

# EL RECONOCIMIENTO AUTOMÁTICO DE LA COMPOSICIÓN EN ESPAÑOL.

## Autores

- Octavio Santana Suárez. Departamento de Informática y Sistemas. Universidad de Las Palmas de Gran Canaria. osatana@dis.ulpgc.es
- Francisco Javier Carreras Riudavets. Departamento de Informática y Sistemas. Universidad de Las Palmas de Gran Canaria. fcarreras@dis.ulpgc.es
- José Rafael Pérez Aguiar. Departamento de Informática y Sistemas. Universidad de Las Palmas de Gran Canaria. jperez@dis.ulpgc.es
- Virginia Gutiérrez Rodríguez. Dpt. Estadística, Investigación Operativa y Computación. Universidad de La Laguna. vgutier@ull.es

## RESUMEN

It deals with computerizing one of the processes of words formation in Spanish: the composition. They will solely be studied those cases in which the compound word has been consolidated like the graphical union of the elements that compose it, in regular or irregular way. The formation rules and the application criteria in each case are deduced, consequently, they allow the automated identification of the compound words. The different compounds are extracted from several lexical sources and the applied mechanisms of recognition will be studied, likewise the grammatical categories of original words and the resultant compound. The found recognition criteria are classified and the detected exceptions and irregularities are considered.

## INTRODUCCIÓN

La creatividad léxica, según Mervin Lang (Lang, 1997), representa una característica fundamental para el habla y la escritura. Los escritores siempre han ideado sus palabras para librarse de las restricciones que les vienen impuestas por el léxico establecido, por lo que utilizan la derivación y la composición como recursos léxicos. Los ejemplos referentes a la formación de palabras se encuentran también en los neologismos de la terminología científica, en la tecnológica, en el comercio, en los medios de comunicación, en el lenguaje creativo de la literatura moderna y en el lenguaje coloquial e innovador del habla actual. Este trabajo centra su estudio en la yuxtaposición y se excluyen otros por no haberse consolidado como palabra el compuesto resultante —se destaca la importancia de la composición constituida por un elemento verbal y un complemento por ser el más caudaloso de los tipos de composición. Se trata, en suma, de procedimientos para crear neologismos —constituyen una alternativa moderna que enriquece la lengua.

## LA COMPOSICIÓN EN ESPAÑOL

La Real Academia de la Lengua Española define la composición como el “procedimiento por el cual se forman palabras juntando dos vocablos con variación morfológica o sin ella —cejijunto, lavavajillas. Se aplica también a las voces formadas con vocablos de otras lenguas, especialmente del latín y el griego —neuralgia, videoconferencia” (RAE, 2001). La composición se sirve de procedimientos para la creación de nuevas palabras, como son: sinapsia, disyunción, contraposición, yuxtaposición, elementos compositivos y acortamiento.

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 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. 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; algunas de tales composiciones pueden llegar a la unión gráfica de sus elementos: caballo de mar→caballo marino, tela de araña→telaraña, agua nieve→aguanieve,... En un grado más alto de unión gráfica está la contraposición, donde los elementos que participan se escriben unidos por un guión que, generalmente, no aparecerá debido a las restricciones del español —coche-bomba→coche bomba, falda-pantalón→falda pantalón—, aunque la Real Academia Española establece que “cuando no hay fusión sino oposición o contraste entre los elementos componentes, se unirán estos con guión” (RAE, 1995) —físico-químico. El más generoso 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, cochitril, hincapié. Aunque la frontera entre derivación y composición no resulta muy clara, sobre todo en el caso del abreviamento —coyotomate— o la acronimia —información automática→informática, poliestar galo→tergal—, muchos autores consideran el acortamiento como un procedimiento de formación de neologismos que por su naturaleza no constituiría una derivación sino que más bien formaría parte de la composición. La utilización de raíces cultas greco-latinas es frecuente en los procesos de generación de nuevas palabras —particularmente en los campos científicos y técnicos—; las voces en cuya formación intervienen podrían, según varios autores, no considerarse propiamente compuestas, ya que la mayoría de sus raíces no pueden aparecer aisladamente, pero tampoco pueden considerarse derivadas, puesto que tienen un comportamiento peculiar —significado léxico— que los aleja de los auténticos afijos. A este tipo de raíces se les da el nombre de elementos prefijales o sufijales —elementos compositivos—, en función de si se anteponen a otra raíz o se posponen.

En el presente estudio se tratan, desde un punto de vista morfológico, los compuestos yuxtapuestos o lexías compuestas, al igual que algunos casos especiales de acortamiento, elementos compositivos 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 inicialmente no se consideran en este trabajo.

## REGLAS DE COMPOSICIÓN

Se parte de una base de unos 4000 compuestos recopilados del Diccionario General de la Lengua Española Vox (Bibliograf, 2003) y del glosario de compuestos del libro “La composición nominal en español” de Eugenio Bustos (Bustos, 1986) —basado en obras de carácter general, DRAE, y en otras de carácter regional o dialectal: hablas leonesas, aragonesas, meridionales, español de América—, además, se han añadido unos 6000 compuestos, que incorporan elementos prefijales, procedentes de diversos diccionarios de español (Clave, 1997; Espasa Calpe, 1991; Casares, 1990; Larousse, 1996; Alvar, 2003; Moliner, 1996; ). Los compuestos analizados se clasifican en grupos según la categoría gramatical de sus constituyentes.

Se busca, a partir del estudio del comportamiento de los vocablos constituyentes del compuesto, las reglas de formación del mismo; algunas coinciden con las tratadas por algunos lingüistas, aunque con una adaptación informática justificada por el comportamiento mayoritario observado —aeriforme → aeri- + -forme, según el

Diccionario General de la Lengua Española Vox (Bibliograf, 2003), sin embargo, el comportamiento mayoritario es aero-. Se define regla de formación a todo comportamiento mayoritario que permita concretar un mecanismo capaz de relacionar los elementos constituyentes del compuesto, para su reconocimiento por medios informáticos. Se estudian además, las reglas fonéticas —cambios gráficos para mantener el sonido de una consonante: anquirredondo → anca + redondo— que se producen como consecuencia de haber aplicado una regla de formación. Se obtiene un conjunto de reglas que, junto a las excepciones encontradas, permiten el reconocimiento automático de las palabras compuestas y en el futuro su generación.

## PROCESAMIENTO DE LAS REGLAS DE COMPOSICIÓN

Se parte de la palabra compuesta y se comprueba que cumpla unas ciertas condiciones —tamaño de la palabra, mayor a cinco caracteres, o bien, número de sílabas, mayor que tres: ñalbo. El proceso de reconocimiento propone partir la palabra hasta que se encuentre un vocablo o ambos, a los que se les aplica la regla correspondiente o bien se tratan como excepción; se pueden obtener múltiples soluciones —algunas o todas incorrectas. La secuencia de cortes permite añadir otro tipo de condicionantes: por ejemplo, las palabras que forman el compuesto no deben ser derivadas, sino constituir una unidad léxica —no contener prefijos, particularmente en el primer elemento del compuesto—, o no admitir la flexión del diminutivo en la segunda palabra del compuesto.

Hay que tener en cuenta que en un estudio cuyo objetivo sea 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, pues 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.

## CONCLUSIONES

Se trata de un trabajo novedoso, ya que han resultado infructuosas las búsquedas de referencias sobre procesamiento automático de la composición en español, a pesar de la presumible trascendencia de tal proceso.

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: movichandal, ciberamor, eurosuelo,... Son imprescindibles procesos automáticos que sean capaces de identificar estas palabras y situarlas en un contexto lingüístico adecuado: morfológico y semántico.

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, 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

## REFERENCIAS

1. Bibliograf, s.a. 2003. “Diccionario General de la Lengua Española VOX” en CD-ROM. Barcelona.
2. Clave SM. 1997. “Diccionario de Uso del Español Actual”. Clave SM, edición en CD ROM. Madrid.

3. David Serrano Dolader. 1995. "Las formaciones parasintéticas en español", Ed. Arco/Libros, S.L.
4. Espasa Calpe. 1991. "Gran Diccionario de Sinónimos y Antónimos", 4ª edic. Espasa Calpe, Madrid.
5. Eugenio Bustos Gisbert. 1986. "La composición nominal en español", Universidad de Salamanca.
6. Jose Alberto Miranda. 1994. "La formación de palabras en español". Ediciones Colegio de España.
7. Julio Casares. 1990. "Diccionario Ideológico de la Lengua Española", 2ª Edición. Ed. Gustavo Gili, s.a. Barcelona.
8. Larousse Planeta, s.a. 1996. "Gran Diccionario de la Lengua Española". Larousse Planeta, s.a., Barcelona.
9. Manuel Alvar Ezquerra. 2002. "La formación de las palabras en español". Cuadernos de lengua española, Ed. Arco/Libros, Madrid.
10. Manuel Alvar Ezquerra. 2003. "Nuevo diccionario de voces de uso actual". Ed. Arco/Libros, Madrid.
11. María Moliner. 1996. "Diccionario de Uso del Español", edición en CD ROM. Gredos, Madrid.
12. Mervyn Francis Lang. 1992. "Formación de palabras en español. Morfología derivativa productiva en léxico moderno". Cátedra, Madrid.
13. Ramón Almela Pérez. 1999. "Procedimientos de formación de palabras en español". Ed. Ariel Practicum.
14. Real Academia Española y Espasa-Calpe. 2001. "Diccionario de la Lengua Española", edición electrónica. 22ª edn. Madrid.
15. Soledad Varela Ortega. 1990. "Fundamentos de Morfología", Ed. Síntesis.
16. Waldo Pérez Cino. 2002. "Manual Práctico de formación de palabras en español I", ed. Verbum.