

**PROCESOS DE CREACIÓN DEL LÉXICO
CON SIGNIFICADO ESPECIALIZADO:
UNIDADES VERBALES EN EL
DISCURSO DEL GENOMA HUMANO¹**

MERCÈ LORENTE CASAFONT

Prefacio

Debo agradecer a Anna M. Thornton, Antonietta Bisetto y, en particular, a mi amigo Claudio Iacobini su gentil invitación para participar en este volumen de homenaje a Sergio Scalise. Comparto con ellos la idea de que la influencia de un maestro no es exclusiva de algunos de sus alumnos directos, sino que a veces trasciende a toda una generación.

Convertido para mí en una referencia obligada ya en mi época de estudiante de Filología Catalana, a partir de la lectura de la traducción española de *Generative Morphology*, las obras de Sergio Scalise han estado presentes en mi trabajo como docente universitaria y en mis investigaciones sobre el léxico. Pero mi contribución está motivada no sólo por el respeto profesional al lingüista, sino también por la estimación a la persona. Mi maestra y amiga M.

¹ Este trabajo se enmarca en los proyectos de investigación TEXTERM(BFF2000-0841) y RICOTERM (TIC2000-1191), ambos financiados por el Ministerio de Ciencia y Tecnología. El análisis de los datos y la redacción se han realizado en parte durante una estancia de investigación en el Istituto de Linguistica Computazionale (ILC-CNR) de Pisa. Agradezco al equipo del proyecto CLIPS, y en particular a Nilda Ruimy y Nicoletta Calzolari, su invitación y su gentil hospitalidad.

Teresa Cabré lo ha invitado diversas veces a Barcelona, y en estos encuentros he podido comprobar, además de su consistencia intelectual, su calidez humana. Auguri, maestro Scalise! Auguri, amico Sergio!

1. Introducción

Este trabajo sobre la formación de verbos del discurso sobre genoma humano en español se inscribe en una línea de investigación más amplia y de equipo, que tiene por objetivo la descripción de la terminología para poder establecer generalizaciones sobre el comportamiento de las unidades léxicas transmisoras de conocimiento especializado, implementables en aplicaciones lingüísticas.

Desde el punto de vista teórico, nos interesan particularmente los procesos de creación terminológicos y los mecanismos de estabilización del léxico que se producen en cada ámbito comunicativo profesional, académico y científico. Consideramos que cada especialidad construye, con los recursos comunes que ofrece la lengua, sus propias terminología, fraseología y tipología discursiva, de manera que antes que generalizar sobre el léxico de especialidad en general debemos trabajar en la descripción de ámbitos temáticos distintos, a fin de establecer patrones y tendencias fiables. La vinculación de la terminología con el resto del componente léxico es evidente tanto en el plano discursivo como en el cognitivo; la gramática interna del hablante no está formada por más de un componente léxico (Cabré 1999), por lo tanto parece fundamental describir en qué consiste el valor especializado que distinguiría la terminología del léxico común. En el período 2001-03, hemos centrado nuestra atención en el discurso sobre genoma humano.

Desde el punto de vista aplicado, tanto la descripción de unidades y procesos como las generalizaciones y tendencias que se deriven de ella sirven de base para la representación de la información en recursos lingüísticos (diccionarios computacionales, bases de datos terminológicas, ontologías) y para la mejora de herramientas de procesamiento del lenguaje natural y para el diseño de sistemas de extracción o de recuperación de información. En esta línea, hemos construido un banco de conocimientos de Genoma Humano, compuesto por un corpus textual, una base de datos terminológica, una ontología y una base de datos documental y factográfica, que nos permite poner en práctica la representación de la información y diseñar nuevas estrategias para mejorar los sistemas de extracción y de recuperación de información existentes.

1.1. Objetivos y marco teórico

En este artículo pretendemos mostrar los resultados del análisis morfológico de las unidades de categoría verbal extraídas automáticamente del corpus textual integrado en el banco de conocimiento. El objetivo principal de este trabajo es detectar los procesos de formación de verbos más productivos en el discurso de la genómica, para ver si se pueden establecer tendencias morfológicas y semánticas que caractericen específicamente las unidades con valor especializado.

Nos situamos dentro de la Teoría Comunicativa de la Terminología (Cabré 1999) basada en una aproximación lingüística a los términos, entendidos como palabras de la lengua que adquieren en contexto un valor semántico y pragmático que consideramos especializado, por cuanto está controlado por un conocimiento adquirido *ad hoc*, no compartido por la totalidad de los hablantes de una lengua.

Desde la perspectiva de la lexicología, partimos de la idea de la entrada plena del lexicón (Jackendoff 1975), de manera que el componente léxico incluiría tanto unidades especializadas como no especializadas, morfemas ligados, palabras simples y palabras construidas, y también todo tipo de relaciones morfológicas, sintácticas, semánticas y enciclopédicas entre unidades.

Para la identificación de verbos con valor especializado entre las unidades de categoría verbal presentes en los textos, nos basamos en Lorente (2002), donde establecemos la distinción entre verbos discursivos y conectores, de uso general, y verbos fraseológicos y verbos-término que vehiculan conocimiento especializado.

1.2. Corpus y metodología de análisis

Los datos de nuestro análisis provienen del Corpus Textual Especializado² del Instituto Universitario de Lingüística Aplicada (IULA) de la Universidad Pompeu Fabra (UPF) de Barcelona, que incluye muestras de textos de medicina, informática, derecho, economía, medioambiente y genoma humano, en español, catalán, inglés, francés y alemán, marcados estructuralmente y procesados lingüísticamente.

El corpus que utilizamos en este trabajo está formado por la totalidad de documentos en lengua española seleccionados para el Corpus Textual de Genoma Humano, seleccionados por expertos del ámbito,

² La consulta del corpus se puede realizar por Internet, mediante Bwananet, herramienta de explotación del corpus construida sobre la base de Corpus Workbench (IMS-Stuttgart), y actualmente en pruebas en <http://brangaene.upf.es/bwananet0/bwananet0a>.

que consta de 126 documentos, con un total de 975.441 palabras lematizadas, etiquetadas y desambiguadas. La interrogación sobre la categoría verbal nos da un resultado de 124.482 ocurrencias, que se corresponden a 2.516 lemas verbales distintos.

A partir del listado de lemas verbales, se ha realizado manualmente una clasificación según procedimientos de formación (conversión, derivación, composición y préstamo), cuyos resultados detallamos más adelante en el análisis cuantitativo. A continuación hemos clasificado por afijos los 326 lemas verbales contruidos por derivación. Los resultados cuantitativos, aunque relativos, nos orientan en la detección de los procesos de formación de verbos más productivos.

Mediante la herramienta Bwananet para la consulta del Corpus, hemos accedido a las listas de concordancias de cada uno de los lemas verbales, que hemos revisado manualmente para dilucidar el significado en contexto de la expresión y extraer ejemplos ilustrativos. En el análisis cualitativo, nos interesa sobre todo observar qué bases y qué afijos son portadores de valor especializado, qué tipo de verbos derivados son los más característicos de la genómica desde el punto de vista morfológico, semántico y sintáctico y qué observaciones pueden reconvertirse en estrategias para la extracción automática de información.

2. Distribución de los procesos de formación

De los 2516 lemas verbales distintos identificados en el corpus, 245 son verbos derivados, 110 compuestos y 2161 presentan tan sólo marca de flexión, por tratarse de verbos simples o derivados por conversión. Véase la distribución con porcentajes en la figura 1.

No obstante debemos tener en cuenta que la distribución según procesos de formación no mantiene proporcionalidad con el número de ocurrencias de cada lema, ya que tan sólo los 6 verbos con mayor número de ocurrencias acumulan el 26,65% del total, oscilando entre el 11% del verbo *ser* (13715 ocurrencias) y el 1,32% de *producir* (1648 ocurrencias). Siguiendo este orden decreciente de lemas por número de ocurrencias, entre los primeros 200 verbos, que presentan de 1202 ocurrencias (0,97% del total) a 124 (0,10%), no localizamos ningún verbo compuesto y únicamente 21 verbos derivados de origen latino y plenamente lexicalizados³.

³ *Utilizar* (1123 ocurrencias), *realizar* (666), *identificar* (551), *describir* (431), *codificar* (413), *representar* (376), *sintetizar* (371), *denominar* (371), *reconocer* (341), *analizar* (318), *localizar* (313), *descubrir* (274), *transmitir* (229), *transcribir* (196).

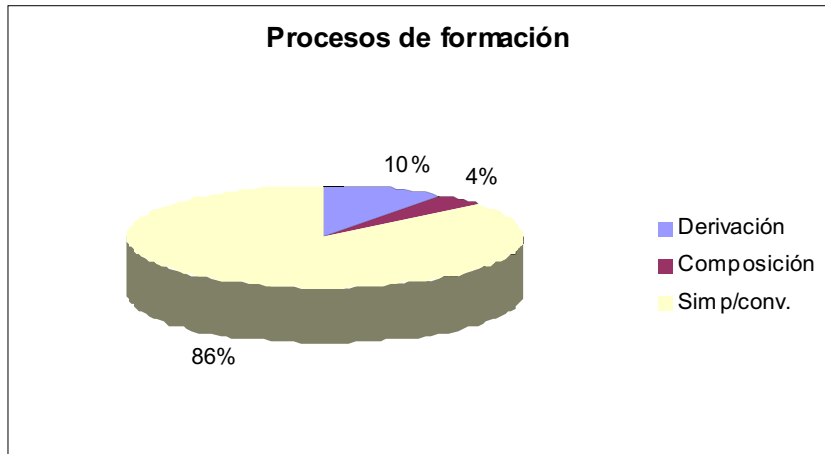


Figura 1. Procesos de formación

Constatamos pues con estos datos que el porcentaje de verbos contruidos por derivación y por composición (14% de lemas) se reduce aun más si tenemos en cuenta las formas reales que ocurren en el corpus textual.

Así el porcentaje más alto de lemas y también de formas flexionadas u ocurrencias corresponde a verbos simples y derivados por conversión. Aunque las herramientas informáticas de consulta no nos permiten dilucidar automáticamente entre unos y otros, nos interesa revisar manualmente la lista de lemas para extraer aquellos que son derivados por conversión de sustantivos o adjetivos y que presentan un significado específico en genómica o en ámbitos vecinos, como la genética, la ingeniería genética, o la farmacogenética. El resultado son 70 verbos derivados, con flexión de la primera conjugación, que analizamos en el apartado 3.

Revisamos también los compuestos en el apartado 4, y los derivados, en el 5, distinguiendo los prefijados de los creados por sufijación, como se muestra en la figura 2.

Aunque las frecuencias de verbos creados por conversión, derivación y composición no sean muy altas en nuestro corpus, si comparamos con el resto de lemas verbales simples usados, consideramos que su análisis puede aportarnos datos interesantes sobre la productividad de los procesos de creación en terminología específica. Recordamos que los análisis de frecuencia en corpus especializados son siempre relativos, ya que al lado de la terminología propia de una especialidad encontramos, con frecuencias mucho más altas, unidades léxicas comunes. En el caso de

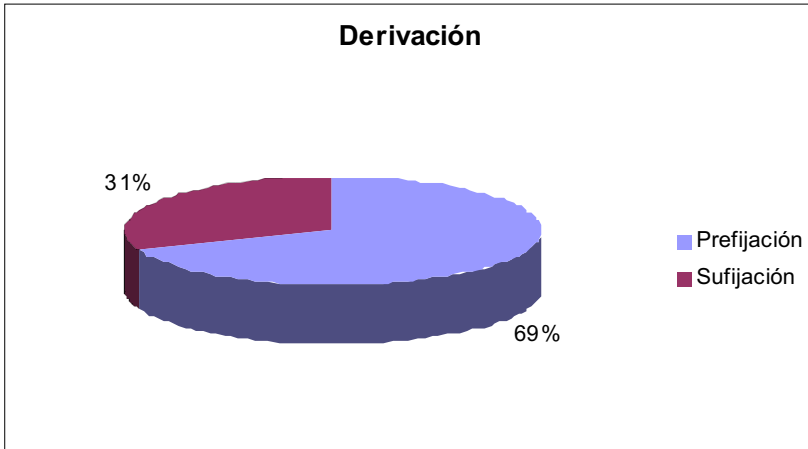


Figura 2. Derivación

los verbos, las frecuencias más altas de formas y lemas se concentran en verbos conectores (copulativos y relacionales) y en verbos discursivos (psicológicos, de ordenación y actos de habla), mientras que los verbos fraseológicos y los verbos-término, aunque con frecuencias de aparición menores, caracterizan, al lado de la terminología nominal y adjetiva, el discurso especializado semántica y pragmáticamente (Lorente 2002).

3. Conversión categorial

La revisión de los 70 verbos derivados por conversión⁴ nos permite observar que, con la excepción de tres bases adjetivas (1), el resto presenta bases de categoría nominal.

- (1) *amarillo/-a* → *amarillar*
azul → *azular*
hibrido/-a → *hibridar*
 Convertir algo en A

Entre las bases nominales podemos distinguir substantivos concretos y usados comúnmente con el mismo significado. Todos estos ejemplos no presentan en los contextos de nuestro corpus diferencias semánticas respecto de la definición general, que en (2a) es ‘poner N en un lugar’,

⁴ Los llamados verbos parasintéticos son tratados en el apartado de prefijación.

en (2b) ‘construir N’, en (2c) ‘excretar N’ y en (2d) diversos significados lexicalizados, aunque los casos de (2b) del corpus de genoma se corresponden con sentidos figurados.

- (2) a. *aceitar, salar, drogar*
 b. *alcantarillar, balizar, orillar*
 c. *orinar, ovular, salivar*
 d. *anillar, barajar, fichar, formular, laminar, parasitar*

Tanto los derivados deadjetivos de (1) como los denominales de (2) están incorporados en diccionarios generales de la lengua con estos mismos significados. Sin embargo, localizamos en el corpus diversos verbos con bases nominales comunes (3) que no han sido entradas en diccionarios. Las de (3a) no presentan formas alternativas, por lo cual si se comprueba que su uso es estable deberían ser incorporadas en diccionarios, generales o especializados. Por otro lado, las de (3b) sí que disponen de formas alternativas (3c), que podrían bloquear su formación, excepto en los casos en que la nueva forma aportara un significado también nuevo.

- (3) a. *horquillar, huevar, orejar, ratonar*
 b. *ladrillar, vinagrar, alcoholar, humanar*
 c. *enladrillar, avinagrar, alcoholizar, humanizar*

Pero la conversión más productiva de nuestro corpus se da con bases nominales referidas a sustancias químicas, algunas de las cuales están incluidas en diccionarios con el mismo significado o con significados cercanos (4a), mientras que otras no constan en repertorios lexicográficos (4b) o presentan formas alternativas (4c).

- (4) a. *azufrar, clorar, fosfatar, hidratar, hidrogenar, oxidar, oxigenar, parafinar, sulfatar, sulfurar; acetilar, metilar*⁵
 b. *alelar, aminor, carbonar*⁶, *clorurar, fosforilar, oxalatar;*
 c. *nitratar* (en el diccionario, *nitrar*)

Nos interesan particularmente los ejemplos de (4b) ya que, con la

⁵ No documentadas en diccionarios generales, pero sí en diccionarios enciclopédicos.

⁶ Aunque en el diccionario podemos localizar *carbonizar* y *carbonear*, referidos a la base *carbón*, y *carbonatar*, referido a la base *carbonato*, el *carbonar* del corpus deriva de la base *carbano*.

excepción de las nominalizaciones *aminación* y *fosforilación* que están documentadas en diccionarios enciclopédicos, no constan en repertorios lexicográficos y su significado no es predecible. Como podemos comprobar mediante los ejemplos de (4a) y también los de (2), la conversión mediante el morfema de flexión de la primera conjugación no presenta resultados sistemáticos semánticamente: mientras que algunos ejemplos son transformaciones o cambios de estado, otros simplemente expresan cambios de lugar.

Parece que las bases de (4) no presentan restricciones fonológicas: en (5a) unidades de acento llano, terminadas en *-e*, *-o*, *-a*, *-eno*; en (5b) sustantivos paroxítonos, terminados en *-o*, *-eno*.

- (5) a. *azufre, cloro, fosfato, hidrato, parafina, sulfato, sulfuro, acetileno, metilo; alelo, amina, carbono, cloruro, oxalato, nitrato*
 b. *hidrógeno, óxido, oxígeno*

Cuando las bases nominales son de carácter eventivo (6), el problema de la direccionalidad de las reglas de conversión categorial o de sufijo cero se manifiesta con más evidencia. Rechazada la comprobación de carácter diacrónico, por las limitaciones de nuestro corpus textual y por la falta de sistematicidad que presentan las fuentes lexicográficas, sobre todo con entradas terminológicas, el criterio de precedencia semántica (Scalise, 1994:272) no nos resuelve el problema, ya que con bases eventivas no parece plausible hipotetizar una de las categorías como previa a la otra. Sólo para algunos de los ejemplos de (6b) parecería adecuado establecer el criterio de adyacencia de sufijos, por el cual podríamos derivar *diseccionar* de *disección*, pero la prelación entre *bloqueo* y *bloquear* no sería clara. Además este criterio no tendría ninguna base semántica o pragmática, únicamente sería de tipo formal. Preferimos para todos los casos de (6) la formulación de reglas de redundancia léxica, que pongan en relación bidireccional formas construidas de categorías distintas.

- (6) a. *anestesiarse, biopsiarse, gangrenarse, herniarse, infartarse, nausearse*
 b. *bloquearse, bombardearse, diseccionarse, distorsionarse, dimensionarse, visionarse, copiarse, seriarse, secuenciarse*

Especialmente interesantes son algunos ejemplos de lemas verbales detectados en el corpus de genoma humano, que a priori se presentaban como depositarios del conocimiento especializado más ligado al proyecto genoma humano. Notemos que la mayoría de ejemplos que hemos

analizado hasta ahora nos remiten a los ámbitos de la medicina (*anestésiar, infartar*) y la química (*carbonatar, sulfurar*), o bien son de uso general (*fichar, salar*). Casos como los de (7) parecían reflejar algunas de los eventos propios de la secuenciación del genoma humano (7a), de la genética (7b) o de la ingeniería genética (7c).

- (7) a. *mapar, mapear, seriar, secuenciar, tipar, zonar*
 b. *heredar, hermanar, ligar*
 c. *modificar, manipular, clonar, copiar*

Lamentablemente, aunque eran verbos posibles, la revisión manual nos ha permitido comprobar que algunos de estos verbos han sido fruto de una lematización incorrecta⁷: son los casos de *hermanar, mapar, tipar* y *zonar*. Aunque se trata de lemas correlacionados con un número importante de ocurrencias⁸, todas y cada una de ellas se corresponden con formas flexionadas sustantivas, en función de núcleo de sintagma nominal o en función de modificador, como vemos en los contextos de (8).

- (8) <doc_codi m00222>: *de recombinación distinta a las anteriores tiene lugar entre cromátides ##hermanas## de un mismo cromosoma .</s>*
 <doc_codi m00190>: <s>*La relativa facilidad con la que se pueden elaborar ##mapas## genéticos bacterianos contrasta con las enormes dificultades con las que*
 <doc_codi m00222>: <s>*El genoma mitocondrial consta de un solo ##tipo## de molécula de DNA localizado en las mitocondrias*
 <doc_codi m00203>: *moscas deben tener células de macho en una pequeña «##zona## de arranque», situada en la parte*

El resto de lemas seleccionados sí que están bien lematizados como verbos, aunque conviene destacar que algunos de ellos se comportan como verbos defectivos en nuestro corpus. Todas las ocurrencias de *clonar, secuenciar, y seriar* son formas de participio pasado e infinitivo.

⁷ Errores derivados de un análisis morfológico con doble etiquetado, que han sido desambiguados inadecuadamente por el desambiguador de base lingüística o por el de base estadística.

⁸ En el caso de *tipar*, 167 ocurrencias, que realmente son formas correspondientes a la categoría nominal.

4. Composición

Los verbos formados por un proceso de composición morfológica, detectados en el corpus de genoma humano, son absolutamente minoritarios. Esta constatación nos ofrece un argumento sólido para rechazar algunas generalizaciones no comprobadas que frecuentemente aparecen en textos de referencia en terminología y discurso de especialidad. Afirmaciones como que la composición de base culta es un recurso léxico muy productivo en los ámbitos discursivos de las ciencias experimentales y biológicas deberían tomarse con mucha precaución si no han sido confirmadas por datos reales extraídos de corpus, ya que a menudo se trata de generalizaciones a partir de ejemplos del discurso médico. Consideramos que cada actividad profesional o científica construye su propio discurso y su terminología según las necesidades comunicativas y de conceptualización, además de basarse en tradiciones distintas y en situaciones comunicativas particulares. Así, dentro de la medicina, una rama estabilizada como la anatomía presenta multitud de ejemplos de composición nominal y adjetival, mientras que otras ramas más recientes como la microcirugía, con un alto componente tecnológico, prefiere la sintagmación como recurso de construcción terminológica (Obrador 1996).

En el discurso de la genómica, la composición parece ser mucho más productiva para los sustantivos referidos a elementos de la bioquímica, pero muy poco productiva en otros casos. Los lemas verbales compuestos extraídos de nuestro corpus se reparten casi al 50% entre ejemplos de composición culta (9a) y calcos del inglés, poco usuales y muy concentrados en algunos documentos del texto, lo que nos hace sospechar que se trata de errores de traducción (9b).

- (9) a. *aminoacilar, carboxilar, catabolizar, fagocitar, hidroxilar, inmunomediar, inmunoprecipitar, inmunorrelacionar*
 b. *anticotón-deber, azúcar-basar, azúcar-fosfatar, azufre-azufrear, base-basar, carbono-carbonar, colín-fosfatar, espermatozoide-surgir, genes-representar, muta-clonar, mutación-derivar*

Dejamos de lado los ejemplos de (9b), en los cuales la expresión de la flexión en el formante de la izquierda nos indica que no se trata de composición morfológica, sino de base sintáctica. Por otro lado, la relación sintáctico-semántica entre el núcleo y el modificador parece clara en los casos de *espermatozoide-surgir* (Sujeto + V) o de *genes-representar* y *mutación-derivar* (Objeto + V), pero no es transparente en el resto de ejemplos. Precisamente al tratarse de calcos de traducción, conven-

dría dilucidar la semántica de estos compuestos para que los organismos de estandarización terminológica propusieran formas alternativas genuinas, compuestas o sintagmáticas.

Los ejemplos de (9a) siguen dos reglas de composición distintas. *Aminoacilar*, *carboxilar*, *catabolizar*, *fagocitar* e *hidroxilar* son verbos formados por conversión de sustantivos compuestos a su vez por dos formantes cultos o semipalabras (10a); en cambio, los compuestos con el formante *inmuno-* contienen como núcleo un verbo libre, de significado común no especializado (10b).

- (10) a. $[N + N]_N + \text{Flex} \rightarrow [[N+N]_N]_V$
 b. $[N] + [V] \rightarrow [N + V]_V$

Otros ejemplos de compuestos, localizados en el corpus de genoma, pero de uso y de significado común, siguen el esquema de (10a), *quin-taesenciar*, *radiografiar*, y el de (10b), *maniatar*.

5. Derivación

5.1. Verbos prefijados

Como hemos visto en la figura 2, la prefijación es el proceso más productivo en los verbos construidos del corpus de genoma humano en castellano, con 245 lemas formados a partir 27 prefijos, incluidos los parasintéticos.

Observemos en la figura 3 la distribución de lemas para cada prefijo. Además de los prefijos propios de la parasíntesis (*a-*, *en-*) y del prefijo de negación (*des-*), los prefijos más productivos son *re-*, *trans-*, *sobre-*, *pre-*, *auto-* y *co-*.

Los formados con los prefijos *re-*, *sobre-*, *co-* y *auto-* (11a) se construyen a partir de bases verbales libres, usuales en el discurso no especializado. En cambio, los prefijos *trans-* y *pre-* (11b) combinan mayoritariamente con bases de categoría verbal ligadas, y sólo hay ejemplos escasos en que las bases sean verbos libres (11c).

- (11) a. *reabsorber*, *redistribuir*, *resecuenciar*; *sobrealimentar*,
sobreexcitar, *sobrevivir*; *coexistir*, *coincubar*; *autoduplicar*,
autopropulsar, *autoreplicar*
 b. *transcribir*, *transmitir*, *transpirar*; *prescribir*, *prescindir*
 c. *transplantar*, *transponer*; *premutar*, *predecir*

Entre estos verbos prefijados hay dos ejemplos, *transmembrar* y

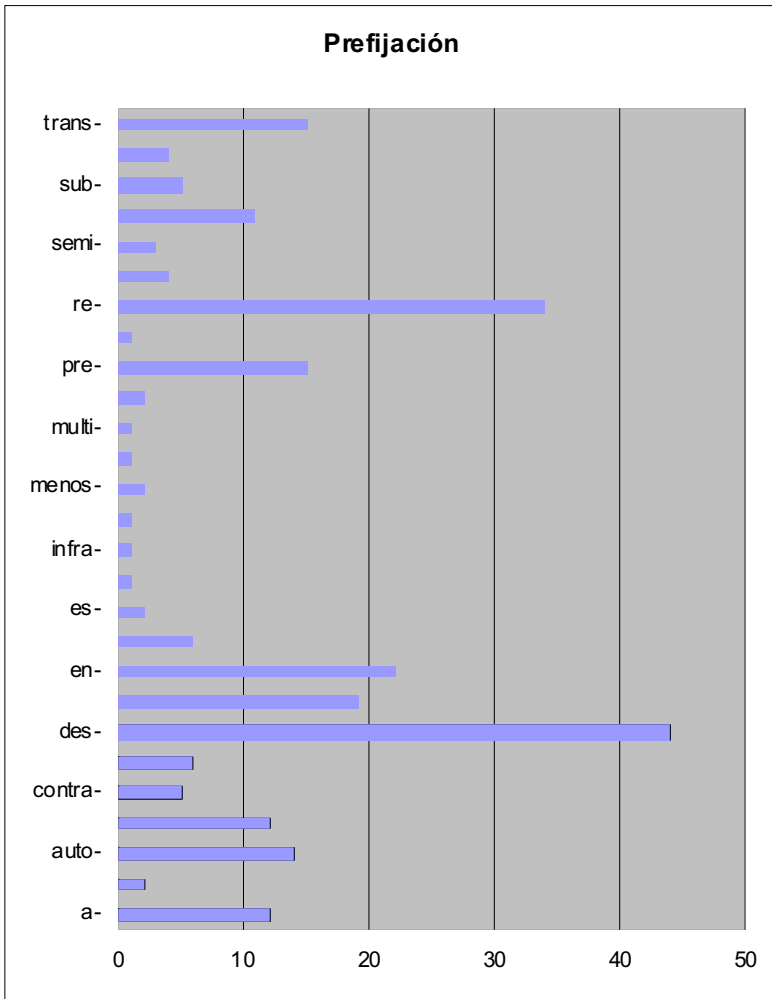


Figura 3. Prefijación

coadyuvar, en los que las bases son aparentemente verbos libres pero que no han sido documentados como tales en el corpus. En el caso de *coadyuvar*, documentado también en diccionarios, la base verbal no es objeto de entrada lexicográfica, por lo que podría tratarse de un derivado defectivo, con base posible pero no usual. Por su parte, no pudiendo documentar el verbo *membrantar* en corpus ni en diccionarios, podríamos pensar que se trata de un caso de verbo prefijado (o parasintético) de base nominal.

Centremos, precisamente ahora, nuestra atención en los llamados parasintéticos. Aunque coincidimos con Scalise (1984) y Corbin (1987) en la utilidad de las bases posibles para explicar algunos procesos de formación, como el ejemplo anterior de *coadyuvar*, consideramos que su uso indiscriminado para mantener un análisis dualista en estos casos es innecesario. Los lemas de nuestro corpus especializado con prefijos *a-* y *en-* son verbos muy usuales en el discurso no especializado. Con muy pocas excepciones de sufijación derivativa en *-ecer* (*enpequeñecer*, *ententecer*, *enriquecer*, *adormecer*), la mayoría presenta marca de flexión de la primera conjugación como los verbos derivados por conversión (12). Compartimos con Jackendoff (1990) el tratamiento de estos derivados sobre la base de marcas de direccionalidad, y consideramos que los prefijos *a-* y *en-* son marcas semánticas preposicionales, que indican dirección y localización, en procesos de formación de unidades léxicas que expresan cambio de estado o de lugar.

- (12) a. *aquietar, abaratar, achicar, adelgazar, afinar; empeorar, enfriar, engordar*
 b. *acompañar, aconsejar, aprovechar, acuñar; empalmar, empaquetar, emparejar, emparentar, encabezar, encadenar, encajar, encaminar, encapsular, encarrilar, encriptar, enganchar, engarzar, enlatar, enraizar*
 c. *adentrar*

Como vemos en (12b) la prefijación con *a-* y *en-* se da de manera más productiva con bases nominales, mientras que es más reducida con bases adjetivas (12a) y puramente testimonial con preposicionales (12c).

La derivación de verbos con prefijos de carácter cuantificador⁹ (*hiper-*, *infra-*, *meno-*, *minus-*, *hipo-*, *sub-*, *super-*, *sobre-*, *semi-*, *multi-*, *pluri-*) son, en términos generales, poco productivos, con la excepción de *sobre-*, *sub-* y *semi-*. Estos tres prefijos, con una presencia en el corpus de genoma que oscila entre tres y once lemas distintos, se caracterizan por combinarse con bases verbales que, aunque son de uso y significado común, están más próximas temáticamente a la actividad de la genética (13a) que el resto de verbos derivados mediante este tipo de prefijos valorativos (13b).

⁹ Algunos de estos prefijos, como *super-*, *sobre-*, *sub-*, pueden tener también un valor locativo, pero los tratamos a parte, siguiendo Gràcia *et al.* (2000).

- (13) a. *semiconservar, semiprocesar, semipurificar; sobreexcitar, sobreexpresar, sobrevivir; subclasificar, subclonar, subcultivar, subdividir*
 b. *hipertrofiar, infravalorar, menoscabar, minusvalorar, multiplicar, superponer*

Otras de las cuestiones que quisiéramos destacar aquí se refiere a los prefijos de contenido espacial (*trans-*, *entre-*, *retro-*, *super-*, *sobre-*)¹⁰. Teniendo en cuenta la limitada productividad que presentan la mayoría de los prefijos de la Figura 3, parece destacable el papel de estos afijos que aportan valores locativos en la globalidad del discurso de la genómica¹¹. Los cinco prefijos se combinan con bases verbales de uso general, en su gran mayoría libres (14a) y sólo en algunos casos se trata de formas lexicalizadas derivadas ya en latín (14b).

- (14) a. *transplantar, transponer, transfundir; entrecruzar, entrelazar, entremezclar, entresacar; retrohibridar, retrotraer, retrotranscribir; sobrepasar, sobresalir; superponer*
 b. *transgredir, transpirar, transducir; retroceder*

5.2. Sufijación

La formación de verbos mediante sufijación se reduce, en los documentos consultados sobre genoma humano, a la combinación de los sufijos *-ificar* e *-izar* con bases nominales o adjetivas¹². El más productivo parece ser *-izar* con un total de 92 lemas en el corpus, mientras que *-ificar* presenta sólo 15 casos de verbos distintos, aunque conviene destacar que dos de ellos, *codificar* y *especificar*, presentan un número muy alto de ocurrencias (413 y 205, respectivamente).

Prácticamente todos los verbos derivados con *-ificar* del corpus son formaciones muy lexicalizadas, usadas frecuentemente en discursos no

¹⁰ Dejamos de lado el caso de *pre-*, ya que en los ejemplos de nuestro corpus tan sólo presenta un valor temporal.

¹¹ Notemos que muchos de los eventos utilizados específicamente en este ámbito discursivo se refieren a relaciones de carácter espacial. El propio Proyecto Genoma Humano (PGH) se ha construido mediante una metáfora: el mapa del ADN, la localización de genes y funciones en una secuencia. Por ello, nos parece interesante para progresar en la descripción de la terminología propia de este ámbito temático analizar en particular todas las unidades léxicas que remiten a estos contenidos espaciales (mapas, direcciones, series, zonas, movimientos, establecimiento de barreras).

¹² Solamente documentamos dos ejemplos de derivados verbales con el sufijo *-itar*, con bases adjetivas: *debilitar, imposibilitar*.

especializados y no especialmente vinculados con temas biológicos. Predominan los casos con bases nominales (15a) sobre los ejemplos del corpus con bases adjetivas (15b).

- (15) a. *acidificar, codificar, cuantificar, descodificar, especificar, momificar, notificar, ramificar, significar, tipificar*
 b. *justificar, semipurificar, simplificar, verificar*

Los derivados con *-izar* presentan mayor diversidad, siendo muchas de sus bases términos de categoría nominal usados preferentemente en el discurso de la química o de la biología (*polímero, hidrólisis, metástesis, inmune*, etc.), al lado de derivados muy usuales comúnmente. Como en el caso del sufijo anterior, los derivados a partir de bases adjetivas tienen poca presencia (16a). Entre las bases nominales, podemos distinguir entre sustantivos concretos (16b) y abstracciones o eventos (16c).

- (16) a. *esterilizar, desnaturalizar, renaturalizar, permeabilizar*
 b. *desespiralizar, espiralizar, despolimerizar, polimerizar, intrumentalizar, hospitalizar*
 c. *resintetizar, sintetizar, metabolizar, esquematizar, hipotetizar*

6. Características sintáctico-semánticas

La mayoría de verbos derivados por conversión del corpus de genoma son verbos transitivos causativos, que expresan cambio de estado o cambio de lugar, tanto si las bases son adjetivas como sustantivas. De todas maneras, conviene resaltar que muchas ocurrencias del corpus textual se corresponden con formas de participio pasado, lo cual dificulta el análisis de la estructura argumental asociada. En muchas ocurrencias, como *avinagrado, orillado, oxalatado, hidrogenado, bloqueado*, se nos plantea ambigüedad sobre si el verbo es usado en cada contexto como causativo o como incoativo. Los casos más claros se presentan con los verbos agentivos del tipo *anestesiarse, diseccionarse*, los verbos con sujeto lógico experimentante *infartarse, herniarse*; o algunos verbos incoativos como *gangrenarse*.

Entre los escasos verbos formados por composición morfológica hay algunos transitivos, como *fagocitar, catabolizar, hidroxilar*, y los que presentan el formante prefijo *inmuno-* son intransitivos, como vemos en los contextos de (17). Sin embargo la dificultad del análisis de la sintaxis en los verbos compuestos del corpus radica en el hecho de que la mayoría de ocurrencias son formas de infinitivo y de participio pasado,

cuando no son errores de lematización¹³.

- (17) <doc_codi m00246>: cuyo grupo 2'- o 3'-OH libre está ##aminoacilado##
 <doc_codi m00413>: de células de levaduras se ##inmunoprecipitaba## una proteína con el peso
 <doc_codi m00673>: <s>Los extractos celulares se ##inmunoprecipitan## con anticuerpos específicos contra Rac-1

Se considera que la prefijación no modifica categoría, pero se han aportado evidencias a favor de que ciertos prefijos, al cambiar el significado de la base, pueden cambiar su comportamiento sintáctico (subcategorización, pronominalización) y su aspecto (Gràcia *et al.* 2000). No es el caso de los prefijos *re-*, *retro-*, *sub-*, que mantienen la sintaxis verbal de las bases (18).

- (18) <doc_codi m00332>: las mutaciones que ##reactivaban## el gen bacteriano defectuoso necesario para utilizar la lactosa ocurrían
 <doc_codi m00435>: <s>Cuando días después se intentó ##reestimar## a los linfocitos T utilizando células presentadoras
 <doc_codi m00361>: y deben ##subcultivar## se a intervalos frecuentes debido a que la sobrepoblación genera
 <doc_codi m00500>: <s>Los enlaces disulfuro ##subdividen## cada cadena en una serie de dominios homólogos

Sin embargo, los verbos formados con *trans-* presentan un cambio sintáctico-semántico con la aparición de un complemento de dirección o meta con las preposiciones *a* y *hasta* (19), común a otros lemas lexicalizados como *transmitir* o *transferir*. Los lemas prefijados con *auto-* y algunos de los prefijados con *entre-*¹⁴ presentan defectividad sintáctica motivada por el cambio semántico, que reafirma el carácter reflexivo de la primera y el recíproco de la segunda (20).

- (19) <doc_codi m00190>: enzimas que ##transcriben## el ADN hasta ARN (transcriptasas) o que hacen
 <doc_codi m00222>: totalidad de l gen , incluidos los intrones , es ##transcrita## a hnRNA (RNA heterogéneo nuclear).</s>

¹³ Se trata de casos de sustantivos etiquetados como verbos, con la misma base o lexema.

¹⁴ Esta defectividad sólo se refiere a los prefijados con *entre-* de valor recíproco, porque hay casos de derivados con *entre-* con valor locativo o de intensidad.

<doc_codi m00572>: Dicho vector se **transfecta** a una línea celular empaquetadora que expresa las proteínas víricas
 <doc_codi m00341>: algunas secuencias cromosómicas de I Y **translocadas** a l brazo corto de l X.

- (20) <doc_codi m00222>: Hay regiones mucho más propensas a **entrecruzar** se , mientras que existen grandes regiones
 <doc_codi m00198>: Dos de estas hebras poliméricas se **entrelazan** entre si para formar la doble hélice de ADN
 <doc_codi m00246>: a tener en cuenta es si cada estructura concreta puede **autoensamblar** se a partir de sus componentes o
 <doc_codi m00332>: ARN tenían propiedades enzimáticas que permitían a l propio ARN **autocortar** se y empalmar se .</s>

La sufijación que, como hemos visto, actúa sobre bases nominales preponderantemente y es de carácter causativo, ya que los derivados expresar acciones o procesos de cambio de estado. En este sentido, no se distinguen semántica ni sintácticamente de los casos de verbos derivados por conversión. El problema reside entonces en la impredecibilidad del proceso de formación morfológica para crear verbos causativos de cambio de estado a partir de bases nominales. Dejaremos para más adelante esta cuestión, con un mayor acopio de datos contextuales y de caracterización semántica de las bases. Como ilustramos en las ocurrencias de (21), todos los verbos sufijados de nuestro corpus comparten la característica de ser verbos transitivos causativos, que pueden funcionar también en su variante intransitiva inacusativa con un clítico¹⁵:

- (21) <doc_codi m00190>: conocer qué tripletes (o "codones") de nucleótidos **codificaban** cada aminoácido de las proteínas
 <doc_codi m00306>: el material genético que constituye los cromosomas se despliega y **desespiraliza** , dando lugar en el interior de l núcleo a
 <doc_codi m00368>: hélice de l DNA separadas como molde , los nucleótidos **polimerizan** en los extremos 3 de las nuevas cadenas
 <doc_codi m00412>: hasta el punto de que podría **despolimerizar** se si su unión al centrosoma no lo mantuviera

¹⁵ Excepto *codificar* y *especificar*, que siempre aparecen como transitivos o con variantes pasivas.

7. A modo de conclusión

Dejando de lado los resultados del análisis cuantitativo, siempre relativos al corpus, podemos concluir que, para el discurso del genoma humano, las unidades verbales que vehiculan conocimiento especializado más frecuentes son verbos fraseológicos, es decir verbos simples o contruidos que no se distinguen formal ni semánticamente de las ocurrencias homónimas usadas en el discurso común o no especializado. El valor especializado de estas unidades hay que buscarlo fundamentalmente en la combinatoria del verbo con un complemento o con su sujeto. Esto abre la puerta a otro análisis, para un futuro muy cercano, sobre fraseología de base predicativa en el discurso de la genómica.

Los afijos no son identificadores del carácter especializado de los derivados, pero conviene destacar que una línea de investigación posible sería la que ponga en relación prefijos, bases y derivados, con unidades léxicas simples y con unidades sintagmáticas, que denoten significados relativos al espacio.

A pesar de estas primeras conclusiones, hemos señalado que en nuestro corpus también se localizan verbos-término, esto es, unidades léxicas de categoría verbal que incorporan morfológicamente el conocimiento especializado. Lemas como *clonar*, *carbonatar*, *polimerizar* son, entre otros, ejemplos paradigmáticos de unidades terminológicas del ámbito del genoma humano. Los verbos-término formados a partir de procesos de conversión, sufijación y composición comparten las siguientes características:

- bases nominales con valor especializado, es decir términos de categoría sustantiva,
- preferentemente del ámbito de la química y de la bioquímica,
- verbos causativos de cambio de estado y de cambio de lugar,
- y transitivos en su mayoría, con algunas excepciones de verbos intransitivos pronominales que no presenten alternancia.

Con esta breve descripción, que debe ser completada en lo referente a aspectos sintácticos y semánticos, nuestra intención es continuar con algunas propuestas aplicadas sobre resolución de problemas de lematización, adquisición automática de léxico y enriquecimiento de bases de datos terminológicas y ontologías. Pero esto será ya en otra ocasión.

REFERENCIAS BIBLIOGRÁFICAS

- Cabré, M. T. (1994), *A l'entorn de la paraula. Lexicologia general (I). Lexicologia catalana (II)*, 2 vols. València: Universitat de València.
- Cabré, M. T. (1999), *La terminología. Representación y comunicación*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Corbin, D. (1987), *Morphologie dérivationelle et structuration du lexique*. Tübingen: Niemeyer.
- Gràcia, Ll., M. Azkárate, M.T. Cabré, S. Varela, M. Lorente, J.C. Odrizola, I. Zabala et al. (2000), *Configuración morfológica y estructura argumental. Léxico y diccionario*. Bilbao: Universidad del País Vasco.
- Jackendoff, R.S. (1975), *Morphological and Semantic Regularities in the Lexicon*. In *Language* 51, 639-671.
- Jackendoff, R. (1990), *Semantic Structures*. Cambridge: The MIT Press.
- Lang, M. (1990), *La formación de palabras en español*. Madrid: Cátedra, 1992.
- Lema. *Diccionario de la Lengua Española*. Barcelona: Spes Editorial, 2001.
- Lorente, M. (2002), *Verbos y discurso especializado*. En *Estudios de Lingüística Española (ELIES)*, 16 [publicación electrónica <http://elies.rediris.es>]
- Lorente, M. (2003), *Construcciones verbales en el discurso de la genómica*. En *Tipología verbal y discurso científico*. Actas del Congreso Internacional de Lingüística El verbo y su entorno: Gramática, léxico, texto (Poznan, Polonia, mayo de 2003) [en prensa].
- Obrador, M.d.M. (1996), *La terminología de las miocardiopatías: Vocabulario y estudio descriptivo*. Facultat de Traducció i Interpretació de la Universitat Pompeu Fabra, tesina de final de llicenciatura (inèdit).
- Scalise, S. (1984), *Morfología generativa*. Madrid: Alianza, 1987.
- Scalise, S. (1994), *Morfología*. Bologna: Il Mulino.

