
PRAGMATIZACIÓN EN LA AUTOMATIZACIÓN
DEL RECONOCIMIENTO DE PALABRAS COMPUESTAS
EN ESPAÑOL

OCTAVIO SANTANA SUÁREZ
Universidad de Las Palmas de Gran Canaria

JOSÉ R. PÉREZ AGUIAR
Universidad de Las Palmas de Gran Canaria

VIRGINIA GUTIÉRREZ RODRÍGUEZ
Universidad de La Laguna

RESUMEN

El objetivo del presente trabajo es analizar desde un punto de vista pragmático uno de los procedimientos de formación de palabras más eficaces de la lengua española –*la composición*– que permita de manera formal y automática tratar neologismos compositivos. Se presentan un conjunto de reglas de carácter léxico y de cambio gráfico para el reconocimiento de compuestos, construidas a partir del comportamiento mayoritario observado en las diversas bases documentales tratadas. Se exponen las normas aplicadas inducidas a partir de la realidad explorada que permiten solventar barreras lingüísticas a través del dominio informático y poner de manifiesto las irregularidades encontradas en su tratamiento. Los mecanismos utilizados para la detección de las reglas atienden a la categoría gramatical del compuesto y a la de sus elementos componentes; con el propósito de garantizar un conocimiento suficiente de la casuística que envuelve a este proceso, se considerarán solo aquellos compuestos consolidados o perfectos, donde el grado de fusión de los elementos del compuesto es total.

PALABRAS CLAVE: composición, neologismos, reglas de formación, procesamiento del lenguaje natural, lingüística computacional.

PRAGMATIC AUTOMATION TO RECOGNITION OF COMPOUND WORDS IN
SPANISH

ABSTRACT

The aim of this study is to analyze from a pragmatic perspective one of the most effective words formation procedures of the language in the field of compounding, that allows us in a regularly and automatically way deal with compounding neologisms. It is introduced a series of lexical and graphic rules for the recognition of

compounds extracted from the behavior shown in the various documentary databases that have been used. There are shown the rules applied which are induced from the reality observed, that let us to solve linguistic barriers through the informatic domain, and, in addition, point out the irregularities found in its treatment. The mechanisms used for the detection of these rules belong to the grammatical category of the compound and to its elements: with the aim of guarantee a basic knowledge of casuistic that is involved in this process, they are considered valid only those solid or perfect compounds where the fusion degree of its elements is complete.

KEY WORDS: composition, neologisms, formation rules, natural language processing, computational linguistic.

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 morfosintácticos, en particular, neologismos compositivos, gracias al influjo provocado por los lenguajes técnicos, publicitarios, sociales, entre otros –hoy en día, la composición es sin duda uno de los procesos de formación de palabras más prósperos que dispone la lengua para la creación de unidades léxicas, vocablos o giros nuevos.

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. En este trabajo se parte de elementos ya presentes en el lenguaje o de otros tomados de fuera para reconocer nuevos vocablos mediante la unión de dos o más de tales elementos. Se estudian aquellos casos de composición y de parasíntesis que se han consolidado como la unión gráfica de los elementos que intervienen¹; se tratan los pseudoprefijoides o pseudosufijoides como elementos compositivos no como morfemas derivativos; se incluye la parasíntesis por composición², desde un punto de vista principalmente morfológico, ni sintáctico ni semántico, lo que no impide que en ocasiones sea necesario hacer referencia a tales aspectos. Se trata, en suma, de procedimientos que pueden crear neologismos, lo que constituye un fructífero manantial para el enriquecimiento de la lengua.

¹ No se estudian aquellos compuestos donde sus elementos componentes han alcanzado una coherencia sintáctica y/o semántica pero sin fusión ortográfica.

² Véase DAVID SERRANO DOLADER, *Las formaciones parasintéticas en español*, Arco/Libros, Madrid, 1995.

1. LA FORMACIÓN DE PALABRAS COMPUESTAS EN ESPAÑOL

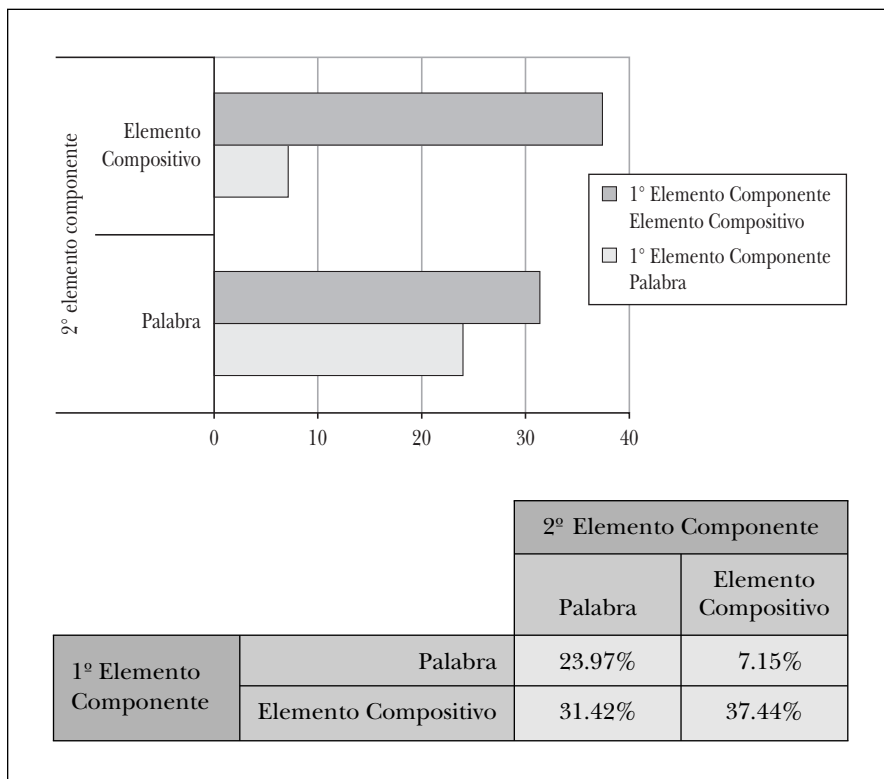
La perspectiva adoptada para el estudio de compuestos se basa en un análisis sincrónico debido a la dificultad que entraña la visión diacrónica ya que es prácticamente inexistente³. En los estudios realizados sobre composición en español, se pone de manifiesto una falta de homogeneidad tanto en el establecimiento de los diferentes tipos de compuestos como en el esclarecimiento de sus características –se remite al lector a las obras de M. Lang, M. Alvar, S. Varela, Alemany Bolufer, Almela Pérez, Bustos Gisbert, Val Álvaro⁴, entre otros. No se ha llegado a dar una definición de forma unánime, sobre todo a la hora de fijar la frontera entre composición y derivación, o bien entre los compuestos sintagmáticos y las colocaciones y locuciones. La definición más próxima a este estudio aparece en el Diccionario de la Real Academia Española –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, quitaipción, cagalaolla, paraguas*. Se añade que consiste en hacer de dos o más elementos una nueva palabra cuyo significado –muchas veces en sentido figurado– suele ser distinto y más extenso que la suma de los de sus componentes. Se considera que los elementos componentes que forman un compuesto pueden ser palabras⁵ “castellanas enteras” o temas cultos de origen grecolatino denominados *elementos prefijales* o *sufijales* –*elementos compositivos*– según se antepongan o se pospongan. La Tabla 1 muestra el porcentaje en la formación de palabras compuestas según las combinaciones posibles de estos dos tipos de unidades léxicas, recogidas en la base documental considerada en este estudio.

³ En la tesis doctoral de CRISTINA BUENAFUENTES DE LA MATA (*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>> [1/10/2013]) 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.

⁴ MERVING LANG, *Formación de palabras en español. Morfología derivativa productiva en léxico moderno*, Cátedra, Madrid, 1992; MANUEL ALVAR ESQUERRA, *La formación de las palabras en español*, Arco/Libros, Madrid, 2002; SOLEDAD VARELA, *Morfología Léxica: La formación de palabras*, Gredos, Madrid, 2005; 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*, Publicaciones de la Universidad de Salamanca, 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.

⁵ Se entiende *palabra* como aquella 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.

TABLA 1. Porcentajes en las combinaciones de distintas unidades léxicas para formar palabras compuestas



En el lenguaje científico y tecnológico es muy habitual la formación de nuevas palabras o neologismos utilizando raíces cultas grecolatinas. 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 derivadas 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 afijos⁶; incluso estos elementos compositivos pueden ocupar tanto la po-

⁶ 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 VARELA (*Morfología Léxica: La formación de palabras, op. cit.*)).

sición inicial como la posición final en el compuesto frente a la posición pre-determinada de los afijos –*cefalópodo, cefalotórax, dollicocéfaló, mesocéfaló*.

La Tabla 3 2 muestra información de los elementos compositivos estudiados, el número de elementos prefijales y sufijales considerados, el número de compuestos encontrados y el total de palabras tratadas –contienen los elementos compositivos fijados.

TABLA 3. Elementos prefijales y elementos sufijales

	Elementos compositivos		Compuestos encontrados	Palabras tratadas
Elementos prefijales	1647	<i>estrati- porta- galo- plamen- porta- mal- hipconco- zoo- tel- ter- malpict- acu- hidr- cinemat-</i>	6952	48493
Elementos sufijales	274	<i>-fero -oide -cito -osis -uro -forme -oma -antometro -logía -mero -ficio -bio -lito -omo -mano -logo -grafo</i>	1650	5005

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, existen casos de tres –*almempena, maldejojo, quetetrncho*–, o incluso cuatro componentes –*correquetecagas, culodemalasiento*–, 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 proviene.

En cuanto a las combinaciones de palabras sin elementos compositivos 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 formas, 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. La composición nominal es la más común –sirve para denominar objetos–, le sigue la composición adjetiva y, en menor medida, otras

categorías gramaticales adicionales, tales como verbos, adverbios, pronombres, conjunciones. En función de la clase funcional de los elementos componentes se obtienen los siguientes esquemas de combinación más típicos:

Sustantivo + Sustantivo

Verbo + Complemento

Sustantivo + Adjetivo

Adjetivo + Sustantivo

Sustantivo + Preposición + Sustantivo

Aunque existe un tipo de composición denominada *sintagmática*⁷, 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*– y entrarían en conflicto con las colocaciones léxicas⁸ –*abanderar lucha, ronda de negociaciones*. Sin embargo, gracias al comportamiento motivado, regular e isomórfico de los elementos constituyentes⁹, el compuesto podría diferenciarse de las colocaciones y las locuciones. Según se recoge en el trabajo de M. Alvar¹⁰, la composición se sirve de procedimientos para la creación de palabras nuevas como 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 –*malsano, agridulce, quitaipción, boquicerrado*. El presente trabajo se restringe a este tipo de composición; deja para estudios posteriores la composición sintagmática y 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.

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 5 –la intensidad del color marca la mayor o menor frecuencia de la combinación.

⁷ 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.

⁸ No pueden adquirir la categoría de compuesto aunque semánticamente estén unidos.

⁹ Véase LEONOR RUIZ GURILLO, “Compuestos, colocaciones, locuciones: intento de delimitación”, en A. VEIGA, M. GONZÁLEZ PERERIRA y M. SOUTO GÓMEZ (eds.), *Léxico y Gramática*, Tris Tram, Lugo, 2002, págs. 327-339.

¹⁰ MANUEL ALVAR EZQUERRA, *Nuevo Diccionario de voces de uso actual*, Arco/Libros, Madrid, 2003.

TABLA 5

CATEGORÍA GRAMATICAL COMPUESTO		CATEGORÍA GRAMATICAL ELEMENTOS COMPONENTES			
		SUSTANTIVO	ADJETIVO	VERBO	ADVERBIO
SUSTANTIVO	SUSTANTIVO	<i>casatienda</i> <i>bocamanga</i> <i>carricoche</i> <i>telaraña</i>	<i>aguardiente</i> <i>hierbabuena</i> <i>artimaña</i> <i>pelirrojo</i>		
	ADJETIVO	<i>ciempiés</i> <i>malasangre</i> <i>mediodía</i> <i>malpaís</i>			
	VERBO	<i>hincapié</i> <i>matamoscas</i> <i>quemasangres</i> <i>saltarrosto</i> <i>quitapenas</i> <i>trotamundos</i>		<i>compraventa</i> <i>duermevela</i> <i>ganapierde</i> <i>tiraiafloja</i> <i>quitaipón</i> <i>correveidile</i> <i>hazmerreír</i>	<i>bogavante</i> <i>atalejo</i> <i>mandamás</i> <i>creique</i> <i>penseque</i> <i>sabelotodo</i> <i>metomentodo</i>
	ADVERBIO	<i>bienandanza</i> <i>menoscuanta</i> <i>bienquerencia</i> <i>cientoemboca</i> <i>milenrama</i>		<i>bienmesabe</i>	<i>porsiacaso</i>
ADJETIVO	SUSTANTIVO		<i>alicaído</i> <i>cejijunto</i> <i>pelirrojo</i> <i>teticiega</i> <i>patidifuso</i>		
	ADJETIVO		<i>tonticiego</i> <i>grandilocuente</i> <i>agridulce</i> <i>hispanohablante</i>		
	VERBO				
	ADVERBIO		<i>bienintencionado</i>		

CATEGORÍA GRAMATICAL COMPUESTO		CATEGORÍA GRAMATICAL ELEMENTOS COMPONENTES			
		SUSTANTIVO	ADJETIVO	VERBO	ADVERBIO
VERBO	SUSTANTIVO			<i>maniatar</i> <i>aliquebrar</i> <i>manuscibir</i>	
	ADJETIVO				
	VERBO				
	ADVERBIO			<i>maldecir</i> <i>malcomer</i> <i>bienpensar</i> <i>menospreciar</i>	
ADVERBIO, PRONOMBRE, CONJUNCIÓN,...	SUSTANTIVO				
	ADJETIVO				<i>otrosi</i>
	VERBO				
	ADVERBIO*		<i>tampoco</i> <i>nosotros</i>	<i>siquiera</i> <i>dondequiera</i> <i>cualquiera</i> <i>quienquiera</i>	<i>anteayer</i> <i>aunque</i> <i>sino</i>

TABLA 5. Clasificación de compuestos atendiendo a la categoría gramatical composicional y las categorías de sus elementos componentes¹¹

- En cuanto a la categoría gramatical del compuesto, destacan por su productividad los que dan lugar a un *sustantivo*. Una de las combinaciones más abundante es la estructura de *Verbo + Sustantivo*, debido a la constitución interna que presenta, así como a la comodidad que le produce a un hablante por la semántica que sus compuestos implican –caracteriza al referente a través de su actividad¹².
- En la composición *adjetiva*, la estructura más profusa es *Sustantivo + Adjetivo*. Existe una marcada relación entre los elementos del com-

¹¹ El asterisco junto al adverbio indica que se incluyen otras categorías gramaticales, menos frecuentes, tales como: pronombre, conjunción, numeral,...

¹² Véase EUGENIO BUSTOS, *La composición nominal en español*, op. cit.

puesto, ya que el segundo miembro predica una cualidad del primero –sustantivo que generalmente designa partes exteriores del cuerpo humano o animal.

- En la composición *verbal* no se ha detectado que combine dos constituyentes de su misma categoría gramatical, es decir, *Verbo + Verbo* –se constata esta misma restricción en otras lenguas.

No se trata la formación del tipo: ADJETIVO + -MENTE → ADVERBIO –*intrínsecamente, fatídicamente*. Aunque algunos autores¹³ la reconocen como un proceso de composición, figura consolidada como una formación derivativa y así lo recoge el Diccionario General de la Lengua Española¹⁴ pues considera formación de adverbios de modo pospuesto a los adjetivos en su forma femenina –*malamente*.

Por otro lado, existen palabras que se forman a partir de una combinación de otras dos más una sufijación –*parasintéticos en composición*– que siguen la estructura *A + B + Sufijación*, según D. Serrano¹⁵. No existe ni la combinación *A + B* –se consideraría derivación de palabra compuesta– ni la palabra *B + Sufijación* –supondría derivado por sufijación–: en caso de aparecer, no serían el origen de la palabra final. Este es el caso de *hurgamandera* –no existe *hurgamanda* ni *mandera*– y *anquimuleño* –no existe *muleño*.

En el presente artículo se procede a estudiar 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, desde un punto de vista morfológico, ya que en los restantes casos se habrían de tener en cuenta factores sintácticos y semánticos para poder justificar formalmente su naturaleza de verdadero compuesto en español. Además, es importante tener presente el *orden fijo* de los elementos constituyentes con respecto a la categoría composicional, es decir, no se han encontrado compuestos verbales o adjetivales formados por un *Verbo + Sustantivo*.

2. AUTOMATIZACIÓN DEL RECONOCIMIENTO DE COMPUESTOS

Las reglas de formación de compuestos que se estudian son de carácter léxico y no sintáctico, semántico o fonológico, lo que no excluye que, en oca-

¹³ WALDO PÉREZ CINO, *Manual Práctico de formación de palabras en español I*, Verbum, Madrid, 2002.

¹⁴ BIBLIOGRAF, S. A., *Diccionario General de la Lengua Española VOX*, Vox, Barcelona, 2003.

¹⁵ DAVID SERRANO DOLADER, *Las formaciones parasintéticas en español*, Arco/Libros, Madrid, 1995.

siones, sea necesario hacer referencia a tales aspectos dado que representan características propias de este proceso de formación de palabras. Ya que las formas libres que constituyen los compuestos poseen 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 **formamiento* por *formación*. La existencia en el lexicón¹⁶ de innumerables particularidades no debería suponer 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 extraen de diferentes diccionarios los distintos compuestos y se estudian los cambios gráficos que han sufrido las diferentes palabras que los forman así como sus categorías gramaticales. Se obtiene así un conjunto de reglas y sus excepciones que permiten el reconocimiento y posterior generación automática de palabras compuestas.

2.1. *Corpus seleccionado*

Se parte de un conjunto de tamaño suficientemente significativo de compuestos, alrededor de 11 000, y se clasifican en grupos según la categoría gramatical de sus constituyentes. Se toma como base documental de partida los compuestos recogidos en los principales repertorios lexicográficos de la lengua española¹⁷ y del glosario de compuestos de E. Bustos¹⁸ –en su elaboración se utilizaron obras de carácter general y otras de carácter regional o dialectal, como hablas leonesas, aragonesas, meridionales y español de América.

¹⁶ 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.

¹⁷ BIBLIOGRAF, S. A., *Diccionario General de la Lengua Española VOX*, op. cit.; REAL ACADEMIA ESPAÑOLA, *Diccionario de la Lengua Española*, Espasa-Calpe, Madrid, 22ª ed., 2001; CLAVE S. M., *Diccionario de Uso del Español Actual*, Clave S. M., Madrid, 1997; LAROUSSE PLANETA, S. A., *Gran Diccionario de la Lengua Española*, Larousse Planeta, S. A., Barcelona, 1996; MARÍA MOLINER, *Diccionario de Uso del Español*, Gredos, Madrid, 1996.

¹⁸ Véase EUGENIO BUSTOS, *La composición nominal en español*, op. cit.

2.2. Criterios en el reconocimiento

A partir de la base que se considera se extraen las reglas de formación de los compuestos. Se parte de la palabra y se prueba que cumpla unas ciertas condiciones tanto respecto a su tamaño como a la relevancia del orden de sus constituyentes. Por norma general, los compuestos presentan una longitud mayor que cinco caracteres o están formados por al menos tres sílabas, como *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*”¹⁹ como las “*Relaciones Morfológicas del español*”²⁰ del *Grupo de Estructura de Datos y Lingüística Computacional de la Universidad de Las Palmas de Gran Canaria*²¹. Durante el proceso se generan cortes que pueden dar lugar a múltiples posibilidades entre las que pueden aparecer algunas que no sean auténticas soluciones; esta aproximación 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.

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, **bocahundido* –falsa composición pues lo correcto sería *boquihundido*– podría considerarse una formación apropiada al no contravenir ninguna regla fonotáctica del lenguaje, ni siquiera la norma de la estructura silábica del español; sin embargo, no sería una so-

¹⁹ <<http://gedlc.ulpgc.es/investigacion/scogeme02/lematiza.htm>> [1/10/2013].

²⁰ <<http://gedlc.ulpgc.es/investigacion/scogeme02/relmorfo.htm>> [1/10/2013].

²¹ OCTAVIO SANTANA SUÁREZ, JOSÉ PÉREZ AGUIAR, ZANÓN HERNÁNDEZ FIGUEROA, FRANCISCO CARRERAS RIUDAVETS, GUSTAVO RODRÍGUEZ RODRÍGUEZ, “FLAVER: Flexionador y lematizador automático de formas verbales”, *Lingüística Española Actual*, XIX-2 (1997), págs. 229-282; OCTAVIO SANTANA SUÁREZ, JOSÉ PÉREZ AGUIAR, FRANCISCO CARRERAS RIUDAVETS, JUAN DUQUE, ZENÓN HERNÁNDEZ FIGUEROA, GUSTAVO RODRÍGUEZ RODRÍGUEZ, “FLANOM: Flexionador y lematizador automático de formas nominales”, *Lingüística Española Actual*, XXI-2 (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, FRANCISCO CARRERAS RIUDAVETS, JOSÉ PÉREZ AGUIAR, VIRGINIA GUTIÉRREZ RODRÍGUEZ, “El Reconocimiento Automático de la Composición en Español”, *Conference Abstracts of the First International Conference of the Alliance of Digital Humanities Organizations*, París, 2006; OCTAVIO SANTANA SUÁREZ, FRANCISCO CARRERAS RIUDAVETS, JOSÉ PÉREZ AGUIAR, JUAN RODRÍGUEZ DEL PINO, “Parasyntetic Morpholexical Relationships of the Spanish: Lexical Search beyond the Lexicographical Regularity”, *Proceedings of the IADIS International Conference, Applied Computing*, 2006, págs. 627-631.

lución de las más probables debido a la invariabilidad formal que presenta el compuesto con respecto a la regla de formación aplicada a través de las categorías gramaticales que intervienen, así como a la frecuencia de uso por los hablantes del término correcto.

En definitiva, se buscan las reglas de formación a partir del estudio del comportamiento de los vocablos constituyentes del compuesto; algunas coinciden con las tratadas por lingüistas, aunque con una adaptación informática justificada por el comportamiento mayoritariamente observado. Asimismo, para un correcto tratamiento informático hay que tener en cuenta que cuando se aplican las reglas de reconocimientos de compuestos hay que considerar los cambios gráficos que se pueden producir como consecuencia de la aplicación de las reglas de formación.

2.3. Reglas de formación

Según sea el grado de modificación que sufran los elementos originales, los procedimientos de formación de palabras compuestas pueden dividirse en dos grandes grupos: o bien por la mera adición de dos o más términos sin que ninguno de ellos se modifique –*cenaoscuros, hullehulle, 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 –*liquidámbar, anquirredondo, claroescuro, cochitril*.

Desde el punto de vista del reconocimiento, se presentan las reglas de formación de palabras compuestas que se ponen de manifiesto en la base de estudio.

Unión sin pérdida ni adición ni cambio. Esta regla se aplica de forma general a todas las categorías gramaticales. Generalmente, 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.

[ADVERBIO + ADJETIVO]_{ADJETIVO}: *malavenido, bienafortunado*.

[ADJETIVO + ADJETIVO]_{ADJETIVO}: *todabuena, todopoderoso, sordomudo*.

VERBO + VERBO: *huelehuele, picapica, chupachupa, pegapega*.

SUSTANTIVO + SUSTANTIVO: *aguasal, aguapié, ajoaceite, bocacalle, pañomanos*.

VERBO + COMPLEMENTO O VICEVERSA: *quetetroncho, portaaviones, guardaespaldas, mataorejas, afilelápices, matahombres, malqueda, abreojos*.

SUSTANTIVO + PREPOSICIÓN + SUSTANTIVO: *maldejojo, pandesapo, rabodejunco*.

SUSTANTIVO + ADJETIVO O VICEVERSA: *tiovivo, retahíla, malalma, avetonto, babasfrías, culocagado, gentilhombre, medialanza, aguafresquera*.

Cuando los elementos componentes terminan y empiezan con la misma vocal *a*, *e* u *o*, se elimina una de ellas –sinalefas o reducciones de los hiatos. En *Verbo + Complemento*, la pérdida que se produce es de la vocal *a*, pues normalmente el elemento verbal aparece en forma imperativa y la flexión de los verbos de la primera conjugación acaba en *a* –los más utilizados en esta regla.

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

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

SUSTANTIVO + SUSTANTIVO: *maestrescuela, telaraña, manobra*.

VERBO + COMPLEMENTO O VICEVERSA: *guardaguas, paraguas, tragaños, saltatrás*.

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

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 ← *quiebra + acho* ← *quiebra + hacha*

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

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

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} hay que tener en cuenta que existen palabras que aplicarían la regla anterior de no ser porque previamente hayan incorporado el infijo compositivo *i*.

[ADJETIVO + ADJETIVO]_{ADJETIVO}: *cojitranco, anchicorto, verdinegro, dulciagrío*.

[SUSTANTIVO + ADJETIVO]_{ADJETIVO}: *alicaído, barbicastano, clarividente, culinegro, sangrigordo, cariancho*.

VERBO + VERBO: *bativoleo, botivoleo, chiticalla, lavatiquevoy, rompirraja*.

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

VERBO + COMPLEMENTO O VICEVERSA: *abriboca, frunciboca, baticola, tentibien*.

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

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*.

ADJETIVO + ADJETIVO: *azuliblanco*.

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

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.

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.

ADJETIVO + ADJETIVO: *eurasiático*.

[SUSTANTIVO + ADJETIVO]_{ADJETIVO}: *papalbo*.

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

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

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

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:

labihendido ← *labio + hendido* (no es solución *labo + hendido*)

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

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*:

cabezcaído ← *cabecicaído* ← *cabeza + caído*

manvacío ← *manivacío* ← *mano + vacío*

Los elementos compositivos pueden estar formados por la agregación a una palabra española de raíces cultas greco-latinas, antepuestas o pospuestas –*acutifolio, nitroglicerina, abulomanía, ecosonda*–; por la combinación de raíces

cultas: raíz prefija y sufija griegas o latinas –*teléfono*, *filicidio*–, raíz prefija griega y sufija latina –*automóvil*– o raíz prefija latina y sufija griega –*hispanófilo*–; o por la incorporación de otro tipo de raíces –*arisblanco*, *eurocheque*, *galiporte*.

2.4. *Tratamiento de las excepciones*

Algunas palabras presentan irregularidades y necesitan un tratamiento especial.

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

<i>botesela</i> : <i>sela</i> ← <i>silla</i> ,	<i>cuerdacarru</i> : <i>carru</i> ← <i>carro</i> ,
<i>capaniú</i> : <i>niú</i> ← <i>nido</i> ,	<i>guadarnés</i> : <i>guad</i> ← <i>guarda</i> ,
<i>pejemuller</i> : <i>peje</i> ← <i>pez</i> ,	<i>rabucocandil</i> : <i>rabuco</i> ← <i>rabo</i> ,
<i>cagarrache</i> : <i>rache</i> ← <i>erraj</i> ,	<i>picatrueu</i> : <i>trueu</i> ← <i>tuero</i> ,
<i>ajilimójili</i> : <i>ajili</i> ← <i>ajo</i> <i>mójili</i> ← <i>moje</i>	

Si el primer elemento acaba en *a*, se sustituye por la vocal *o* y se lleva a cabo la unión –*ceromiel* ← *cera* + *miel*–; si el primer elemento acaba en *o*, se sustituye por la vocal *a* y se procede a la unión –*cerrapan* ← *cerro* + *pan*. Se conserva el plural en los casos en que se produzca –*ajaspajas* ← *ajos* + *pajas*. De forma similar ocurre con la segunda palabra –*aguagriero* ← *agua* + *agriera*, *batiporte* ← *bate* + *porta*.

El sintagma nominal ha sufrido un proceso de aglutinación sin pérdida de la preposición, del artículo, de la conjunción (copulativa, condicional,...) o del pronombre personal clítico o enclítico presentes.

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

VERBO + COMPLEMENTO: *saltaembarca*, *besalamano*, *bienmesabe*, *ahimelollevas*.

SUSTANTIVO + PREPOSICIÓN + SUSTANTIVO: *rabodejunco*, *juanencueros*, *patiporsuelo*.

Unión con pérdida de la vocal del segundo elemento o nexa: hay casos en los que se descarta alguno de los fonemas del nexa presente en el compuesto, o bien su final vocálico o bien su principio vocálico, ante la posibilidad de formación de un diptongo o de una contracción –generalmente se elimina la vocal inicial *e*.

VERBO + COMPLEMENTO: *montambanco*, *tentempie*, *tentenelaire*, *pontentodo*, *guardalagua*.

SUSTANTIVO + PREPOSICIÓN + SUSTANTIVO: *trampantojo*, *fijodalgo*, *hidalgo* ← *hijo de algo*.

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

- 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 ← *vuelta + luego* ← *vuelta + luego*
tragafees ← *traga + fees* ← *traga + fe*

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

cascorvo ← *casco + corvo*

- 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*
peciluengo ← *peciolo + luengo*
cejunto ← *ceja + junto*

- 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*

Por último, los elementos constituyentes del compuesto pueden presentar una construcción paranomásica –creados por semejanza de sonidos con otros vocablos.

sabidondo, sabilongo ← *sabihondo*
*pedabobo*²²
balsopeto ← *bolso + peto*

²² Véase, MARÍA PILAR ORTEGA (“Neología y Prensa: Un binomio eficaz”, *Espéculo*, 18 (Madrid, 2001), <<http://pendientedemigracion.ucm.es/info/especulo/numero18/neologism.html>> [1/10/2013]): “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*.

2.5. Reglas fonéticas o de cambio gráfico

Hay que tener en cuenta que cuando se aplican las reglas anteriores pueden producirse cambios gráficos en el compuesto final; por tanto, se han de tener en cuenta las siguientes reglas para lograr un correcto tratamiento informático en el reconocimiento de tales palabras. Se muestran desde el punto de vista de la generación de los compuestos.

Si el segundo elemento empieza por *r*, se debe sustituir por *rr* –*aguarrás*, *alirrojo*, *andarríos*, *alzarrabos*, *arrancarrabia*, *buscarruidos*, *barbirrostro*, *boquirroto*, *carirraído*, *grecorromano*, *matarratas*, *pararrayos*, *pejerrey*, *yerberreuma*–, salvo cuando va precedida de “*l*” o “*n*”: *milenrama* ← *mil + en + rama*.

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.

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

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.

<i>cabeciancho</i>	←	<i>cabeza + ancho</i>
<i>cabeciduro</i>	←	<i>cabeza + duro</i>

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.

<i>almaempena</i> ,	<i>balompié</i> ,	<i>mampuesto</i> ,
<i>montambanco</i> ,	<i>saltambanco</i> ,	<i>tentempié</i>

Se sustituye la *y* copulativa que funciona como elemento conjuntivo por la vocal *i* y se lleva a cabo la unión.

<i>correvidile</i>	←	<i>corre + ve + y + dile</i>
<i>quitaipón</i>	←	<i>quita + y + pon</i>
<i>tiraiafloja</i>	←	<i>tira + y + afloja</i>

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

<i>picajuyendo</i>	←	<i>pica + huyendo</i>
<i>bocajacha</i>	←	<i>boca + haya</i>

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

cabizcaído ← *cabeza + caído*
cabizmordido ← *cabeza + mordido*

Hay que tener en cuenta la acentuación del compuesto. Por lo general, el acento recae en la segunda palabra del compuesto –son monoacentuados²³– debido a la existencia de un mayor número de palabras llanas en español.

aguarrás ← *agua + ras*
balompié ← *balón + pie*
ahimelolleva ← *ahí + me + lo + llevas*
espíritusanto ← *espíritu + santo*
ganapán ← *gana + pan*
catetómetro ← *cateto + -metro*

Por último, puede darse que cualquiera de los constituyentes del compuesto presente una formación atípica de diptongos o alteraciones de los mismos:

<i>quebra</i>	←	<i>quiebra,</i>	<i>tenta</i>	←	<i>tienta,</i>
<i>vola</i>	←	<i>vuela,</i>	<i>roda</i>	←	<i>rueda,</i>
<i>escorna</i>	←	<i>escuerna,</i>	<i>escola</i>	←	<i>escuela,</i>
<i>torce</i>	←	<i>tuerce,</i>	<i>troca</i>	←	<i>trueca,</i>
<i>toste</i>	←	<i>tueste</i>			

CONCLUSIONES

Pese a que la composición sea uno de los procesos de formación de palabras de mayor importancia y actualmente en auge, se observan importantes carencias tanto desde el punto de vista su procesamiento automático como en sus referencias bibliográficas. Este trabajo contribuye a paliar el vacío existente en el tratamiento informático de palabras compuestas, a partir del conjunto de reglas que se han expuesto; se han solventado las barreras lingüísticas que dificultarían una actuación correcta y formal gracias a proyectar sus mecanismos de reconocimiento al dominio informático, justificado por el comportamiento mayoritariamente observado en la base de referencia.

Se destacan la composición de *Verbo + Sustantivo*, la de *Sustantivo + Sustantivo* y la de *Sustantivo + Adjetivo* como muy frecuentes; constituyen los procedimientos más productivos de la composición del español actual, sea por su

²³ Salvo en casos excepcionales como *portaaviones*.

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 –actúan 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.

REFERENCIAS BIBLIOGRÁFICAS

- 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, *La formación de las palabras en español*, Arco/Libros, Madrid, 2002.
- ALVAR EZQUERRA, MANUEL, *Nuevo Diccionario de voces de uso actual*, Arco/Libros, Madrid, 2003.
- BIBLIOGRAF, S. A., *Diccionario General de la Lengua Española VOX*, Vox, Barcelona, 2003.
- BUSTOS GISBERT, EUGENIO, *La composición nominal en español*, Publicaciones de la Universidad de Salamanca, Salamanca, 1986.
- 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>> [1/10/2013].
- CLAVE S. M., *Diccionario de Uso del Español Actual*, Clave S. M., Madrid, 1997.
- ESTORNELL PONS, MARÍA, *Neologismos en la prensa: criterios para reconocer y caracterizar las unidades neológicas*, *Quaderns de Filología de la Universidad de Valencia, Anejos nº 70*, Publicaciones de la Universidad de Valencia, Valencia, 2009.
- FELIÚ ARQUIOLA, ELENA, “Palabras con estructura interna”, en ELENA DE MIGUEL (eda.), *Panorama de la lexicología*, Ariel, Barcelona, 2009, págs. 51-82.
- LANG, MERVING, *Formación de palabras en español. Morfología derivativa productiva en léxico moderno*, Cátedra, Madrid, 1992.
- LAROUSSE PLANETA, S. A., *Gran Diccionario de la Lengua Española*, Larousse Planeta, S. A., Barcelona, 1996.
- 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 y JOSÉ FRANCISCO VAL ÁLVARO (eds.), *Morfología y español como lengua extranjera (E/ELE)*, Prensas Universitarias de Zaragoza, Zaragoza, 2009, págs. 21-66.
- MOLINER, MARÍA, *Diccionario de Uso del Español*, Gredos, Madrid, 1996.
- ORTEGA, MARÍA DEL PILAR, “Neología y Prensa: Un binomio eficaz”, *Espéculo*, 18 (Madrid, 2001), <<http://pendientedemigracion.ucm.es/info/especulo/numero18/neologism.html>> [1/10/2013].
- PÉREZ CINO, WALDO, *Manual Práctico de formación de palabras en español I*, Verbum, Madrid, 2002.

- REAL ACADEMIA ESPAÑOLA, *Diccionario de la Lengua Española*, Espasa-Calpe, Madrid, 22ª ed., 2001.
- 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 A. VEIGA, M. GONZÁLEZ PERERIRA y M. SOUTO GÓMEZ (eds.), *Léxico y Gramática*, Tris Tram, Lugo, 2002, págs. 327-339.
- SANTANA SUÁREZ, OCTAVIO; PÉREZ AGUIAR, JOSÉ; HERNÁNDEZ FIGUEROA, ZENÓN; CARRERAS RIUDAVETS, FRANCISCO y RODRÍGUEZ RODRÍGUEZ, GUSTAVO, "FLAVER: Flexionador y lematizador automático de formas verbales", *Lingüística Española Actual*, XIX-2 (1997), págs. 229-282.
- SANTANA SUÁREZ, OCTAVIO; PÉREZ AGUIAR, JOSÉ; CARRERAS RIUDAVETS, FRANCISCO; DUQUE, JUAN; HERNÁNDEZ FIGUEROA, ZENÓN y RODRÍGUEZ RODRÍGUEZ, GUSTAVO, "FLANOM: Flexionador y lematizador automático de formas nominales", *Lingüística Española Actual*, XIX-2 (1999), págs. 253-297.
- SANTANA SUÁREZ, OCTAVIO; PÉREZ AGUIAR, JOSÉ; CARRERAS RIUDAVETS, FRANCISCO y 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; 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*, París, 2006, págs. 190-193.
- SANTANA SUÁREZ, OCTAVIO; CARRERAS RIUDAVETS, FRANCISCO; PÉREZ AGUIAR, JOSÉ y RODRÍGUEZ DEL PINO, JUAN, "Parasynthetic Morpholexical Relationships of the Spanish: Lexical Search beyond the Lexicographical Regularity", *Proceedings of the IADIS International Conference, Applied Computing*, 2006, págs. 627-631.
- SERRANO DOLADER, DAVID, *Las formaciones parasintéticas en español*, Arco/Libros, Madrid, 1995.
- VAL ÁLVARO, JESÚS, "La composición", en IGNACIO BOSQUE y VIOLETA DEMONTE (eds.), *Gramática descriptiva de la Lengua Española*, Espasa-Calpe, vol. 3, Madrid, 1999, págs. 4757-4841.
- VARELA ORTEGA, SOLEDAD, *Fundamentos de Morfología*, Síntesis, Madrid, 1990.
- VARELA ORTEGA, SOLEDAD, *Morfología Léxica: La formación de palabras*, Gredos, Madrid, 2005.