



El uso de los compuestos clásicos en bioinformática

Eva San Evaristo Pascual
C.A. "F. Tomás y Valiente"
U.N.E.D.
Valencia, España
esanevaristo@valencia.uned.es

Keywords: Bioinformatics, word formation, neo-classical compound words, terminology.

Abstract

Greek and Latin have been used to coin new terms in Science, especially since the Renaissance. However, many experts think that in the last sixty years their use in Spanish is moving back in benefit of English both for the specialised communication and the information exchange. The present paper sets out a short theoretical approach to Spanish word-formation focused on the neo-classical compounds because most eminent linguists do not agree about their status. Secondly, we will try to analyze the quantitative and qualitative use of neo-classical compound words in the Spanish discourse of a new discipline: Bioinformatics, as well as to confirm or not their decreasing occurrence in new fields of study with less classical tradition than other ones, such as Medicine or Botany.

1 Introducción

De todos es sabido que desde la Edad Media, pero especialmente desde el Renacimiento, el griego y el latín han servido de cantera para la acuñación de nuevos términos primero en el ámbito científico (del que destacamos la medicina o la botánica) y luego en el técnico por ser en su momento las lenguas de la cultura y del saber. Sin embargo, muchos estudiosos coinciden en afirmar que desde la segunda mitad del siglo XX su presencia en la lengua española está retrocediendo en beneficio del inglés en casi todos los campos, pero especialmente en el de la informática (Martín Zorraquino, 1997:324). No obstante, ello no significa que muchos autores no pongan de relieve la importancia de las lenguas clásicas como pilares de la creación terminológica (Val Álvaro, 1999:4777; Vivanco, 2006:21; Cabré & Rigau, 1984:154, López Piñero & Terrada, 2005:viii-ix) y el hecho, también significativo, de que muchos de estos términos especializados, traspasando las fronteras de su "original especialización" (Lang, 1997:240) se integren en la lengua general y sean asimilados por los hablantes.

1.1 Objetivos de nuestro estudio

El presente trabajo pretende analizar cualitativa y cuantitativamente los compuestos clásicos en la bioinformática; comprobar si el griego y el latín, a través de los compuestos, siguen ocupando el puesto relevante de antaño en el discurso científico moderno y así, confirmar o rebatir su disminución en la ciencia y la técnica contemporáneas. Se ha tomado como campo de estudio esta joven disciplina porque es una ciencia emergente en España y porque sus bases se apoyan tanto en la ciencia como en la técnica. También abordaremos, aunque



brevemente, las posturas acerca del tipo de formación que los compuestos clásicos constituyen al existir una cierta disparidad de opiniones entre los lingüistas.

1.2 Aporte teórico

1.2.1 La Bioinformática

Aunque el término “bioinformática” induce a pensar en una disciplina que pone el acento sobre la vertiente informática, en cambio si observamos su sinónimo *biología computacional*, calco de *computational biologics*, el peso se desplaza hacia el lado de la biología, en cuyo seno nace. Toma auge a partir de la publicación de la estructura del ADN (Watson & Crick, 1953), si bien podríamos hacer remontar sus orígenes incluso a las primeras décadas del siglo XX con la creación de la base de datos con la primera información relativa al ADN y a las proteínas (Margaret O. Dayhoff, 1947). Ha alcanzado un hito importantísimo al publicarse la secuencia completa del genoma humano (febrero de 2001) dando paso a una nueva etapa, denominada por los expertos “era post-genómica” en la que indudablemente quedan todavía muchos logros que alcanzar.

Existe, pues, desde su nacimiento una estrecha vinculación a la biología molecular y a la genética; sin embargo, también está presente en otros ámbitos de la biología, como la vegetal, orientada a la agricultura o la conservación de especies amenazadas. No obstante, la perspectiva más extendida acota su campo de acción al estudio de la genética humana (genoma, proteínas). *Sensu lato*, la bioinformática es una disciplina que combina la biología, la informática, las matemáticas, la estadística y la teoría de la información para almacenar, gestionar, interpretar e integrar el ingente volumen de información¹ que la biología molecular, la genómica (rama de la biología que estudia el genoma humano) o la proteómica (centrada en el estudio y predicción del plegamiento o estructura tridimensional de las proteínas) producen. Para tal fin desarrolla aplicaciones informáticas y métodos de análisis estadístico. Se organiza en varias áreas, de las que destacamos el trabajo de los científicos del Proyecto Genoma Humano. En este proyecto multinacional ha sido fundamental el desarrollo de aplicaciones para el manejo de bases de datos, el estudio de las secuencias del ADN y de las proteínas o el alineamiento múltiple de sus secuencias para la reconstrucción de árboles filogenéticos (es decir, el parentesco de las especies, establecido a partir de las semejanzas y diferencias entre sus secuencias de proteínas o ácidos nucleicos).

Asimismo tiene una aplicación directa en la biomedicina (para algunos, una rama diversa de la bioinformática) con el fin de estudiar qué genes están ligados a ciertas enfermedades, hereditarias o no, y así conocer mejor sus bases moleculares, realizar un diagnóstico precoz o desarrollar mejores tratamientos para los pacientes. Son precisamente estas aplicaciones las que han atraído el interés de numerosas empresas (*bioempresas*), tanto farmacéuticas como informáticas, cuyo impacto económico es destacado y creciente en países como Estados Unidos, Canadá y Japón, que, junto con Rusia, son los que más invierten en bioinformática y mayor impulso le están dando gracias a la creación de centros reales y virtuales donde se llevan a cabo investigaciones muy avanzadas.

1.2.2 La composición culta: estado de la cuestión

De entre los diversos mecanismos de formación de palabras (en adelante FP), creemos que la composición culta es el más problemático. Los manuales y monografías consultados tienen en común el hecho de reconocer velada o abiertamente que la composición con bases

¹ Sirva de ejemplo el proporcionado por Guigó & Valencia (2000:8): un solo experimento con arrays (matrices) de expresión genera 60Mb de información y se llevan a cabo cientos o miles repetidamente en todo el mundo.



grecolatinas es difícilmente clasificable en un mecanismo de FP concreto, no pudiéndosele atribuir los mismos rasgos que a las formaciones con afijos y/o con bases patrimoniales. Así, Lang (1997:221) declara que “el estatuto de estas formas resulta problemático”; Pensado (1999) y Val Álvaro (1999) reconocen que no se pueden emplear los mismos criterios para la formación nativa y para la culta de modo que habrían de explicarse separadamente. Es más, ciertos autores incurren en contradicciones en sus propias obras, lo que demuestra la dificultad de esta formación. Y sin embargo, se le dedica muy poca atención en las gramáticas generales y en las monografías de FP. Salvo el capítulo de Val Álvaro en la *Gramática descriptiva de la Lengua Española* (1999) dedicado a los compuestos, no hemos encontrado un estudio pormenorizado ni siquiera en las anteriores gramáticas de la Real Academia Española (RAE): En el *Esbozo de una nueva gramática de la lengua española* (1973) se encuentran someras referencias a la composición y la derivación dispersas en varios capítulos, como también sucede en la *Gramática de la lengua española* de Alarcos (1994).

No existe tampoco un acuerdo acerca del tipo de formación a que dan lugar ni una denominación común a los temas que intervienen: unos las consideran raíces (pero con funcionamiento diferente al de las raíces nativas) y dan lugar a “compuestos clásicos” o “neoclásicos”, “compuestos cultos”, “compuestos eruditos o científicos” (*Esbozo*, 1973:77-78) etc. Si se consideran más próximas a los prefijos o sufijos (teniendo presente que tampoco comparten los mismos rasgos que los afijos patrimoniales), encontraremos un largo listado de etiquetas absolutamente insuficientes: “raíces prefijas y raíces sufijas” (Marcos Marín, 1980:430 citando a Seco; García et al., 2004:275), “elementos cultos”, “afijoides-prefijoides” y “seudoafijos, seudoprefijos, seudosufijos (Alvar Ezquerro, 1993:49), “pseudoprefijos latinos o griegos” (Serrano Dolader, 1995:202), “temas”, “raíces ligadas” o “bases no autónomas” (Val Álvaro, 1999:4775-6) o la amplia gama de Lang (1997:55, 97 y 269): “afijos cultos internacionales”, “lexemas latinos o griegos que, sincrónicamente funcionan como prefijos o sufijos”, “afijos neoclásicos constituidos por los sufijoides y prefijoides”.

La cuestión fundamental es, pues, ¿en cuál de los procesos de FP podríamos englobar lo que estamos designando “compuestos clásicos”? La bibliografía oscila entre dos polos: composición y prefijación (en ocasiones mediante un paso intermedio de transformación de tema a prefijo) aunque nadie se postula taxativamente a favor de un mecanismo u otro, pues, como expone Lang (1997) o Serrano Dolader (1995), dada la proximidad de los procesos de composición y de prefijación, no es de extrañar que haya quien no establezca fronteras entre ambos mecanismos.

Predomina la consideración de composición en estos autores y obras: el *Esbozo* (1973) distingue compuestos eruditos o científicos y compuestos con prefijo o prefijación, en la que incluye prefijos de origen grecolatino llamados “preposiciones impropias”, entre las que hay palabras de varias categorías gramaticales con “poder reproductivo en la lengua española” (1973:76) como *auto-*, *bi-*, *multi-*, *semi-*, *tele-*. En cambio, se enuncian entre las formadoras de compuestos palabras de igual categoría (p. ej. *mono-*, *omni-*, *seudo-* *uni-*). Por lo tanto, parece que el hecho determinante para que una palabra de origen grecolatino dé lugar a un compuesto erudito o a un prefijado dependerá de la productividad o no de sus formantes. Marcos Marín (1980:430) recoge la opinión de M. Seco (*Gramática esencial del español*, 1972) del que dice trata las “raíces prefijas” y “raíces sufijas”, que son “adaptaciones de algunas palabras que toman una forma especial al unirse a otras” y que no son otra cosa que palabras grecolatinas (se sabe por los ejemplos) de las que unas se han especializado en aparecer antepuestas (p. ej. *aero-*) y otras pospuestas (como *-algia*). García et al. (2004:312),



por un lado, afirman que “hay también otra modalidad <de composición> consistente en tomar vocablos (especialmente del griego o del latín) y emparejarlos como partes de compuestos: *cefalgia*, *endogamia*, *omnívoro* [...]” que van siempre “ligadas a otra de su especie”; pero por otro, señalan que mediante la parasíntesis (un subtipo de derivación que se da cuando actúan simultáneamente un prefijo y un sufijo, condición compartida por Serrano Dolader, 1995:8) se constituyen “ciertas” palabras de origen grecolatino (no se dicen cuáles ni por qué) que en español no pueden aparecer de forma independiente y proponen el ejemplo de *hemeroteca*. Pero a nuestro parecer ¿qué diferencia hay entre *cefalgia* y *hemeroteca*? Justifican, pues, dos tipos de formaciones distintas con ejemplos que son, para nosotros, iguales. No obstante, su opción final es la de considerarlos compuestos ya que concluyen diciendo: “Este tipo de compuestos suele darse en nomenclaturas científico-técnicas” (García et al. 2004:312).

Los siguientes autores se muestran en cambio más proclives a la prefijación y a la derivación (entendiendo por tal la sufijación): Lang (1997:97) reconoce que palabras como *agorafobia*, *dactilógrafo*, *piscifactoría* se forman por prefijación; otras, por sufijación, pues son “lexemas latinos o griegos que sincrónicamente funcionan como prefijos o sufijos” (*ibidem*). Es en posición inicial, en su opinión (sin aducir datos concretos o estudios de frecuencia), la de mayor productividad en español (*cf.* apdo. 3.2). Para Alvar Ezquerro (1993) la formación mediante prefijos y sufijos cultos es derivación ya que es condición de la composición que sus elementos integrantes aparezcan libres (opinión unánime entre los autores consultados) y la mayoría de los temas cultos no lo son. Sin embargo, cuando se unen dos “afijoides”, Alvar Ezquerro (1993:50) habla de dos “elementos” cultos cuya resultante está muy próxima a la composición; pero no se atreve a incluirlos en dicha formación. En último lugar, la nueva gramática de la RAE (*Gramática descriptiva de la Lengua Española*, 1999), al no estar redactada por un mismo autor, ofrece diversas consideraciones aunque en líneas generales se tiende a hablar en términos parecidos. De todos los capítulos, el de Val Álvaro aborda en profundidad la composición culta y, adoptando el estudio de Marchand aplicado al inglés (*The Categories and Types of Present-Day English Word-Formation*, 1969:132 y ss) distingue la composición con base neolatina en la que ambos integrantes del compuesto pueden ser griegos o latinos y la prefijación con base nativa con un primer elemento grecolatino que ha de ser ligado (los llamados “compuestos híbridos”). Así pues, un mismo elemento podrá formar parte de un compuesto o de un prefijado: todo dependería del segundo formante. Para Varela & Martín (1999) no es admisible hablar de “prefijos cultos”, “prefijos grecolatinos” etc. y descartan la consideración de las raíces clásicas como una forma de prefijación “atípica” por tres criterios:

- a) Distributivo: porque la mayoría de ellos pueden aparecer en primer lugar (*filólogo*), como los prefijos, pero también en segundo (*bibliófilo*), algo vedado a los prefijos y porque dos raíces clásicas se pueden combinar entre sí, mientras un prefijo más un sufijo, nunca. Comparten este criterio Val Álvaro (1999), Lang (1997) y Adams (1973).
- b) Formal: la variación de algunos temas depende de su posición (inicial o final); en cambio, los alomorfos de un prefijo nada tienen que ver con su posición (siempre inicial).
- c) Funcional: los auténticos prefijos no son núcleo de la palabra sino que son “adjuntos que modifican el significado de la palabra completa de manera 'circunstancial’” (Varela, 1999:4998), no así los temas clásicos.



Varela & Martín incluyen entre los prefijos españoles los de origen culto, así como algunos adjetivos/pronombres (*auto-*, *neo-*, *mega-*, *hetero-*), cuantificadores (*multi-*, *pluri-*) o adverbios (*endo-*) que funcionalmente caracterizan semánticamente a la base léxica del mismo modo que los prefijos, por lo que creemos entender que consideran compuestos cultos aquellos en que intervengan temas nominales, verbales o calificativos de uso menos extenso que los citados, salvo que sufran un acortamiento (*vid. infra*). Rainer (1999) considera sufijos de origen grecolatino toda una serie de raíces verbales (*-ícola*, *-ífico*, *-ífero*, *-ígero*, *-oide*) y designa como "otros" los que no son fáciles de enmarcar dentro de las etiquetas morfológicas establecidas, pues algunos tienen un uso lexicalizado y otros son de uso marginal (*-icida*, *-íloquo*, *-ívoros*). En cambio, todos ellos son para Val Álvaro (1999) temas grecolatinos formadores de compuestos adjetivos. Por último, Serrano Dolader (1999) considera que las antiguas raíces clásicas ya no constituyen palabras existentes, en consecuencia funcionan actualmente como prefijos y sufijos. Sin embargo, con anterioridad (1995:202) declaraba que las formaciones con "semipalabras o pseudoafijos" (entre la que se hallan algunas que originalmente no lo eran: *tele-*, *foto-*, *hidro-*), por tener un carácter "de evidentes cultismos", no habían sido tenidas en cuenta en su estudio sobre la parasíntesis puesto que la diferencia entre composición y prefijación en formaciones cultas no está bien "resuelta".

Si analizamos los autores de formación o *impostazione* clásica – Echeverri (1979), Quintana Cabanas (1989), Martínez Hernández (1992), López Piñero & Terrada (2005), Estébanez (1998) y Pingarrón (1998) – observamos: a) que no tratan los mecanismos de FP, salvo Martínez Hernández y López Piñero & Terrada; simplemente se limitan a definir (cuando lo hacen) y a listar cuáles son los prefijos y sufijos. b) Que cuantificadores (*multi*, *pluri*, *oligo*...), calificativos que indican tamaño (*mega*, *micro*), indefinidos (*hetero*, *homo*, *omni*, *pan(t)* etc.) no son prefijos, de manera que cuando van unidos a otro/s tema/s constituirán compuestos.

No queremos concluir este somero aporte teórico sin comentar otros tres procesos: Primero, los híbridos o hibridismos: Aunque hemos hecho mención a esta formación considerada ya sea composición ya sea prefijación, muy pocos (Val Álvaro, 1999) usan este término. La composición que mezcla griego y latín es apenas tratada por las gramáticas y monografías de la lengua española, pero sí por dos autores "periféricos": Echeverri (1988:8), quien considera "hibridismos" la combinación de una raíz griega y otra latina o bien de una griega y otra española (nativa)² y para el inglés, Adams (1973:129), que deja bien clara esta posibilidad de mezcla de lenguas clásicas entre sí o con otra nativa sin limitar o determinar la posición del elemento clásico, como hacen Val Álvaro y otros, al primer lugar. Segundo, los procesos de acortamiento, cuyo estudio ha ido progresivamente en aumento: desde la única alusión del *Esbozo* (1973:178) al hablar de los sustantivos femeninos en -o como *moto(cicleta)* hasta ser considerado por sí solo un procedimiento independiente de FP (García et al., 2004:272). Pero ¿hasta qué punto constituyen palabras distintas respecto a la forma completa? No creemos que *kilo*, *tele* u *otorrino* respecto a *kilogramo*, *televisión* u *otorrinolaringólogo* sean una nueva palabra, como propugnan García et al. En nuestra opinión no son más que compuestos clásicos que en determinados contextos sociolingüísticos (no siempre coloquiales), al igual que sucede con palabras patrimoniales (*Toni*, *chuches* etc.), se acortan mediante determinados fenómenos morfofonéticos. Así opinan también Cabré & Rigau (1984) para el catalán, que es perfectamente aplicable a nuestro caso. Lo que sí es importante es el proceso que sigue a muchos de los acortamientos, bien sea de un compuesto (con el significado global de éste y no con el del propio tema o raíz), de un derivado o de una palabra simple: la creación de

2 Dado que es un diccionario de helenismos, no contempla los hibridismos con el latín.



nuevos compuestos. Ahora bien, cabría preguntarse si realmente dichos compuestos se han formado a partir de un acortamiento o si se han acuñado directamente sobre el tema clásico. Dando por válida la primera hipótesis, los temas clásicos se convertirían en prefijos en opinión de Lang (1997:112) y Pensado (1999:4458). El tercer procedimiento, siglas y acrónimos en la ciencia y la técnica, ha sido poco tratado por nuestros autores cuando son precisamente de amplio uso, especialmente en el ámbito de la informática, como bien destaca Adams (1973:136-138).

2 Materiales y método

2.1 Materiales

Para llevar a cabo la parte aplicada de este estudio, se ha creado un corpus cuyas fuentes debían responder a una serie de criterios previamente sopesados: En primer lugar, documentos escritos (los *corpora* orales no creemos sean útiles a nuestros propósitos), originales, no traducidos. En segundo lugar, en formato electrónico ya que agilizaba su tratamiento mediante una aplicación informática específica. En tercer lugar, dada la diversidad de áreas que abarca la bioinformática, consideramos que el *corpus* sería más homogéneo si lo acotábamos a un único campo de estudio, decantándonos por el estudio del genoma humano y su aplicación a la biomedicina porque sus retos nos parecen los más interesantes. Por último, pretendíamos un equilibrio entre textos escritos por especialistas en informática y en biomedicina tanto por su número como por su volumen y fecha de publicación con el fin de comprobar si existían diferencias de frecuencia, formantes etc. en cada subcorpus. Sin embargo, de la lectura de los documentos se desprende que no puede hacerse una distinción neta entre los pertenecientes al ámbito biológico (o biomédico) y al informático ya que ambas disciplinas tienen parecida presencia en todos ellos.

Con tales premisas ha resultado verdaderamente difícil hallar materiales originales con que constituir nuestro corpus, fundamentalmente porque libros y revistas especializadas están escritos mayoritariamente en inglés para facilitar su difusión, por lo que podemos desgraciadamente afirmar que el español no es una lengua de divulgación científico-técnica en el campo de la bioinformática. Sirva de ejemplo el hecho de que mediante el buscador de revistas electrónicas de la Biblioteca de la Universidad Politécnica de Valencia (UPV) no se haya podido extraer ningún documento sobre bioinformática en nuestro idioma o que el buscador especializado “SpringerLink” no permita seleccionar artículos en español.

Finalmente, hemos conseguido constituir un corpus de 99.005 palabras (tokens) cuyos lemas totales (types) ascienden a 8.372 extraídos a partir de catorce documentos en soporte informático, fechados entre 2000 y 2007, relativos al campo de la bioinformática aplicada a la genómica y la biomedicina, pero de distinta naturaleza: textos académicos (documentos nº 10, 11, 12 y 13 reseñados en la sección “Corpus constituido” apdo. Referencias)³; estudios científicos (doc. nº 3), un e-book y un capítulo de un libro (docs. nº 14 y 4 respectivamente), artículos (docs. nº 1, 5, 7), algunos de divulgación (docs. nº 6, 8, 9) y un manual de uso de un programa informático (doc. nº 2). Sobre la posible influencia de esta variedad textual en los resultados finales hablaremos en el apartado 4.

2.2 Método

Antes de usar el software de análisis léxico WordSmith Tools, hemos eliminado del corpus

³ Agradecemos a M. A. Peiró y a J. M^a Silla su permiso para la utilización desinteresada de sus Proyectos Final de Carrera inéditos, así como también a sus directores Dr. D. Juan Císcar, D. Juan M. García Gómez y D. Antonio Gabaldón.

los abstracts en inglés, las direcciones URL, fórmulas, combinaciones de nucleótidos y secuencias génicas. Hemos decidido no tomar en consideración un grupo problemático, pero muy numeroso por razones obvias: los elementos y compuestos químicos puesto que históricamente la terminología de la química es poco deudora del mundo grecorromano, como mucho a través del árabe (por ej. *alcohol*, *colirio*) y la mayoría de los afijos de la química y bioquímica se han extraído del latín y del griego a veces de forma artificial y con un significado específico que no tenían en origen o era muy difícil de relacionar⁴.

Hay que hacer constar que, a pesar de las condiciones de selección, el corpus no está íntegramente en español. Se utiliza bastante el inglés tanto al incluir citas textuales como en la propia terminología. Puesto que superaba el ámbito de este trabajo, hemos descartado el estudio de sus compuestos e híbridos, pero sí los hemos contabilizado a efectos meramente contrastivos (tabla I y apdos. 3.3, 3.4 y 4).

En suma, el objeto de nuestro análisis ha tenido en cuenta estas formaciones: compuestos clásicos en sentido estricto y otras formaciones que en construcciones nativas serían calificadas como parasíntesis o derivación con cualquiera de estas posibilidades combinatorias: [raíz+raíz+sufijo] como *fenotípico*; [prefijo+raíz+raíz], como *enciclopedia* o incluso [prefijo+raíz+raíz+sufijo] aunque sin simultaneidad ni exigencia mutua entre prefijo y sufijo, como *epidemiológico* o *hiperleucocitosis*. Bajo esa misma etiqueta incluiremos los híbridos de todo tipo y los acortamientos, siglas y acrónimos acuñados sobre un compuesto clásico o híbrido.

3. Análisis

3.1 Datos estadísticos

3.1.1 Por frecuencia

Los datos totales, sin discriminar el tipo de información, son los siguientes:

	Español	Inglés
Compuestos clásicos ⁵	264	51
Porcentaje sobre total lemas (types)	3,15%	0,06%
Ocurrencias	1789	534
Porcentaje sobre total palabras (tokens)	1,8%	0,53%
Ocurrencias formas plenas	1754	268
Porcentaje ocurrencias/total ocurrencias	1,77%	0,27%
Compuestos en formas abreviadas (siglas, apócopos)	11	35
Ocurrencias formas abreviadas	35	266
Porcentaje ocurrencias/total ocurrencias	0,03%	0,26%

Tabla I. Número de formaciones y ocurrencias

Salta a la vista que la frecuencia de uso de los compuestos clásicos es mínimo en relación con el total de palabras, lo cual supone una gran decepción aun sumando los compuestos del inglés (total 2,33%). Si restamos del total de palabras las preposiciones, conjunciones

4 Como –amina, unión de *-am*(oníaco), cuya etimología parece derivar de Amón o de la palabra “arena” (ἄμνος) y del sufijo –ina que indica procedencia, pertenencia. O –asa, de alta aparición en nuestro corpus, por ej. en las siglas ADNasa y ARNasa, evolución del sufijo griego –sis a través del francés.

5 Los dobles gráficos o acentuales (*auto-organizativo/autoorganizativo*, *video/vídeo*, *colorectal/colorrectal* etc.) se han considerado como un mismo término.

coordinantes más frecuentes –y, o, pero, sino– y artículos (31.421 oc., 31,74% del total), las cifras no aumentan espectacularmente: los compuestos en español supondrían el 2,64% del total del corpus reducido (2,59% las formas plenas y 0,05% las abreviadas) y un 3,43% añadiendo las formas en inglés. Por número de formaciones, ni siquiera se alcanza el 5% de lemas. Sin embargo, no se ha de perder de vista a la hora de ponderar los datos la relativa presencia de términos químicos que engruesan nuestro corpus y, por lo tanto, hacen disminuir el porcentaje de compuestos clásicos. No obstante, con estos resultados, no podemos defender la frecuencia de uso en la lengua de la bioinformática de los compuestos cultos, lo cual no implica que no abunden las raíces grecolatinas, pues si este trabajo hubiera versado sobre su presencia en la prefijación y la derivación (sufijación), pensamos que los resultados seguramente no serían tan negativos.

Para valorar la frecuencia de uso de los compuestos clásicos, hemos establecido seis rangos:

$f \geq 100$	4	$50 > f \geq 25 >$	6	$10 > f \geq 5$	26
$100 > f \geq 50$	2	$25 > f \geq 10$	20	$f < 5$	206

Tabla II. Número de compuestos por cada rango de frecuencia

Como no podía ser menos, el compuesto más repetido es *bioinformática*, con 219 ocurrencias (0,221% del total de tokens). El salto al segundo puesto, con casi un tercio menos de apariciones, lo ocupa una palabra que no habríamos esperado inicialmente: *filogenético* (145 oc., 0,146%) ya que quizás se pensaría que, debido al tema, fuera más lógico que siguiera la palabra *biología* (que ocupa el cuarto puesto con 131 oc., 0,132%) o quizás el adjetivo *bioinformático*, que ocupa en cambio el octavo lugar, con 53 oc. (0,053%). Sin embargo, si tenemos presente cuál es uno de los objetivos primordiales de la bioinformática en su aplicación a la genómica, se comprende que este adjetivo sea el segundo en uso en colocaciones como *paquetes filogenéticos*, *árboles filogenéticos*, *reconstrucción filogenética*, etc. El tercer lugar se encuentra el adjetivo *biológico* (140 oc. 0,141%). Ahora bien, el dato más significativo es la elevada cifra de compuestos de reducida aparición (menos de 25 ocurrencias) que aumenta sensiblemente en el rango menor a 5 oc., que es más del doble de los otros juntos y supone la mayor variedad léxica dentro de la escasez de uso señalada.

3.1.2 Por categorías gramaticales

Sabemos de antemano (Cabré&Rigau, 1984:155; Martín Zorraquino, 1997:323 Cabré, 2004:25) que en los lenguajes de especialidad predominan las formaciones de base nominal, como confirman los datos: sustantivos, 180 (68,18%); adjetivos, 77 (29,16%); verbos, 3 (1,19%) y adverbios, 4 (1,51%). Los sustantivos superan en dos terceras partes el conjunto de términos. Cerca de un tercio corresponde a los adjetivos sobre una base nominal, generalmente presente también en el corpus, como *bionformática > bioinformático*, *cromosoma > cromosómico*, *ribosoma > ribosomal/ribosómico*, *topología > topológico*. Ello se puede hacer extensible a los verbos: *hidrólisis > hidrolizar* y a tres de los cuatro adverbios, todos en –mente: *biológicamente*.

3.2 Temas y tipos de formación

Es en este punto (tablas III y IV) donde podemos apreciar mejor la variedad de los temas grecolatinos en el lenguaje de la bioinformática y sus posibilidades combinatorias.

	Griego	Latín	Total
Temas en posición inicial (compuestos de dos bases)	56	23	79
Temas en posición final (compuestos de dos o más bases)	29	11	40
Temas en posición inicial y final (compuestos de dos bases)	11	0	11
Temas en posición inicial y media (compuestos de dos o tres bases)	5	0	5
Temas en posición inicial, intermedia y/o final (compuestos de dos, tres y cuatro bases)	5	0	5
Temas sólo en posición intermedia	1	0	1
Total	107	34	141

Tabla III. Clasificación de los temas por lengua y posición.

A continuación, mostraremos el número y combinación de temas en los compuestos clásicos e híbridos (a excepción de siglas y acrónimos):

Compuestos				Híbridos							
g+g	g+g+g	g+g+g+g	l+l	g+l	g+g+l	l+g	l+g+g	g+e	g+g+e	e+g	l+e
155	8	1	15	11	1	18	1	32	1	2	20

Tabla IV. Clasificación de los compuestos por número de bases y combinación de lenguas.

A la vista de todos estos datos cabe afirmar:

a) La preeminencia absoluta de los temas griegos (71,328%) sobre los latinos (28,671%) y que aquéllos prefieren combinarse consigo mismos o con el español.

b) Respecto a su distribución, la mayor variedad de temas se sitúa en la posición inicial (hecho que defendía Lang, 1997:237) y donde se concentra también el mayor número de temas latinos, siendo la segunda posición ocupada tanto por temas clásicos como por bases patrimoniales.

c) Los temas griegos tienen mayor capacidad combinatoria (hemos encontrado compuesto de hasta cuatro temas: *nanobiotecnología*), si bien el número de temas con tal capacidad disminuye a medida que se crean compuestos largos, en los que nunca interviene una base nativa.

d) Los datos negativos respecto al latín serían aún inferiores si viéramos como palabras nativas algunos términos que en nuestra opinión no lo son, como *molécula*, *espectro* y forman parte de compuestos cultos, como *biomolécula* o *espectrometría*.

e) Por el número de compuestos a que dan origen, también el griego supera al latín: en primer lugar, un sólo tema, de extensísimo uso, larga tradición y de múltiples combinaciones, está presente en 50 compuestos (incluidos híbridos), *logo*: *biológicamente*, *hematológico*, *histopatología*, *logaritmo*, etc. En segundo lugar, naturalmente, *bio* (32 formaciones), que puede aparecer en posición inicial e intermedia: *biofísico*, *biocomputador*, *microbiología* etc. Por último, el tema *geno* (origen), con 17 formaciones, cuya presencia es evidente por la selección del contenido de los textos que integran nuestro corpus, tal como hemos indicado en el punto 2.1; por lo tanto, este tema griego es, junto con *bio*, un elemento básico en compuestos, derivados y parasintéticos de la bioinformática, genómica y biomedicina. El



resto de temas se distribuye en estos rangos de frecuencia:

20> f ≥ 10	5	Además de <i>geno</i> : <i>auto</i> , <i>grama/grafó</i> , <i>micro</i> , <i>multi</i> .
10> f ≥ 5	13	<i>nomía/-o</i> , <i>nano</i> , <i>metría/o</i> etc.
f < 5	121	<i>onco</i> , <i>fero</i> , <i>plasia</i> , <i>lito</i> etc.

A la vista de estos datos podemos concluir que sólo dos temas generan muchos compuestos (en total, 82 términos diferentes, 58,15% del total), mientras que la mayoría tiene reducida capacidad combinatoria o, al menos, no se usan excesivamente en el lenguaje de la bioinformática. Sin embargo, cabría preguntarse si son esos mismos tres primeros temas (incluimos *geno* por estar estrechamente ligada a los pilares de la bioinformática) los que se hallan en las palabras de más alto índice de frecuencia y si crean todo tipo de palabras.

tema	nº términos	frecuencia	categorías gramaticales
logo	<u>50</u>	641	nom., adj., adv.
bio	32	<u>704</u>	nom., adj.,
<u>geno</u>	17	<u>231</u>	nom., adj.
total.....	99.....	1576	

Tabla V. Relación entre los tres temas con mayor capacidad productiva, frecuencia de sus compuestos y categorías gramaticales a que dan lugar.

En líneas generales, los tres primeros temas suponen casi la mitad de los compuestos y prácticamente la totalidad de la frecuencia de aparición, por lo que podemos afirmar que en conjunto los temas más productivos van ligados a las palabras de mayor uso. Sin embargo, el tema que más apariciones presenta es el segundo en cuanto a variedad de términos, hecho que se puede considerar justificado por el propio objeto de estudio de la bioinformática, mientras que *logo* es un tema que se puede encontrar en cualquier disciplina humanística, artística o científico-técnica.

3.3 Acortamientos

Tres son los procedimientos de acortamiento o abreviación que afectan a los compuestos clásicos en nuestro corpus:

1. Apócope, como en *alelo* (2 oc.) por *alelomorfo*, siempre en su forma truncada; *fotos* (2 oc.) aparece alternando con *fotografía* (4 oc.) y *video* (1 oc.), que alterna con *videocámara* (2 oc.) y *cámara de vídeo* (1oc.). En cambio las formas *clado* (“rama”) y *grafo*, aparentemente acortadas, son en realidad formas libres, lo cual desmontaría la teoría de quienes sostienen la imposibilidad de que las raíces clásicas lo sean y, por lo tanto, sean funcionalmente afijos: *clado* significa “OTUs <Operational Taxonomic Units> que comparten un ancestro común” (Silla, 2007:38) y no es equivalente a *cladograma* (en términos sencillos, representación en forma de diagrama del parentesco evolutivo de varias especies que comparten un ancestro común e incluye tanto a dicho ancestro como a sus descendientes vivos o ya extinguido; todos ellos constituyen un *clado*). Y *grafo*, sólo o en la sigla GAD (*grafo acíclico dirigido*) o DAG⁶, se usa con su sentido básico, es decir, “dibujo”, “trazo” etc.

6 Grafo acíclico dirigido es una representación gráfica “similar to hierarchies but differ in that a more specialized term (child) can be related to more than one less specialized term (parent)”, Gene Ontology (<http://geneontology.org/GO.doc.shtml>).

2. Compuestos formados tras el acortamiento de otros términos⁷ (algunos, neologismos): *fenograma* (fenotipo+grama), *filograma* (filogenia+grama), *inmunoteñido*, pero especialmente con *biología*: *biocomputación*, *bioinformática/-ico*, *biomedicina*, *biomédico*, *biometría*, *biométrico*, *biotecnología* o *bioempresas* (calco del inglés *biocompanies*). Dudosas son en cambio algunas formaciones con *tele-*, que se podrían analizar como apócope de *teléfono* sobre el que se forma un nuevo compuesto híbrido o bien directamente sobre *telo* (lejos): *telecomunicación*, *teleconsulta* o *telemedicina*. Aunque escapa a nuestro estudio, pensamos que éste es el procedimiento que subyace a numerosos híbridos del inglés presentes en el corpus, especialmente a partir de *biology/biological* o *genetics*, usados mayoritariamente para designar nombres comerciales (empresas de biotecnología, software etc.), bases de datos etc.: *Global Telegenetics*, *bioPYTHON*, *bionavigator*, *LexGen*, o dispositivos especiales: *biochip* (40 oc, 0,04%), *microarray* (85 oc., 0,08%) etc.

3. Por lo que respecta a siglas y acrónimos en los que intervenga un compuesto clásico o híbrido, en líneas generales su variedad y frecuencia en español es insignificante en comparación con los del inglés (*cfr.* tabla I). Destacaremos sólo las 17 ocurrencias de C.N.I.O. (Centro Nacional de Investigaciones *Oncológicas*). En el resto predominan los compuestos con *bio* (*bioempresas*, *biomédicas*, *biotecnología*).

3.4 Neologismos

Tomando en consideración las recomendaciones de Lerat (1997) para quien el diccionario de la lengua es el punto de partida y la consagración de la lengua especializada, nuestro diccionario de referencia ha sido la última edición (la vigésimo segunda) del *Diccionario de la Real Academia Española* (DRAE), versión online; además, hemos intentado cruzar la información recurriendo a un diccionario especializado (González, 2005), en el que, no obstante, también existen lagunas que se explican porque “las fronteras de las disciplinas extienden sus ramas conceptuales continuamente de modo que sólo la lectura de la actualidad científica permite vislumbrar el significado de los nuevos conceptos. Aunque hoy en día proliferan los diccionarios especializados, inmediatamente quedan carentes de actualidad y, en muchos casos, obsoletos” (Marco Stiefel, 2000:204). Por último, puesto que uno de los objetivos de la ciencia y la técnica es la transmisión del conocimiento y es precisamente a través de los medios de comunicación como se hacen llegar al gran público (Marco Stiefel, *ibid.*), ha resultado de gran utilidad el banco de neologismos del Observatori de Neologia (BOBNEO) de la Universitat Pompeu Fabra de Barcelona, constituido por textos de la prensa oral y escrita y en continua actualización.

Hemos contabilizado un total de 75 neologismos, casi un tercio del total de compuestos cultos, que, salvo *bioinformática* y otros dos, presentan un bajo número de ocurrencias (185, 0,18% sobre el total). La primera datación de algunos de ellos se remonta a finales de los '80 o principios de los '90 del siglo XX, según datos de BOBNEO, tiempo suficiente para que la RAE los hubiera incorporado a su diccionario, como *milisegundo*, *bioinformática* y *biomolécula* (todos de 1989) o *inmunohistoquímica* (1991). En cambio, un grupo significativo (31 términos) aún no ha sido recogido ni por el diccionario técnico consultado ni en BOBNEO: *bicatenario/monocatenario*, *biotecnológico*, *citogenético*, *filograma*, *inmunoteñido* (existe *inmunotinción*), *micromatriz*, *nanodispositivo*, *nulisomía*, *organomegalia*, *pseudo-energía*, *ribosomal*... De otros dos tampoco se han hallado

⁷ No hemos aludido en nuestra contribución al término *genoma* y derivados (ni a otras formaciones similares) ya que pensamos está formado con el sufijo griego *-ωμα* (de valor colectivo) y no acuñado por su creador H. Winkler (1920), como se cree, como un acrónimo de *gen(e)+(chromos)ome*. *Cfr.* Lederberg-McCray (2001:8).



referencias puesto que *biovisualizador* es el nombre dado por su autor (J.C. Calvo) a una herramienta informática y el otro, *nonanosistema*, podría ser un *hapax legomena* o bien simplemente una errata.⁸

Estos neologismos se pueden agrupar en cuatro campos:

a) Componentes de la célula, funciones moleculares o procesos biológicos (18 términos), como *citogenético*, *mitocondrial*, *biomolécula*, *biomolecular*, *bidireccional*, *biomacromolecular*, *centrómero*, *monoclonal*, *semiconservativo*, *pseudogén*.

b) Nuevas disciplinas, ramas técnicas o de la biología y sus relaciones con otras disciplinas o con el mundo empresarial (15 términos): *biotecnológico* (10 oc., 0,03%), *bioempresa*, *nanobiología*, *nanobiotecnología*, *nanobioingeniería*.

c) Nuevas técnicas de análisis, de laboratorio, de representación, criterios de valoración, dispositivos, “herramientas” etc. (33 términos), tales como *auto-estimación* (21 oc., 0,02%), *bidimensional*, *biomarcador*, *filograma*, *espectromería*, *fluorocromo*, *genotipado/genotipaje*, *inmunoteñido*, *inmunohistoquímica*, *multietiqueta*, <algoritmo> *multiobjetivo*, *micromatriz*, *nanodispositivo*, *pseudo-secuencia*, *uni-/multivariante*.

d) Enfermedades y su tratamiento, alteraciones genéticas relacionadas con éstas, servicios al paciente (9 términos): *organomegalia*, *colorrectal*, <leucemia> *linfoblástica*, <leucemia> *linfocítica*, <enfermedad> *multifactorial*, *farmacogenético*, *telemedicina*, *teleservicio* etc.

Por el tipo de composición, destacan las formaciones híbridas griego+español (27 términos, 36% del total de neologismos) en las que hemos incluido *biorrobótica* y *microrrobot* ya que, aunque el término *robot* procede del checo a través del inglés, está incorporado al español desde hace tiempo, así como los formados con *ribo-*, acortamiento del inglés *ribose*, del alemán *Ribose*, adaptación libre de *Arabinose* o “arabinosa, azúcar de la goma arábiga” (DRAE), procedente a su vez de *Ἀραβικός*, arábico). Les siguen las compuestas sólo por temas griegos, ya sean dos o tres (24 términos, 32%). En cambio no hay ningún neologismo integrado por dos bases latinas: éstas se combinan con griegas (13 términos, 17,33%) o con patrimoniales (11 términos, 14,66%). Todo ello confirma lo dicho en el punto 3.2. Sin embargo, a todos esos datos subyace nuestra creencia de que una buena parte de estos neologismos son préstamos del inglés, lengua que los acuñó recurriendo a fuentes clásicas dado que las fechas de su primera aparición escrita son anteriores (y a veces, con muchos años de diferencia) a las formas españolas. He aquí algunos ejemplos:

Español	Fuente	Año	Inglés	Fuente	Año
<i>Autorreplicación</i>	González	2005	<i>self-replication</i>	BNC	1992
<i>Multidimensional</i>	---	----	<i>multidimensional</i>	BNC	1989
<i>biomolécula</i>	BOBNEO	1989	<i>biomolecule</i>	OED	1901
<i>bioinformática</i>	BOBNEO	1989	<i>bioinformatics</i>	OED	1978
<i>kilobase</i>	González	2005	<i>kilobase</i>	BNC	1975
<i>micromatriz</i>	----	----	<i>microarray</i>	OED	1967
<i>microrrobot</i>	BOBNEO	2001	<i>microrobot</i>	OED	1982
<i>nanotecnología</i>	BOBNEO	1998	<i>nanotechnology</i>	OED	1974
<i>telemedicina</i>	BOBNEO	2006	<i>telemedicine</i>	OED	1967

Tabla VI. Neologismos españoles con la datación de sus correlatos ingleses.

⁸ Hemos intentado recabar confirmación del término del autor del texto, pero no se nos ha facilitado.



4 Conclusiones

Los datos expuestos en el apartado 3 son el resultado de la postura teórica que hemos adoptado y que creemos ha quedado patente a lo largo del artículo: y es que descartamos la consideración como prefijación y sufijación de todas aquellas palabras formadas por raíces o temas que en las lenguas clásicas eran palabras plenas y no meros prefijos o sufijos, así como la terminología “remedial” que se intenta aplicar a estas situaciones. Pensamos que la composición clásica debe constituir un mecanismo diferente de la composición patrimonial en cuanto a los requerimientos mínimos formales, distribucionales o funcionales. Para quien tiene una formación clásica es inadmisibles que *bio*, *logo*, *multi*, *poli* sean considerados simples prefijos o sufijos. En cambio, los estudiosos con una visión exclusivamente sincrónica los ven como unos formantes más que por el hecho de que no se hayan integrado en la lengua como formas autónomas, como sí ha sucedido a otras (*génesis*, *ritmo*, *música*, *gen*) son equiparadas a afijos españoles. Precisamente, hemos citado dos ejemplos de temas libres (*clado* y *grafo* que, incluso, presenta una forma prefijada, *subgrafo* con 3 oc.) que habitualmente aparecen en palabras complejas, por lo que en potencia una forma grecolatina ligada podría convertirse en forma libre.

Por otra parte, a pesar de nuestra consideración amplia de “compuesto clásico”, contraria a las posturas de reputados estudiosos, reconocemos que cuantitativa y cualitativamente su uso es muy reducido, más incluso de lo que esperábamos. Ésta es la conclusión fundamental de nuestro estudio. Era indudable que la bioinformática tenía que usar en su lenguaje de especialidad los compuestos clásicos (incluidos parasintéticos, acortamientos etc.). Creíamos que, puesto que es una confluencia de varias disciplinas, al menos la biología arrastraría consigo una buena cantidad de compuestos (cfr. Quintana, 1989:15 y ss., Estopà et al., 2000:855-6, López Piñero & Terrada, 2005:ix), aunque no es menos cierto que la informática ha aportado al español muchos híbridos ya integrados en la lengua común y que paradójicamente apenas cuentan con representación en el corpus: *megabytes/MB* (9 oc.), *terabytes/TB* (3 oc.) y *teraflops* (2 oc.). Pero no ha sido así. Por lo tanto, algunos mitos sobre la preeminencia de las lenguas clásicas en el lenguaje científico-técnico (uno de cuyos rasgos es justamente la elevada presencia de compuestos cultos así como de compuestos sintagmáticos, siglas y abreviaciones) no se sostienen aquí si hablamos en términos absolutos. ¿Sería éste motivo suficiente para considerar que nuestro corpus está más cerca del lenguaje general que del discurso de especialidad? Basándonos sólo en los datos obtenidos, tendríamos que responder que sí. Pero este lenguaje se define también por rasgos discursivos o gramaticales, entre otros, que no han sido abordados en esta contribución. Por ello, deberíamos ser cautos a la hora de caracterizarlo y ponderar los resultados negativos a los que podrían haber contribuido tres hechos:

1. El propio corpus: de pequeño tamaño, integrado por pocos textos y de naturaleza variada, si bien a grandes rasgos se dividirían en dos tipos: divulgativos (tres) y especializados (once). Cabría pensar que los primeros han hecho disminuir su especialización y de ahí la baja incidencia de los compuestos cultos. No creemos, sin embargo, que esos tres documentos hayan sido determinantes en la escasa presencia de la composición culta ya que, aun constituyendo aproximadamente la cuarta parte del número de textos, del volumen del corpus (26,81% de palabras) y de las ocurrencias de los compuestos (427 oc., 23,86%), contienen 104 compuestos diferentes, es decir, casi la mitad del total, entre los que hay que incluir varios neologismos (por ej. formados *pseudo* y *nano*) y el *hapax legomena*. Pensamos que la razón se debe, en primer lugar, a la caracterización de estos artículos, especializados por la



temática y de media o media-alta densidad ya que están escritos por expertos para un público no general, sino con una cierta formación e intereses. En segundo lugar, el otro grupo de textos está destinado a especialistas y contiene numerosos diagramas, secuencias, muestras de tejidos, etc. incomprensibles para los profanos. Para hacer llegar parte de esa información al lector interesado, pero no experto, ésta se ha de transformar en palabras. De entre todas ¿qué mejor que los compuestos clásicos, muchas de cuyas raíces son conocidas por un amplio sector de lectores (en este sentido es muy acertada la denominación de Lang (1997:221) de vocabulario “de carácter internacional”), que históricamente se han empleado en la terminología de la ciencia y que son económicos, pues un solo compuesto proporciona una información que, de otro modo, se tendría que expresar con más palabras, y por ende comportan una “compresión semántica elevada” (Cabré & Domènech, 2001:546)? Eso justificaría, en nuestra opinión, al menos la variedad de compuestos en el subcorpus divulgativo.

2. La no inclusión de compuestos de la terminología química y bioquímica (del tipo *desoxirribonucleico*, *aminoácido* etc.), con una considerable presencia en el corpus.

3. Nuestro trabajo se ha efectuado con un solo corpus, no tomándose como referencia ningún estudio sobre la composición clásica en otros ámbitos o ni contrastándolo con otros de la misma materia. Tampoco se ha llevado a cabo un cálculo de las formaciones prefijadas o sufijadas con una base clásica, como *análisis*, *parámetro* etc. que sin duda habría aumentado el peso de términos de origen grecolatino y habrían atenuado el pesimismo antes manifestado.

Centrándonos ya en nuestros compuestos, hemos visto la relativa poca variedad de temas en los términos de mayor frecuencia. En cambio, los de escasa o mínima aparición suponen prácticamente la totalidad y además ofrecen una amplia gama de temas: usuales en los lenguajes de especialidad y en su mayoría en términos incorporados a la lengua general (*arquía*, *grafía/grama/grafó*, *filo* –“amor”–, *poli*, *nomía*, *pan(t)*, *multi* etc.), menos “conocidos”, pero muy variados, los propios⁹ de la biología o la medicina (*cito*, *clado*, *geno*, *filo* –“raza”–) y de la técnica (*nano*, *micro*, *tecno*), que es la vertiente, en nuestra opinión, más parca en temas o raíces específicos, como por otra parte parece ser habitual (Estopà et al., 2000:855, Vivanco, 2006:21). Siguiendo precisamente a Vivanco, sí se confirma la presencia mayoritaria del griego en las ciencias a causa de lo que llama “tradición milenaria”. Según esta autora, su presencia, aunque poca en la “tecnología dura”, se explica por el trasvase de información de la ciencia a la técnica, cuyo vocabulario es más joven y más moderno y donde predominan relativamente las formaciones híbridas del inglés con temas latinos. Pero nuestro corpus no cumple absolutamente sus previsiones: por una parte, es cierta la juventud del léxico de la bioinformática en lo que a compuestos cultos se refiere y de hecho casi una tercera parte de éstos son neologismos, algunos recogidos por vez primera en 2007 (BOBNEO: *biomacromolecular*, *nanobiología*), otros incluso ni siquiera están atestiguados en las dos fuentes de referencia. Sin embargo, pertenecen casi más al ámbito científico que al técnico, por más que una buena parte corresponde a técnicas, procesos y herramientas de análisis. Por otra parte, el número de compuestos latinos es menor, de híbridos con el inglés mínimo (con *data* y *semi*) y nulo en el caso de los neologismos. Otro de los tópicos, la progresiva infiltración del inglés en el español de la mano de las innovaciones tecnológicas anglosajonas (Vivanco, 2006:37) y especialmente en la informática (Martín Zorraquino,

9 Teniendo siempre presente que un mismo término no es exclusivo de una disciplina, sino que adquiere significados o matices específicos según el campo en que se use. Tal es el caso de *ontología*, *cristalografía* u *ortólogo*, con significados diferentes en bioinformática y en filosofía, geología o fonética, respectivamente.



1997:324) tampoco parece ser aplicable aquí (al menos en cuanto al tipo de formación que estamos tratando), salvo en algunos casos donde se prefiere la utilización del término inglés (*microarray*) frente al español (*micromatriz*). Podrán quedarse tranquilos Gili Gaya y D. Alonso para quienes, según Martín Zorraquino (1997:330-331), “el aumento de los tecnicismos en inglés, en detrimento de los que reflejan una composición de elementos grecolatinos, constituía un peligro para la estabilidad lingüística”. En este sentido, no parece que el inglés desestabilice el español ni por número de formas (al menos, plenas) ni por su frecuencia ni por el rechazo de los formantes grecolatinos que están, en última instancia, presentes en ellos (“neo-classical compounds” y “splinters”): *biochip*, *cytoskeleton*, *microarray*, *Gene Ontology (GO)*, etc. Hay que admitir la importancia de las siglas y acrónimos, uno de los rasgos caracterizadores del lenguaje científico-técnico, en inglés ya que triplican en número a las españolas y sus ocurrencias son casi ocho veces más que las del español. Además, detrás de ciertos compuestos e híbridos es casi indudable que está el inglés o están fuertemente influidos por él. No obstante, puesto que recurren al léxico grecolatino, que no nos es precisamente ajeno, su incorporación al español con los mínimos retoques fonéticos y/o morfológicos hace que tales formas sean sentidas como propias. En resumen, esas formas, ya sean plenas o acortadas, se alzan sobre bases clásicas. Dicho esto, en nuestra opinión el inglés, lengua generadora de neologismos, es al menos en este campo la heredera de las lenguas clásicas en detrimento de sus “naturales” herederas (las lenguas romances). En ese sentido, cabría afirmar que, por el momento, griego y latín siguen *vivos* gracias a la capacidad lexicogenética del inglés desde donde se incorporan al español.

Quizás nuestro análisis haya estado sesgado por nuestra formación académica y nuestra posición suponga un retroceso o ir contracorriente, tal como Martínez Hernández (1992) critica acusando, tal vez con razón, a los estudiosos de las lenguas clásicas de que estamos fuertemente marcados por el componente histórico. Pero defendemos el recurso a los formantes clásicos en la terminologización y pensamos que siguen sirviendo de fuente para la creación de nuevos términos ya sea en español, ya en inglés.

Referencias bibliográficas

- Adams, V. (1973). *An Introduction to Modern English Word-formation*. Longman: London.
- Alarcos Llorach, E. (1994). *Gramática de la lengua española*. Espasa Calpe: Madrid.
- Alvar Ezquerro, M. (1993). *La formación de palabras en español*. Arco Libros: Madrid
- Banc de l'Observatori de Neologia (BOBNEO), Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. <http://obneo.iula.upf.edu/bobneo/index.php>
- Bel, G. & Jiménez, M^a D. (2006). Código genético y lenguaje verbal. *Revista Española de Lingüística*, 36: 285-318.
- British National Corpus (BNC). <http://www.natcorp.ox.ac.uk/>
- Buehler, L.K. & Rashidi, H.H. (2005). *Bioinformatics Basics*. Taylor & Francis: Boca Raton.
- Cabré, M^a.T. (2004). ¿Lenguajes especializados o lenguajes para propósitos específicos? *Foro Hispánico - Revista Hispánica de Flandes y Holanda*, 26(1):19-34.
- Cabré, M^a.T. (1999). Hacia una teoría comunicativa de la terminología: Aspectos metodológicos. En *La terminología. Representación y comunicación* (pp. 129-150). Institut Universitari de Lingüística Aplicada-Universitat Pompeu Fabra: Barcelona.
- Cabré, M^a T. & Domènech, M. (2001). Terminologia i tipologia textual: com establir el nivell d'especialització d'un text científic tècnic. En C. Muñoz (coord.) *Trabajos en Lingüística Aplicada* (pp. 543-548). AESLA/Univerbook: Barcelona.
- Cabré, M^a T. & Rigau, G. (1984). *Lexicología i semàntica*. Enciclopèdia Catalana: Barcelona.



- Conklin, D. (1998): La Bioinformática: una perspectiva de la estructura de proteínas. *Temas de Ciencia y Tecnología*, 6: 3-10.
- Ernout, A. & Meillet, A. (1985). *Dictionnaire Étymologique de la langue latine. Histoire des mots*. Klincksieck: Paris.
- Eseverri Hualde, C. (1979). *Diccionario etimológico de helenismos españoles*. Aldecoa: Burgos.
- Estébanez, F. (1998). *Étimos griegos. Monemas básicos del léxico científico*. Octaedro: Barcelona.
- Estopà, R., Cabré, M^a T. & Vivaldi, J. (2000). Use of Greek and Latin forms for term detection. *Proceedings of Second International Conference on Language Resources and Evaluation*, vol. 3 (pp. 855-859). National Technical University of Athens: Atenas.
- Feliu, J., Vivaldi, J. & Cabré, M^a T. (2002). *Ontologies: a Review*. Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra: Barcelona.
- García, S., Meilán, A. & Martínez, H. (2004). *Construir bien en español. La forma de las palabras*. Editorial Nobel y Ediuono (Ediciones de la Universidad de Oviedo): Oviedo.
- Gene Ontology: <http://geneontology.org/>
- Genomic Glossaries. <http://www.genomicglossaries.com>
- González y Pozo, V. (2005). *Diccionario Técnico-Científico español inglés/inglés español*. RC&Associates: Toronto.
- Klug, W.S., Cummings, M.R. & Spencer, Ch. (2006). *Conceptos de genética*. Pearson Educación: Madrid. Originalmente publicado en inglés como W. S. Klug, M.R. Cummings & Ch. Spencer (2006). *Concepts of Genetics*. Prentice Hall International: Upper Saddle River.
- Lang, M. F. (1997). *La formación de palabras en español. Morfología derivativa productiva en el léxico moderno*. Cátedra: Madrid. Originalmente publicado en inglés como M. F. Lang (1990). *Spanish word formation: productive derivational morphology in modern lexis*. Routledge: London.
- Lederberg, J. & McCray, A. T. (2001). ‘Ome sweet ‘Omics: A Genealogical Treasury of Words. *The Scientist*, 15(7): 8. <http://www.the-scientist.com/article/display/12335/>
- Lerat, P. (1997). *Las lenguas especializadas*. Ariel: Barcelona. Originalmente publicado en francés como P. Lerat (1995). *Les langues spécialisées*. Presses Universitaires de France: Paris.
- López Piñero, J. M. & Terrada, M.L. (2005). *Introducción a la terminología médica*. Masson: Barcelona.
- Marco Stiefel, B. (2000). La decodificación del mensaje en la alfabetización científica. *Escuela abierta*, 4:199-214.
- Marcos Marín, F. (1980). *Curso de gramática española*. Cincel: Madrid.
- Martín Zorraquino, M.A (1997). Formación de palabras y lenguaje técnico. *Revista Española de Lingüística*, 27(2):317-339.
- Martínez Hernández, M. (1992). La formación de palabras en griego antiguo desde el punto de vista semántico. *Cuadernos de Filología Clásica*, 2:95-122.
- Oxford English Dictionary* (OED). <http://www.oed.com>
- Pena, J. (1999). Partes de la morfología. Las unidades del análisis morfológico. En I. Bosque & V. Demonte (eds.): *Gramática descriptiva de la Lengua Española*, 3 vol. (pp. 4305-4366). Espasa-Calpe: Madrid.
- Pensado, C. (1999). Morfología y Fonología. Fenómenos Morfofonológicos. En I. Bosque & V. Demonte (eds.): *Gramática descriptiva de la Lengua Española*, 3 vol. (pp. 4423-4504). Espasa-Calpe: Madrid.



- Pingarrón, E. (1998). *Étimos latinos. Monemas básicos del léxico científico*. Octaedro: Barcelona.
- Quintana Cabanas, J. M. (1989). *La terminología médica a partir de sus raíces griegas*. Dykinson: Madrid.
- Rainer, F. (1999). La derivación adjetival. En I. Bosque & V. Demonte (eds.): *Gramática descriptiva de la Lengua Española*, 3 vol. (pp. 4595-4644). Espasa-Calpe: Madrid.
- Real Academia Española (2001). *Diccionario de la Lengua Española*. <http://buscon.rae.es/draeI>.
- Real Academia Española (1973¹⁶). *Esbozo de una nueva gramática de la lengua española*. Espasa Calpe: Madrid.
- Rojo, G. (2006). Informática y Lingüística: Las lenguas en la sociedad del conocimiento. *Boletín RedIRIS*, 74-75: 1-8. <http://www.rediris.es/rediris/boletin/74-75/ponencia1.pdf>
- Santiago, R. & Bustos, E. (1999). La derivación nominal. En I. Bosque & V. Demonte (eds.): *Gramática descriptiva de la Lengua Española*, 3 vol. (pp. 4505-4594). Espasa-Calpe: Madrid.
- Sebastián Yarza, F.I. (1984). *Diccionario Griego Español*. Ramón Sopena: Barcelona.
- Serrano Dolader, D. (1999). La parasíntesis. En I. Bosque & V. Demonte (eds.): *Gramática descriptiva de la Lengua Española*, 3 vol. (pp. 4682-4756). Espasa-Calpe: Madrid.
- Serrano Dolader, D. (1995). *Las formaciones parasintéticas en español*. Arco Libros: Madrid.
- Val Álvaro, J. F. (1999). La composición. En I. Bosque & V. Demonte (eds.): *Gramática descriptiva de la Lengua Española*, 3 vol. (pp. 4757-4842). Espasa-Calpe: Madrid.
- Varela, S. & Martín, J. (1999). La prefijación. En I. Bosque & V. Demonte (eds.): *Gramática descriptiva de la Lengua Española*, 3 vol. (pp. 4493-5040). Espasa-Calpe: Madrid.
- Vivanco, V. (2006). *El español de la ciencia y la tecnología*. Arco Libros: Madrid.

Corpus constituido por estas obras:

1. Barbosa, D. (2007). *Desarrollo de Workflows en Bioinformática utilizando el potencial de los Web Services y XML. Seminario de investigación 2007 II*. http://seminario2012107.googlepages.com/Estado_de_Arte_Primer_Entrega.pdf
2. Calvo, J. C.: *Biovisualizador- Manual del usuario*. <http://atc.ugr.es/~jccalvo/download/manualusuario.pdf>
3. Centro Nacional de Investigaciones Oncológicas (2004). Diagnóstico molecular. *Investigación de Transferencia en Cáncer: del Laboratorio a la Clínica 2005-2006*. (pp. 19-24). http://cnio.es/es/servicios/transferencia-cancer/docs/06/03_diagnostico-molecular.pdf
4. Dopazo, J. & Valencia, A. (2001). *Bioinformática y genoma*. <http://bioinfo.cipf.es/docus/courses/filogenias/BioinfoGenomica.pdf>
5. Fichas informativas BET (2006). http://www.upv.es/crib/docs/ficha_bet_bioinformatica.pdf
6. Guigó, R. (2000). Bioinformática ¿una ciencia sin científicos? *Quark* (18). <http://www.prbb.org/quark/18/default.htm>
7. Guigó, R. & Valencia, A. (2006). Bioinformática en la transición de genes y proteínas a genomas y proteomas. *Boletín Sociedad Española de Bioquímica y Biología Molecular*, 130: 6-9.
8. Joyanes Aguilar, L.: Bioinformática y Salud Un nuevo paradigma de la sociedad. http://www.fpablovi.org/publicaciones/estudios/03estudios/bioinformatica_ljoyanes.pdf
9. Martín Sánchez, F. (2000). Genoma, bioinformática y comercio electrónico. *Quark*, (18). <http://www.prbb.org/quark/18/018056.htm>



10. Peiró Sancasto, J. M. (2006). *Clasificación multietiqueta de genes basada en la Gene Ontologies*. Pfc dirigido por Dr. D. Alfonso Juan Císcar y D. Juan Miguel García Gómez.
11. Sánchez, A. *Introducción a los microarrays*.
http://www.ub.es/stat/docencia/Biologia/introbioinformatica/MicroArrays/Microarrays_de_DNA.pdf
12. Silla Martínez, J. M. (2007). *TrimAl: limpieza de alineamientos múltiples de secuencias de ADN o de proteínas para mejorar la reconstrucción filogenética*. Pfc dirigido por los profs. D. Juan Miguel García Gómez (IBIME, Física Aplicada, UPV) y D. Antonio Gabaldón (CIPF).
13. Tríptico Informativo curso Bioinformática Universidad SEK.
14. Zepeda García, O. (2005). *Bioinformática*. Ilustrados.com: Perú.
<http://site.ebrary.com/lib/universvaln/Doc?id=10092364>
