir las vocales átonas finales *a* y *o* correspondientes al primer constituyente por el infijo compositivo *i* acaba en la sílaba *zi*, se sustituye por *ci* y se lleva a cabo la unión.

cabeciancho ← *cabeza* + *ancho*

cabeciduro ← *cabeza* + *duro*

peciluengo ← *pezón* + *luego*

Regla 19 (-n+m)

Si el primer constituyente acaba en *n* y el segundo comienza por *b* o *p* se sustituye la *n* por *m* y se lleva a cabo la unión.

almaempena, *balompié*, *mampuesto*, *montabanco*, *saltabanco*, *tentempié*

Regla 20 (copulativa)

Se sustituye la *y* copulativa que funciona como elemento conjuntivo por la vocal *i* y se lleva a cabo la unión. Esta regla se trata, desde el punto de vista computacional, de igual forma que la regla 4 (+i); se aprecia la producción de compuestos en la Figura 18.

correveidile ← *corre* + *ve* + *y* + *dile*

quitaipón ← *quita* + *y* + *pon*

tiraiafloja ← *tira* + *y* + *afloja*



Figura 18.- Producción de la regla 4 (+i) y regla 20 (copulativa) en base a la categoría gramatical de los constituyentes.

Regla 21 (-h+j)

Si el segundo elemento del compuesto empieza por *h*, se cambia por *j* y se lleva a cabo la unión.

picajuyendo ← *pica* + *huyendo*

bocajacha ← *boca* + *haya*

Regla 22 (metafonía)

Elisión de la vocal *a* final y metafonía⁷⁶ en la vocal central tónica del primer miembro.

cabizcaído ← *cabeza* + *caído*

cabizmordido ← *cabeza* + *mordido*

Regla 23 (tilde)

Hay que tener en cuenta la acentuación del compuesto. Por lo general, el acento recae en la segunda palabra del compuesto —son monoacentuados⁷⁷— debido a la existencia de un mayor número de palabras llanas en español.

aguarrás ← *agua* + *ras*

balompié ← *balón* + *pie*

abimelolleva ← *abí* + *me* + *lo* + *llevas*

espíritusanto ← *espíritu* + *santo*

ganapán ← *gana* + *pan*

catetómetro ← *cateto* + *-metro*

⁷⁶ Metafonía (de meta-, entre; del griego fono, voz; e -ía) f. Fon. Cambio de timbre que la vocal tónica sufre por el influjo de la vocal final o de un sonido vecino.

⁷⁷ Salvo en casos excepcionales como *portaaviones*.

Regla 24 (diptongo)

Por último, puede suceder que cualquiera de los constituyentes del compuesto presente una formación atípica de diptongo o alteración del mismo. Excluir de estos casos a los elementos compositivos:

- quebra* ← *quiebra*
- vola* ← *vuela*
- escorna* ← *escuerna*
- torce* ← *tuerce*
- toste* ← *tueste*
- tenta* ← *tienta*
- roda* ← *rueda*
- escola* ← *escuela*
- troca* ← *trueca*

Modelo de datos

En la Figura 19 se muestra el grafo relacional de la Base de Datos BARGCoS que refleja cómo está organizada y clasificada la información de las palabras compuestas, así como las reglas de formación y cambio gráfico definidas.

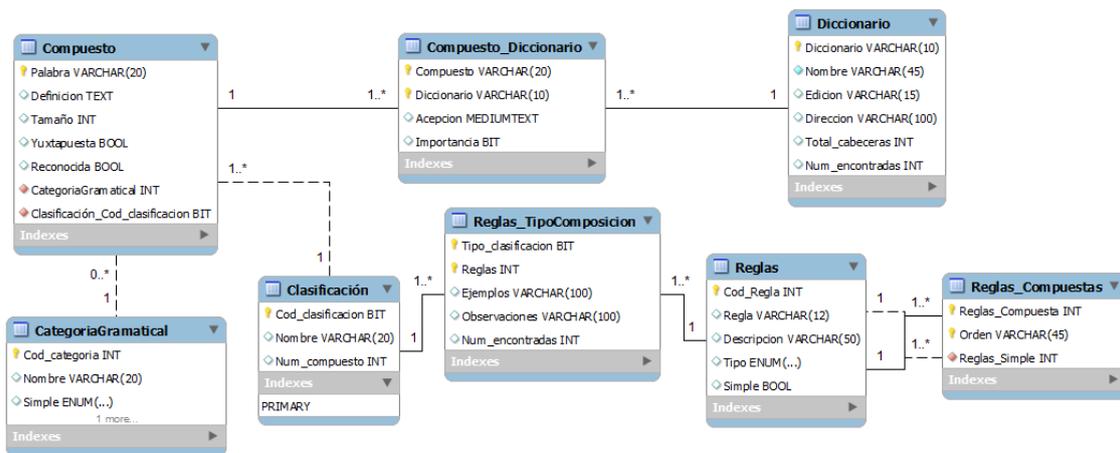


Figura 19.- Modelo de lógico global de la BBDD ARGCoS.

Se describe a continuación las entidades que se añadieron con sus propiedades correspondientes:

Entidad
Reglas
Descripción
Entidad que contiene las reglas de formación, cambio gráfico y excepciones estudiadas
Atributos
Cod_Regla: Código de regla Regla: Nombre identificativo de la regla Ejemplos: Ejemplos significativos de la regla Descripción: Definición de la regla de composición Tipo: Tipo de regla: {Formación, Cambio Gráfico, Excepción} Simple: Determina si es una regla es simple o compuesta Num_encontrados: Cantidad de palabras compuestas encontradas a las que se le aplica la regla correspondiente

Conclusiones del capítulo → Resulta inexplicable la casi total carencia de estudios que se acerquen reflexivamente a las diversas cuestiones que atañen al tratamiento automático de nuevos vocablos compositivos. En el presente capítulo se presenta una base documental de alrededor de 11 000 compuestos recopilada de diversas fuentes lexicográficas y se buscan las reglas de formación a partir del comportamiento de los vocablos constituyentes del compuesto. Se usan tales reglas para definir procesos automáticos que sean capaces de identificar y crear palabras compuestas y situarlas en un contexto lingüístico idóneo.

CAPÍTULO 4.- RECONOCIMIENTO DE COMPUESTOS

RESUMEN: Este capítulo expone el procedimiento de reconocimiento de palabras compuestas en español mediante la individualización de las formas simples —palabras o elementos compositivos— que la forman; se presta especial cuidado en el establecimiento de su comportamiento morfológico-semántico.

Introducción

Se hace necesario definir procesos automáticos que sean capaces de identificar palabras compuestas y situarlas en un contexto lingüístico idóneo que permita de manera formal y automática tratar neologismos compositivos. Resulta inexplicable la casi total carencia de estudios que se acerquen reflexivamente a las diversas cuestiones que atañen al reconocimiento automático de nuevos vocablos compositivos. La escasez de estudios monográficos sobre la composición en nuestra lengua, hizo que se llevara a cabo un estudio en profundidad sobre este tema.

Se exponen las normas aplicadas para el reconocimiento de palabras compuestas yuxtapuestas en español inducidas a partir de la realidad explorada, que permita solventar barreras lingüísticas a través del dominio informático y se manifiestan las irregularidades que se encuentran en el tratamiento automatizado.

Se parte de elementos ya presentes en el lenguaje para reconocer nuevos vocablos mediante la unión de dos o más de tales elementos; se toma como base documental BARGCoS definida en epígrafe *Modelo de datos* —aproximadamente 11 000 compuestos recopilados de diversos repertorios lexicográficos y del glosario de compuestos de Bustos Gisbert⁷⁸, entre otras—; se incluyen compuestos procedentes de diversas fuentes neológicas y nuevos diccionarios gracias a ARGCoS⁷⁹ —*Automatic Recognition and Generation Compounds in*

⁷⁸ Véase EUGENIO BUSTOS, *La composición nominal en español*, op. cit. págs. 367-463.

⁷⁹ ARGCoS es una utilidad lingüística que permite el reconocimiento y la generación automática de palabras compuestas en español desarrollada en el seno del *Grupo de Estructuras de Datos y Lingüística Computacional (GEDLC)* de la *Universidad de Las Palmas de Gran Canaria (ULPGC)*.

Spanish. Tal herramienta facilita el señalamiento de dos o más formas simples que funcionan como una unidad léxica compositiva para establecer su comportamiento morfológico-semántico por medio del grafo de dependencias de las reglas de formación y de cambios gráficos —ya se estudiaron en el capítulo anterior—; se logran plasmar así las casuísticas que se detectan en la automatización del proceso compositivo del español.

Automatización del reconocimiento de compuestos

El proceso de reconocimiento de una palabra compuesta permite identificar dos o más formas simples —palabras o elementos compositivos— que funcionan como una unidad léxica; se examina con cuidado para establecer el comportamiento morfológico-semántico, ya que no se puede deducir únicamente a partir de las propiedades de las formas simples constituyentes.

Se reconocen aquellos casos de composición y parasíntesis por composición, *Regla 15 (parasíntesis)*, que se han consolidado como la unión gráfica de los elementos que intervienen —objeto de estudio a lo largo de la presente tesis—; se obvian aquellos casos que contengan guion o nexos —espacio o preposición— y se tratan algunos de acortamiento, *Regla 13.4 (acronimia)* y *Regla 14 (abreviamento)*. Los elementos constituyentes del compuesto pueden ser palabras o elementos compositivos prefijales o sufijales, *Regla 6 (elemento compositivo)*, y estar formados por preposiciones, artículos, conjunciones o pronombres clíticos o enclíticos que no han perdido en el proceso de aglutinación, *Regla 7 (@preposición@)*.

Se parte de un conjunto de tamaño suficientemente significativo de compuestos —la base de datos BARGCoS— recogidos de los principales repertorios lexicográficos de la lengua española y del glosario de compuestos de Bustos Gisbert (ver definición de la base de referencia en los epígrafes *Corpus seleccionado* y *Modelo de datos*), y clasificados en grupos según la categoría gramatical de sus constituyentes. Para examinar si una palabra es identificada como compuesta, implica llevar a cabo alguno de los siguientes procesos: o bien se consulta en la base y se dictamina que ya ha sido tratada como tal —en cuyo caso presenta una única solución— o, en caso contrario, se procede a la aplicación de las reglas de formación y cambio gráfico que sean precisas para su reconocimiento —se puede presentar una, muchas e incluso ninguna solución. En un estudio cuyo objetivo sea la automatización de la morfología con medios informáticos, los aspectos formales o teóricos no tienen por qué

coincidir con los estrictamente lingüísticos. Así, **bocabundido* —falsa composición pues lo correcto sería *boquibundido*— podría considerarse una formación apropiada al no contravenir ninguna regla fonotáctica del lenguaje, ni siquiera la norma de la estructura silábica del español.

La estrategia diseñada conlleva que el reconocedor compositivo actúe sobre una palabra de entrada —posiblemente un neologismo compositivo— y la segmente en posibles pares (*elemento constituyente1*, *elemento constituyente2*) contruidos a partir de las reglas de formación del compuesto adecuadas —a partir de ahora, se considera una solución como la pareja formada por (*pal1*, *pal2*). De esta forma, el reconocimiento de *compuestos yuxtapuestos*, *parasintéticos por composición*, *neologismos compositivos* y *compuestos con pseudoprefijos o pseudosufijos* —entre otros procesos de lexicalización— de manera automática es una labor que aglutina procedimientos que implican diferentes tareas: (véase Figura 20):

- Módulo 1: Tratamiento de errores ortográficos o posibles neologismos.
- Módulo 2: Reglas de formación y de cambios gráficos o fonéticas.
 - ✓ Ordenación de las soluciones atendiendo a las categorías gramaticales del compuesto y de los constituyentes según la regla de formación aplicada.
- Módulo 3: Búsquedas en internet para confirmar y ordenar estadísticamente las posibles soluciones
- Módulo 4: Procedimientos semánticos que engloban a los diccionarios en la búsqueda del significado más próximo.

En el reconocimiento de palabras compuestas en español, se sigue el tratamiento de errores ortográficos, se consideran inclusiones o formaciones de neologismos, compuestos parasintéticos por composición y derivación de compuestos, entre otros —se prueba si la palabra compuesta cumple unas ciertas condiciones. Claro está que se pueden encontrar palabras que no sean compuestos, pero esta primera aproximación permite ver la necesidad de añadir otro tipo de condicionantes que funcionen como excepciones, con el objetivo de encontrar una lista completa de compuestos mediante la aplicación a los diferentes diccionarios de las reglas finales obtenidas. A través de la herramienta ARGCoS se procuran inclusiones de compuestos procedentes de diversas fuentes neológicas o nuevos diccionarios.

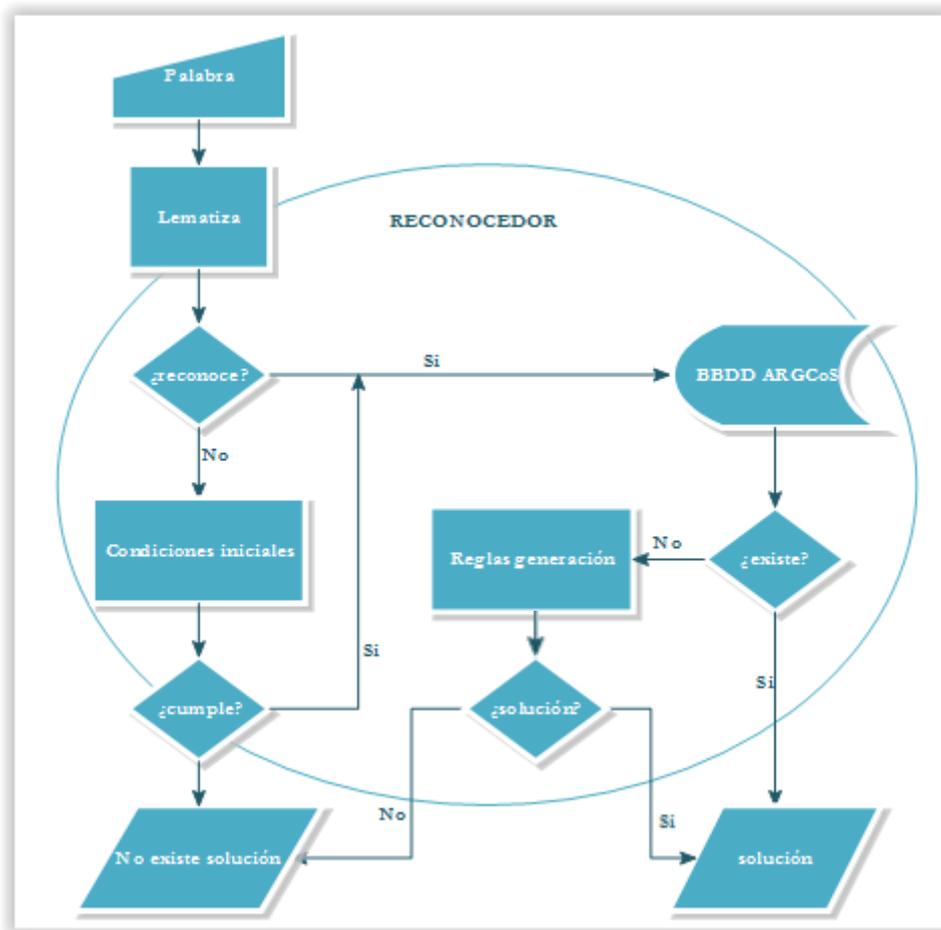


Figura 20.- Diagrama de flujo del proceso de reconocimiento de compuestos.

Si se quiere reconocer los elementos que forman la composición *boquihundido*, se segmenta la palabra hasta encontrar pares de soluciones (pal1, pal2):

Solución 1: (boquí, hundido)

Solución 2: (boca, hundido)

¿Cuál de las soluciones tiene mayor probabilidad de ser correcta?

La propuesta primera no sería la solución más verosímil debido a la invariabilidad formal que presenta el compuesto con respecto a la regla de formación aplicada según las categorías gramaticales que intervienen.

La frecuencia de uso por los hablantes del término correcto fija otra condición: se busca la coocurrencia⁸⁰ entre (boquihundido, boqui) y (boquihundido, boca) que posibilita una ordenación —la aparición más elevada permite discriminar la solución.

Por último, se busca en el significado de los elementos cuál es el que utiliza para definir el compuesto. Se tiene:

boquihundido (DRAE):

1. adj. Dícese de la caballería que tiene muy altas las comisuras de los labios.

boqui (DRAE)

1. m. Especie de enredadera de Chile, de la familia de las vitáceas, cuyo tallo, que es muy resistente, se emplea en la fabricación de cestos y canastos.

boca (DRAE)

1. f. Abertura anterior del tubo digestivo de los animales, situada en el extremo anterior del cuerpo, o sea en la cabeza. Sirve de entrada a la cavidad bucal. También se aplica a toda la expresada cavidad en la cual está colocada la lengua y los dientes cuando existen.

Se concluye que la solución segunda es la candidata: *boquihundido* → *boca* + *hundido*.

Sin embargo, si el reconocedor compositivo trata de identificar la palabra *bocabundido** —falsa composición pues lo correcto sería *boquihundido*—, encuentra que se trata de un posible neologismo y como única solución posible se obtiene: (boca, hundido) y no hace falta aplicar ninguno de los otros procesos.

boquihundido → *boca* + *hundido*

Se pasa a definir cada una de las fases del reconocedor compositivo, se hace especial hincapié en el módulo de las reglas de formación compositivas para el reconocedor.

⁸⁰ En lingüística general, se define coocurrencia a la utilización conjunta de dos unidades léxicas —en nuestro caso se toman la palabra compuesta y el primer elemento constituyente— en una unidad superior —la comunidad de hablantes, es decir, la web. Existe la suposición de que los términos son mutuamente dependientes cuando su utilización conjunta es muy común. Existen tests estadísticos que proveen medios para la supuesta dependencia, pero se considera únicamente en el presente estudio la frecuencia de aparición debido a que interesa discriminar las soluciones entre sí para un mismo compuesto, lo cual posibilita una ordenación de las mismas. Como estas coocurrencias pueden deberse tanto a razones gramaticales como semánticas, se hace necesario añadir otros tipos de condicionantes.

Fases en el reconocedor de palabras compuestas

En la Figura 20 se muestra un esquema del funcionamiento del proceso de reconocimiento de palabras compuestas. En caso de errores tipográficos, se utiliza la distancia de Levenshtein para solventarlos. Si se trata con un neologismo, se piden su definición y su categoría gramatical.

Se lematiza⁸¹ la palabra y se identifican sus diversas formas canónicas⁸², categorías gramaticales y las flexiones o derivaciones que las producen, por medio del “*Flexionador*”^{83 84} y

⁸¹ La lematización es un proceso lingüístico que consiste en, dada una forma flexionada, hallar el lema correspondiente. El lema es la forma que por convenio se acepta como representante de todas las formas flexionadas de una misma palabra: singular para sustantivos, masculino singular para adjetivos, infinitivo para verbos. <<http://lexicoon.org/es>> [1/10/2015]

⁸² Se define como forma canónica a todo vocablo con identidad propia susceptible de aplicársele o de habersele aplicado en su formación algún mecanismo de derivación.

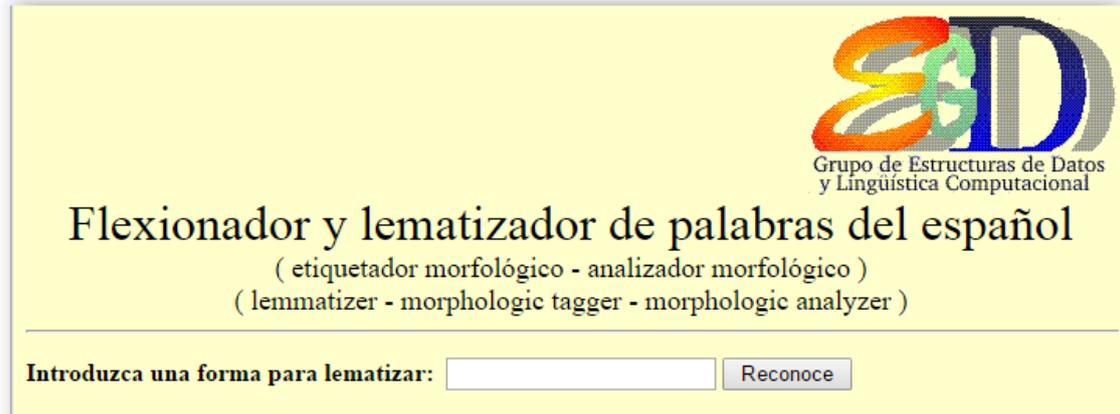
⁸³ FLANOM: *Flexionador y Lematizador Automático de Formas Nominales* y FLAPE *Flexionador y Lematizador Automático de Palabras del Español* (1999) son proyectos realizados en el seno del GEDLC. Estos trabajos representan la morfología flexiva del español actual que a través de una aplicación informática permite resolver y responder a aspectos de las palabras en el reconocimiento de las mismas como elemento funcional y morfológico: formas canónicas a las que pertenece, flexión respecto de cada forma, categoría gramatical de las formas, información semántica en el caso de los verbos e incorpora pronombres enclíticos y prefijos. La aplicación informática que se presenta lematiza cualquier palabra del español al identificar su forma canónica, categoría gramatical y la flexión o derivación que la produce, y obtiene las formas correspondientes a partir de una forma canónica y de la flexión o derivación solicitada. A partir de 151 103 formas canónicas (incluye 14 859 nombres de personas y apellidos), se obtienen algo más de 4 900 000 formas flexionadas y derivadas (sin sumar la ampliación inherente a los prefijos y a los pronombres enclíticos) y se establecen unas 90 000 relaciones morfológicas. El sistema incluye todas las entradas del Diccionario de la Real Academia, del Diccionario General de la Lengua Española Vox, del Diccionario de Uso del Español de María Moliner, del Gran Diccionario de la Lengua Española de Larousse Planeta, del Diccionario de Uso del Español Actual Clave SM, del Diccionario de voces de uso actual dirigido por Manuel Alvar Ezquerro, del Gran Diccionario de Sinónimos y Antónimos de Espasa-Calpe y del Diccionario Ideológico de la Lengua Española de Julio Casares.

<<http://gedlc.ulpgc.es/investigacion/scogeme02/lematiza.htm>> [01/10/2015]

⁸⁴ Los proyectos anteriores se integran para ofrecer un *Sistema Computacional de Gestión Morfológica del Español* (SCOGEME). Este sistema es capaz de resolver y responder a cualquier aspecto morfológico de una palabra del español: consiste en el reconocimiento, la generación y la manipulación de las relaciones morfológicas a partir de cualquier palabra, incluye la recuperación de toda su información lexicogenética hasta llegar a una primitiva, la gestión y control de los afijos en el tratamiento de sus relaciones, así como la regularidad en la

Lematizador de palabras del español” y de las “*Relaciones Morfológicas del español*” del Grupo GEDLC de la ULPGC (Santana et al., 1997, 1999, 2004,2006)⁸⁵.

Por lo general, si se trata de una palabra compuesta, las diversas formas canónicas⁸⁶ se reducirían a una, coincidiendo con la palabra en sí. Además, la categoría gramatical suele ser simple: sustantivo —en la mayoría de los casos—, adjetivo o verbo (ver Figura 22).



relación establecida. Proporciona una visión global del comportamiento y productividad de las palabras del español en los principales procesos de formación (sufijación, prefijación, parasíntesis, supresión, regresión, modificación-cero, apócope, metátesis y otros no clasificables que generan grafías alternativas).

<<http://gedlc.ulpgc.es/investigacion/scogeme02/reльмоrfo.htm>> [01/10/2015]

⁸⁵ OCTAVIO SANTANA SUÁREZ, FRANCISCO CARRERAS RIUDAVETS, JOSÉ PÉREZ AGUIAR, JUAN C. RODRÍGUEZ DEL PINO, “Parasyntetic Morpholexical Relationships of the Spanish: Lexical Search beyond the Lexicographical Regularity”, *Proceedings of the LADIS International Conference. Applied Computing*, 2006, págs. 627-631; OCTAVIO SANTANA SUÁREZ, JOSÉ PÉREZ AGUIAR, FRANCISCO CARRERAS RIUDAVETS, JUAN DE DIOS DUQUE, ZENÓN HERNÁNDEZ FIGUEROA, GUSTAVO RODRÍGUEZ RODRÍGUEZ, “FLANOM: Flexionador y lematizador automático de formas nominales”. *Lingüística Española Actual XXI*, 2, Arco/Libros, S.L., 1999, págs.253-297; OCTAVIO SANTANA SUÁREZ, JOSÉ PÉREZ AGUIAR, FRANCISCO CARRERAS RIUDAVETS, GUSTAVO RODRÍGUEZ RODRÍGUEZ, “Suffixal and Prefixal Morpholexical Relationships of the Spanish”. *Lecture Notes in Artificial Intelligence*, 3230, Springer-Verlag, 2004, págs. 407-418; OCTAVIO SANTANA SUÁREZ, JOSÉ PÉREZ AGUIAR, ZENÓN HERNÁNDEZ FIGUEROA, FRANCISCO CARRERAS RIUDAVETS, GUSTAVO RODRÍGUEZ RODRÍGUEZ, “FLAVER: Flexionador y lematizador automático de formas verbales”. *Lingüística Española Actual XIX*, 2, Arco/Libros S.L., 1997.

⁸⁶ En el corpus de referencia del GEDLC, se consideran como tales las palabras resultantes de la unión de todas las entradas de la fuentes consultadas —119 959 formas no verbales y 14 150 infinitivos—, siempre que posean un significado institucionalizado, independientemente de que en su formación entrara sufijo flexivo —*colchón*, *cafetucho*; pero no se consideran *mundonón*, *animalucho*— y que sean reconocidas por la fuente como una entrada distinta de la proporcionada por un afijo como elemento dependiente del léxico.



Figura 21.- Portal web del Flexionador y Lematizador Morfológico y de Relaciones Morfológicas del Grupo de Investigación GEDLC de la UPLGC.

Se analiza si la palabra es compuesta sin prefijo y derivada —en cuyo caso se trataría de derivación de composición y se pasaría a reconocer su forma canónica— o pudiera ser parasintética por composición. Si no existiera ninguna forma canónica incluida en la base de estudio de compuestos, se aplicarían los *Criterios en el reconocimiento de nuevos compuestos* para intentar nuevas soluciones; de no ser posible, no cabría solución en el reconocimiento.

Criterios en el reconocimiento de nuevos compuestos

El reconocedor compositivo actúa sobre una palabra de entrada: primero probaría que cumpla unas ciertas condiciones, tanto respecto a su tamaño como a la relevancia del orden de sus constituyentes y luego llevaría a cabo su segmentación en posibles pares de soluciones: (pal1, pal2).

Por norma general, los compuestos presentan una longitud mayor que cinco caracteres o están formados por al menos tres sílabas, como *uñaalbo* o *abrojo* —habitualmente los vocablos en español son bisilábicos o trisilábicos, por lo que se puede decir que los compuestos nominales contienen de cinco a seis sílabas.

El reconocimiento opera mediante particiones de la palabra hasta que se localice un vocablo íntegro o ambos, a los que se les aplica la regla correspondiente o se trata como

excepción —se utilizan tanto el “*Flexionador y Lematizador de palabras del español*” como las “*Relaciones Morfológicas del español*” del GEDLC que se muestra en la Figura 21.

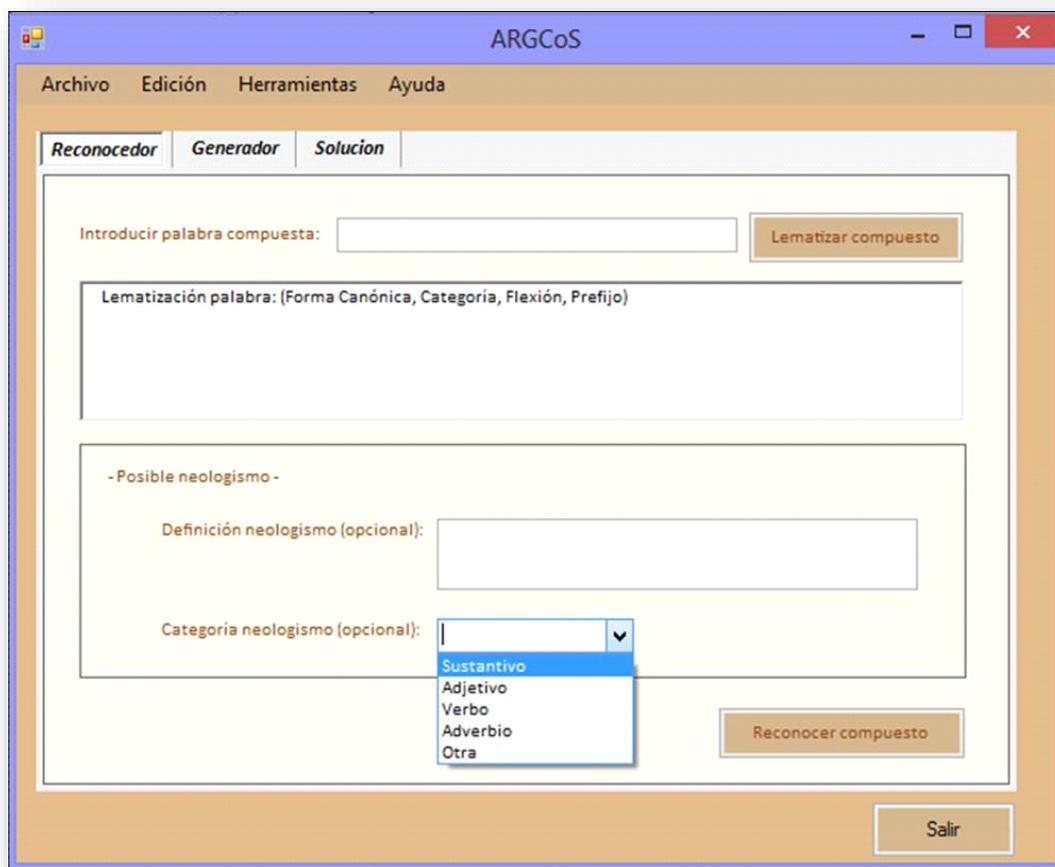


Figura 22.- Interfaz de la aplicación ARGCoS para el reconocimiento de compuestos.

Durante el proceso se generan cortes que pueden dar lugar a múltiples posibilidades entre las que pueden aparecer algunas que no sean auténticas soluciones o, incluso, todas pueden ser incorrectas. Un conjunto de 3 197 palabras compuestas se sometieron al reconocedor del aplicativo ARGCoS y se obtuvo 1,56 como media del número de soluciones que podría tener una palabra compuesta, es decir en torno a 2 y cómo máximo 5 —resultado alentador por reconocer un número considerable de soluciones posibles. Los datos recogidos se muestran en la Tabla 10. Se obtuvo al menos una solución para cada compuesto, pues se partía de palabras que eran compuestas, formadas a partir de una de las reglas definidas en el Capítulo 3. Se observa que se genera un conjunto de soluciones donde algunos de los elementos del compuesto no eran palabras sin prefijo ni marca derivativa. Esta observación permite añadir como condicionante la tendencia a la supresión de marcas flexivas:

- ✓ las palabras que forman el compuesto no son derivadas sino que forman una única unidad léxica, sin prefijos básicamente en el primer constituyente ni marca de apreciativo en el segundo componente.

Número soluciones	Frecuencia
1	2 256
2	676
3	177
4	80
5	8
Total compuestos tratados:	3 197
Cantidad soluciones encontradas:	4 999

Tabla 10.- Número de soluciones para un conjunto de compuestos.

Obtención de las reglas de formación para el reconocimiento

En el Capítulo 3 se plantea que a partir de la base documental BARGCoS, se investigan las reglas de formación de compuestos. Hay que tener en cuenta que en un estudio cuyo objetivo sea la automatización de la morfología con medios informáticos, los aspectos formales o teóricos no tienen por qué coincidir con los estrictamente lingüísticos. Así, la falsa composición *albonegro** —debería ser *albinegro*— no tendría por qué tratarse de una formación errónea ya que no contraviene ninguna regla compositiva del español: se segmenta la palabra hasta encontrar como solución el par (albo, negro) obtenida de aplicar la *Regla 1 (unión)*, para un compuesto adjetival formado por: [Adjetivo+Adjetivo]_{Adjetivo}.

Se establecen las reglas de formación para el reconocimiento a partir del estudio del comportamiento de los vocablos constituyentes del compuesto; en general, coinciden con las tratadas por los lingüistas, aunque con alguna adaptación informática justificada por el comportamiento mayoritario observado. Hay que tener en cuenta que cuando se aplican las reglas para reconocer los compuestos hay que considerar los cambios gráficos que intervengan para su correcto tratamiento (véase epígrafe *Reglas fonéticas o de cambio gráfico*).

Las reglas definidas en el capítulo anterior operan en sentido inverso: recorrerlas en sentido contrario implica que funcionen en la modalidad de reconocimiento. A continuación

se presenta, en la Tabla 11, la correspondencia entre las reglas según el proceso que se aplique: reconocimiento o generación.

Tipo Regla	Número	Generación	Reconocimiento
		Nombre	Nombre
Formación	1	unión	unión
	2	-a2	+a2
		-e2	+e2
		-o2	+o2
		-ha	+ha
		-ho	+ho
3	-a+i	-i+a	
	-o+i	-i+o	
	-e+i	-i+e	
	-eo+i	-i+eo	
4	+i	-i	
5	-a	+a	
	-o	+o ⁸⁷	
	-e	+e	
6	Elemento compositivo	posponer guion a pal1 → pal1- anteponer guion a pal2 → -pal2	
Excepción	10	Excepción: género	-o+a -a+o
	13.1	2ª palabra plural	-s -es
Fonética	16	+rr	-rr
	17	-ca+qui	-qui+ca
		-co+qui	-qui+co
	18	-za+ci	-ci+za
		-zo+ci	-ci+zo
		-zón+ci	-ci+zón
	19	-n+m	-m+n
	20	copulativa	-i
21	-h+j	-j+h	
23	tilde	tilde	

Tabla 11.- Tabla de correspondencia entre reglas de generación y reglas de reconocimiento.

⁸⁷ Existen casos en los que parecería que más aplicar esta regla, se tratara de la regla -i+o pues la palabra pal1 acaba en i —se observa que sólo se ha dado para los casos +a y +o.

Existen reglas en la generación que reciben un tratamiento especial en el reconocimiento debido a su implementación. Tal es el caso de las reglas de parasíntesis, acortamiento, variantes lexemáticas, preposiciones, paranomasia,... que se aplican en caso de no encontrar solución alguna en el reconocimiento.

Integrar los procesos automáticos de reconocimiento y generación de compuestos en un aplicativo permite definir las reglas de composición de forma única: recorrer todas las reglas en un sentido o en el contrario implicaría que la herramienta funcionara de una forma u otra: reconocimiento o generación.

Grafo de dependencias de las reglas de formación y de cambio gráfico

Un grafo de dependencias de reglas de composición es un modelo de computación en el que se considera un conjunto finito de nodos posibles —en el presente caso se trataría de las reglas de formación y de cambio gráfico— y una función de transición que establece cómo se cambia de un nodo a otro según las entradas que se dispongan en cada instante. Se establece una relación entre las reglas de ambos nodos con un sentido —grafo dirigido— que fija el paso entre reglas con las condiciones de tránsito, si se especifican en la correspondiente etiqueta.

Se parte de un corte inicial de la palabra compuesta, generalmente se corresponde con el primer elemento componente del compuesto. A partir de esta entrada, se atraviesa el grafo y se aplica en cada punto la regla de formación o fonética correspondiente. En caso de ser reconocida la transformación del elemento constituyente entrante, se registran los resultados obtenidos durante el recorrido secuencial. Se transita el grafo de dependencias de las reglas hasta que se hayan recorrido todos los caminos posibles.

En la Figura 23 se muestra una parte del esquema total de representación de un grafo de dependencias de las reglas, aplicado a un estado inicial —un corte determinado de la palabra a reconocer—, para el que se quiere ejecutar un conjunto de reglas de formación de compuestos. Se recorre el grafo de forma que a una palabra se le apliquen las reglas de formación y fonéticas que cumplan las condiciones de tránsito entre las reglas definidas.

❖ Si se consideran las palabras:

labibendido, calicanto, agridulce, cochifrito, azuliblanco, liquidámbar, peciluengo

cuyos estados iniciales respectivos, en un momento o corte determinado, son:

labi, cali, agri, cochi, azuli, liquid, peci

Para cada caso, se recorre el grafo por todos los caminos posibles en función de las transiciones que se encuentren. Se obtienen las soluciones:

labi → labio, labe

cali → cali, cali-, cal⁸⁸, cal, cala, calo

agri → agri agri- agra agro agre agrio

cochi → cochi coch⁸⁹ cocha coche cocho

azuli → azul

liquid → liquido, líquido

peci → pezón

El recorrido del grafo de dependencias define el orden y las reglas que se aplican en cada momento para todas las soluciones obtenidas. Se consigue definir las dependencias y comportamientos entre las reglas de composición de forma precisa. De esta forma, es posible añadir de manera eficaz nuevas reglas de reconocimiento que no hubieran sido contempladas originalmente.

⁸⁸ No se considera solución pues proviene de una flexión de verbo que no se corresponde en composición.

Forma canónica: caler

Categoría: verbo intransitivo (palabra antigua o desusada)

Flexión: 1ª per. sing. pret. indefinido

⁸⁹ Esta solución no se considera pues se trata de una forma prefijada, se muestra el resultado que devuelve el lematizador:

Forma canónica: chi

Categoría: sustantivo femenino

Flexión: forma canónica (femenino singular)

con prefijo: co-

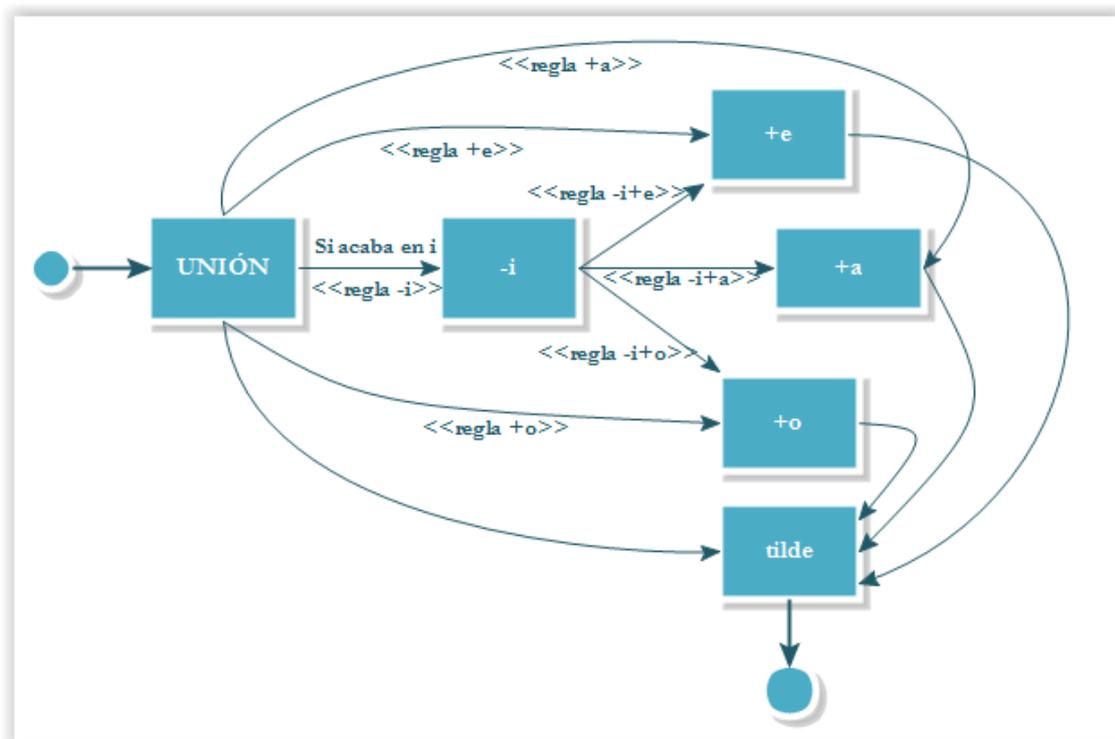


Figura 23.- Extracción del grafo de dependencias de las reglas de formación y cambio gráfico para el reconocimiento.

Para los ejemplos anteriores se obtienen los pares (pal1, pal2) de soluciones:

✓ *labibendido*

Solución 1: (labio, hendido) Reglas: +o

Solución 2: (labe, hendido) Regla: -i+e

✓ *calicanto*

Solución 1: (cal, canto) Reglas: -i

Solución 2: (cali, canto) Regla: unión

Solución 3: (cala, canto) Reglas: -i+a

Solución 4: (calo, canto) Reglas: -i+o

Solución 5: (cali-, canto) Reglas: elemento_prefijal

✓ *agridulce*

Solución 1: (agri, dulce) Reglas: unión

Solución 2: (agra, dulce) Regla: -i+a

	Solución 3: (agre, dulce)	Reglas: -i+e
	Solución 4: (agro, dulce)	Reglas:-i+o
	Solución 5: (agrio, dulce)	Reglas: +o
	Solución 6: (agri-, dulce)	Reglas: elemento_prefijal
✓	<i>cochifrito</i>	
	Solución 1: (cochi, frito)	Reglas: unión
	Solución 2: (cocha, frito)	Regla:-i+a
	Solución 3: (coche, frito)	Reglas: -i+e
	Solución 4: (cocho, frito)	Reglas:-i+o
✓	<i>azuliblanco</i>	
	Solución 1: (azul, blanco)	Reglas: -i ⁹⁰
✓	<i>liquidámbar</i>	
	Solución 1: (líquido, ámbar)	Reglas: +o/tilde
	Solución 2: (liquido, ámbar)	Reglas: +o
✓	<i>peciluengo</i>	
	Solución 1: (pezón, luengo)	Reglas: -ci+zón

Se registra en la base de datos BARGCoS la información de las posibles soluciones obtenidas durante el proceso de reconocimiento compositivo; se marca, si procede, cuál es la solución recomendada como posible.

Base de datos de las soluciones

Se amplía la base de estudio BARGCoS para incluir las soluciones al reconocimiento de compuestos y se tiene en cuenta las inclusiones de nuevos compuestos procedentes de nuevas fuentes neológicas o nuevos diccionarios. En la Figura 24 se muestra el grafo relacional de la base de datos que refleja cómo está organizada y clasificada la información.

⁹⁰ Como sólo existe una solución, se propone como tal.

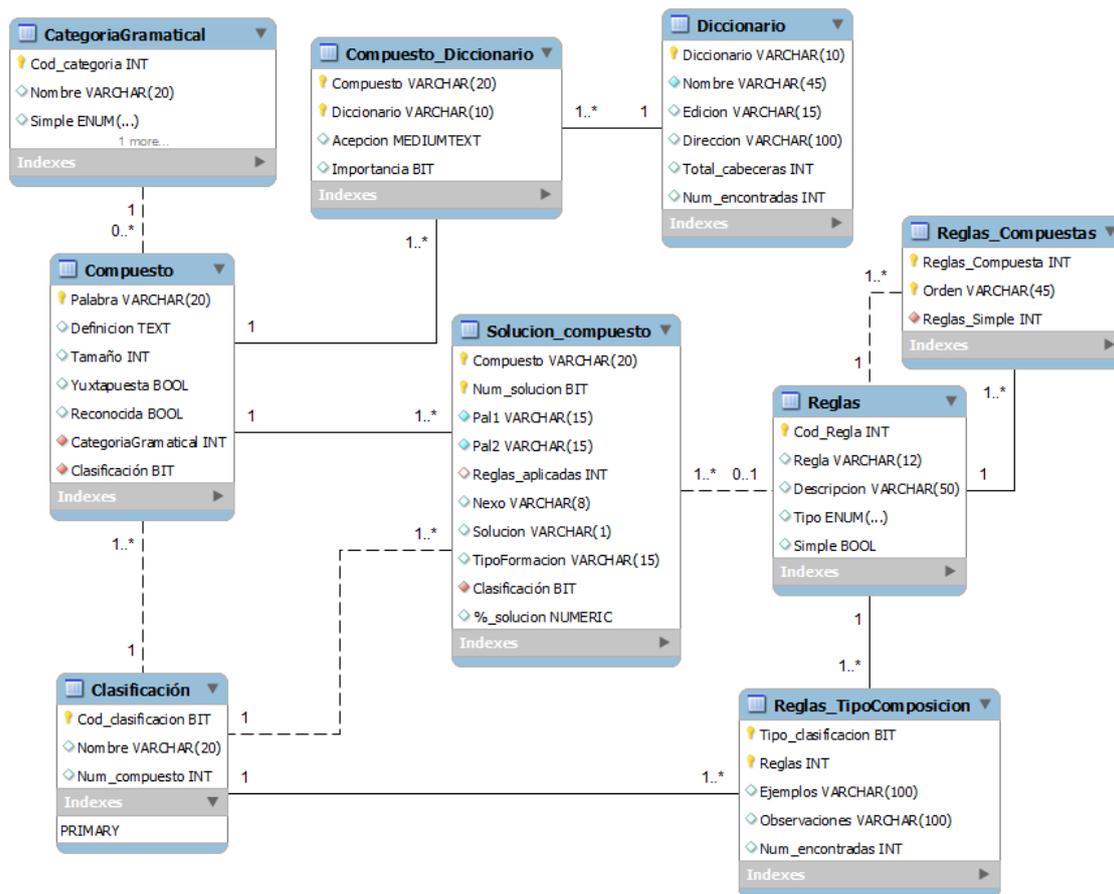


Figura 24.- Modelo lógico global de la Base de Datos BARGCoS.

Se describe a continuación la entidad añadida y sus propiedades correspondientes:

Entidad
Solucion_compuesto
Descripción
Entidad que contiene las diferentes soluciones de las palabras compuestas.
Atributos
Compuesto: Palabra compuesta reconocida
Num_solucion: Número de solución asignada {1..10}
Pal1: Primer elemento constituyente del compuesto
Pal2: Segundo elemento constituyente del compuesto
Reglas_aplicadas: Reglas de formación y/o fonéticas aplicada a la solución

Nexo: Indica si existe o no un nexo en la solución: {"+" sin espacio, "-" guion, "@" espacio, "+preposición+, @preposición@,..}

Solucion: Indica si la solución es correcta o no {V Verdadero, F Falso, Q Quizás, ? Duda}

TipoFormacion: Tipo de formación de palabra utilizado {Composición, Culta, Híbrida, Parasíntesis, Acortamiento,..}

Clasificacion: Clasificación del compuesto según las categorías gramaticales de los elementos constituyentes (pal1, pal2): {*Adjetivo+Adjetivo, Sustantivo+Sustantivo,..*}

%_solucion: Peso o importancia de la solución con respecto al resto.

Una vez recogido el corpus, se estudian los elementos que lo forman y las alteraciones que han sufrido los mismos para poder deducir las reglas de formación y los cambios gráficos que permitan la identificación automatizada de palabras compuestas yuxtapuestas. Se obtiene así un conjunto de reglas y sus excepciones que facultan el reconocimiento y generación automática de palabras compuestas.

Ordenamiento de las soluciones según categorías

Según la categoría gramatical del compuesto y la de los elementos que constituyen la solución propuesta (pal1, pal2), se ordenan las soluciones con las mismas características atendiendo a la regla de formación compositiva aplicada; se considera la solución más probable aquella que sea la más abundante en dicho tipo de composición. Se muestra la ordenación para cada categoría gramatical según las preferencias de la regla con un porcentaje que indica la relevancia de la regla (véanse los gráficos de barras y radiales correspondientes, Figura 25, Figura 26, Figura 27 y Figura 28).

Se muestra un procedimiento algoritmo para la ordenación de las reglas:

- Se buscan las soluciones que devienen de aplicar la regla más fructífera según la categoría gramatical de la palabra compuesta.

En caso de dos o más soluciones procedentes de aplicar la misma regla:

- Se analiza el tipo de composición que forman los pares (pal1, pal2) y se escoge el más abundante.

En caso de que dos o más soluciones sean del mismo tipo:

- ✓ Se aplica la misma ponderación.

Este proceso se repite hasta ordenar todas las soluciones.

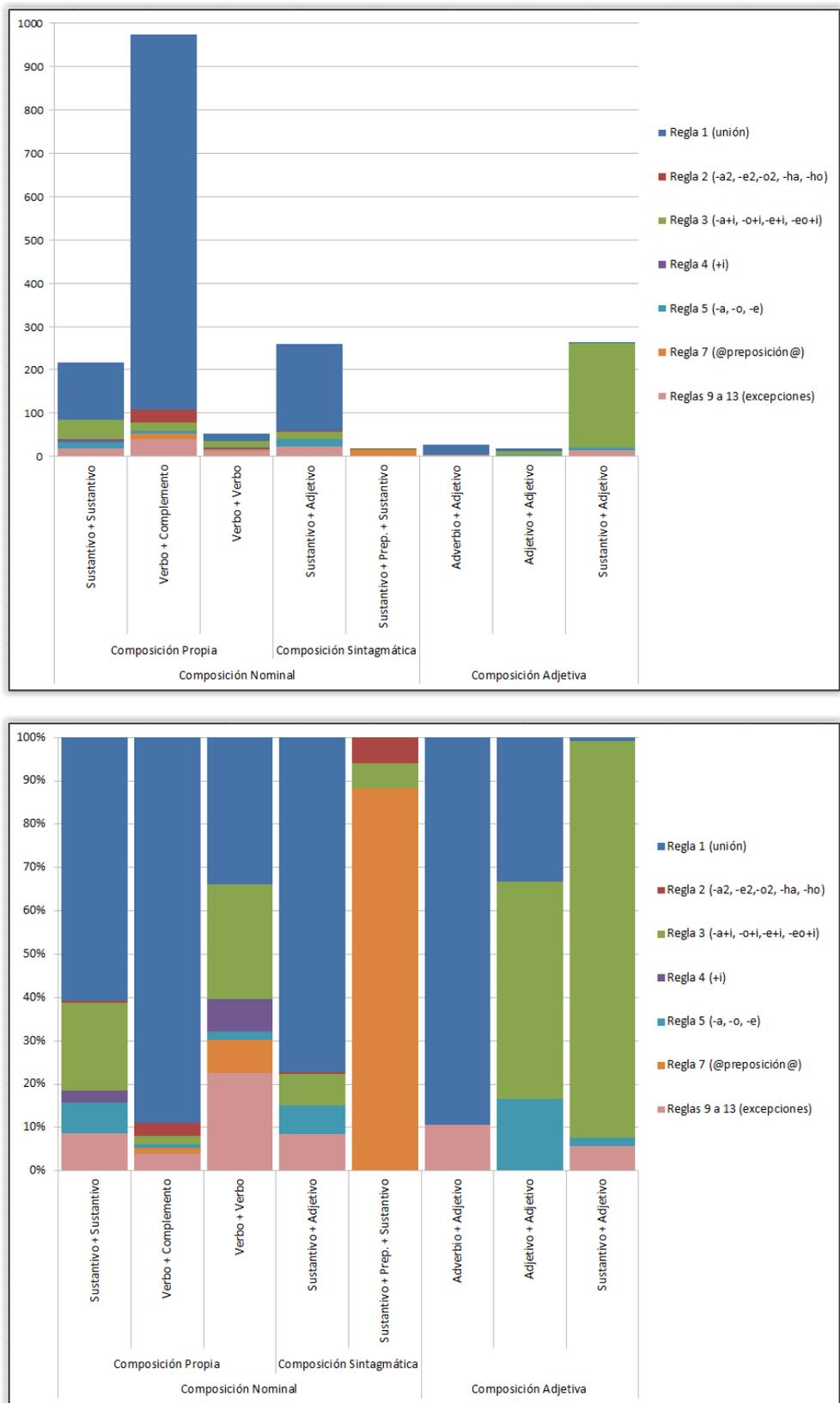


Figura 25.- Distribución del total de cada tipo composición según regla de formación en valores absolutos —imagen superior— y relativos —imagen inferior.

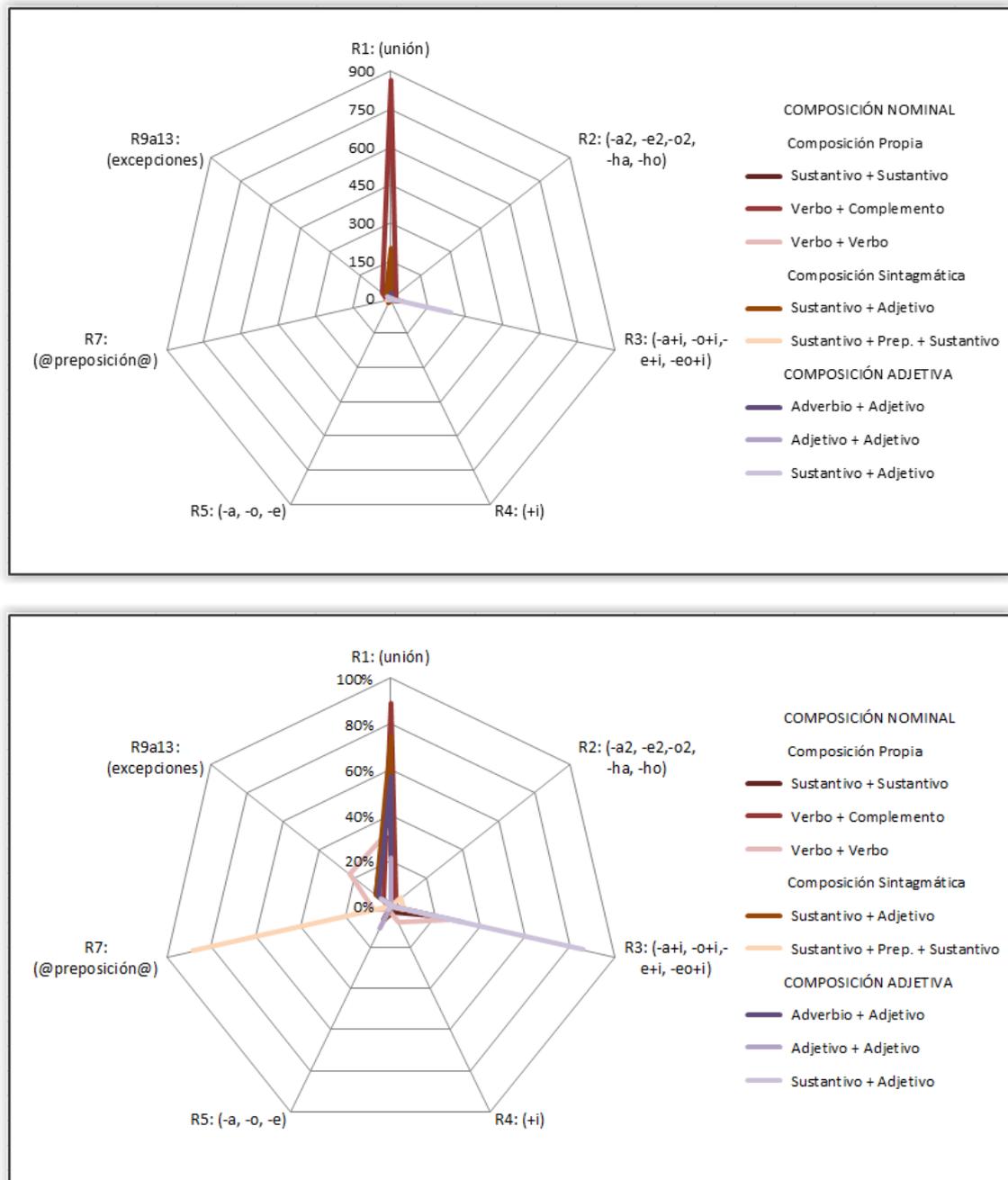


Figura 26.- Gráfico radiales para la distribución del total de cada tipo compositivo según la regla de formación en valores absolutos —imagen superior— y relativos —imagen inferior.

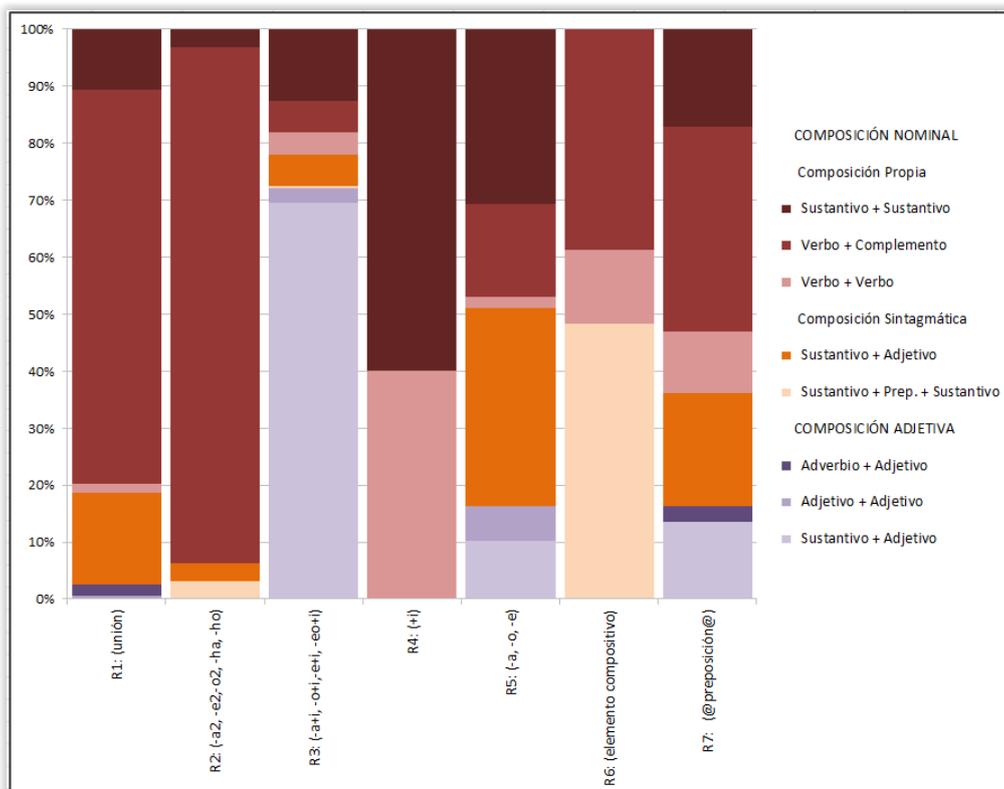
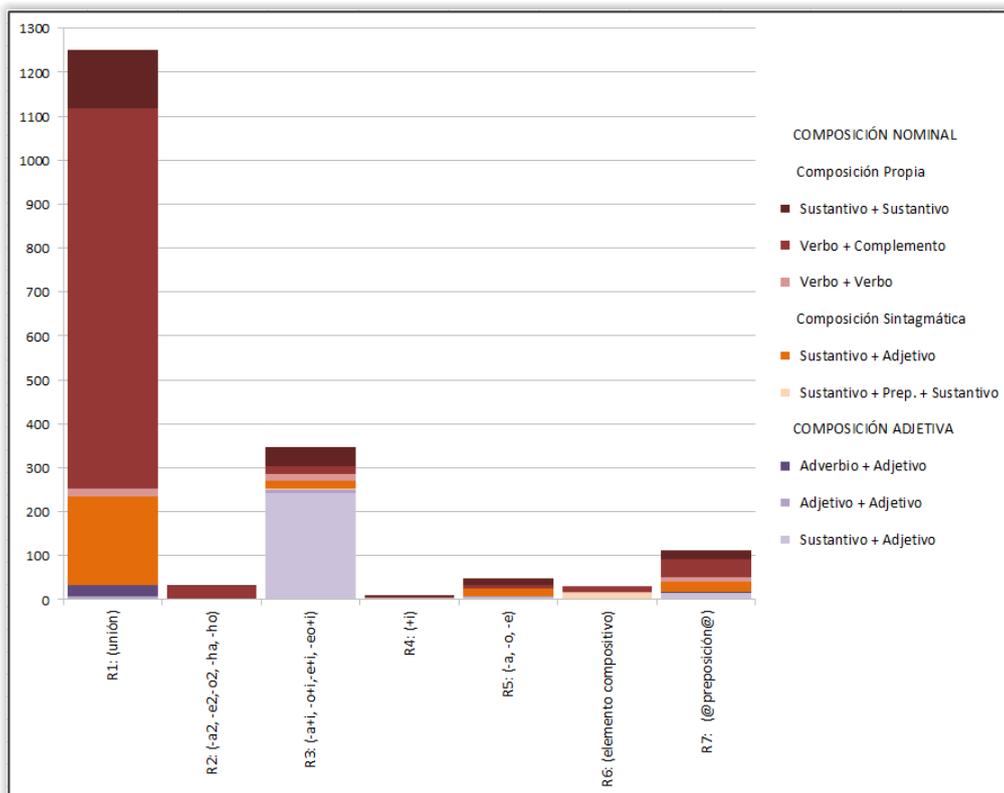


Figura 27.- Distribución del total para cada regla de formación según tipo composición en valores absolutos —imagen superior— y relativos —imagen inferior.

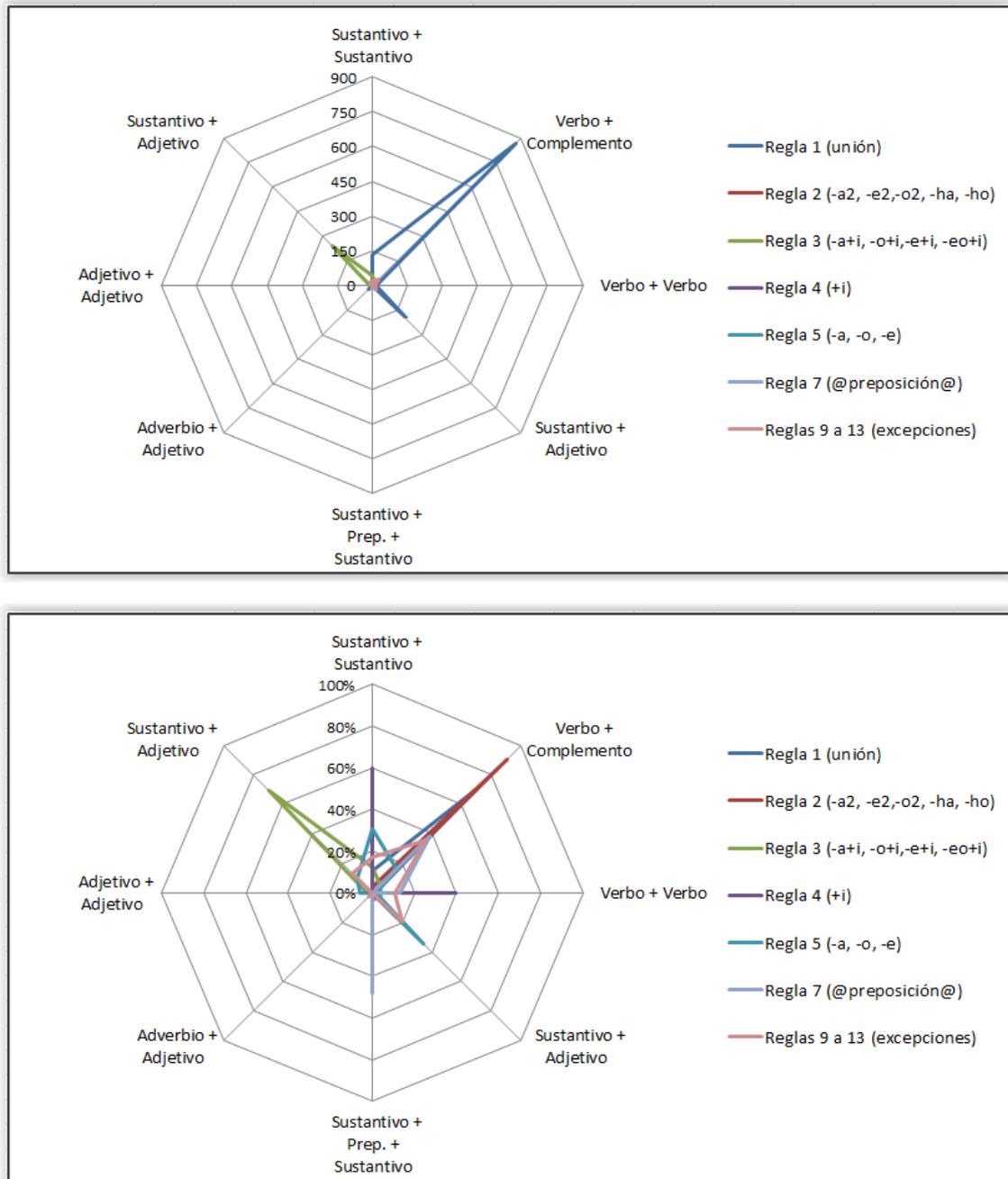


Figura 28.- Gráficos radiales para cada regla de formación según tipo composición en valores absolutos y relativos, respectivamente.

Para la palabra compuesta *cochifrito*, las soluciones quedan ordenadas de la siguiente forma:

cochifrito → Composición Nominal

Solución 1: (cochi, frito)	Reglas: unión		50%
Solución 2: (cocha, frito)	Reglas: -i+a	Tipo: <i>Sustantivo+Adjetivo</i>	16,6%
Solución 3: (coche, frito)	Reglas: -i+e	Tipo: <i>Sustantivo+Adjetivo</i>	16,6%
Solución 4: (cocho, frito)	Reglas: -i+o	Tipo: <i>Sustantivo+Adjetivo</i>	16,6%

Obtención de la solución según significado

Finalmente, se analizan las soluciones obtenidas para poder garantizar que los elementos constituyentes de cada solución se aproximan semánticamente a la palabra compuesta reconocida, de manera que se pueda garantizar mínimamente su aproximación, debido a que se entiende por composición al proceso que:

«consiste en hacer de dos o más elementos —llamados elementos componentes— una nueva palabra cuyo significado —muchas veces en sentido figurado— suele ser distinto y más extenso que la suma de los significados de sus componentes⁹¹»

Para poder llevar a cabo esta fase, se recurre a procedimientos semánticos que engloban a los diccionarios en la búsqueda del significado más próximo. Se hace difícil encontrar en los compuestos exocéntricos su significado ya que no se desprende de la suma de sus integrantes, sino que ostenta un significado añadido, imprevisible y de forma figurada generalmente.

Para los ejemplos anteriores se obtiene la información semántica para las soluciones de la palabra compuesta *cochifrito*:

⁹¹ La definición de la palabra compuesta debe verificar que, en caso de ser endocéntrica:

$$\begin{aligned}
 & \text{palabra} \in (\text{definición_pal1} \cup \text{definición_pal2}) \text{ OR} \\
 & \text{grupo_semántico_palabra} \in (\text{definición_pal1} \cup \text{definición_pal2})
 \end{aligned}$$

cochifrito (VOX)

cochifrito

(de *cocho I* + *frito*)

1 *Sustantivo masculino*. Guisado de cabrito o cordero medio cocido y después frito y condimentado.

❖ Solución 1: (cochi, frito)

cochi

1 *Sustantivo masculino*. Voz con que, pronunciada repetidamente, se llama a los cerdos.

frito, -ta

1 pp. irreg. de *freír*.

2 *Sustantivo masculino*. Fritada.

3 Manjar frito.

❖ Solución 2: (cocha, frito)

cocha

(quechua *kocha*, laguna)

1 *Sustantivo femenino*. En el beneficio de los metales, estanque que se separa del lavadero principal con una compuerta.

2 *Chile, Ecuador*. Laguna, charco.

3 *Perú*. Espacio grande y despejado, era, pampa.

❖ Solución 3: (coche, frito)

II) coche

(de *cocho II*)

1 *Sustantivo masculino*. Cerdo.

2 *América*. *coche de monte*, puerco de monte, pécarí o jabalí de las tierras cálidas.

I) coche

(checo *kochi*)

1 *Sustantivo masculino*. Carruaje, gralte. de cuatro ruedas, con una caja, dentro de la cual hay asientos para dos o más personas: *coche de camino*, el destinado para hacer viajes; *coche de rúa*, el que no era de camino; *coche de estribos*, el que tenía asientos en las portezuelas.

2 por extensión. Vehículo movido por un motor propio, con tres o más ruedas; esp. el de pequeño tamaño destinado al transporte de seis personas o menos: *coche celular*, vehículo acondicionado para transportar personas detenidas por la autoridad; *coche escoba*, el que va recogiendo a los corredores que abandonan la carrera; *coche de bomberos*, el equipado con los medios necesarios para la extinción de incendios; *coches de choque*, diversión de feria en la que coches eléctricos chocan entre sí; *coche de línea*, el que lleva pasajeros entre dos o más poblaciones; *coche de plaza o de punto*, el matriculado y numerado con destino al servicio público por alquiler, y que tiene un punto fijo de parada; *coche fúnebre*, el construido para la conducción de cadáveres al cementerio; *coche patrulla*, el de la policía.

3 por extensión. Tranvía o vagón de ferrocarril para pasajeros: *coche cama*, el que lleva camas para los viajeros.

❖ Solución 4: (cocho, frito)

I) cocho, -cha

1 p. irreg. de *cocer*.

II) cocho, -cha

(voz imitativa)

1 *Sustantivo masculino, sustantivo femenino*. *Asturias, Galicia, Logroño*. Cerdo.

2 Sucio, asqueroso.

La solución propuesta para la palabra compuesta *cochifrito* es: *cocho* + *frito* (solución 4).

Conclusiones del capítulo → A pesar de la presumible trascendencia que existe en el proceso formativo compositivo del español, han resultado infructuosas las búsquedas de referencias sobre procesamiento automático de la composición en español. Se hace necesario definir procesos automáticos que sean capaces de identificar palabras compuestas y situarlas en un contexto lingüístico adecuado, permitiendo de manera formal y automática tratar neologismos compositivos. Al mecanizar las reglas de formación, justificadas por su comportamiento mayoritario observado y su frecuencia de uso, se ha conseguido definir las dependencias y los comportamientos entre las reglas de composición de forma precisa, lo que permite de manera eficaz añadir nuevas reglas no contempladas originalmente.

CAPÍTULO 5.- GENERACIÓN DE COMPUESTOS

RESUMEN: Este capítulo explica el procedimiento de generación compositiva en español que consiste en interaccionar específicamente dos o más formas simples —palabras o elementos compositivos— para originar una determinada palabra compuesta o un neologismo compositivo.

Introducción

En el capítulo anterior se introduce la importancia de definir procesos automáticos que sean capaces de identificar palabras compuestas —*reconocimiento compositivo*— y situarlas en un contexto lingüístico idóneo que permita de manera formal y automática tratar neologismos compositivos. En un siglo como el actual en el que se vive una acelerada innovación en muchos aspectos de la vida, debido en gran parte a la informática y a la rapidez de las comunicaciones⁹², se ha producido un importante dinamismo lingüístico en el léxico que hace necesario determinar procesos que permitan crear neologismos compositivos a partir de elementos de la lengua.

Muchas discusiones hay al respecto a la hora de considerar si los neologismos, en nuestro caso compositivos, no se consideren palabras que se hayan consolidado en la lengua al no estar recogidas en los principales reportorios lexicográficos aunque no todos los resultados posibles se recogen en los diccionarios; en particular según el DRAE(2014):

«sino únicamente aquellos que han alcanzado cierta reiteración en el uso. No es necesario señalar que la ausencia de un derivado o compuesto posible no implica su legitimidad»

No se considera en la presente Tesis analizar la productividad de la innovación léxica propuesta, si a proveer de un mecanismo compositivo que permita relacionar dos o formas simples para originar una determinada palabra compuesta o un neologismo compositivo, de forma que verifique las reglas de generación compositiva definidas en Capítulo 4.

⁹² Véase MANUEL CASADO VELARDE, *La innovación léxica en el español actual*, op.cit.

Automatización de la generación de compuestos

El proceso de generación de palabras compuestas consiste en interaccionar específicamente dos o más formas simples —palabras o elementos compositivos— para dar origen a una determinada palabra compuesta, mediante la aplicación de las reglas de generación correspondientes según las categorías gramaticales de los elementos simples constituyentes así como la categoría composicional.

Se trata de generar palabras compuestas en español, teniendo en cuenta las inclusiones o formaciones neológicas, parasintéticos por composición, entre otros procesos de lexicalización, de forma que los elementos constituyentes se hayan consolidado gráficamente —se obvian aquellos casos que contengan guion o espacio— y que cumpla unas ciertas condiciones. Los elementos constituyentes del compuesto son palabras o elementos compositivos prefijales o sufijales —pseudoprefijoides o pseudosufijoides—, pero nunca erróneas o inventadas; pueden estar formados por preposiciones, artículos, conjunciones o pronombres clíticos o enclíticos que se quieran unir en la generación.

Anexar dos o más elementos para generar una palabra compuesta, implica llevar a cabo alguno de los siguientes procesos: o bien se consulta en la base documental si ya ha sido generada con anterioridad —en cuyo caso se presenta una única solución— o, en caso contrario, se procede a la aplicación de las reglas de formación y cambio gráfico que sean precisas para su generación; en ambos casos, se presenta al menos una solución.

Hay que tener en cuenta que los aspectos formales o teóricos no tienen por qué coincidir con los estrictamente lingüísticos, en un estudio cuyo objetivo sea la automatización de la generación de compuestos con medios informáticos. Así, si se quiere componer los adjetivos:

claro + vidente [Adjetivo + Adjetivo]_{Adjetivo}

se obtiene, entre las soluciones con mayor probabilidad de ser composición, el resultado de aplicar la regla del morfema compositivo “i”:

claro + vidente → *clarividente*

No obstante *clarovidente** —resultado de aplicar la regla de “unión”— también podría serlo.

claro + vidente → *clarovidente*

Sin embargo, si ahora se quisiera fusionar *cocho* —cocer— y *frito* —fritada—, la regla más usual que se aplicaría en este caso sería “*unión*” obteniendo *cochofrito* —semánticamente corresponderían los significados de los elementos que lo componen y el compuesto en sí. También aparecería como posible solución *cochifrito*, aunque la semántica de los grafismos empleados —*cochi* existe y significa voz de cerdo— no coincide con su significado.

La estrategia diseñada conlleva que el generador compositivo actúe sobre dos elementos de entrada y un posible nexos (*pal1*, *pal2*, *nexo*) para llevar a cabo su fusión ortográfica mediante la aplicación de las reglas de generación de compuestos adecuadas. De esta forma, la generación de manera automática de *palabras compuestas* —origina un sustantivo, un adjetivo o, incluso, verbal—, mediante compuestos *cultos o híbridos*, *parasintéticos compositivos* o *acortamientos*, entre otros procesos lexicográficos, es una labor que engloba procesos que implican diferentes tareas, Figura 29:

- Módulo 1: Tratamiento de errores ortográficos
- Módulo 2: Reglas de formación y cambios gráficos o fonéticas
 - ✓ Ordenación de las soluciones según la categoría gramatical del compuesto, de sus elementos constituyentes y de las reglas de formación aplicada.
- Módulo 3: Procedimientos semánticos, en caso de dar la definición de los elementos constituyentes del compuesto.

Para generar palabras compuestas se tratan los errores ortográficos —los elementos constituyentes del compuesto deben de ser palabras reconocidas—, se incluye o no de preposiciones o partículas que funcionen como nexos y se realiza el tratamiento de la parasíntesis por composición. El resultado puede no ser una palabra reconocida en el léxico, de aquí la innovación que producen los hablantes cuando crean expresiones y formas de nuevo cuño, factores representativos de esta genuina modalidad de creación léxica. Para aplicar las reglas de formación y cambio gráfico en la generación se atiende al tipo compositivo —nominal, adjetival, verbal—, a las categorías gramaticales de los

constituyentes y a la proliferación de la regla de formación.

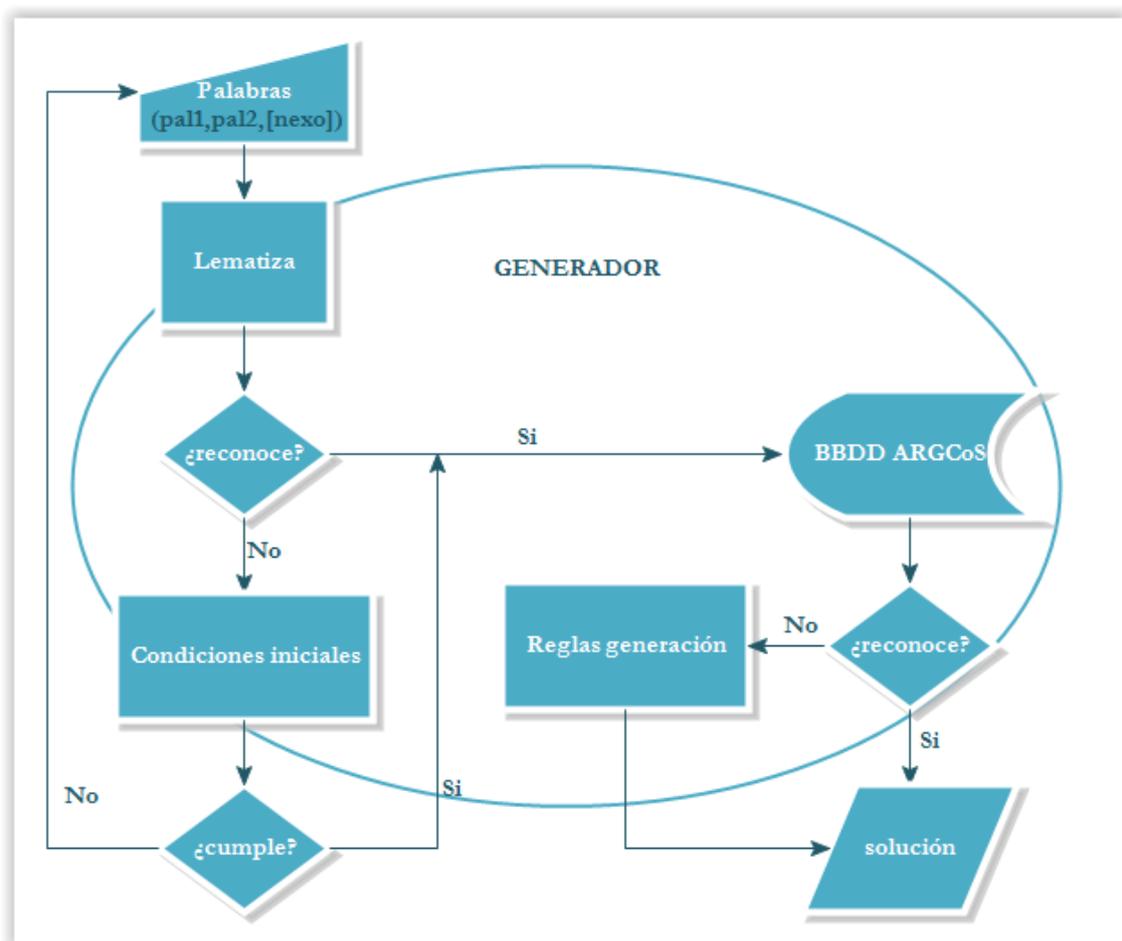


Figura 29.- Diagrama de flujo del proceso de generación de compuestos

Se definen a continuación las distintas fases del generador compositivo⁹³.

Fases en el generador de palabras compuestas

En la Figura 29 se muestra un esquema del funcionamiento del proceso de generación de palabras compuestas:

⁹³ No se hace especial hincapié en el módulo de las reglas de formación para la generación de compuestos pues fueron tratadas extensamente en el Capítulo 3.

- Se comprueba si los elementos componentes están presentes en el lenguaje o se ha producido algún tipo de error ortográfico, en cuyo caso se permite su corrección —se plantea dar opciones posibles a partir de la base de conocimiento BARGCoS o se posibilita introducirla nuevamente—; se debe tener en cuenta que en un procedimiento de generación de compuestos no puede partir de neologismos —un compuesto debe estar formado por elementos presentes en el léxico.

Se parte de los elementos componentes (pal1, pal2) a partir de los cuales se quiere generar la correspondiente forma compuesta, Figura 30. Habría que definir una serie de parámetros para obtener unas ciertas condiciones de entrada: el número de elementos que participan y sus categorías gramaticales, especificar si importa el orden de colocación de los elementos componentes, si existe algún tipo de nexo de unión y si se desea construir un compuesto parasintético. A continuación se lematizan (pal1, pal2) —se identifican sus diversas formas canónicas, categorías gramaticales, flexiones y derivaciones que las producen a través del “*Flexionador y Lematizador*” del GEDLC—, se hace el producto cartesiano de las formas canónicas o literales y se determina con cuales de estos pares de formas (pal1, pal2) se trabaja en el proceso de generación.

Si alguno de los elementos constituyentes es un verbo o forma flexionada verbal: se revisa el tipo compuesto que se genera —nominal o verbal— y el orden de los constituyentes, pues interesa considerar: o bien su forma canónica verbal o bien su flexión de 2ª persona del singular del imperativo o 3ª persona del singular del presente indicativo —únicas formas que se permiten en la composición nominal cuando alguno o los dos constituyentes son verbos.

Una vez definidos los posibles pares de elementos, se determina un orden en función de la productividad de las categorías gramaticales del compuesto y de los constituyentes. En primer lugar se sitúan los que producen un resultado *nominal* por ser los más frecuentes en la formación de compuestos; se colocan al principio las composiciones propias formadas por las categorías gramaticales *Verbo+Complemento*, luego las *Sustantivo+Sustantivo*, y así sucesivamente según el orden de las frecuencias que se aprecia en la Figura 31. En segundo lugar, se plantea que la solución sea un compuesto *adjetivo* y se repite el proceso anterior, para finalizar con los compuestos *verbales*.

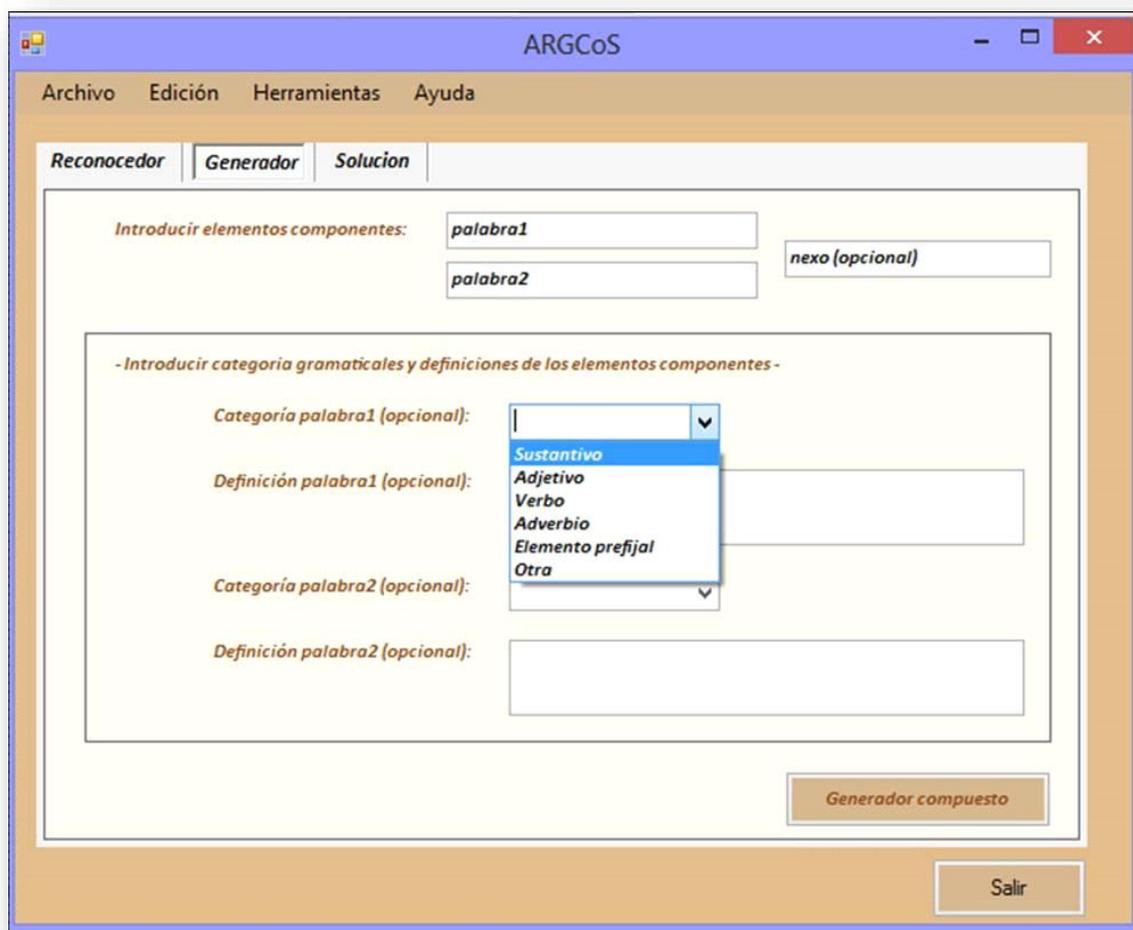


Figura 30.- Interfaz de la aplicación ARGCoS para la generación de compuestos.

Una vez identificados y ordenados los pares de elementos (pal1, pal2) que forman la composición, se busca la solución en la base documental de estudio de compuestos BARGCoS; en caso de no existir, se procede a la aplicación de las reglas de formación y cambio gráfico que sean precisas para su generación —tratadas en el epígrafe *Las reglas de formación de compuestos*. Durante el proceso se pueden crear compuestos homógrafos con diferentes categorías gramaticales, cada uno se sitúa en el orden que le corresponda.

Grafo de dependencias de las reglas de generación

Se define el grafo de dependencias para las reglas de formación y cambio gráfico para la generación de compuestos, donde cada nodo representa una regla. Una relación entre dos

nodos con un sentido establece el paso entre ellas con las condiciones de tránsito que se pudieran especificar en la correspondiente etiqueta, Figura 32.

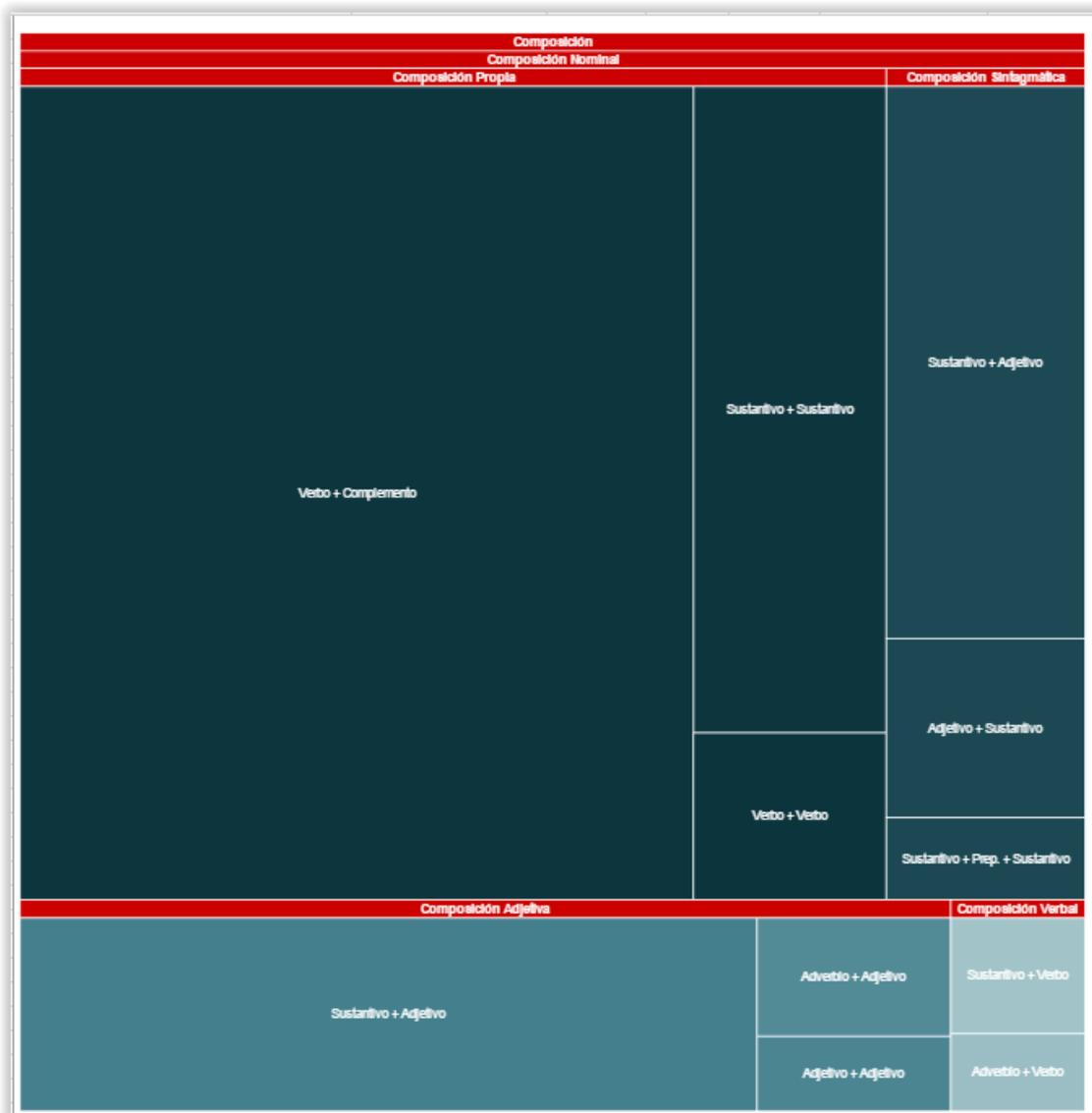


Figura 31.- Clasificación de la composición en español atendiendo a las categorías gramaticales del compuesto y de los elementos componentes.

Se parte de la forma canónica o literal del primer constituyente. Se recorre el grafo y se aplica en cada punto la regla de formación o fonética correspondiente. En caso de ser reconocida la transformación del elemento constituyente entrante, se registra el resultado obtenido tras el recorrido secuencial. Se transita el grafo de dependencias de las reglas hasta que se hayan recorrido todos los caminos posibles. Se itera el proceso para cada uno de los pares (pal1, pal2).

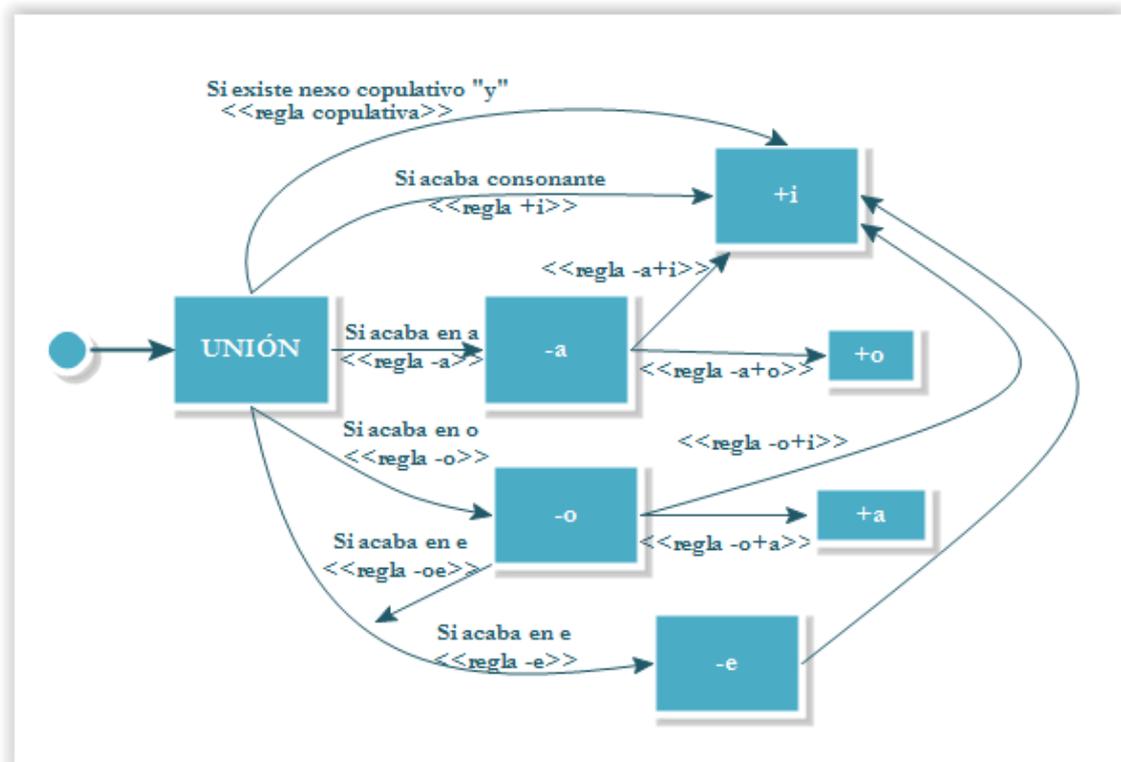
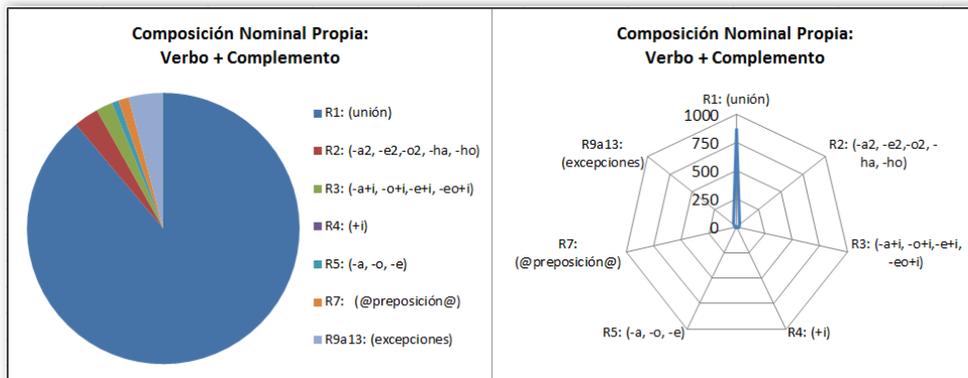
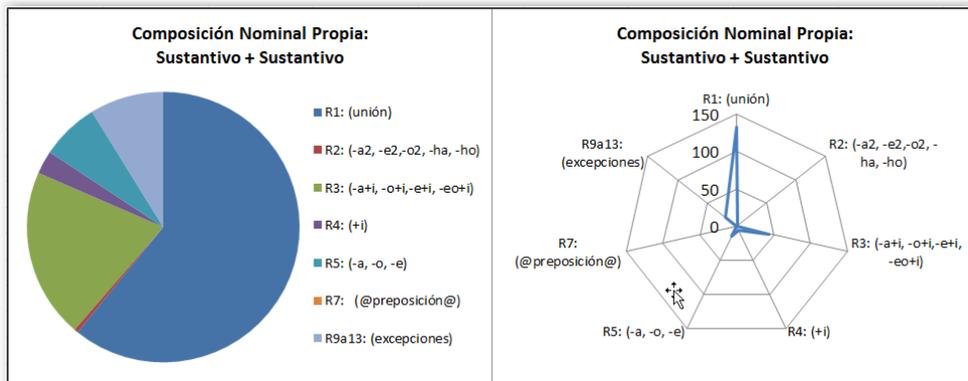
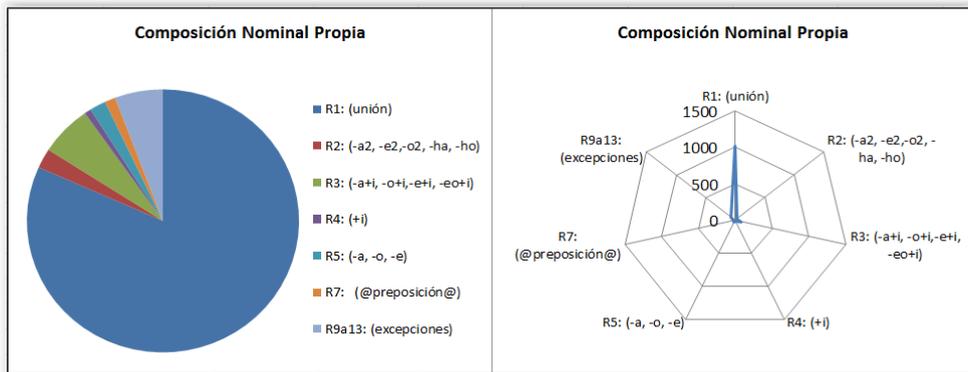
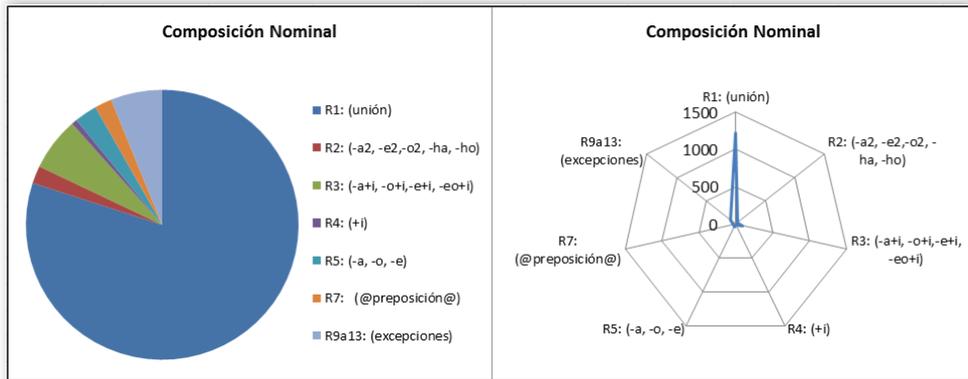


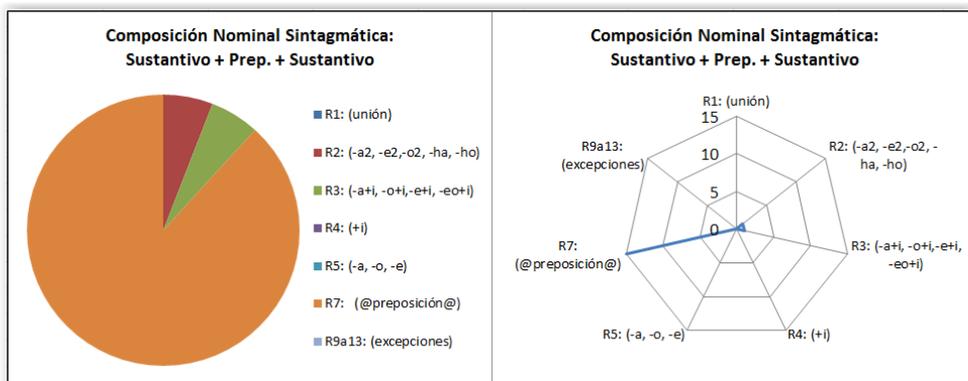
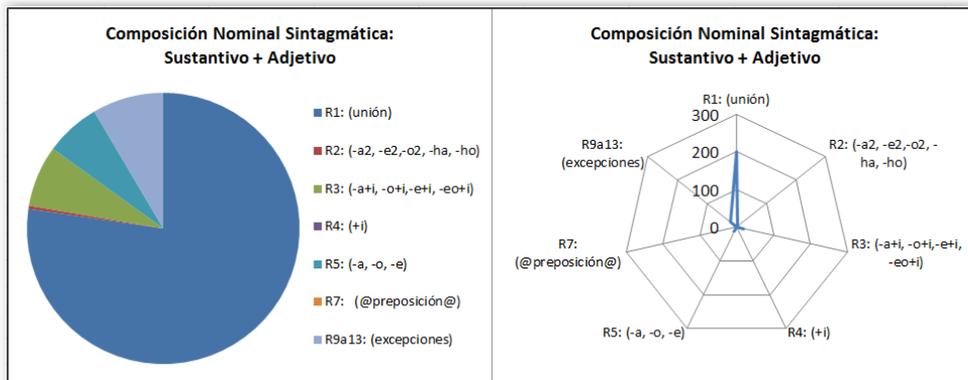
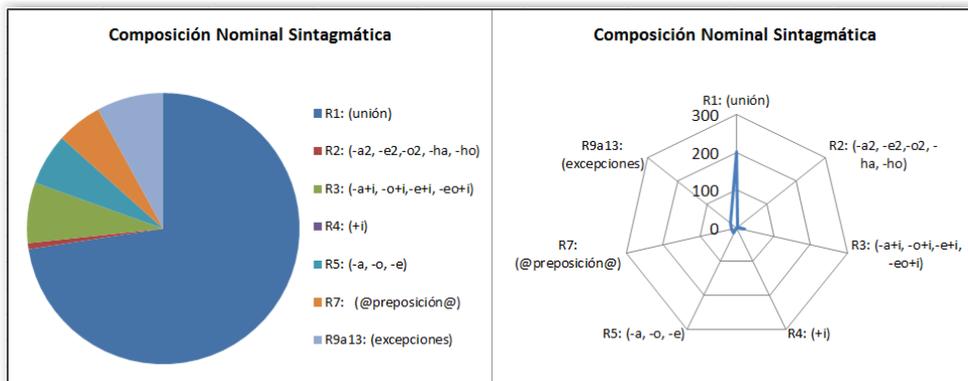
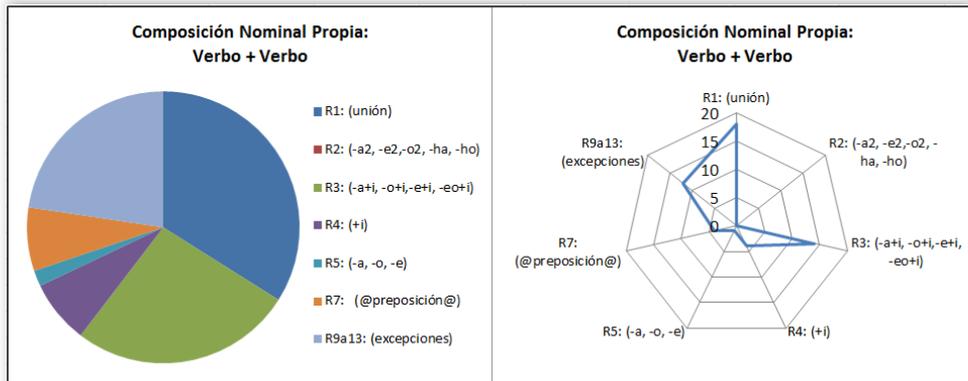
Figura 32.- Extracción del grafo de dependencias de las reglas de formación y cambio gráfico para la generación.

El recorrido del grafo de dependencias define el orden y las reglas que se aplican en cada momento para cada una de las soluciones. De esta forma, es posible añadir nuevas reglas de generación no contempladas originalmente.

Ordenación de la solución según significado

Finalmente, en caso de que se provea de las definiciones de los elementos componentes, se localiza la categoría gramatical de los pares (pal1, pal2) y se toma como principal solución aquella que resulte de aplicar la regla de generación más prolífica en el tipo compositivo correspondiente a las categorías de los constituyentes. A continuación se muestran en gráficos circulares y radicales la producción de las reglas de composición para cada tipo de compuesto —nominal o adjetivo— y para cada tipo de composición —propia o sintagmática— definidas según las categorías gramaticales de los elementos componentes según correspondan —*[Sustantivo + Adjetivo]_{Sustantivo}*, *[Verbo+Verbo]_{Sustantivo}*, *[Sustantivo+ Adjetivo]_{Adjetivo}*, entre otras.





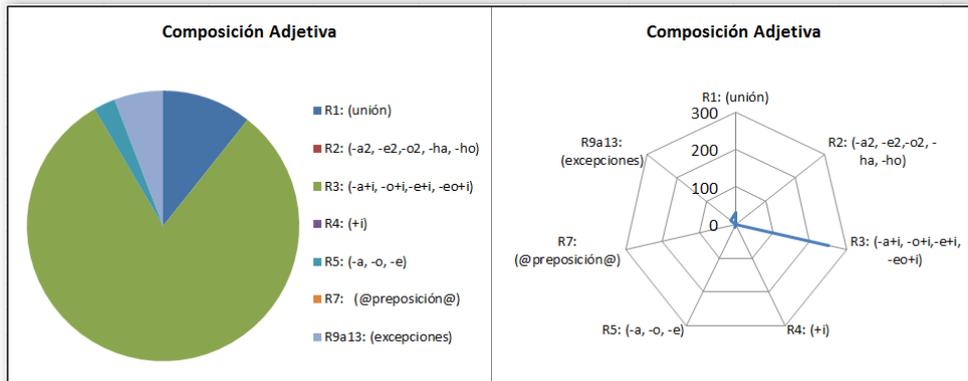
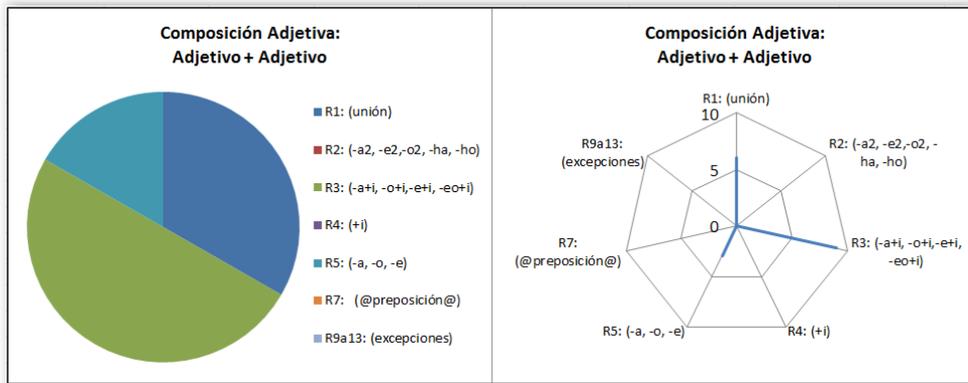
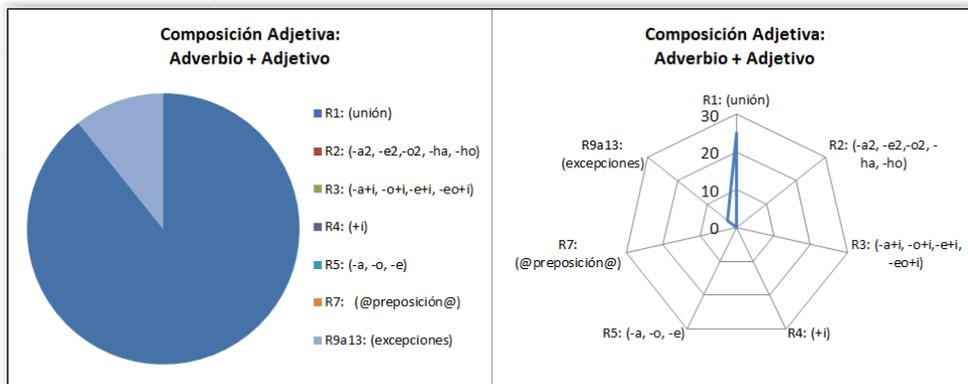
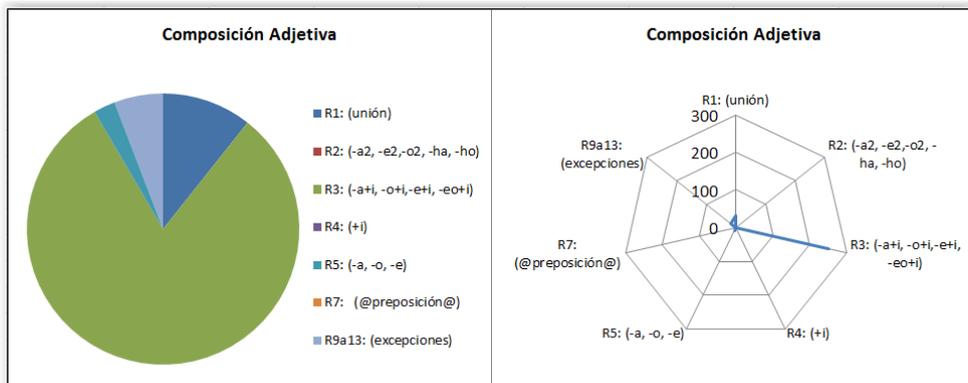


Figura 33.- Distribución de las reglas de formación para compuestos sustantivos según la categoría gramatical de sus elementos constituyentes.



❖ Los pares de candidatos para aplicar las reglas de generación son:

Candidato 1: (cocho, frito): SI se considera
 [Sustantivo + Sustantivo]
 [Sustantivo + Adjetivo]
 [Adjetivo + Sustantivo]
 [Adjetivo + Adjetivo]

Candidato 2: (cocho, freír): NO se considera
 La flexión verbal del segundo componente no se corresponde con las permitidas en composición⁹⁴.

Candidato 3: (cocho, fritar): NO se considera
 La flexión verbal del segundo componente no se corresponde con las permitidas en composición.

Candidato 4: (cochar, frito): NO se considera
 La flexión verbal del primer componente no se corresponde con las permitidas en composición.

Se ordenan las posibles soluciones a partir de las categorías gramaticales de los elementos constituyentes:

Par de componentes: (cocho, frito)
 Posibilidad 1: [Sustantivo + Sustantivo]_{Sustantivo}
 Posibilidad 2: [Sustantivo + Adjetivo]_{Sustantivo}
 Posibilidad 3: [Adjetivo + Sustantivo]_{Sustantivo}
 Posibilidad 4: [Sustantivo + Adjetivo]_{Adjetivo}
 Posibilidad 5: [Adjetivo + Adjetivo]_{Adjetivo}

Si *cocho* funciona como adjetivo —participio irregular del verbo “cocer”— y frito como sustantivo masculino que significa “fritada”, las soluciones serán de la forma:

[Adjetivo + Sustantivo]_{Sustantivo}.

Se recorre el grafo de dependencias de las reglas de generación y se obtiene:

⁹⁴ En todo caso, se considera en composición *verbal* para formar un compuesto [*Sustantivo + Verbo*]_{Verbo}.

Solución 1: <i>cochofrito</i>	[Adjetivo + Sustantivo] _{Sustantivo}	Regla: unión
Solución 2: <i>cochifrito</i>	[Adjetivo + Sustantivo] _{Sustantivo}	Regla: -o+i
Solución 3: <i>cochafrito</i>	[Adjetivo + Sustantivo] _{Sustantivo}	Regla: -o+a

La regla más usual para este tipo de composición es la “*unión*”, seguida de *-o+i*; sin embargo, la solución es:

cocho+ frito → *cochifrito*

Conclusiones del capítulo → Los avances científicos, tecnológicos y la propia dinámica social, hace que los hablantes creen expresiones y formas de nuevo cuño, factores representativos de una genuina modalidad de creación léxica. Disponer de procesos automáticos que permitan crear neologismos compositivos a partir de elementos existentes favorece el dinamismo lingüístico, reflejo de los cambios que tienen lugar en la lengua española.

CONCLUSIONES

En este trabajo se ha profundizado en el estudio de los compuestos del español, con un enfoque orientado al problema del reconocimiento y generación automática de estos vocablos.

1. Internet y el lenguaje periodístico recogen con frecuencia neologismos compositivos debido a la rápida evolución de los acontecimientos y su inmediata transcripción al mundo de las tecnologías de la información. Pese a que la composición es uno de los procesos de formación de palabras de mayor importancia —actualmente en auge— se observan importantes carencias tanto desde el punto de vista del procesamiento automático del reconocimiento y generación de compuestos, como en sus referencias bibliográficas.
2. Se destacan, por muy frecuentes, la composición de *Verbo+Sustantivo*, la de *Sustantivo+Sustantivo* y la de *Sustantivo+Adjetivo*; constituyen los procedimientos más productivos de la composición del español actual, sea por su construcción interna sintáctica, sea por su simplicidad —en la mayoría de los casos es nulo el grado de modificación que sufren los elementos originales—, sea porque se interpretan y se forman con naturalidad, o sea por el amplio uso que se hace en los nuevos medios de comunicación social, en el lenguaje periodístico, en el publicitario, entre otros factores que han actuado de catalizadores en tales procesos. Todas estas circunstancias hacen imprescindibles procesos automáticos que sean capaces de identificar los compuestos, situándolos en un contexto lingüístico adecuado, tanto desde un punto de vista morfológico como sintáctico y semántico.
3. Se ha delimitado las unidades del estudio considerando el tratamiento de los compuestos yuxtapuestos o lexías compuestas en composición nominal y adjetiva, principalmente, al igual que algunos casos especiales de acortamiento, elementos compositivos y parasintéticos por composición.

4. Se analizan los compuestos desde un punto de vista fónico tratándolos como unidades monoactuadas e integrando las dos unidades léxicas, se obvian aquellas que contengan guion o nexos (espacio o preposición). Se ha tenido en cuenta el aspecto morfológico—considerando el número de unidades que formen el compuesto, generalmente dos. Por último se ha considerado su naturaleza teniendo en cuenta el comportamiento de los elementos componentes con respecto a su flexión.
5. Se determinan los rasgos fundamentales que se exigen a las palabras de interés en este estudio, correspondiendo a factores sintácticos —se comprueba que exista un *orden* fijo de constitución de los compuestos específicamente— y semánticos —se tratan como una unidad en el proceso de creación de nuevas palabras de *sentido único*— para poder justificar formalmente su naturaleza de verdadero compuesto en español.
6. La revisión de los trabajos más relevantes sobre la composición en español ha permitido recopilar un corpus, en principio, de 3 000 compuestos que han sido analizados minuciosamente, para la extracción de reglas lingüísticas de formación y de cambio gráfico. A partir de esta información se construye una base de conocimiento de reglas que permiten buscar la existencia de marcas de composición —inclusión del interfijo *i*, entre otras— lo que facilita la identificación automática de dos o más formas simples que funcionan como una unidad léxica compositiva o bien permita interaccionarlas y crear neologismos compositivos según el problema al que se apliquen.
7. En una segunda fase del trabajo de investigación se elaboró un corpus más extenso formado por 11 000 vocablos sobre los que se aplican las reglas establecidas en la fase previa. Aunque la composición se considera un mecanismo poco productivo, el estudio aquí presentado demuestra su clara incidencia en la creación de nuevas palabras en la actualidad, hecho contrastado en el estudio de los neologismos compositivos.
8. Se hace obvio que en el enfoque orientado a la automatización de la composición con medios informáticos, los aspectos formales o teóricos no tienen por qué coincidir con los estrictamente lingüísticos. Así, *clarovente** —falsa composición, lo correcto

sería *clarividente*— no tendría por qué tratarse de una mala formación al no contravenir ninguna regla fonotáctica del lenguaje, ni siquiera la norma de la estructura silábica del español.

9. Se diseña una metodología en la que se tiene en cuenta el tipo de compuesto que se está procesando —sustantivo, adjetivo, principalmente— así como la categoría gramatical de los elementos componentes, orientada a establecer un orden en las soluciones según la regla de formación aplicada.

Aportaciones originales

Para finalizar, esta investigación contribuye a cubrir una parte del vacío tanto informático como bibliográfico existente en el tratamiento de la composición —resultan infructuosas las búsquedas de referencias sobre procesamiento automático de la composición en español, a pesar de la presumible trascendencia de tal proceso.

Por otro lado, consideramos que la presente Tesis realiza un exhaustivo y detallado análisis de la composición en español, al analizar no sólo los compuestos léxicos yuxtapuestos o propios, sino también los tipos compositivos que se han visto relegados a un segundo plano hasta el momento, como son la composición culta o la parasíntesis por composición. Se incluye un banco de datos con los neologismos compositivos recopilados debidamente catalogados.

Reconocer y generar palabras compuestas en español es útil en aplicaciones para el procesamiento automático del lenguaje natural, debido a que lleva implícito vínculos semánticos, sobre todo en los compuestos endocéntricos. Asimismo, potencia las búsquedas en internet al ampliar el abanico de relaciones morfológicas deducidas de los compuestos estudiados, sus derivaciones y flexiones.

Futuras Líneas de Investigación

En definitiva, esta Tesis contribuye a conseguir el reconocimiento y generación automática de la composición en español. Sin embargo, su complejidad semántica⁹⁵ hace que se convierta en un aspecto lingüístico que deja todavía mucho terreno por explorar. En conexión con relaciones de especialización, determinativas o con valores connotativos o de cambio semántico que puede tener un compuesto, resulta un entresijo curioso por resolver *cómo crear compuestos según reglas semánticas que permitan diferenciar automáticamente el endocentrismo y exocentrismo que existen en la composición*. Por otro lado, resultaría de interés comprobar si un compuesto, desde un punto de vista sintáctico, revela un comportamiento peculiar al poderse sustituir alguno de sus elementos componentes por otro sinónimo o por otro término de la misma serie léxica, y estudiar su relación y frecuencia de uso en el lenguaje cotidiano.

⁹⁵ Se deja para posteriores estudios el tratamiento semántico de los compuestos endocéntricos —su automatización resulta artificiosa por su construcción—, frente a los compuestos exocéntricos que escapan actualmente al dominio informático — su significado no se desprende de la suma de sus integrantes, sino que ostenta un significado añadido, imprevisible y de forma generalmente figurada.

BIBLIOGRAFÍA

Bibliografía lingüística

- ALARCOS LLORACH, EMILIO, "Consideraciones sobre el neologismo", *El neologismo necesario*, Fundación EFE, Madrid, págs. 19-29, 1992.
- ALEMANY BOLUFER, JOSÉ, *Tratado de formación de palabras en la lengua castellana. La derivación y la composición. Estudios de los sufijos y prefijos empleados en una y otra*, Victoriano Suárez, Madrid, 1920.
- ALMELA PÉREZ, RAMÓN, *Procedimientos de formación de palabras en español*, Ariel, Barcelona, 1999.
- ALVAR EZQUERRA, MANUEL, "El elemento tele formante de palabras en español", *Estudios de Lingüística Hispánica. Homenaje a Félix Monge*, Gredos, Madrid, págs. 55-64, 1995.
- ALVAR EZQUERRA, MANUEL, *La formación de las palabras en español*, Cuadernos de lengua española, Arco/Libros, Madrid, 2002.
- ÁLVAREZ DE MIRANDA, PEDRO, "Neología y pérdida Léxica", en ELENA DE MIGUEL (eda.), *Panorama de la lexicología*, Ariel, Barcelona, 2009, págs. 133-158.
- ARRIETA DE MEZA, BEATRIZ, MEZA CEPEDA, RAFAEL DANIEL, BATISTA OJEDA, JUDITH, "Procedimientos morfológicos para la creación de neologismos en el discurso académico", *Revista de Investigación Lingüística*, nº 13, págs. 219-240, 2010.
- BOOIJ, GEERT E., "Compounding and derivation: Evidence for Construction Morphology", *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 264, págs. 109-132, 2005.
- BOSQUE, IGNACIO, GUTIÉRREZ-REXACH, JAVIER, *Fundamentos de Sintaxis Formal*, Akal (1ª edición), Madrid, 2009.
- BUENAFUENTES DE LA MATA, CRISTINA, *Procesos de gramaticalización y lexicalización en la formación de compuestos en español*, Tesis Doctoral, Universidad Autónoma de Barcelona, 2007, <<http://dialnet.unirioja.es/servlet/tesis?codigo=5621>> [1/10/2015].

- BUENAFUENTES DE LA MATA, CRISTINA, "Procesos de gramaticalización en el estudio de los compuestos del español: el caso del elemento \"mal/a\"". *Anuari de Filologia*, 23/14/F/11-12, págs. 17-30, 2001.
- BUSTOS GISBERT, EUGENIO, *La composición nominal en español*, Pub. Universidad, Salamanca, 1986.
- CABRÉ CASTELLVI, MARÍA TERESA, "La clasificación de los neologismos: Una tarea compleja". *Alfa*. São Paulo, 50 (2), págs. 229-250, 2006.
- CASADO VELARDE, MANUEL, *La innovación léxica en el español actual*, Síntesis S.A., Madrid, 2015.
- COSERIU, EUGENIO, "Los procedimientos semánticos en la formación de palabras", *Odisea* n°3, págs. 179-189, 2003.
- CHUMACEIRO, IRMA, "Morfología". *Español Actual*, vol. 69, págs.51-66, Madrid, 1998.
- ESTORNELL PONS, MARÍA, *Neologismos en la prensa: criterios para reconocer y caracterizar las unidades neológicas*, Quaderns de Filología de la Universidad de Valencia, Anejos n° 70, Pub. Universidad, Valencia, 2009.
- FELÍU ARQUIOLA, ELENA, "Notas sobre un tipo de compuestos N+N en español actual". *Español Actual*, 75, págs. 27-32, 2001.
- FELÍU ARQUIOLA, ELENA, "Palabras con estructura interna", en ELENA DE MIGUEL (eda.), *Panorama de la lexicología*, Ariel, Barcelona, 2009, págs. 51-82.
- HIDALGO DE LA TORRE, RAFAEL, *Sugerencias sugestivas con las palabra*, Bolsillo Octaedro, 2010.
- GARCÍA LOZANO, FRANCISCO, "Los compuestos de sustantivo + adjetivo de tipo \"pelirrojo\"", *Iberoromania*, 8, págs. 82-89, 1978.
- GARCÍA PLATERO, JUAN MANUEL, "Los compuestos sustantivos en español actual", *Español Actual*, 64, págs. 99-100, 1995.
- GARCÍA PLATERO, JUAN MANUEL, "Procedimientos lexicogenésicos en el discurso periodístico" en J. A. MOLINA REDONDO. y J. DE D. LUQUE DURÁN, *Estudios de lingüística general. Actas del II Congreso Nacional de Lingüística General*, Granada, 1996. Servicio de Publicaciones de la Universidad de Granada, vol. 2, págs. 139-145, 1997.

- GÓMEZ DE ENTERRÍA, JOSEFA, “Últimas tendencias neológicas en la prensa económica” en M^a TERESA CABRÉ, J. FREIXA y E. SOLÉ (eds.), *La neología en el tombant de segle*, Observatorio de Neología IULA Barcelona, págs. 75-84, 2000.
- GUERRERO RAMOS, GLORIA, *Neologismos en el español actual*, Arco/Libros, Madrid, 2010 [1995].
- JIMÉNEZ RÍOS, ENRIQUE, “Tratamiento lexicográfico de los compuestos nominales de verbo+nombre”. *Lingüística Española Actual*, 24/2, págs. 253-276, 2002.
- MARTÍN CAMACHO, JOSÉ CARLOS, “Los procesos neológicos del léxico científico. Un esbozo de clasificación”, *Anuario de Estudios Filológicos*, vol. XXVIII, págs.157-174, 2004.
- MIRANDA, JOSE ALBERTO, *Formación de palabras en español*, Ediciones Colegio de España, 1994.
- LANG, MERVING, *Formación de palabras en español. Morfología derivativa productiva en léxico moderno*, Cátedra, Madrid, 1992.
- LICERAS, JUANA, “La morfología léxica del español y el llamado problema lógico de la adquisición del lenguaje no nativo”, en DAVID SERRANO DOLADER, MARÍA ANTONIA MARTÍN ZORRAQUINO & JOSÉ FRANCISCO VAL ÁLVARO (eds.), *Morfología y español como lengua extranjera (E/ELE)*, Prensas Universitarias de Zaragoza, Zaragoza, 2009, págs. 21-66.
- MARTINET, ANDRÉ, *Gramática funcional del francés*, Barcelona, Ariel, 1984 [1979].
- OBNEO, *Protocolo de vaciado de textos de prensa escrita*, IULA, Universidad Pompeu Fabra, Barcelona, 2003, <<http://obneo.iula.upf.edu/bobneo/index.php>> [01/10/2015].
- ORTEGA, MARÍA DEL PILAR, “Neología y Prensa: Un binomio eficaz”, *Especulo n° 18*, Universidad Complutense de Madrid, (2001) <<https://pendientedemigracion.ucm.es/info/especulo/numero18/neologism.html>> [1/10/2015].
- PÉREZ CINO, WALDO, *Manual Práctico de formación de palabras en español I*, Verbum, Madrid, 2002.
- PÉREZ LAGOS, MANUEL FERNANDO, “Composición de verbo más nombre en el "DRAE"”, *Lingüística Española Actual*, VIII-1, págs. 21-58, 1986.
- PÉREZ LAGOS, MANUEL FERNANDO, "¿Entre composición y afijación? Naturaleza de los elementos de formación culta" En MOLINA REDONDO, J. A. Y LUQUE DURÁN J DE D.

- (eds.), *Estudios de lingüística general. Actas del II Congreso Nacional de Lingüística General, Granada, 1996*. Servicio de Publicaciones de la Universidad de Granada vol. 3, págs. 361-369, 1997.
- PÉREZ LAGOS, MANUEL FERNANDO, “Euro-: un nuevo elemento compositivo”, *Español Actual*, 65, págs. 74-75, 1996.
- PÉREZ LAGOS, MANUEL FERNANDO, “Nuevos elementos compositivos en el “Diccionario de la Real Academia””, *Analecta Malacitana*, 24/2, págs. 439-450, 2001.
- PÉREZ VIGARAY, JUAN MANUEL, “Locuciones y compuestos nominales. Aportaciones de Julio Casares al estudio de la formación de palabras”, *Philologia Canariensis*, 2-3, págs. 295-310, 1997.
- PÉREZ VIGARAY, JUAN MANUEL, *La composición nominal en español*, Tesis Doctoral, Universidad Las Palmas de Gran Canaria, 1994, <<http://hdl.handle.net/10553/2242>>.
- RAINER, FRANZ, VALERA ORTEGA, SOLEDAD “Compounding in Spanish”, *Rivista di Linguistica*, 4/1, págs. 117-142, 1992.
- ROMERO GUALDA, M^a VICTORIA, *Hacia una tipología del neologismo literario*, Anales de la Universidad de Murcia, Murcia, 1978.
- ROSSOWOVÁ, LUCIE, *Las palabras compuestas en español*, Baka lárská diplomová práce, Masarykova Univerzita, Brno, 2007.
- RUIZ GURILLO, LEONOR. “Compuestos, colocaciones, locuciones: intento de delimitación”. En VEIGA A., GONZÁLEZ PERERIRA M., & SOUTO GÓMEZ M. (eds.), *Léxico y Gramática*. Tris Tram, Lugo, 2002, págs. 327-339.
- SANTANA SUÁREZ, OCTAVIO, GUTIÉRREZ RODRÍGUEZ, VIRGINIA Y PÉREZ AGUIAR, JOSÉ, “Pragmatización en la automatización del reconocimiento de palabras compuestas en español”. *LEA Lingüística Española Aplicada*, XXXV/2, págs. 181-201, 2013.
- SANTANA SUÁREZ, OCTAVIO, GUTIÉRREZ RODRÍGUEZ, VIRGINIA, PÉREZ AGUIAR, JOSÉ Y SÁNCHEZ BERRIEL, ISABEL, “Reglas de formación de palabras compuestas en español para la automatización de su reconocimiento”, *Procesamiento del Lenguaje Natural*, vol. 51, págs. 75-82, Jaén, 2012.
- SANTANA SUÁREZ, OCTAVIO, CARRERAS RIUDAVETS, FRANCISCO, PÉREZ AGUIAR, JOSÉ Y GUTIÉRREZ RODRÍGUEZ, VIRGINIA, “El Reconocimiento Automático de la Composición

- en Español”, *Conference Abstracts of the First International Conference of the Alliance of Digital Humanities Organizations*, Paris, 2006.
- SANTANA SUÁREZ, OCTAVIO, CARRERAS RIUDAVETS, FRANCISCO, PÉREZ AGUIAR, JOSÉ, RODRÍGUEZ DEL PINO, JUAN. “Parasyntetic Morpholexical Relationships of the Spanish: Lexical Search beyond the Lexicographical Regularity”, *Proceedings of the LADIS International Conference. Applied Computing*, 2006, págs. 627-631.
- SANTANA SUÁREZ, OCTAVIO, PÉREZ AGUIAR, JOSÉ, CARRERAS RIUDAVETS, FRANCISCO, DUQUE, JUAN, HERNÁNDEZ FIGUEROA, ZENÓN, RODRÍGUEZ RODRÍGUEZ, GUSTAVO, “FLANOM: Flexionador y lematizador automático de formas nominales”, *Lingüística Española Actual XXI*, 2, Arco/Libros, S.L., 1999, págs.253-297.
- SANTANA SUÁREZ, OCTAVIO, PÉREZ AGUIAR, JOSÉ, CARRERAS RIUDAVETS, FRANCISCO, RODRÍGUEZ RODRÍGUEZ, GUSTAVO, “Suffixal and Prefixal Morpholexical Relationships of the Spanish”. *Lecture Notes in Artificial Intelligence*, 3230, Springer-Verlag, 2004, págs. 407-418.
- SANTANA SUÁREZ, OCTAVIO, PÉREZ AGUIAR, JOSÉ, HERNÁNDEZ FIGUEROA, ZENÓN, CARRERAS RIUDAVETS, FRANCISCO, RODRÍGUEZ RODRÍGUEZ, GUSTAVO, “FLAVER: Flexionador y lematizador automático de formas verbales”. *Lingüística Española Actual XIX*, 2, Arco/Libros S.L., 1997.
- SERRANO DOLADER, DAVID, *Las formaciones parasintéticas en español*, Arco/Libros, Madrid, 1995.
- SERRANO DOLADER, DAVID, “Sobre los compuestos (para)sintéticos ¿en español?”, *Los límites de la morfología: estudios ofrecidos a Soledad Varela Ortega*, Universidad Autónoma de Madrid. págs. 427-442, 2012.
- VAL ÁLVARO, JESÚS, “La composición”. En IGNACIO BOSQUE Y VIOLETA DEMONTE (eds.), *Gramática descriptiva de la Lengua Española*, Espasa Calpe vol. 3, Madrid, págs. 4757-4841, 1999.
- VARELA ORTEGA, SOLEDAD, *Fundamentos de Morfología*, Síntesis, Madrid, 1990.
- VARELA ORTEGA, SOLEDAD, *Morfología Léxica: La formación de palabras*, Gredos, Madrid, 2005.

Obras utilizadas para la elaboración del corpus. Diccionarios

ALCALÁ VENCESLADA, ANTONIO, *El Vocabulario Andaluz*, Madrid, 1951.

ALVAR, MANUEL, *Palabras y cosas en la Aezcoa*, Estación de Estudios Pirenaicos del Consejo Superior de Investigaciones Científicas, Zaragoza, 1947.

ALVAR EZQUERRA, MANUEL, *Nuevo Diccionario de voces de uso actual*, Arco/Libros, Madrid, 2003.

COROMINES, JOAN, *Breve diccionario etimológico de la lengua castellana*, Editorial Gredos, Madrid, 2010, [1961].

COROMINES, JOAN, PASCUAL JOSÉ A., *Diccionario crítico etimológico castellano e hispánico*, Editorial Gredos, Madrid, 1991, [1980].

Diccionario de neologismos de la lengua española, Larousse, S.A., Barcelona, 1998.

Diccionario de Uso del Español Actual, Clave S.M., Madrid, 1997.

Diccionario de Uso del Español de América y España, VOX., Spes Editorial S. L. Barcelona, 2003.

Diccionario General de la Lengua Española VOX, Barcelona, 2003.

Gran Diccionario de la Lengua Española, Larousse Planeta, S.A., Barcelona, 1996.

MOLINER, MARÍA, *Diccionario de Uso del Español*, Gredos, Madrid, 1996.

REAL ACADEMIA ESPAÑOLA, *Diccionario de la Lengua Española*, Espasa-Calpe, Madrid, edición electrónica. 22ª, 2001.

REAL ACADEMIA ESPAÑOLA Y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA, *Diccionario de la Lengua Española*, Espasa, Madrid, 23ª edición, 2014.

Sitios web

BANCO DE NEOLOGISMOS DEL CENTRO VIRTUAL CERVANTES
<http://cvc.cervantes.es/obref/banco_neologismos/> [Octubre/2015].

CENTRO VIRTUAL CERVANTES <<http://cvc.cervantes.es/>> [Octubre/2015].

INSTITUTO UNIVERSITARIO DE LINGÜÍSTICA APLICADA, Universidad Pompeu Fabrá
<<http://www.iula.upf.es>> [Octubre/2015].

OBNEO Observatorio de Neologismos, Instituto Universitario de Lingüística Aplicada,
<<http://obneo.iula.upf.edu/bobneo/index.php>> [1/10/2015].

FLEXIONADOR Y LEMATIZADOR DE PALABRAS DEL ESPAÑOL DEL ESPAÑOL, Grupo de Estructuras de Datos y Lingüística Computacional, Universidad de Las Palmas de Gran Canaria, <<http://gedlc.ulpgc.es/investigaciobn/scogeme02/lematiza.htm>> [Octubre/2015].

RELACIONES MORFOLÓGICAS DEL ESPAÑOL, Grupo de Estructuras de Datos y Lingüística Computacional, Universidad de Las Palmas de Gran Canaria <<http://gedlc.ulpgc.es/investigaciobn/scogeme02/relnorfo.htm>> [Octubre/2015].

APENDICES

APÉNDICE A: DIAGRAMAS DE FLUJO

Diagrama de flujo ampliado para proceso de reconocimiento de palabras compuestas

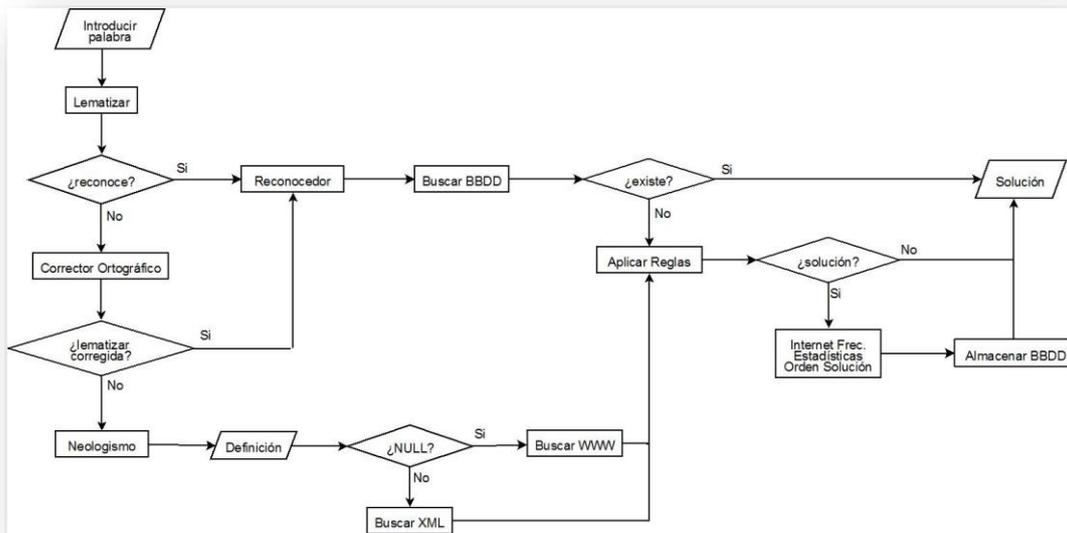
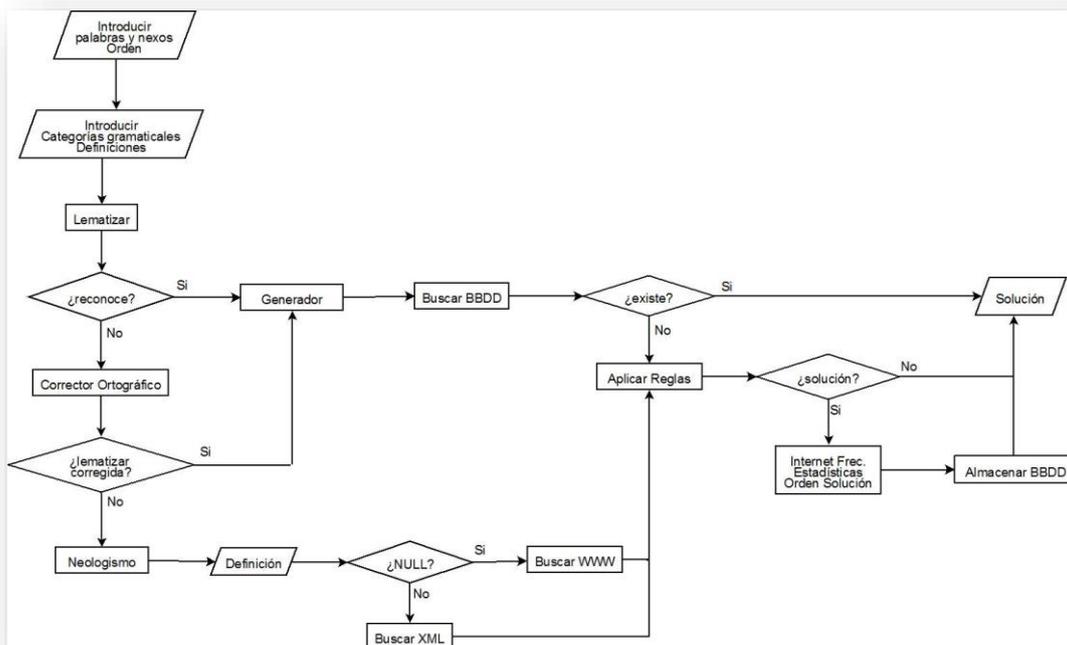
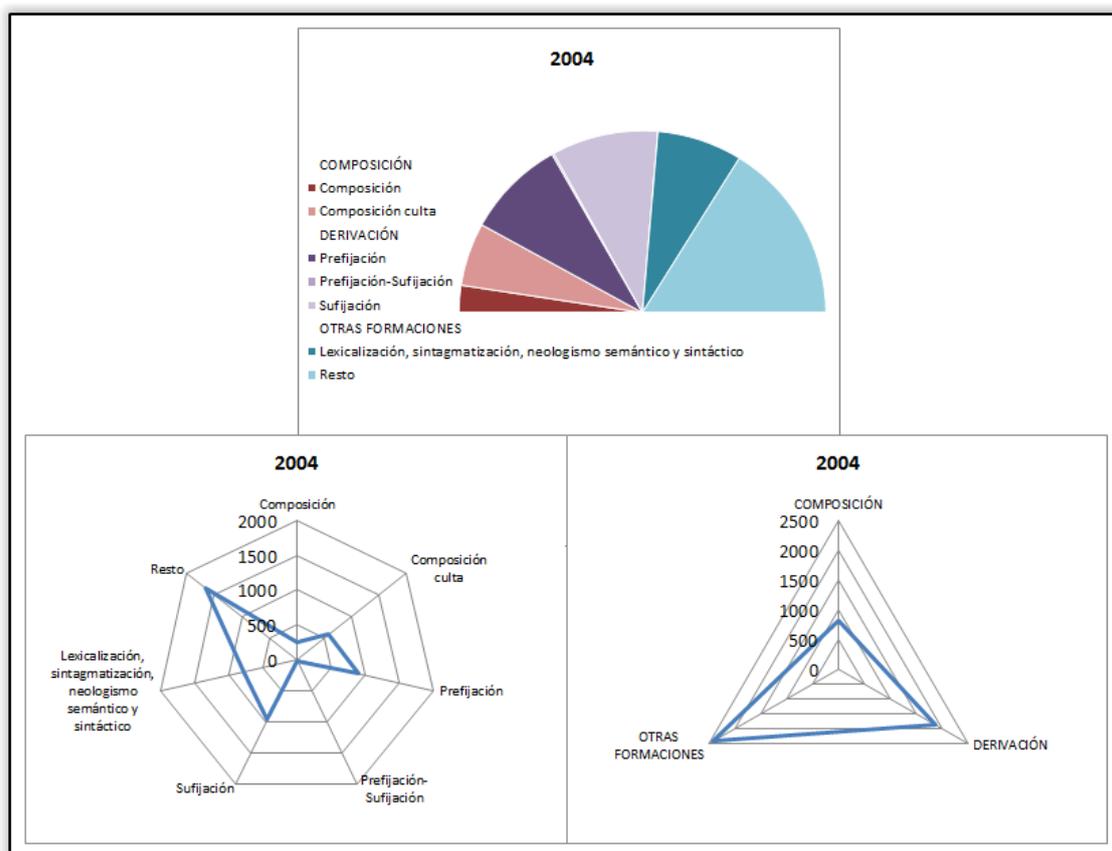


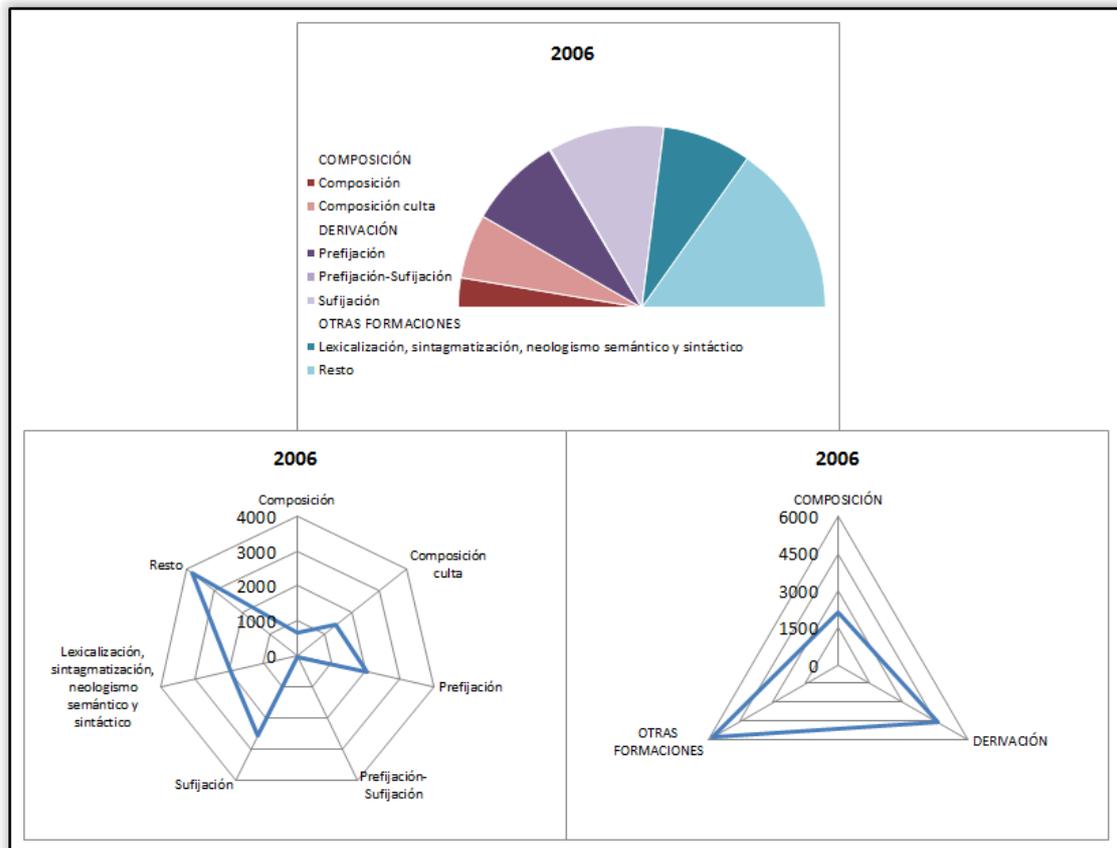
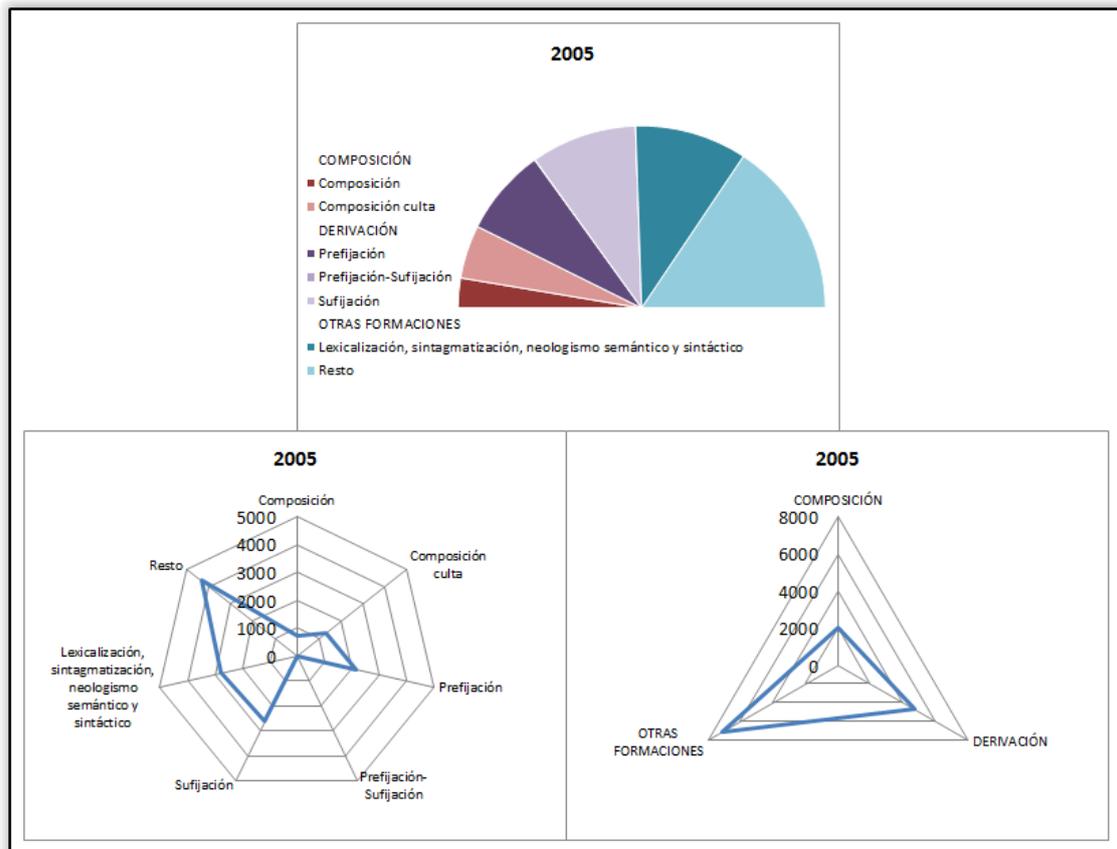
Diagrama de flujo ampliado para el proceso de generación de palabras compuestas

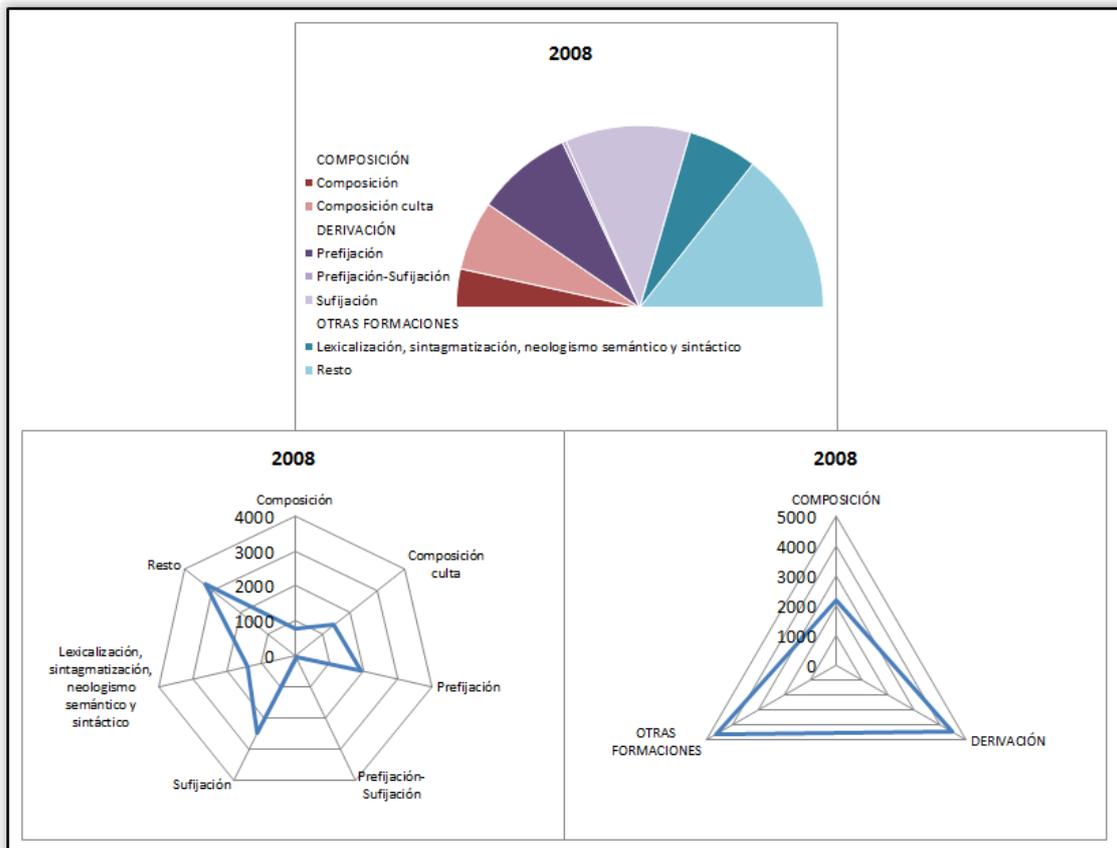
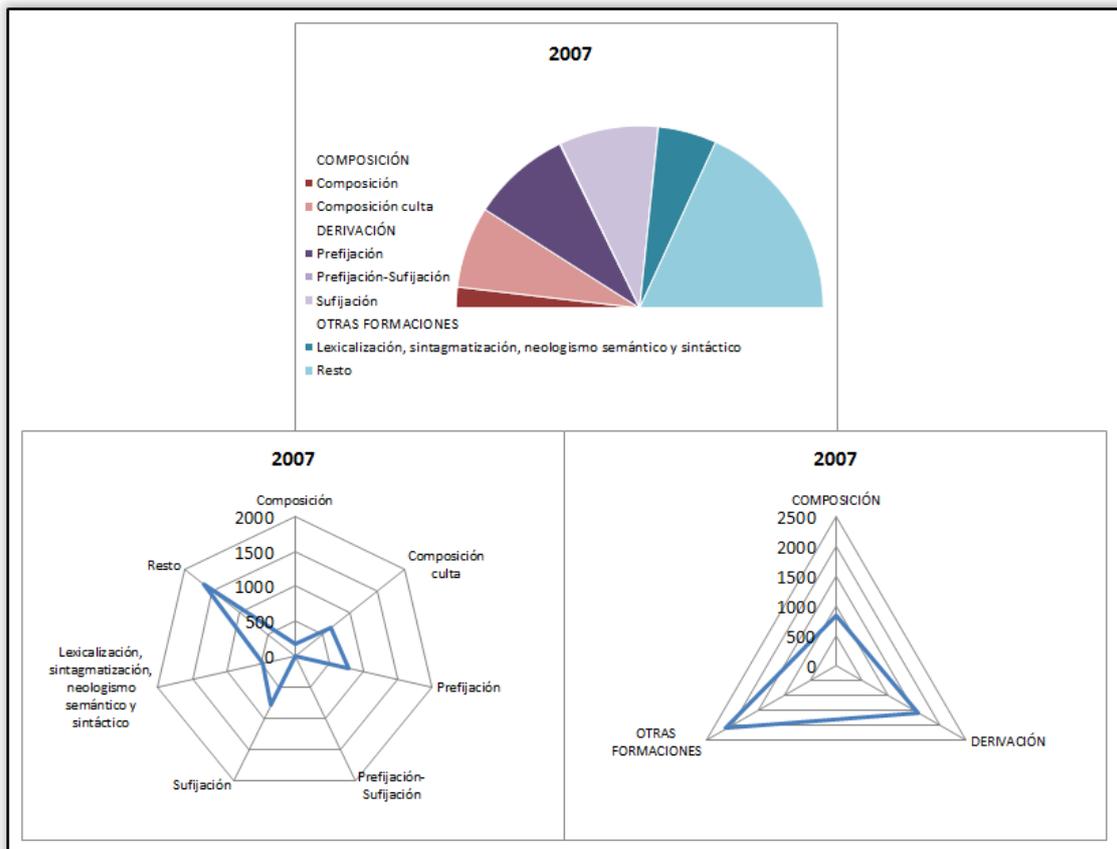


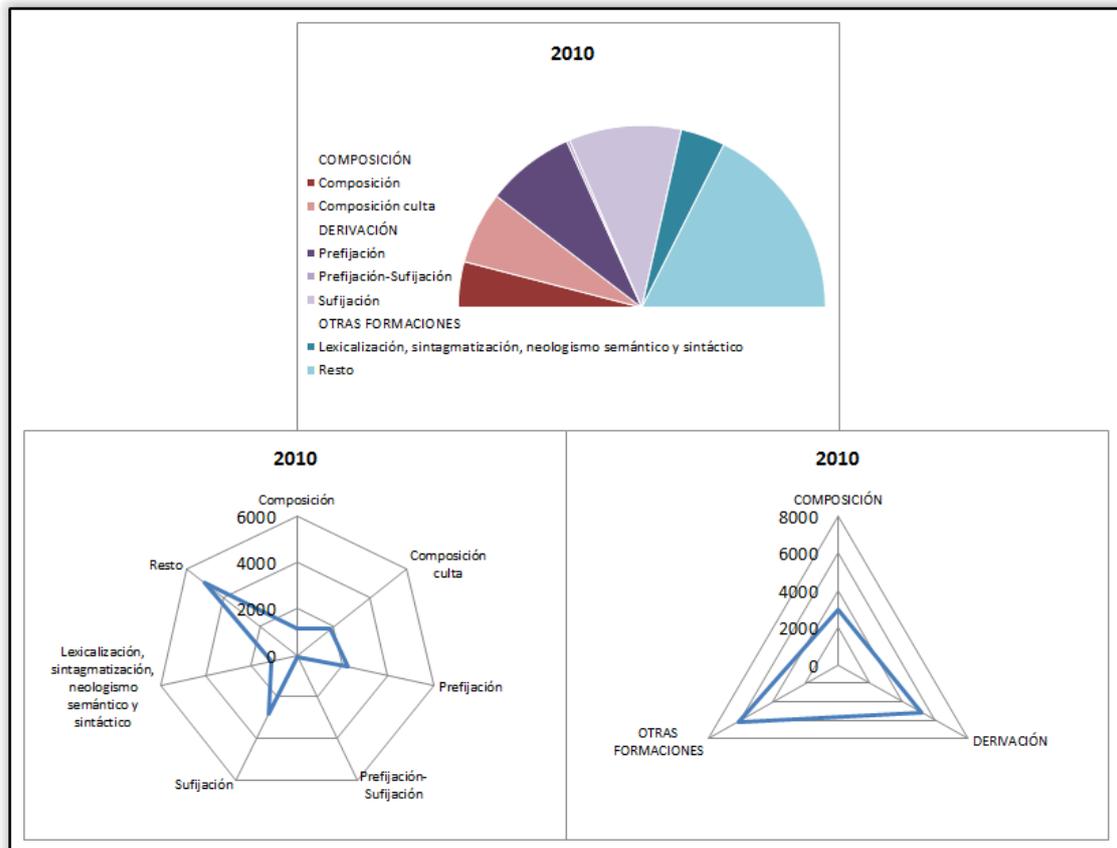
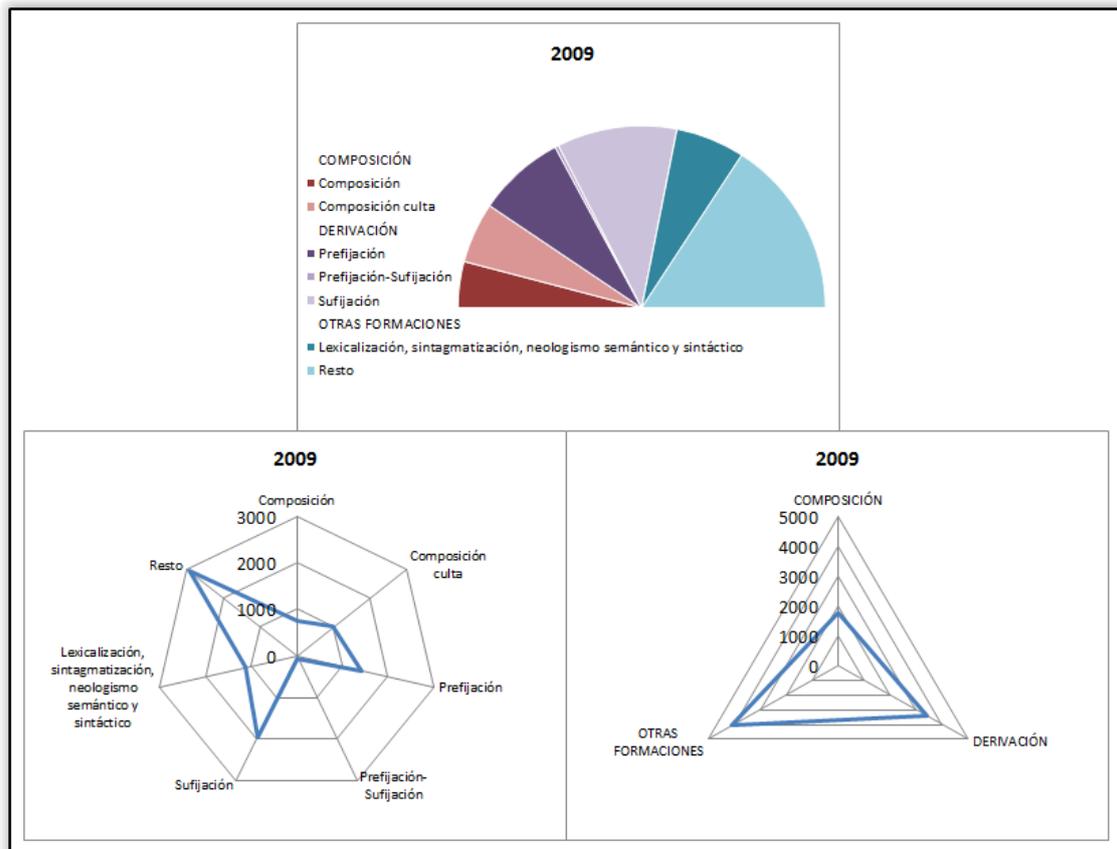
APÉNDICE B: GRÁFICOS PARA LOS PROCESOS DE FORMACIÓN FORMALS DE NEOLOGISMOS.

Gráficos circulares y radiales en el periodo 2004-2010.









LISTA DE ACRÓNIMOS Y ABREVIATURAS

ARGCoS	Automatic Recognition and Generation of Compounds in Spanish (Software)
BBDD	Bases de Datos
BARGCoS	Base de Datos <i>Reconocimiento y Generación Automática de Compuestos en eSpañol</i>
BOBNEO	Base de Datos del Observatorio de Neología
CLAVE	Diccionario de Uso del Español Actual
CVC	Centro Virtual Cervantes
DRAE, RAE	Diccionario de la Real Academia Española
GDLE	Gran Diccionario de la Lengua Española (Larousse)
IULA	Instituto Universitario de Lingüística Aplicada
MOLINER	Diccionario de Uso del Español María Moliner
OBNEO	Observatorio de Neología
ULPGC	Universidad de Las Palmas de Gran Canaria
UPF	Universidad Pompeu Fabra
VOX	Diccionario General de la Lengua Española
VOX (2003)	Diccionario de Uso del Español de América y España

GLOSARIO DE TÉRMINOS

Abreviatura (VOX 2003): Procedimiento que se utiliza para representar palabras escribiendo sólo una o varias de sus letras.

Acortamiento (VOX 2003): Efecto de acortar o acortarse (disminuir la longitud, la duración o la cantidad de algo).

Acronimia (VOX 2003): Procedimiento para la formación de palabras mediante la unión de letras o sílabas del principio y el fin de dos o más palabras que forman una expresión.

Colocaciones (Wikipedia): (Del lat. *collocatio*, *-ōnis*). Designa combinaciones frecuentes de unidades léxicas fijadas en la norma o una combinación de palabras que se distinguen por su alta frecuencia de uso —*cierre hermético, maraña inextricable, desear ardientemente, negar categóricamente*.

Composición (RAE 2001): Procedimiento por el cual se forman palabras juntando dos vocablos con variación morfológica o sin ella —*cejiunto, lavavajillas*. Se aplica también a las voces formadas con vocablos de otras lenguas, especialmente del latín y el griego —*neuralgia, videoconferencia*.

Contraposición (VOX 2003): Acción de contraponer o contraponerse (poner junta dos cosas, una al lado de la otra o frente a frente, para encontrar parecidos y apreciar diferencias entre ellas).

Derivación (RAE 2001): Procedimiento por el cual se forman vocablos ampliando o alterando la estructura o significación de otros que se llaman primitivos; v. gr.: cuchillada, de cuchillo; marina, de mar.

Disyunción (Rae 2001): Acción y efecto de separar y desunir.

Gramática (Rae 2001): Ciencia que estudia los elementos de una lengua y sus combinaciones

Homógrafo (Vox 2003): (lingüística) Que se escribe y se pronuncia exactamente igual que otra pero tiene distinto significado y distinta etimología.

Locución (Rae 2001): Grupo de palabras que forman sentido, frase. Combinación estable de dos o más palabras, que funciona como oración o como elemento oracional, y cuyo sentido unitario no siempre se justifica, como suma del significado normal de los componentes.

Metáfora (Vox 2003): RETÓRICA. Figura que consiste en trasladar el sentido recto de las voces en otro figurado, en virtud de una comparación tácita: *la primavera de la vida*; metáfora continuada, alegoría en que unas palabras se toman en sentido recto y otras en sentido figurado.

Metonimia (Vox 2003): RETÓRICA. Tropeo que consiste en designar una cosa con el nombre de otra tomando el efecto por la causa o viceversa, el signo por la cosa significada, etc.: las canas por la vejez.

Neologismos (Vox 2003): Vocablo, acepción o giro nuevo de la lengua; provienen de neo- (palabra) y del gr. logismós (razonamiento).

Parasíntesis (Rae 2001): Formación de vocablos en que intervienen la composición y la derivación; como encañonar.

Sinapsia (Wikipedia): En lingüística se refiere a la unión de dos lexemas escritos en dos palabras distintas anexados por medio de una preposición.

Yuxtaposición (Vox 2003): Acción de yuxtaponer o yuxtaponerse (poner [una cosa] junto a otra). GRAMÁTICA. Sucesión de oraciones sin palabras que expresen el enlace.

LISTA DE VOCES CITADAS

A

- abeja reina*, 6
abrazafarolas, 7
abrelatas, 54
abreojos, 55
abriboca, 57
abrojo, 59, 80
abulo-, 60
abulomanía, 60
aclaraaguas, 55
acuametría, 8
acuti-, 60
acutifolio, 60
adipogénesis, 8
aerobio, 8
afilalápices, 55
agorafobia, 8
agridulce, 6, 16, 20, 54, 58, 59, 85
agua de nieve, 27
agua nieve, 20
agua sal, 27
aguafresquera, 55
aguagriero, 63
aguamanil, 25, 66
aguamano, 25
aguanafa, 55
aguanieve, 20
- aguapié*, 55
aguardiente, 15, 27
aguarrás, 67, 69
aguasal, 55
aguariva, 7, 55
abimelollea, 69
abimelollevas, 11, 61
ajaceite, 58
ajaspajas, 63
ajiaceite, 57
ajilimójili, 62
ajipuerro, 57
ajoaceite, 55
ajolio, 56
ajoqueso, 54
alapivot, 7
albasol, 7
*albonegro**, 82
alborotapueblos, 7
alicaído, 16, 57
aliquebrar, 17
alirrojo, 57, 67
almaempena, 61, 67
almempena, 10
alta mar, 20
altibajo, 57
aluminografía, 8
alzarraños, 67
anchicorto, 57
- andarríos*, 67
andrófobo, 8
anorexígeno, 8
anquimuleño, 24, 66
anquirredondo, 54
anteayer, 17
antibacteriano, 24
*añista**, 25
apagafuegos, 7
archiconocido, 8
ardivejas, 57
aris-, 61
arisblanco, 61
arrancarrabia, 67
artimaña, 15
atalejo, 15
aunque, 18
auto-, 60
autobús, 21, 66
automóvil, 60
ave fría, 20
avefría, 20
avetonto, 55
azulcrema, 7
azuliblanco, 11, 58, 85

B

- babasfrías*, 55
balompié, 54, 67, 69

- balsopeto*, 64
barbicastaño, 57
barbirrostro, 67
bases de datos, 6
batiaguas, 57
baticola, 57
batiporte, 63
bativoleo, 57
bellaluisa, 55
besalamano, 11, 61
bienandanza, 15
bienafortunado, 54
bienhablado, 16
bienintencionado, 16, 54
bienmesabe, 11, 15, 16, 61
bienpensante, 15
bienpensar, 17
bienquerencia, 15
bienteveo, 11, 61
bienvenir, 9
bocabajo, 56
bocacalle, 55
bocacalles, 15
*bocabundido**, 77
bocajacha, 68
bogavante, 15, 56
boquicerrado, 6, 20
boquihundido, 76, 77
boquirroto, 67
botaguas, 56
botesela, 62
botifuera, 57
botivoleo, 57
brazalbo, 59
bucbeamarrillo, 55
burriciego, 57
buscarruidos, 67

C

cabeciancho, 67
cabeciduro, 67
cabezcaído, 58, 59
cabizcaído, 69
cabizmordido, 69
cagalaolla, 5, 11
cagarrache, 62
caja negra, 6
calicanto, 58, 85
*cantano**, 25
cantautor, 21
capaniu, 62
capipardo, 57
capisayo, 57
-cardia, 60
cardimuelle, 65
cardioprotector, 8
cardocuco, 16
cariacontencido, 55
cariancho, 57
caridura, 57
carininfo, 57
cariparejo, 16
carirraido, 67
carnicol, 20, 57
carricoche, 15, 57
cartapel, 65
casatienda, 15
cascalbo, 65
cascorvo, 65
catetómetro, 69
catricofre, 57
cefalópodo, 9
cefalotórax, 9
cejijunto, 16
cejunto, 65
cenaoscuroas, 11, 54, 61
ceromiel, 63
cerrapan, 63

Ch

chambeculos, 59
chiticalla, 57
chupachupa, 55

C

cibercampana, 8
-cidio, 60
ciempiés, 15
cientoemboca, 15
cinematógrafo, 21, 66
clarividencia, 57
clarividente, 57, 119
claroscuro, 54, 55
clarovidente, 100
*clarovidente**, 100, 118
cochafrito, 112
coche bomba, 21
coche-bomba, 21

cochifrito, 57, 85, 94, 95,

112

cochitril, 20, 54, 65*cochofrito*, 101, 112*cojitranco*, 57*coliblanca*, 57*coliflor*, 54, 58*colinabo*, 58*colirrojo*, 67*comicalla*, 57*compraventa*, 15*contrachapado*, 23*contradanza*, 23*contradecir*, 23*contrarreformista*, 24*cornidelantero*, 8*correiverás*, 58*correquetecagas*, 10, 61*correvidile*, 10, 16, 58, 68*corriverás*, 57*cortiancho*, 57*coyotomate*, 5, 21, 65*creique*, 15*cualesquiera*, 64*cualquiera*, 18*cuatrojos*, 56*cubrecamas*, 15*cuellilargo*, 16*cuerda sin fin*, 20*cuerdacarru*, 62*culinegro*, 57*culocagado*, 55*culodemalasiento*, 10*cultipicaño*, 16**D***dedodedama*, 55*dedodedama*, 11*desmochollas*, 59*diostedé*, 11*disanto*, 59*doblescudo*, 56*doceañista*, 24, 25*doceaño*, 25*dolicocéfalo*, 9*dondequiera*, 18*duermevela*, 15*dulciagrío*, 57*duraluminio*, 59**E***eco-*, 60*ecosistema*, 8*ecosonda*, 60*enticonejuno*, 8*entreacto*, 23*entremedio*, 23*entresacar*, 23*esperalaultima*, 11, 61*espiritusanto*, 69*estrellamar*, 55*eurasiático*, 58*euro-*, 61*euroasiático*, 8*eurocheque*, 61*europa*, 22*eurovisión*, 21**F***falda pantalón*, 21*falda-pantalón*, 21*faldicorto*, 16*fatídicamente*, 23*-fero*, 60*fibromialgia*, 8*fijodalgo*, 62*filántropo*, 8*fili-*, 60*filicidio*, 60*-filo*, 60*físico-químico*, 22*flor de la abeja*, 20*-fobia*, 60*-fobia*, 60*-fono*, 60*fotografía*, 21, 66*franjirojo*, 7*frunciboca*, 57**G***gali-*, 61*galicursi*, 57*galiparla*, 61*ganapán*, 69*ganapierde*, 15*gargantiazul*, 55*gatuña*, 58*gentilshombres*, 63*gentilhombre*, 55, 63

golpe de estado, 22

grandilocuente, 16

grecorromano, 67

guadarnés, 62

guardaespaldas, 55

guardaguas, 56

guardalagua, 11, 62

guerra civil, 22

H

hazmerreír, 16

heroificar, 17

hidalgo, 14, 62

hierbabuena, 15

hijo de algo, 14, 62

hijodalgo, 11, 62

hincapié, 15, 20

hispanohablante, 16

hoja de lata, 14

hojalata, 7, 14, 55

huelehuele, 55

hullebulle, 54

*burgamanda**, 24

burgamandera, 24

I

iberoamericano, 55

indogermánico, 8

informática, 21

intrínsecamente, 23

itañol, 21

J

juanencueros, 11, 61

juntarrimas, 7

L

labibendido, 59, 85

lavatiquevoý, 11, 57, 61

lavavajillas, 27

lengüilargo, 16

libro de cocina, 6

liquidámbar, 54, 58, 85

logopeda, 8

luci-, 60

lucífilo, 60

M

maestrescuela, 7, 56

malacara, 7

malaconsejado, 16

malalma, 55

malasangre, 15

malaslenguas, 63

malastripas, 63

malavenido, 54

malcasar, 17

malcomer, 17

maldecir, 17

maldejojo, 10, 11, 55

maleducado, 15

malgastar, 7

malpaís, 15

malqueda, 20, 55

malsano, 6, 20

malsufrido, 54

mampuesto, 67

mancuadra, 58

mandamás, 15

*mandera**, 24

mandoble, 59

maniatar, 17

manivacio, 57

manobra, 56

manuscribir, 17

manvacío, 58, 59

manzorda, 59

mariposa, 54, 63

mariposas, 63

mariposita, 25

matabombres, 55

matalbahaca, 56

matambre, 56

mataorejas, 55

matarratas, 67

matemáticas, 22, 66

mediagua, 56

medialanza, 55

mediodía, 15, 54

menoscuanta, 15

menospreciar, 17

mesa camilla, 5, 6

mesocéfalo, 9

metomentodo, 11, 16

milenrama, 15

*misacanta**, 25

misacantano, 25, 66

montabanco, 11, 62, 67

-móvil, 60
muleño*, 24

N

necrodactilar, 8
nitro-, 60
nitroglicerina, 60
nosotros, 18

O

ojítruco, 16
otrosí, 17

P

pájaro mosca, 20
pan de azúcar, 20
pandesapo, 11, 55
paniqueso, 58
pañomanos, 55
papalba, 58
paraguas, 5, 56
paraguazo, 25
pararrayos, 7, 67
parlaembalde, 61
paso a nivel, 20
paticoyo, 57
patidifuso, 16
patiporsuelo, 57, 61
peciluengo, 67, 85
pedabobo, 64
pedabobos, 64
pegapega, 55
pejemuller, 62

pejerrey, 67
pelirrojo, 15, 16
penseque, 15
perrogorda, 63
pesamedello, 56
peso pluma, 20
petrolífero, 60
picabejero, 56
picajuyendo, 68
picapedrero, 24
picapica, 55
picatruco, 62
pinchauva, 7
pinchaúvas, 7
pintamonas, 7
pontentodo, 11, 62
porsiacaso, 11, 15
portaaviones, 55
profesor, 22, 66
puntapié, 7

Q

quebracho, 56
quemasangres, 15
quetetroncho, 10, 11, 55
quebrarado, 56
quienesquiera, 64
quienquiera, 18
quincañera, 24
quincañero, 25, 26
quitaipón, 5, 6, 11, 16, 20,
58, 68
quitapenas, 15

R

rabicorto, 57
rabiguana, 58
rabodejunco, 11, 55, 61
rabucocandil, 62
rascacielos, 63
rectinervio, 16
retahíla, 55
ricadueña, 63
ricahembra, 63
ricasdueñas, 63
ricashembras, 63
rojiblanco, 54
rojinegro, 57
rompeolas, 54
rompesquinas, 56
rompirraja, 57
ropavejero, 24, 25, 66
rosadelfa, 56

S

sabelotodo, 16
sabidondo, 64
sabibondo, 59, 64
sabilongo, 64
sabiondo, 56
salipez, 58
salsipuedes, 11, 55, 61
saltaembarca, 61
saltambanco, 67
saltarrostro, 15
saltatrás, 56
saltigalga, 57

sangrigrordo, 57

siempreviva, 54

sietemesino, 24

sillonball, 7

sin embargo, 18

sino, 18

siquiera, 17

sobreabundante, 23

sobredimensionar, 23

sobredosis, 23

sordomudo, 55

subibaja, 57

T

tampoco, 17

taqui-, 60

taquicardia, 60

tecnotrónica, 21

tela de araña, 20, 26

telaraña, 15, 20, 26, 56

telarañas, 7

tele-, 60

teléfono, 60

televisor, 22

tenconten, 11, 55, 61

tentempie, 62

tentempié, 11, 67

tentenelaire, 11, 62

tentibien, 57

tergal, 21

teticiega, 16

tiovivo, 55

tiraiafloja, 11, 16, 58, 68

todabuena, 55

todopoderoso, 55

tonticiego, 16

tragaavemarias, 55

tragafees, 64

tragaños, 56

trampantojo, 10, 56, 62

transistor, 21

trotamundos, 15

U

ubriciega, 57

uñalbo, 7, 55, 80

V

vaivén, 11, 15

veinticinco, 18

*vejero**, 25

vendepatrias, 7

verdinegro, 57

vitrocerámica, 8

vuelvaluegos, 64

X

xeno-, 60

xenofobia, 60

Y

yerbarreuma, 67

Z

zoológico, 22

zooplankton, 8