

Reglas de formación de palabras compuestas en español para la automatización de su reconocimiento

Formation rules of compound words in Spanish to automate recognition

Octavio Santana Suárez, José Pérez Aguiar
 Universidad de Las Palmas de Gran Canaria
 Edificio Departamental de Informática y
 Matemáticas
 Campus Universitario de Tafira
 35017 Las Palmas de Gran Canaria
 {osantana, jperez}@dis.ulpgc.es

**Virginia Gutiérrez Rodríguez, Isabel
 Sánchez Berriel**
 Universidad de La Laguna
 Edificio de Física y Matemáticas
 Campus Universitario Anchieta
 C/Astrofísico Francisco Sánchez s/n
 38271 La Laguna
 {vgutier, isanchez}@ull.es

Resumen: En el presente trabajo se recogen las reglas de formación y los criterios de aplicación que se deberían llevar a cabo en cada situación para permitir la identificación automatizada de uno de los procesos de formación de palabras que tiene el español: la composición, se estudian sólo aquellos casos en los que se ha producido dicho fenómeno mediante la unión gráfica de los elementos que participan, con miras a su automatización. A tal fin, se extraen de diferentes diccionarios los distintos compuestos con el propósito de garantizar un conocimiento suficiente de los diferentes casos que se pueden prever de la materia y se estudian los mecanismos de unión aplicados según la categoría gramatical del compuesto y las de sus elementos componentes.

Palabras clave: Composición, reglas de formación, neologismos, procesamiento del lenguaje natural, lingüística computacional

Abstract: The following work presents the formation rules and application criteria that should be carried out in each situation in order to obtain the automated identification of one of the word's formation processes in Spanish. In the composition are studied only those cases in which this phenomenon has been produced through the graphic link of the elements involved, aimed to automation. At this aim, different compounds are extracted from dictionaries with the only purpose of guarantee a basic knowledge of the cases that are provided by the materia. The linking mechanisms are also studied depending on the grammatical category of the compounds and the elements which form part of it.

Keywords: Composition, formation rules, neologisms, natural language processing, computational linguistic

1 *Introducción*

Uno de los medios con los que cuenta la lengua española para ampliar el conjunto de voces del idioma es utilizar mecanismos de tipo morfológico —también llamados neologismos morfosintácticos— para formar nuevas palabras como son la composición, la derivación o la parasíntesis, entre otros. Sin duda, la composición es uno de los procesos de formación de palabras con mayor importancia que dispone la lengua para la renovación y enriquecimiento de su léxico.

En estos mecanismos morfológicos se parte de elementos ya presentes en el lenguaje o de otros tomados de fuera para crear nuevos vocablos; en la composición mediante la unión de dos o más de estos elementos, en contraposición a la derivación donde 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 —afijos. Por último, en la parasíntesis se combinan los mecanismos anteriores, bien por afijación que simultanea

dos procedimientos derivativos —sufijación y prefijación— o por combinación de elementos de la composición y de la derivación como trata Serrano Dolader (1995). En el presente trabajo se estudian aquellos casos de compuestos que se han consolidado como la unión gráfica de los elementos que intervienen¹, además se tratan los pseudoprefijoides o pseudosufijoides como elementos compositivos no como morfemas derivativos, y se incluye la parasíntesis por composición, desde un punto de vista morfológico. Se trata, en suma, de procedimientos que pueden crear neologismos, lo que constituye una alternativa moderna para el enriquecimiento de la lengua.

2 La composición en español

En los estudios realizados sobre composición en español, no se ha llegado a dar una definición de forma unánime, sobre todo a la hora de establecer la frontera entre composición y derivación. Para M. Lang (1997) 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 M. Alvar (2002) donde en la composición participan dos o más unidades léxicas que pueden aparecer libres en la lengua; pero la más próxima a nuestro estudio es la definida por el Diccionario de la Real Academia de Lengua Español —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*. Se considera que los elementos componentes que forman una palabra compuesta pueden ser de dos tipos: palabras² “castellanas enteras” (P) y temas cultos de origen grecolatino llamados elementos compositivos prefijales o sufijales (EC). Al fijar que los elementos compositivos son “temas cultos” se establece la diferencia con lo que se entiende por verdaderos afijos,

¹ Hay que tener en cuenta que la composición puede dividirse en composición de tipo ortográficamente unidos y composición de tipo sintagmático, donde los componentes han alcanzado una coherencia semántica sin fusión ortográfica. El presente trabajo sólo se centrará en el primer tipo de composición.

² Entendiéndose como palabras aquellas que tienen un solo lexema o un lexema unido a morfemas flexivos o las que están formadas por un morfema libre o unido a morfemas flexivos.

tanto sean prefijos como sufijos³. La Tabla 1 muestra ejemplos de palabras compuestas formadas por las combinaciones posibles de estos dos tipos de unidades léxicas.

Composición	Palabra	Elemento compositivo
Palabra	<i>hojalata</i> <i>malgastar uñalbo</i>	<i>germanófilo</i> <i>timbrología</i>
Elemento compositivo	<i>ecosistema</i> <i>cardioprotector</i>	<i>logopeda</i> <i>filántropo</i>

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

La utilización de raíces cultas greco-latinas es frecuente 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 les da el nombre de *elementos prefijales* o *sufijales* —*elementos compositivos*—, en función de si se anteponen o se posponen a otra raíz.

Hay que tener en cuenta que aunque, por norma general —incluido en muchas definiciones de composición— el número de constituyentes que forma una palabra compuesta es de dos, existen casos de tres —*tiraiafloja, bienteveo, cenaaoscuras*—, o incluso cuatro componentes —*correvidile, ahimelollevas, tentenaire*—, aunque, en la mayoría de ellas, suelen insertarse elementos monosilábicos átonos como preposiciones, conjunciones copulativas, pronombres, artículos, entre otros. Las palabras compuestas 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 (P+P), estas pueden crearse a partir de

³ Un reconocimiento automatizado de compuestos formados por elementos compositivos tiene que considerar la distinción entre elementos compositivos y afijos: un elemento prefijal no equivale a la derivación por prefijación de una palabra (*bienvenir* es una palabra compuesta, no derivada del verbo *venir* con los prefijos *bi-* sino proviene del elemento compositivo *bien-*).

conceptos que estén relacionados, o mejor dicho, 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. La composición más común es aquella que sirve para denominar objetos como la composición nominal, seguida de la composición adjetiva y, en menor medida, la composición verbal. 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

donde E. Bustos (1986) plantea una clasificación según las categorías gramaticales de los elementos componentes, así como la categoría del compuesto como se indica en la Tabla 2.

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 nominales⁴, sino como unidades léxicas pues el compuesto se utiliza como una única palabra en la que han quedado cohesionados todos sus componentes; donde la composición *adjetiva* y la *propia* sustantiva —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 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* → *hojalata*, *hidalgo* o *aguanafa*. Desde un punto de vista sincrónico, el no reconocimiento de grupos sintácticos en alguno de estos tipos de compuestos es un problema de

⁴ No pueden adquirir la categoría de compuesto aunque semánticamente estén unidos como *farola de jardín* o *casa hospital* frente a palabras que sí lo son como *libro de cocina* o *restaurante espectáculo*.

tradición lingüística o de realidad de los hechos del lenguaje. 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 y por lo tanto no se van a considerar en el presente estudio.

Composición Adjetiva	Adverbio+Adjetivo
	Adjetivo+Adjetivo
	Sustantivo+Adjetivo
Composición Nominal	Composición Propia
	Sustantivo+Sustantivo
	Verbo+Verbo
Composición Sintagmática	Verbo+Complemento
	Sustantivo+Adjetivo o
	Adjetivo+Sustantivo
	Sustantivo+Preposición
	+Sustantivo

Tabla 2. Clasificación de la composición según E. Bustos Gisbert

Según la frecuencia de aparición de la categoría gramatical del compuesto, se pueden detallar las siguientes combinaciones:

- ❖ Uno de los compuestos más productivos son los que dan como resultado un *sustantivo*. Dentro de este tipo, los más caudalosos son los de *Verbo+Complemento* 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 E. Bustos (1986). En la Tabla 3 se aprecian diferentes formaciones de compuestos nominales.

Existen compuestos 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: *porsiacaso*.

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

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+PRON.PERSONAL+VERBO: *bienmesabe*.

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

VERBO+PRON.PERSONAL+VERBO: *hazmerreir*.

VERBO+PRON.PERSONAL+PREPOSICIÓN+

PRON. IMPERSONAL: *metomentodo*.

Categoría	Sustantivo	Adjetivo
Sustantivo	<i>casatienda bocamanga</i>	<i>aguardiente</i>
	<i>carricoche telaraña</i>	<i>hierbabuena</i>
Adjetivo	<i>ciempiés malasangre</i>	<i>artimaña pelirrojo</i>
	<i>mediodía malpais</i>	
Verbo	<i>hincapié matamoscas</i>	
	<i>quemasangres</i>	
	<i>saltarostro quitapenas</i>	
Adverbio	<i>trotamundos</i>	
	<i>bienandanza</i>	
	<i>menoscuenta</i>	
	Verbo	Adverbio
Verbo	<i>compraventa</i>	<i>bogavante atalejo</i>
	<i>duermevela</i>	<i>mandamás</i>
	<i>ganapierde</i>	
Adverbio	<i>bienmesabe</i>	

Tabla 3. Compuestos nominales según la categoría gramatical de sus elementos componentes

- ❖ El segundo caso en productividad son los compuestos *adjetivos*. Los más caudalosos son los de *Sustantivo+ Adjetivo* donde existe una marcada relación entre los elementos del compuesto ya que el segundo miembro predica una cualidad del primero, sustantivo⁵ que generalmente designa partes exteriores del cuerpo humano o animal.
- ❖ 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.

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.

SUSTANTIVO+VERBO: *maniatar, aliquebrar*.

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

- ❖ Que dan lugar a *adverbio*
 CONJUNCIÓN+VERBO: *siquiera* (adverbio o conjunción).
 ADVERBIO DE CANTIDAD+ADJETIVO O ADVERBIO: *tampoco*.
 ELEMENTO COMPOSITIVO+ADVERBIO DE TIEMPO: *anteayer*.

⁵ Existe una marcada tendencia a colocar en el primer miembro sustantivos bisílabos

ADJETIVO+ADVERBIO DE MODO: *otrosí*.

Raros: ADVERBIO+VERBO: *dondequiera*.

- ❖ 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*

Categoría	Adjetivo (o Participio)
Sustantivo	<i>alicaído, cejijunto,</i>
	<i>pelirrojo, teticiega, patidifuso</i>
Adjetivo	<i>tonticiego, grandilocuente,</i>
	<i>agridulce, hispanohablante</i>
Adverbio	<i>bienintencionado,</i>
	<i>malaconsejado</i>

Tabla 4.- Compuestos adjetivales según la categoría gramatical de sus elementos componentes

Se tiene en cuenta no tratar la formación del tipo: ADJETIVO + -MENTE → ADVERBIO — *rápidamente, compulsivamente*— debido a que, aunque algunos autores (Pérez, 2002) la reconocen como un proceso de composición, está consolidada como una formación derivativa; así lo recoge el Diccionario General de la Lengua Española (VOX, 2003) ya que considera -mente como un sufijo, no elemento sufijal, que entra en la formación de adverbios de modo pospuesto a los adjetivos en su forma femenina⁶ —*buenamente*.

Por otro lado, según se recoge en el trabajo de M. Alvar (2003), la composición se sirve de procedimientos para la creación de palabras nuevas, como son: sinapsia, disyunción, contraposición, yuxtaposición, prefijos vulgares y acortamiento. El más caudaloso de los procesos de composición es la *yuxtaposición o lexías compuestas*, aquí la fusión gráfica de los elementos participantes en el compuesto es total, así como su lexicalización y su gramaticalización —*carnicol, malqueda, cochitrit, hincapié*. Sin embargo, la unión de los miembros en la *sinapsia* es de naturaleza sintáctica, no morfológica, por lo que es difícil determinar si se ha producido lexicalización o no. Suele

⁶ El adjetivo adopta siempre la forma femenina, si la tiene, pues -mente es femenino en latín y conserva el acento.

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*—, correspondiendo este compuesto con la clasificación sintagmática que hace E. Bustos Gisbert (1986) de *Sustantivo + Preposición + Sustantivo*. Por otro lado, 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*— correspondiéndose con diversos compuestos *Sustantivo+Sustantivo* o *Sustantivo+Adjetivo* o *viceversa* de E. Bustos. Se puede incluso llegar, en las dos composiciones anteriores, a la unión gráfica de los elementos —*tela de araña*→*xelaraña*, *agua nieve*→*aguanieve*, *ave fría*→*avefría*.

En un grado más alto de unión gráfica que los dos anteriores, está la *contraposición* donde los elementos que participan en ella se escriben unidos por un guión, aunque generalmente el resultado aparecerá sin él debido a las restricciones del uso del guión que tiene el español —*falda pantalón* → *falda-pantalón*—, considerándola como uno de los compuestos anteriores.

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 S. Varela (1990) en su libro *Fundamentos de Morfología*, aunque M. 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.

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 como parte de la composición, pero la frontera entre ellos no está todavía muy clara sobre todo en el caso de abreviamento.

Existen palabras que se forman a partir de una combinación de dos palabras más una sufijación, según D. Serrano (1995)

—siguiendo la estructura *A+B+Sufijación*—, pero no existe la combinación *A+B* ni *B+Sufijación* —se considerarían derivación de palabras compuestas o derivados por sufijación, respectivamente—, y en caso de existir no son el origen de la palabra final. A este tipo especial de palabras se le conoce con el nombre de parasintéticos en composición.

misacantano —no existe *misacanta* ni *cantano*—, *ropavejero* —no existe *vejero*—, *doceañista* —no existe *añista*—, *aguamanil* —no existe *mano+il*

Es, por todo lo anterior, que en el presente artículo se procede a estudiar los compuestos yuxtapuestos o lexías compuestas en composición nominal y adjetiva, según la clasificación de E. Bustos (1986), 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 tipos se han de tener en cuenta factores sintácticos y semánticos para poder justificar que es un verdadero compuesto en español y esto se escapa actualmente del dominio informático.

3 Reglas de formación de palabras compuestas

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 ocasiones, sea necesario hacer referencia a dichos aspectos dado que son características inseparables de este proceso de formación de palabras.

Las primitivas, a las cuales se le aplican las reglas de formación de palabras compuestas que se buscan, deben ser palabras consolidadas en nuestra lengua o neologismos, pero nunca palabras incorrectas —ortografía irregular o palabras bloqueadas como **grabamiento* por *grabación*. La existencia en el lexicon⁷ de innumerables particularidades no debería suponer una barrera para la automatización de este 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

⁷ Los ejemplos utilizados pueden ser formaciones neológicas, no necesariamente documentadas en la base de referencia.

de diferentes diccionarios los distintos compuestos y se estudian los cambios gráficos que han sufrido las diferentes palabras que los forman así como las categorías gramaticales de éstas. Se obtiene un conjunto de reglas que, junto a las excepciones encontradas, permitan el reconocimiento y posterior generación automática de palabras compuestas.

3.1 Obtención de la base de estudio

Se parte de un conjunto de tamaño suficientemente significativo de compuestos, alrededor de 11000, y se clasifican en grupos según la categoría gramatical de sus constituyentes. Para ello, se toma como base documental de partida los compuestos recogidos en los principales repertorios lexicográficos de la lengua española (Bibliograf, 2003; RAE, 2001; Larousse, 1996; Clave, 1997; Moliner, 1996) y del glosario de compuestos del libro *La composición nominal en español* de E. Bustos (Bustos, 1986), donde para su elaboración se utilizaron obras de carácter general —DRAE— y obras de carácter regional o dialectal —hablas leonesas, aragonesas, meridionales, español de América.

3.2 Obtención de las reglas de formación

A partir de los conjuntos obtenidos se buscan las reglas de formación de los mismos. Se parte de la palabra y se prueba que cumpla unas ciertas condiciones —tamaño de la palabra, por norma general, mayor a cinco caracteres, o formado por al menos tres sílabas⁸— *uñalbo*. El proceso de reconocimiento empieza haciendo cortes a la palabra hasta que se encuentra algún vocablo íntegro o bien ambos, a los cuales se les aplica la regla correspondiente o bien se trata como excepción, utilizando para ello el “*Flexionador y Lematizador de palabras del español*” del Grupo de Estructura de Datos y Lingüística Computacional (Santana et al., 1997, 1999, 2006). Al ir generando cortes se pueden obtener múltiples soluciones entre las que se pueden encontrar varias que no lo sean, pero esta primera aproximación permite añadir otro tipo de condicionantes⁹.

⁸ Generalmente los vocablos en español son bisilábicos o trisilábicos, por lo que se puede decir que los compuestos nominales contienen de cinco a seis sílabas.

⁹ Las palabras que forman el compuesto no son derivadas sino forman una única unidad léxica, es decir, no tienen prefijos, en especial el primer constituyen del

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, **altobajo* —falsa composición pues lo correcto sería *altibajo*— 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.

En la Figura 1 muestra el funcionamiento del procedimiento de reconocimiento de compuestos, donde una vez superada las condiciones de partida, como no contener ningún error ortográfico, se lematiza la palabra identificando sus diversas formas canónicas, categorías gramaticales y las flexiones o derivaciones que las produce. Por lo general, si se tratara de una palabra compuesta, las diversas formas canónicas 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. Se analiza si la palabra es compuesta “pura”, derivada —en cuyo caso se estaría hablando de derivación por composición, pasándose a reconocer la forma canónica de la misma—, o pudiera ser parasintética por composición (Serrano, 1995). A continuación, se aplican las reglas de formación de compuestos si no pertenece a la base de estudio, se comprueba cuáles son buenas, ordenados por criterios de idoneidad, en caso contrario, no existe solución en el reconocimiento. Por todo lo anterior, 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 mayoritario observado. Además, se debe tener en cuenta que cuando se aplican las reglas hay que considerar los cambios gráficos que se pueden producir como consecuencia de aplicarla, siendo esto necesario para un correcto tratamiento informático.

3.3 Reglas de formación

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

compuesto, o bien no se admite la flexión de diminutivo en la segunda palabra del compuesto, entre otras.

originales: o bien por la mera adición de dos o más términos sin que ninguno de ellos se modifique —*rompeolas, abrelatas, mediodía*— o 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 de adición al resultado final —*agridulce, rojiblanco, coliflor, balompié*.

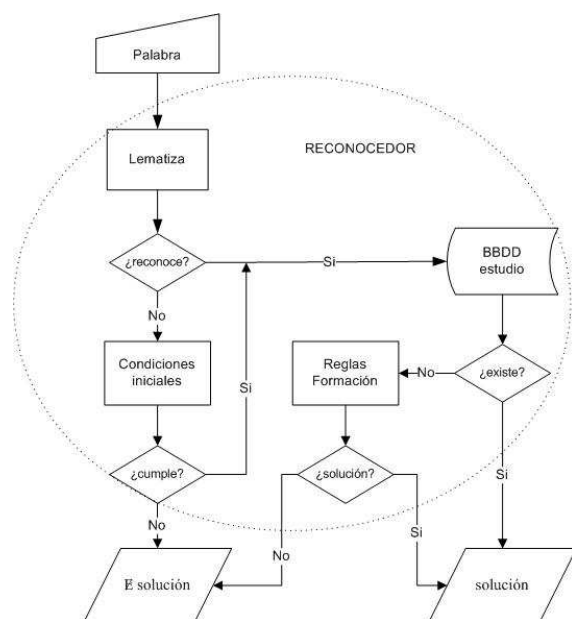


Figura 1.-Diagrama para el reconocimiento de palabras compuestas

Se presentan las reglas de formación de palabras compuestas detectadas 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. La más productiva de la lengua española es la composición nominal de *Verbo + Complemento*, coincide en que ésta es, precisamente, la regla que más se utiliza para este tipo de compuesto —*dedodedama, bienintencionado, iberoamericano, tragaavemarias*.
- ❖ Cuando las palabras a unir 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—*claroscuro, uñalbo, caridelantado, ajolio, ceracate, aguardiente*.

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

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

- ❖ Sustitución de las vocales átonas finales *a, e* y *o* correspondientes al primer constituyente por el infijo compositivo *i*, o en el caso de *Verbo+Complemento* sustituir 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 llevar a cabo la unión sin pérdida. Generalmente, en la composición *Adjetivo+Adjetivo* se utiliza con adjetivos de color. Por la construcción del tiempo verbal en los compuestos *Verbo+Verbo*, la mayoría de los casos son con la vocal átona *e*. Sin embargo, son más los que hay que sustituir el morfema *o* en los compuestos de tipo *Sustantivo+Adjetivo* o *Adjetivo+Sustantivo*. Debido a la estructura peculiar del compuesto [*Sustantivo+Adjetivo*]_{Adjetivo} hay que tener en cuenta que existen palabras que aplicarían la regla anterior si no es porque previamente han incorporado el infijo compositivo *i* —*altibajo, cortiancho, rojinegro, dulciagrío, ardiviejas, batiaguas*.
- ❖ Si el primer elemento acaba en consonante y el segundo empieza en consonante, introducir la *i* en medio y llevar 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* —*azuliblanco, calicanto, paniqueso*.
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 ellos no se atisba claramente ninguna razón de tipo semántico o fonológico.
- ❖ Se elimina la última vocal fuerte del primer constituyente, si acaba en vocal, o bien cuando es final vocálico el primer

componente y principio vocálico el segundo y tienen la posibilidad de formación de un diptongo o de una contracción, se procede a la unión con pérdida de la vocal del primer elemento —*eurasiático, papalbo, liquidámbar*..

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

agridulce ← *agrio* + *dulce* (E *agro* + *dulce*)

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

manvacío ← *manivacío* ← *mano* + *vacío*

- ❖ Los elementos compositivos pueden estar formados por la agregación de raíces cultas greco-latinas a una palabra española, antepuesta o pospuesta a la misma —*cornialto, denticonejudo, petrolífero, carnívoro*— o por la combinación de raíces cultas: raíz prefija y sufija griegas o latinas —*teléfono, filiforme*—, raíz prefija griega y sufija latina —*automóvil*— o raíz prefija latina y sufija griega —*hispanofilia*—, o por otro tipo de raíces —*arisblanco, euroasiático, galicursi*.

4 Conclusiones

A pesar de ser la composición, sin duda, uno de los procesos de formación de palabras de mayor importancia —pese al escaso tratamiento recibido por parte de la bibliografía— se ha contribuido a llenar en parte el vacío informático existente en su tratamiento, ya que han resultado infructuosas las búsquedas de referencias sobre procesamiento automático de la composición en español.

Uno de los procesos de formación de palabras compuestas más productivos, con respecto a los demás tipos compositivos, son las formaciones de *Verbo*+*Sustantivo* debido a la gran expresividad, simplicidad —el grado de modificación que sufren los elementos originales, en la mayoría de los casos, es nulo— y al frecuente uso que se hace en el lenguaje publicitario de este sistema compositivo. El auge en el uso de nuevos medios de comunicación social, el lenguaje periodístico, entre otros factores, han hecho que se creen neologismos compositivos debido a la rápida evolución de los acontecimientos y a su inmediata incorporación al mundo de las Tecnologías de la Información. Por ello, son imprescindibles procesos automáticos que sean capaces de identificar estas palabras situándolas en un contexto lingüístico adecuado, tanto

desde un punto de vista morfológico como semántico.

Bibliografía

- Alvar Ezquerro, M. 2002. “La formación de las palabras en español”. Cuadernos de lengua española, Ed. Arco/Libros, Madrid.
- Bibliograf, S.A. 2003. “Diccionario General de la Lengua Española VOX” Barcelona.
- Bustos Gisbert, E. 1986. “La composición nominal en español”, Universidad de Salamanca.
- Clave SM. 1997. “Diccionario de Uso del Español Actual”. Clave SM, Madrid.
- Lang, M. 1992. “Formación de palabras en español. Morfología derivativa productiva en léxico moderno”. Cátedra, Madrid.
- Larousse Planeta, S.A. 1996. “Gran Diccionario de la Lengua Española”. Larousse Planeta, S.A., Barcelona
- María Moliner. 1996. “Diccionario de Uso del Español”, Ed. Gredos, Madrid.
- Pérez Cino, W. 2002. “Manual Práctico de formación de palabras en español I”, Ed. Verbum.
- Real Academia Española y Espasa-Calpe. 2001. “Diccionario de la Lengua Española”, edición electrónica. 22ª edn. Madrid.
- Santana, O.; Carreras, F.; Pérez, J.; Gutiérrez, V. 2006. “El reconocimiento automático de la composición en español”. Conference Abstracts of the First International Conference of the Alliance of Digital Humanities Organizations.
- Santana, O.; Pérez, J.; Carreras, F.; Duque, J.; Hernández, Z.; Rodríguez, G. 1999. “FLANOM: Flexionador y lematizador automático de formas nominales”. *Lingüística Española Actual* XXI, 2, Ed. Arco/Libros, S.L.
- Santana, O.; Pérez, J.; Hernández, Z.; Carreras, F.; Rodríguez, G. 1997. “FLAVER: Flexionador y lematizador automático de formas verbales”. *Lingüística Española Actual* XIX, 2, Ed. Arco/Libros, S.L.
- Serrano Dolader, D. 1995. “Las formaciones parasintéticas en español”, Ed. Arco/Libros.