

UNIDADES DE CONOCIMIENTO ESPECIALIZADO, CARACTERIZACIÓN Y TIPOLOGÍA

M. Teresa Cabré Castellví (teresa.cabre@upf.edu)
Rosa Estopà Bagot (rosa.estopa@upf.edu)

Resumen

En este informe nos proponemos analizar la noción y tipología de las Unidades de Conocimiento Especializado (UCE) para establecer los parámetros pertinentes para su clasificación y determinar qué elementos son relevantes para caracterizar las Unidades Terminológicas (UT) y diferenciarlas de otros tipos de UCE que aparecen con ellas en el discurso especializado. Nos centramos en el ámbito de Genómica, ámbito que ha sido desarrollado específicamente en este proyecto en forma de un recurso digital : la Base de conocimiento GENOMA.

Para realizar completamente este trabajo nos proponemos tomar dos corpus de datos: un corpus lexicográfico especializado del que extraemos un corpus de UT reconocidas como tales por los especialistas, y un corpus textual sobre Genómica, incluido en el banco de textos de la Base de conocimiento GENOMA. En este informe se trabaja únicamente el primer corpus de UT con la finalidad de disponer de las estructuras más representativas de las unidades de Genómica reconocidas explícitamente por los expertos como unidades terminológicas. El corpus lexicográfico de partida es el diccionario de KAUFMANN, U. y BERGENHOLTZ, H. (1998) Diccionario enciclopédico de Ingeniería genética. 2 vols. Toronto, Lugus Libros LatinAmerica Inc.

1. Conocimiento y discurso especializado

Muchos son los autores que se han ocupado de describir las características que convergen en las situaciones que convenimos en denominar *especializadas*. Entre los factores relevantes cabe destacar las condiciones del emisor, del receptor, de la situación, de la temática y de la función comunicativa básica.

Cada uno de los elementos anteriores requiere una breve precisión. De entrada hay que aclarar que en general no se trata de que elementos de unas determinadas características impongan a sus producciones la condición de especializadas, aunque en algún caso la correlación sea más que evidente, sino de constatar que en las situaciones comunicativas que se reconocen externamente como especializadas los elementos que se dan en todo esquema de comunicación presentan características peculiares constantes.

El emisor o emisores son siempre de individuos con un conocimiento importante del tema de comunicación, conocimiento que por regla general han aprendido conscientemente en una situación explícita de aprendizaje. Este conocimiento es el que asegura que un tema especializado sea tratado como tal, como veremos a continuación.

El receptor o receptores de la comunicación especializada son muy diversos y sería absurdo intentar clasificarlos de manera precisa, pero existe una condición que se da en todos ellos: el hecho de tener expectativa de recibir información, de ser informados, expectativa que no se da en otras situaciones comunicativas cuyo propósito general es de otra índole.

En relación al conocimiento que se transmite, podemos decir que solo será especializado si se representa y se transmite respetando escrupulosamente la estructura conceptual del campo en cuestión, ello significa que cada concepto de una disciplina será transmitido sin sobrepasar los límites de sus características, y además salvaguardando las fronteras nítidas existentes entre los conceptos en el contexto de una especialidad. La propiedad de precisión semántica de los términos solo puede respetarse en la comunicación si quien comunica controla bien la estructura conceptual del ámbito (Cabré 2002b).

Este control conceptual, necesario para comunicar un conocimiento especializado, condiciona y explica que en ese tipo de transmisión la función comunicativa primordial sea la informativa y ello a pesar de que en un análisis del texto puedan encontrarse pistas de otras funciones.

Finalmente las situaciones en las que se produce y transmite conocimiento especializado suelen ser profesionales, esencialmente producidas por expertos y dirigidas a un público también experto o semi-experto. Una excepción parece constituir la comunicación científica o técnica con intención divulgativa, pero no es tal si observamos que la transmisión de información especializada siempre cumple la condición de ser el experto o su mediador quien transmite el conocimiento con la finalidad constante de aumentar la competencia especializada en una temática del destinatario (en este caso el lego). En estas situaciones, que no serían las más significativas dentro del esquema de la comunicación especializada, el entorno situacional adquiere rasgos de “profesional” por cuanto quien produce el conocimiento actúa como un experto, aunque sea sólo un mediador actuando como un experto en la transmisión de formas adecuadas de comunicación según las condiciones discursivas (Cabré 2002a).

Todo este conjunto de características permite diferenciar la comunicación especializada en general de otros tipos de comunicación (lúdica, literaria, emotiva, etc.) y encontrar en los textos producidos características estructurales específicas.

2. Los textos especializados y su triple estructuración

La observación de los textos especializados (orales y escritos), considerados especializados por las condiciones discursivas que los caracterizan, nos lleva a establecer en ellos, siguiendo la propuesta de Cabré (2002b), por lo menos la conjunción de tres estructuras:

- a) Una estructura formal, relacionada con el género textual de cada texto
- b) Una estructura informativa, mejor denominada cognitiva, que representa el conocimiento transmitido por el texto
- c) Una estructura gramatical, conformada por la interrelación de las distintas estructuras lingüísticas que conforman el texto.

Las tres estructuras que constituyen un texto están íntimamente imbricadas entre sí de forma que la aparición relevante de determinadas Unidades de Conocimiento Especializado (UCE) o su mayor o menor concentración depende de la estructura formal y de la estructura cognitiva.

Cognitivamente, las Unidades Terminológicas (UT) están todas ellas relacionadas, más o menos explícitamente, más o menos directamente, y se agrupan en conjuntos que pueden ser secuenciales y no secuenciales. El conjunto de las UT y sus relaciones puede representarse topográficamente en forma de mapa cognitivo. El mapa de un texto corresponde a su representación cognitiva. El mapa consensuado sobre una materia corresponde a la estructuración general de un ámbito especializado o a una escuela de pensamiento o acción.

Finalmente la interrelación entre el formato y la gramática condiciona el tipo de estructuras sintácticas, los marcadores, la mayor o menor densidad de léxico especializado, etc.

Veamos a continuación brevemente algunas características de las tres estructuras.

La estructura formal del texto, asociada al género textual al que pertenece, distribuye el texto en partes relevantes que lo caracterizan como perteneciente a un tipo textual determinado. Así, un *paper* científico, por ejemplo, suele estar estructurado según el Convenio de Vancouver, que establecieron en 1978 un grupo de editores y que actualmente apoya el Comité Internacional de Directores de Revistas Médicas (CIDRM), en los siguientes apartados (Junyent, 2003):

Título
Autor o autores y organismo de referencia
Resumen
Introducción o justificación
Materiales y Métodos
Resultados obtenidos
Discusión o conclusiones
Bibliografía

La estructura de conocimiento de una materia, así como la que presenta un texto especializado, se podría representar como un mapa de estructuras de nodos y relaciones.

Los nodos pueden tener distinto alcance y de acuerdo con este criterio corresponder lingüísticamente a fragmentos textuales (por ejemplo las partes de un texto perteneciente a un determinado género textual o un párrafo), a unidades lingüísticas

oracionales, sintagmáticas, léxicas o morfológicas. Las unidades mínimas de un nodo son las léxicas, aunque en determinadas materias las unidades morfológicas marquen el carácter especializado de estas unidades léxicas y constituyan por tanto el foco de su valor especializado. En cada nodo mínimo pues figura una UCE de carácter léxico. Las relaciones están expresadas lingüística o topográficamente.¹

A partir de un fragmento de un texto especializado extraído de una revista médica, podemos observar representaciones de las UCE de distintos niveles con sus relaciones. En la primera ilustración hemos marcado con un círculo los nudos de conocimiento y hemos subrayado las relaciones entre estos nudos. Constatamos además que la relación entre fragmentos textuales se establece normalmente (aunque no exclusivamente) sobre la base de los denominados marcadores textuales. La ilustración 2 presenta la estructura textual mínima de conocimiento y, en ella, podemos observar que los nodos están ocupados por unidades de conocimiento que corresponden a unidades del léxico. La última ilustración es una abstracción del conocimiento que transmite el texto.

La Esclerosis Múltiple (EM) es una enfermedad del Sistema Nervioso Central (SNC) que afecta de forma focal a la sustancia blanca. La lesión característica consiste en una desmielinización segmentaria de la sustancia blanca del SNC, asociada a un infiltrado inflamatorio perivasculare en las fases activas (o agudas) de la enfermedad. Si bien la causa de la enfermedad sigue siendo desconocida, hay numerosos datos que indican que el sistema inmunológico juega un papel importante.

Este fragmento puede descomponerse en las siguientes proposiciones:

La Esclerosis Múltiple (EM) es una enfermedad del Sistema Nervioso Central (SNC)

(La Esclerosis Múltiple (EM)) afecta de forma focal a la sustancia blanca (del SNC)

La lesión característica consiste en una desmielinización segmentaria de la sustancia blanca del SNC

(La desmielinización segmentaria de la sustancia blanca del SNC) (está) asociada a un infiltrado inflamatorio perivasculare en las fases activas (o agudas) de la enfermedad

La causa de la enfermedad sigue siendo desconocida

Hay numerosos datos que indican que el sistema inmunológico juega un papel importante en (la causa de la enfermedad desconocida)

¹ Para profundizar en este punto, véase Cabré (2002b).

La representación conceptual general de este fragmento teniendo en cuenta todas las informaciones del texto podría representarse:

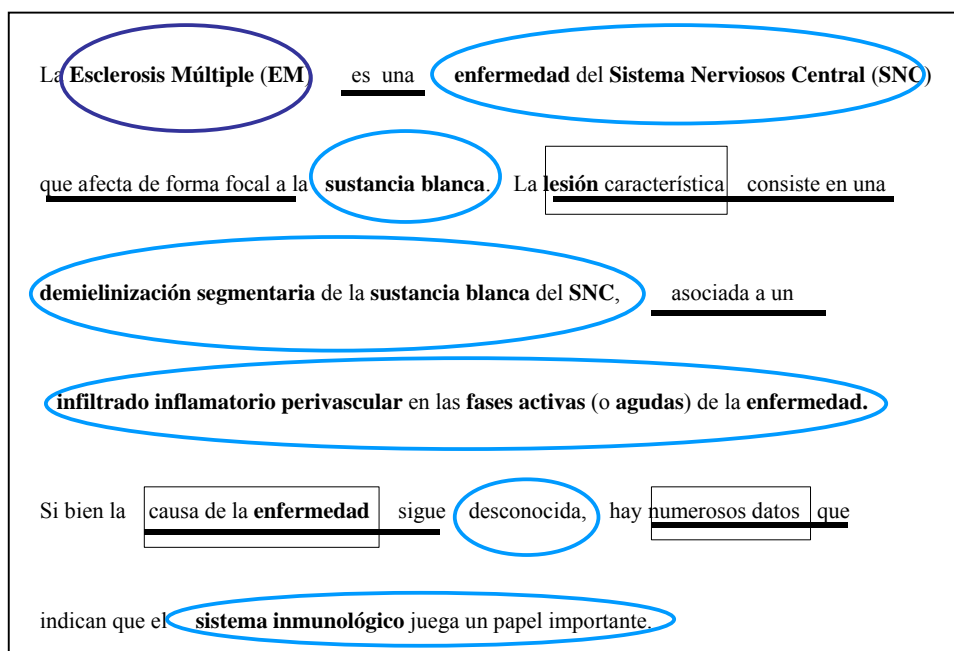


Ilustración 1

Esta representación, si prescindimos de los datos no esenciales para la caracterización del objeto *esclerosis múltiple*, podría reducirse:

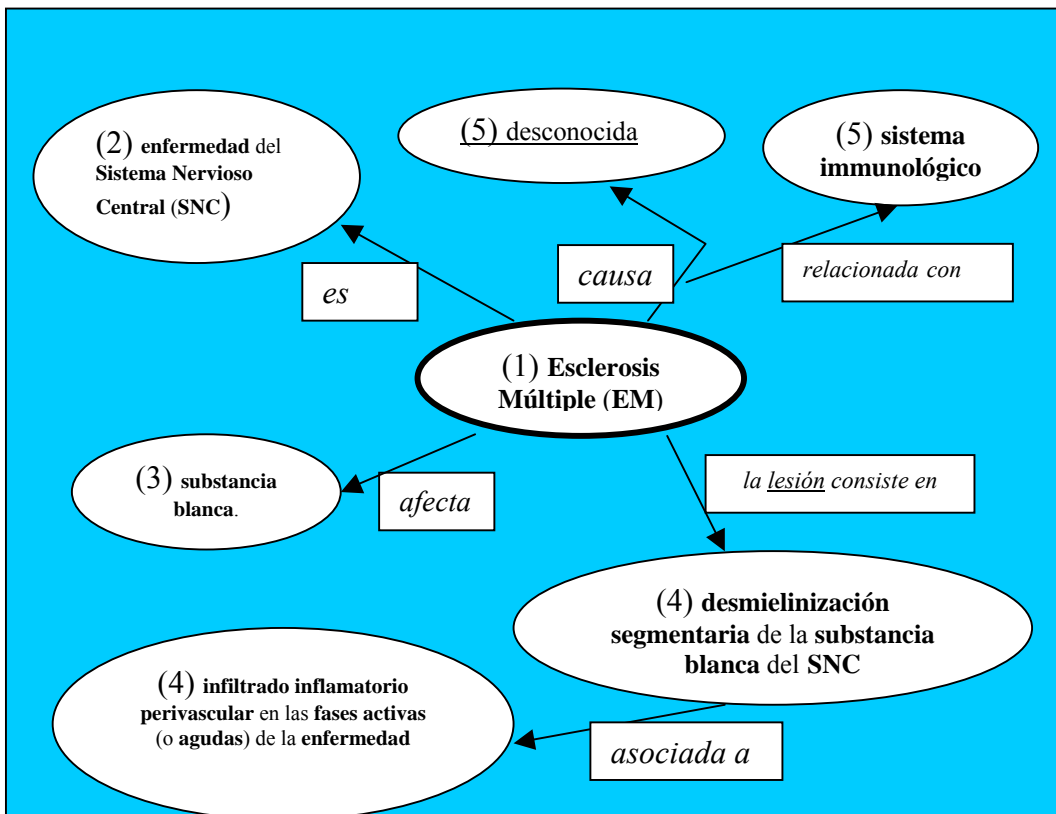


Ilustración 2

En esta representación reducida pueden observarse que la información se organiza en torno a distintas clases de información, clases que corresponden a diferentes aspectos (o dimensiones) del objeto conceptual *esclerosis múltiple* como se muestra en la ilustración 3:

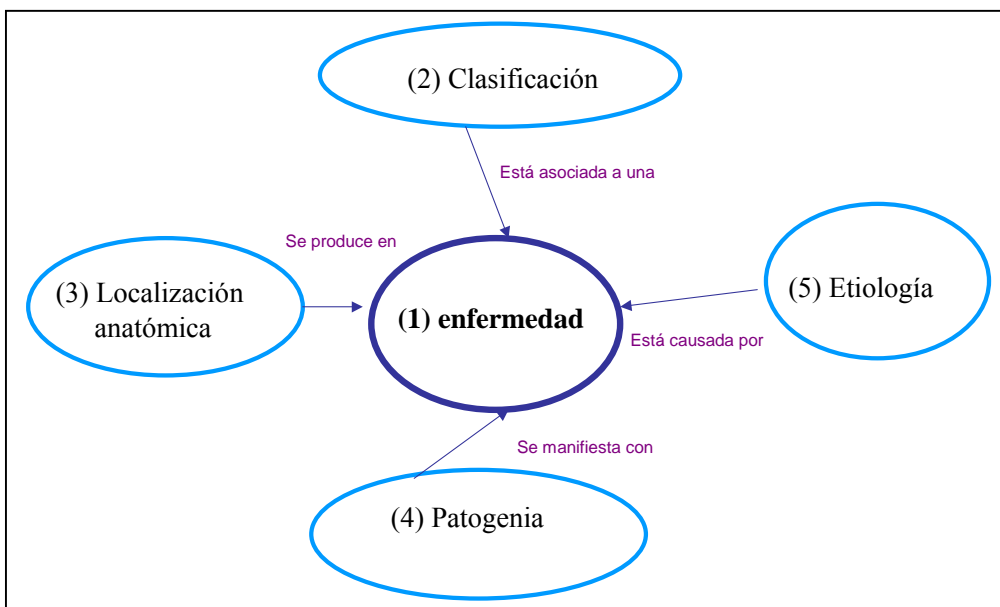


Ilustración 3

Pero si de este fragmento de texto extraemos, no todos los conocimientos que el texto transmite, sino solo los conceptos esenciales en los que se basa el conocimiento sobre el objeto *esclerosis múltiple* (EM), nos damos cuenta que estos conceptos forman parte de la estructura básica del objeto conceptual *esclerosis múltiple*. Se observa también que además todos los conceptos coinciden con unidades terminológicas:

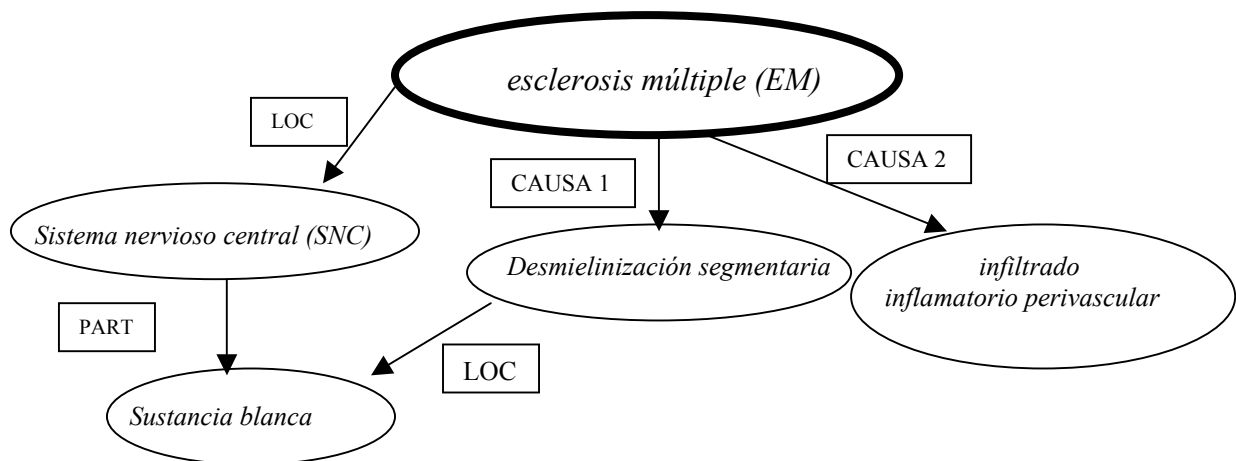
Conceptos específicos clave

Sistema nervioso central (SNC)
Sustancia blanca
Desmielinización segmentaria
Infiltrado perivascular

Conceptos genéricos

Lesión
Fase activa o aguda
Enfermedad
Sistema inmunológico

Los conceptos clave representados como nodos de conocimiento se relacionan de la siguiente manera (en donde LOC significa una relación de locativa, PART una relación partitiva i CAUSA una relación de causa-efecto):



Dejando momentáneamente a un lado la estructura conceptual o cognitiva y observando su estructura gramatical, podemos ver que la estructura gramatical de este fragmento se organiza en distintos niveles lingüísticos:

Un nivel textual (que en este caso correspondería a la totalidad del texto del que procede este fragmento), en el que aparecen distintas estructuras multioracionales y oracionales interrelacionadas por marcadores discursivos².

Un nivel sintáctico, en el que las oraciones se componen de constituyentes estructurales o sintagmas que ocupan un lugar y ejercen una función en la oración.

En el nivel sintagmático infraoracional tenemos combinaciones estructuradas jerárquicamente de unidades del léxico que, en conjunto, pueden corresponder a unidades del nivel léxico inferior o a otras unidades que son simples combinaciones sintagmáticas de discurso. En el léxico tenemos una serie de unidades que estructuralmente pueden proceder de la sintaxis (sintagmas lexicalizados) o ser originariamente lexemáticas.

Todas estas unidades poseen una estructura morfológica, aunque algunas se compongan solo de un elemento morfológico. Por lo tanto, en el nivel de la estructura léxica pueden detectarse además unidades morfológicas de valor especializado.

Aparte del criterio compositivo en la organización estructural de la gramática del texto especializado, se dan muy sistemáticamente algunas características gramaticales que, de modo orientado según los casos, permiten caracterizar específicamente el texto especializado en todos los niveles lingüísticos, o, mejor aun, que permiten contrastarlo con las características de los textos no especializados.

En el nivel textual existe un control bastante sistemático de la distribución de la información, que aparece más o menos concentrada y es de uno u otro tipo en función del género del texto. También existe una organización espacial precisa de los contenidos más rigurosa que en los textos no especializados. Y por encima de todo puede observarse en los textos especializados se da una insistencia en presentar formalmente los conceptos distintos bien diferenciados. Por ello aparecen a menudo gráficos, cuadros, esquemas, y otros recursos organizativos de la información. También presentan, sobre todo en los textos de materias técnicas y de algunas materias científicas de ciencias naturales, médicas, experimentales y exactas (menos en ciencias humanas y sociales), símbolos y abreviaturas de carácter internacional, fórmulas, nomenclaturas científicas y otros signos (pocas veces icónicos) representativos de los conceptos simples o complejos del ámbito. Se observa también en este nivel una tendencia a la impersonalización o bien a la asunción del yo individual por parte del yo colectivo (nosotros), a la presentación de los datos de forma que se perciban como objetivos, a la cita de voces que corroboren la información (polifonía discursiva), etc.

En el nivel sintáctico se detecta en los textos especializados de carácter descriptivo una tendencia a estructuras poco complejas y a oraciones de menor longitud. Determinados tipos oracionales no aparecen, o si lo hacen no son representativos, como, por ejemplo, frases exclamativas o interrogativas. La longitud de las oraciones, sin embargo, suele variar en función del tipo de texto y, por ejemplo, en los de tipo argumentativo la sintaxis es más compleja y las oraciones de mayor longitud. Cabe tratar también como caso excepcional el discurso estereotipado propio de determinados ámbitos, como podría ser el jurídico, de mucha mayor complejidad sintáctica.

² Es en este nivel textual en que puede detectarse el tipo o clase textual a través de la organización de la información y la presencia de marcadores textuales.

En el nivel léxico, que es el nivel que junto con el textual más diferencia el discurso especializado del no especializado, se observa una selección del léxico dependiente de dominio, de léxico ya sea formal y semánticamente, o sólo semánticamente. Formalmente y semánticamente, porque existen vocablos que solo aparecen en el discurso de una determinada especialidad. Sólo semánticamente, porque, aunque se trate de vocablos utilizados también en otras materias o en la comunicación general, estos adquieren un sentido específico dentro de cada dominio especializado, sentido que se ubica en el conjunto de sentidos que conforman el significado global de una unidad léxica³. Desde el punto de vista del léxico, se observa en el texto especializado una mayor proporción de siglas y acrónimos, y a veces de estructuras poco naturales morfológicamente (por ejemplo, el nombre de algunas proteínas como *I-amilasa*, *B-amilasa*, en las que se usa el guión como indicador de un falso prefijo). Finalmente es preciso recalcar que en todos los textos especializados, al lado de las unidades léxicas específicas del ámbito en cuestión, aparece el vocabulario general necesario para trabar el discurso.

En el nivel semántico detectamos una tendencia hacia la evitación de la redundancia en los textos de mayor nivel de especialidad, tendencia que desaparece gradualmente a medida que disminuye el nivel de especialización del texto. Dentro de la redundancia, sin embargo, destaca en los textos especializados el uso de repeticiones explicativas o parafrásticas del tipo UT seguida de definición, explicación o paráfrasis o a la inversa explicación o definición seguida de UT. Estos recursos ya sean UT, ya sean paráfrasis o definiciones, suelen figurar entre paréntesis, aparecer conectadas mediante conjunciones o formar parte de cláusulas explicativas, oracionales o no. Es de destacar también en los textos especializados, como consecuencia de la búsqueda de la mayor precisión la introducción de marcadores denominativos explícitos del tipo Unidad parafrástica + Marcador metalingüístico + UT. La redundancia se manifiesta también en el uso de ilustraciones y gráficos que refuerzan (sitúan, precisan o complementan) determinados conceptos o estructuras conceptuales expresados en el texto.

En el nivel morfológico aparecen morfemas específicos de un ámbito, aunque no en todos los ámbitos con el mismo grado. Podemos observar también una tendencia a la complejidad morfológica por parte de los términos, mayor que en el vocabulario clave de los textos no especializados, y una tendencia mayor a la nominalización de los verbos dando al texto un carácter más designativo y aportando un ingrediente mayor de abstracción.

Y finalmente, aunque la ortografía de los textos especializados es la común de la lengua en la que están elaborados, cabe notar en algunas materias algunas grafías no corrientes, como la repetición o alternancia de mayúsculas y minúsculas en las siglas, contraviniendo la ortografía de las palabras (*anti-VIH*, *ADN circular*), el uso de letras de otros alfabetos (*rayos γ*) o la presencia simultánea de letras y números en un mismo vocablo (*I-amilasa*, *ácido C terminal*, *acetil-CoA*).

3. Unidades de Conocimiento Especializado: Tipología general

³ Ha sido en esta acepción que en algunos trabajos, como en Estopà (1999, 2002), se ha utilizado la sigla USE (unidades de significación especializada) para referirse a estas unidades.

La observación del texto priorizando el aspecto de conocimiento que transmite permite establecer elementos de distinta naturaleza en la estructura de conocimiento y sus elementos, los nodos y las relaciones.

Cuando el foco de nuestra observación se ubica en la relación entre los elementos de conocimiento que conforman la estructura cognitiva del texto y las unidades que vehiculan este conocimiento entramos ya en el terreno de la lingüística. En este plano, forma y significado devienen indisociables.

Desde este punto de vista, observamos que tanto las zonas de conocimiento del texto como los nudos cognitivos se corresponden con unidades lingüísticas, sean grupos de proposiciones relativas a un punto o núcleo de información o elementos que constituyen estas proposiciones con el rol de predicados o argumentos.

3.1 Cuestiones de terminología

Denominamos genéricamente *UCE (Unidades de Conocimiento Especializado)* a las unidades de distinto nivel descriptivo que constituyen los nudos de conocimiento de un texto o forman parte de ellos. La condición esencial para considerar que una unidad es una UCE es el tipo de conocimiento que transmite (condiciones cognitivas y semánticas) y su uso en discurso (condiciones pragmático-discursivas).

Presentamos a continuación los términos y siglas más relevantes de este texto, tanto los referidos a la clasificación como a los elementos de análisis:

Unidad terminológica (UT): unidad léxica, cuya estructura corresponde a una unidad léxica de origen o producto de la lexicalización de un sintagma, que posee un significado específico en el ámbito al que se asocia y es necesaria en la estructura conceptual del dominio del que forma parte. Las condiciones que cumple una unidad para ser una UT son tres:

- a) estructura
- b) especificidad semántica
- c) necesidad en la estructura conceptual

Unidad sintagmática o unidad poliléxica: unidad compleja de estructura sintáctica, que en función de su grado de lexicalización, puede corresponder a una unidad léxica o a una unidad fraseológica (David, 1993). Si esta unidad léxica cumple los requisitos de especificidad semántica y necesidad, corresponde a una *unidad terminológica*. Si cumple el requisito de especificidad, pero no el de necesidad, corresponde a una *unidad fraseológica especializada*, que contiene siempre una UT.⁴⁵

⁴ Véase en este mismo proyecto el informe correspondiente a las unidades fraseológicas (Bevilacqua 2004).

⁵ Corbin (1992, 1997) utiliza el término *unidad polilexémica* para denominar la unidad léxica formada por más de un lexema, e incluye en este concepto todos los tipos de unidades compuestas, tanto las compuestas cultas, como las patrimoniales y mixtas, y las unidades sintagmáticas lexicalizadas. En este texto no utilizamos esta denominación.

Unidad monoléxica: unidad léxica formada por una sola unidad léxica. Las unidades monoléxicas pueden ser simples o construidas, formadas por un solo lexema o por varios lexemas.

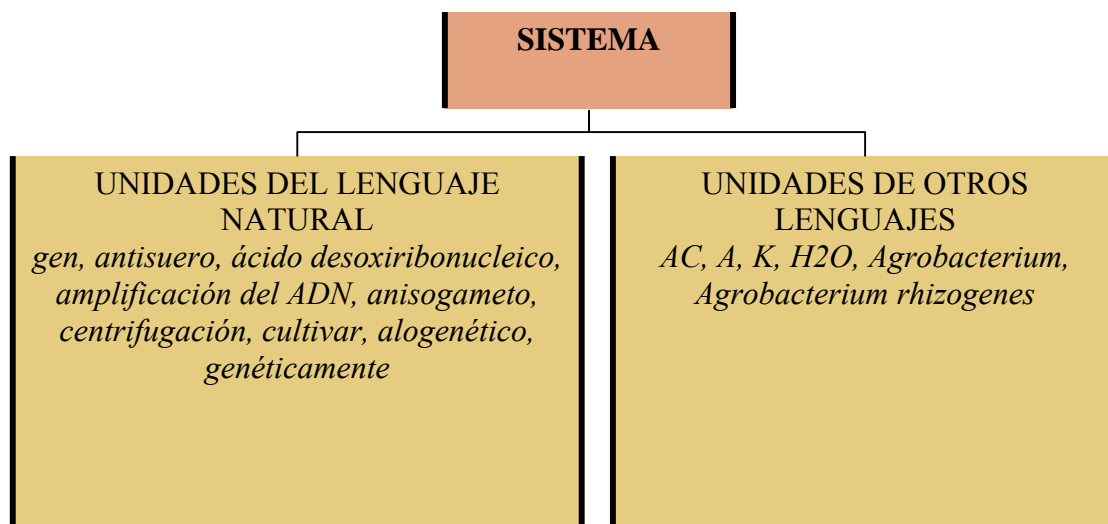
3.2 Criterios de clasificación y tipología de las UCE

Las UCE pueden clasificarse por diferentes criterios. Los seis que en nuestra opinión son relevantes para distinguir distintos tipos de UCE son:

- el sistema al que pertenecen
- la estructura
- la categoría gramatical
- la clase conceptual
- la cohesión interna
- el papel lingüístico-conceptual que ejercen en el discurso especializado

3.2.1 Clasificación por el sistema al que pertenecen

Desde el punto de vista del sistema al que pertenecen, las UCE pueden ser unidades del lenguaje natural o pertenecer a una gran multiplicidad de sistemas artificiales.



Hay que llamar la atención sobre el hecho de que las unidades especializadas representadas por signos distintos de los del sistema del lenguaje natural equivalen en la comunicación a las mismas categorías gramaticales del sistema lingüístico natural, conceptualmente pueden ser descritas a través de las mismas clases semánticas, su estructura puede corresponder a la tipología de las unidades del sistema lingüístico, pueden también presentar distintos grados de cohesión interna si se trata de unidades sintagmáticas y cumplen los mismos roles en el discurso que las unidades lingüísticas.

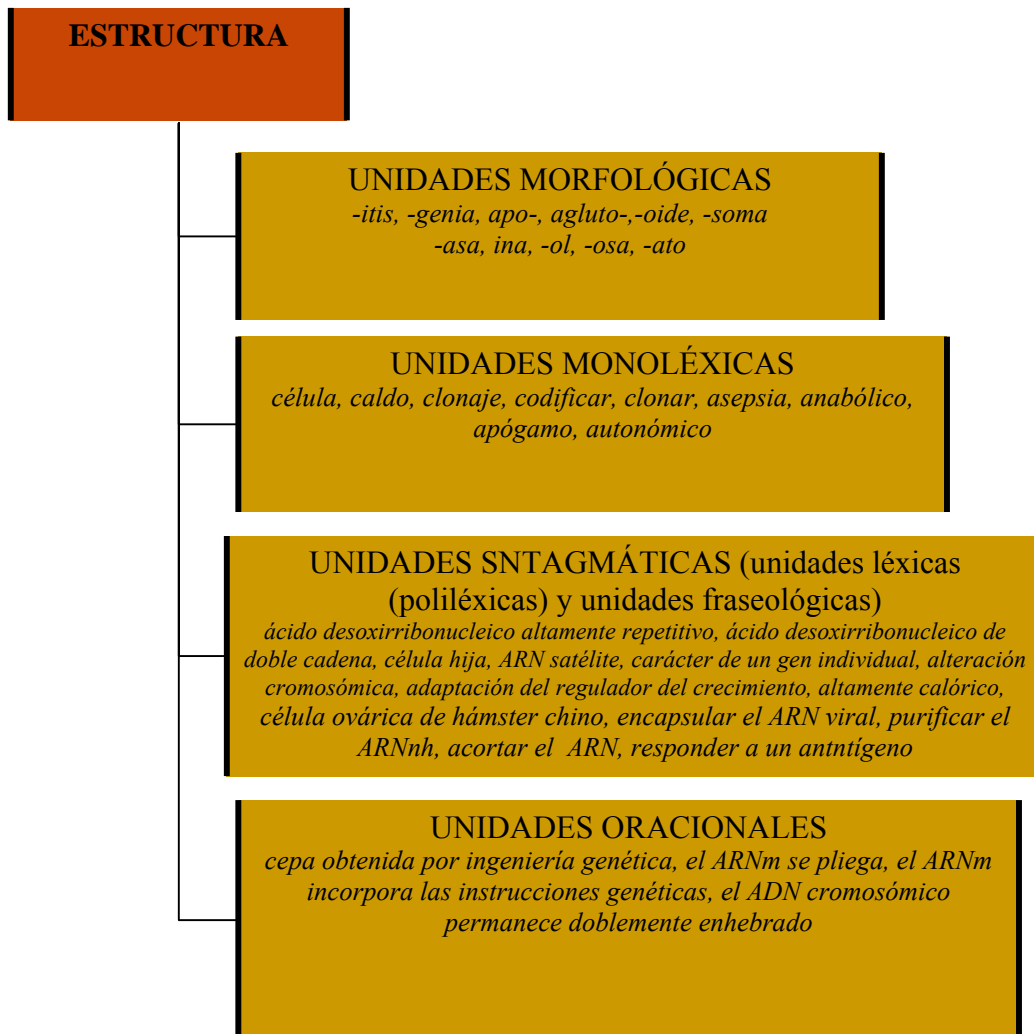
A menudo encontramos en los textos especializados unidades híbridas en cuya formación intervienen elementos de distintos sistemas de signos, combinándose unidades del lenguaje natural con unidades de lenguajes artificiales diversos; sería el caso de los elementos químicos: los sufijos *-oso, -asa, -ico, -ina*, etc. que por su aspecto formal parecen sufijos del lenguaje natural, y, en cambio, son de creación totalmente artificial (*acetato, adenina, adenosina, azarosa, alanina, alcohol, amilasa, aminopeptidasa, anión, aplicón, ampilicina*). Otros ejemplos de unidades híbridas son las combinaciones de símbolos (números y letras) con unidades léxicas (*ácido C terminal, acetil-CoA, ARNm, ADN A, actinomicina D, hepatitis*) o las combinaciones de unidades léxicas de dos sistemas distintos (*I-amilasa, B-amilasa, Agrobacterium mediated*).

Las siglas merecen en el discurso especializado una mención explícita, por tres razones. En primer lugar por su importancia cuantitativa, creciente en la sociedad contemporánea. En segundo lugar por la variación que introducen en los textos cuando estos no son originales sino traducidos. La diferencia estructural de las lenguas, asumida totalmente en la traducción cuando se refiere a la morfología y la sintaxis, se resuelve con vacilación en el caso de las siglas (*ADP, ARN, ADN* junto a *PDA, RNA* o *DNA*). Y en tercer lugar, porque las siglas, en tanto que unidades a caballo entre el léxico de una lengua natural y el de un sistema artificial se combinan con unidades léxicas del sistema lingüístico siguiendo las reglas combinatorias propias de la lengua en la que se usan, tanto si corresponden a unidades propias como a préstamos (*ADN circular, ADN copia ADN nativo, ADN pasajero, ADN polimerasa dependiente del ADN, amplificación del ADN, amortiguador TEB*). Y finalmente, en relación a las unidades abreviadas que aparecen en los textos, podemos observar que, aunque este tipo de recurso parece tener una gran importancia en los textos especializados, un análisis más a fondo sobre las unidades aparentemente abreviadas determina la distinción de dos tipos distintos: por un lado las formas abreviadas propiamente dichas, de carácter léxico y fruto de uno de los tipos de formación de palabras regular en las lenguas; por otro lado, las abreviaturas, formas aparentemente gramaticales que aparecen en las lenguas por consenso y no admiten variación.

3.2.2 Clasificación por la estructura

Desde el punto de vista de su estructura interna, las UCE pueden coincidir con morfemas (*-itis, -ico, apo-, agluto-, -asa, -ina*), con unidades léxicas, simples, derivadas y compuestas (*virus, hepatitis, ocular, autonómico, anabólico, apógamo, cruzar, coagular, amplificar, genético, antisuero, Agrobacterium*), con sintagmas, ya sean terminológicos o fraseológicos (*esclerosis múltiple, nervio alveolar inferior, secuenciación del genoma, altamente calórico, amplificación del ADN, Agrobacterium tumefaciens*), o con oraciones (que son muy escasas).

En el caso de las UCE lingüísticas (morfemas, palabras, sintagmas u oraciones), las categorías o clases estructurales que se le atribuyen deben interpretarse literalmente, de acuerdo con la terminología gramatical. Cada unidad corresponde a un morfema, a un lexema, a una combinación de lexemas o a una oración. Sin embargo, cuando nos referimos a las UCE no lingüísticas, la clase estructural debe interpretarse como "correspondiente a" las propias de las unidades lingüísticas: cada unidad se asocia a una clase que corresponde a un morfema, un lexema, un sintagma o una oración.



Además de clasificarse por su estructura, las UCE léxicas se pueden organizar en subgrupos en función de los elementos internos que componen su estructura.

Por ejemplo, las UCE monoléxicas pueden pertenecer a tres grandes grupos: a unidades simples, a unidades derivadas y a unidades compuestas; y aun entre las derivadas se pueden distinguir las unidades derivadas por prefijación, las derivadas por sufijación y las integradas por sufijos y prefijos conjuntamente⁶. Y las unidades derivadas y compuestas pueden clasificarse de acuerdo a sus modos de formación en derivadas y compuestas cultas, formadas por morfemas o bases léxicas del fondo greco-latino, y derivadas y compuestas patrimoniales, constituidas por morfemas o bases léxicas del fondo actual de las lenguas.

⁶ Esta composición no presupone que la formación se haya realizado adjuntando simultáneamente un prefijo y un sufijo al mismo tiempo.

La clasificación estructural de las unidades léxicas sintagmáticas se lleva a cabo en función de la categoría de las unidades léxicas y gramaticales que las constituyen y las describen como estructuras sintácticas. La mayoría de ellas corresponden a sintagmas nominales, aunque encontramos también sintagmas verbales y adjetivales (y potencialmente sintagmas adverbiales).

A pesar de esta formulación de posibilidades, es preciso aclarar que la mayoría de los sintagmas verbales cuyos componentes co-aparecen con una cierta frecuencia en los textos de una especialidad pueden corresponder, no a unidades terminológicas, sino a unidades fraseológicas propias de esta especialidad. Más problemático es aun el caso de los sintagmas nominales cuyo núcleo es un nombre deverbal relacionado con un verbo semánticamente “fuerte” en el ámbito. En estos casos, para diferenciar lo que correspondería a un fraseologismo o a una unidad terminológica, nos remitimos a Lorente (2003) cuyo trabajo de clasificación de los verbos y nombres deverbales en relación a su carácter fraseológico o terminológico forma parte de los resultados de este Proyecto de investigación.⁷

Pero hay un dato que nos parece incontrovertible: en la representación cognitiva de la estructura conceptual de un ámbito, los nudos centrales están “ocupados” por estructuras nominales (nombres o sintagmas nominales), ya sean de origen nominal, adjetivo o verbal, hasta el punto que cuando una noción que corresponde a un proceso o a una acción pueda representarse verbalmente o nominalmente, es la forma nominal la que “ocupa” el nudo correspondiente a la estructura conceptual de un ámbito. Este dato permitiría formular dos cuestiones: La primera tiene que ver con los modos de representación conceptual en terminología y con los modos de conceptualización de los expertos (no osamos decir, por falta de datos empíricos, de los modos de conceptualización de los hablantes en general). La segunda nos llevaría a cuestionarnos si es efectivamente a través de unidades de categoría lingüística cómo el experto organiza la estructura de su ámbito⁸; o más finamente, si es universal que sea a través de unidades lingüísticas. El análisis empírico sobre estas cuestiones será objeto de trabajos posteriores.

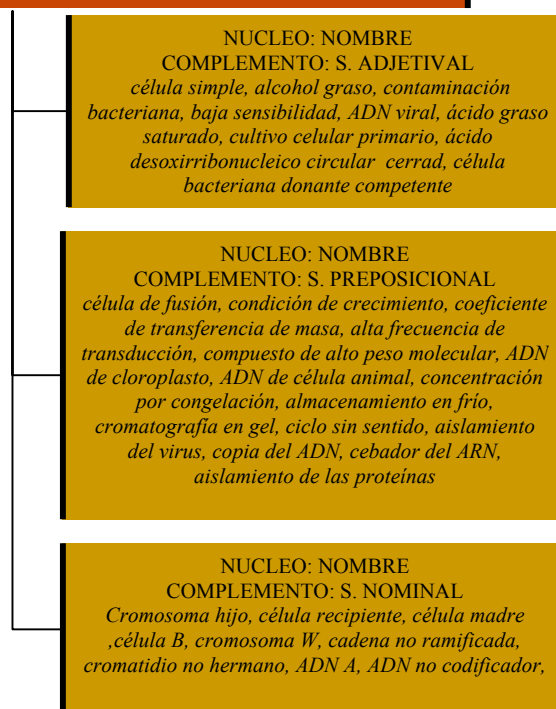
Recuperando ahora las unidades sintagmáticas nominales en la clasificación de las unidades terminológicas, las unidades sintagmáticas nominales pueden corresponder a tres estructuras:

- a) las formadas por un núcleo nominal complementado por un adjetivo o sintagma adjetivo
- b) las formadas por un núcleo nominal complementado por un sintagma preposicional
- c) las formadas por un núcleo nominal complementado por un sintagma nominal.

⁷ Para una descripción de los verbos que parecen en un corpus textual de Genómica, puede verse Lorente 2003.

⁸ En este sentido puede verse la tesis de doctorado de Cleci Bevilacqua, incluida en este proyecto y cuya presentación tendrá lugar en julio de 2004, que muestra cómo en un corpus de tratamientos medioambientales el campo en cuestión se estructura conceptualmente al entorno de nodos verbales.

ESTRUCTURA DE LAS UCE SINTAGMÁTICAS NOMINALES

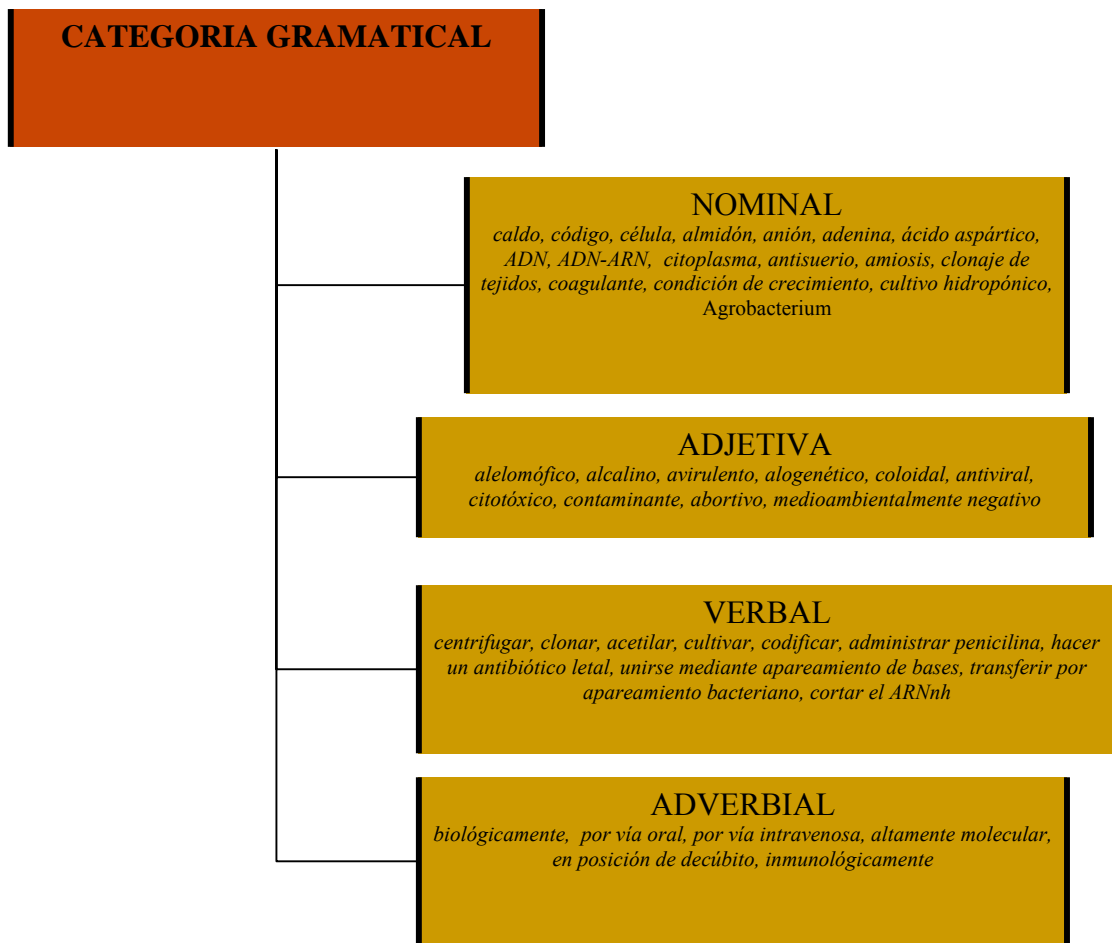


Cada una de estas estructuras puede recibir expansiones, en forma subordinada o a través de la coordinación de alguno de sus elementos, o incluir entre sus componentes esenciales otras unidades, básicamente determinantes o cuantificadores (*alta frecuencia de transducción, actividad fijadora del nitrógeno, adaptación del regulador del crecimiento, ADN circular cerrado covalentemente, ADN moderadamente repetitivo, análisis de inmunosorbentes enlazados a enzimas, ARN nuclear pequeño, cromatografía en capa delgada, cultivo en agar blando*).

3.2.3 Clasificación por la categoría gramatical

Desde el punto de vista morfosintáctico, las UCE léxicas o UT pueden pertenecer a cuatro categorías gramaticales: nominal, verbal, adjetival y adverbial. De entre todas ellas, las de categoría nominal son más representativas, a gran distancia de las de categoría verbal y adjetival. Las UCE adverbiales son relativamente pocas y suelen tener un correlato adjetival en el ámbito.⁹

⁹ Es preciso especificar que aunque en este párrafo hablamos de categoría intrínseca nominal, verbal, adjetival y adverbial, cabría añadir, por un lado, las estructuras sintácticas lexicalizadas y, por el otro, las

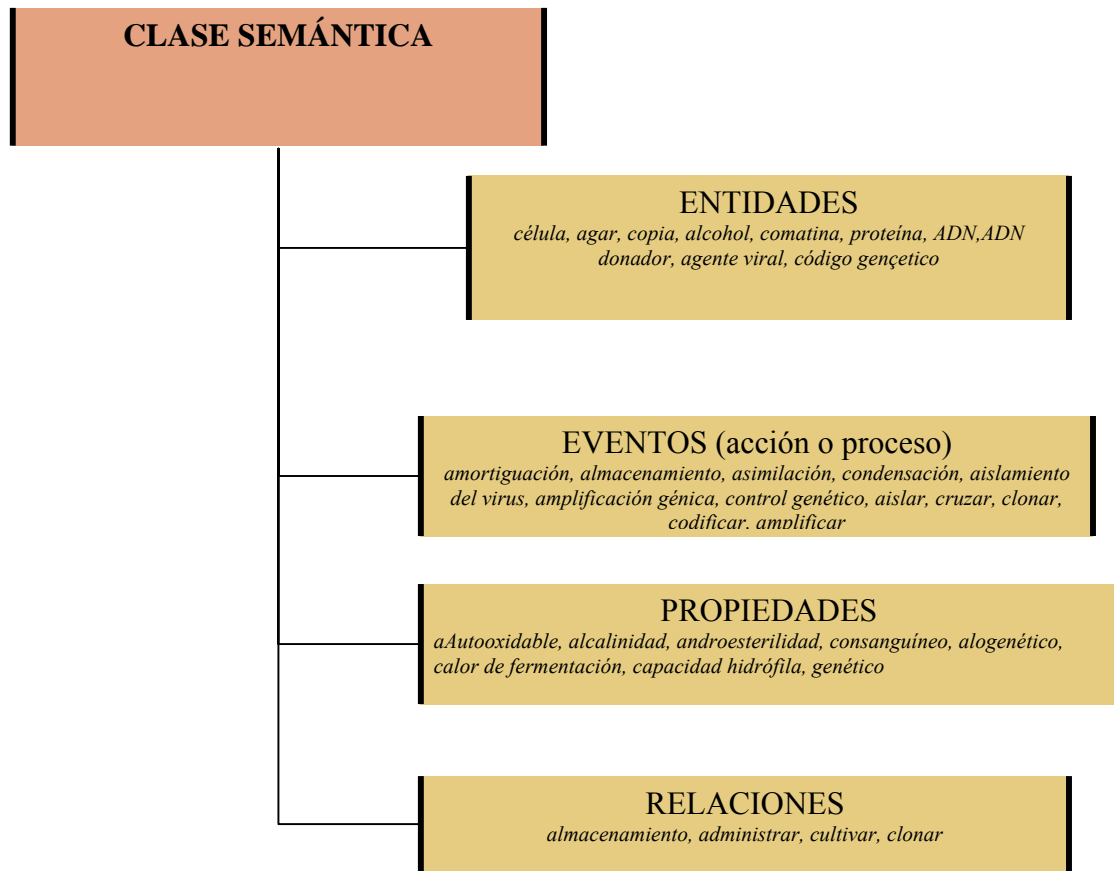


3.2.4 Clasificación por la clase conceptual

Semánticamente, las UCE pueden reducirse a cuatro grandes clases conceptuales, si seguimos la clasificación de Vendler (1967): entidades, eventos, propiedades y

unidades que a pesar de no pertenecer a una clase gramatical pasan a funcionar como propias de esa clase (*por vía oral (adv), analizador (m), antifúngico (m), coagulante (m), concentrado (m)*).

relaciones.



Estas cuatro clases no pueden concebirse ni como una lista absoluta (existen otras clasificaciones que amplían las clases conceptuales básicas), ni como una propuesta discreta. Por un lado estas clases pueden subespecificarse; lo muestra la propuesta del propio Vendler (1967) que, refiriéndose a los verbos, establece cuatro clases aspectuales basadas en los criterios de procesividad y telicidad: estados, actividades, realizaciones y logros¹⁰.

Por otra parte, estas clases se resisten a describir muchas veces aquellas unidades que están a caballo entre dos clases y participan por tanto de las características de ambas, a veces con el predominio de la una sobre la otra, pero en algunos casos sin elementos que permitan decantarse claramente hacia una u otra opción, ya que ésta puede depender del punto de vista que el analista priorice. Así, por ejemplo, unidades como *cáncer*, *aborto bacteriano*, *clonaje forzado*, *clonaje pasmídico*, *clonación in vitro*,

¹⁰ Ciercia & McConnell-Ginet (1990: 353) definen cada uno de estos subtipos:

States are like snapshots of the world at a given instant. They lack a natural culmination or end point, and their subject is perceived not as an agent (as doing something) but as an experiencer (as experiencing something). (...) Activities share with states the property of lacking a natural culmination. Yet they are agentive in that they typically involve a subject doing something. (...) Telic eventualities have a natural end point or culmination.

agroinfección pueden considerarse eventos (acciones o procesos) o entidades (resultados) en función del punto de vista que se priorice; *actividad génica*, *cambio evolutivo*, *cambio genético* pueden ser considerados eventos y propiedades, y *centrífuga refrigerada* podría ser entidad y propiedad).

Cada clase conceptual puede desplegarse además en subclases que *heredan* todas o algunas de las propiedades de la clase y subclases superiores jerárquicamente. Este principio constituye la base de las clasificaciones jerárquicas y de las ontologías especializadas. Veamos a continuación algunas muestras:

a) Sager i Kageura (1994/1995) formularon a principios de la década de los noventa una propuesta de clasificación jerárquica de los conceptos especializados a partir de cuatro categorías muy básicas: entidades, actividades, cualidades y relaciones:

| ENTIDADES | ACTIVIDADES | RELACIONES | PROPIEDADES |
|--|--|--|---|
| <u>Materiales:</u> animales (gente, organismos). <u>Clasificadores:</u> elementos simples, elementos complejos, conceptos separados, parte de un término complejo <u>Abstractos:</u> estructura de conocimiento (general, específico, lingüístico, matemático); entidades abstractas (según la temática); entidades subjetivas (roles, evaluativas, psicológicas). <u>Representacionales:</u> Entidades de documentación (tipos, partes). | <u>Unidades de actividad:</u> actividades específicas (acción/proceso, transferencia, cambio de estado, producción) | <u>Etiquetas de tipos de relaciones:</u> Generales, según el área temática, medida en el área temática <u>Valores de relaciones:</u> Físicas (espaciales, temporales, funcionales) No físicas (modo de relación, comparación, situación relativa). | <u>Etiquetas del tipo de cualidades:</u> general, según el área temática, medida en el área temática <u>Valores de cualidades:</u> Físicas (espaciales, temporales, funcionales, de situación, de forma, de color, cuantitativas, escalares, otras) No físicas (generales, especiales, naturales, ratificales, simples, complejas, negativas, de manera según el área temática). |

b) Una de las clasificaciones semánticas de la unidades léxicas más utilizadas en proyectos de extracción de información actualmente es la ofrecida por el proyecto WORDNET [<http://www.cogsci.princeton.edu/~wn/index.shtml>]. A continuación ofrecemos una muestra de esta clasificación y la búsqueda en su base léxica de la unidad *cell*:

LEXNAMES(SWN) manual page - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

← Atrás → Búsqueda Favoritos Multimedia

Dirección http://www.cogsci.princeton.edu/~wn/man/lexnames.SWN.html

Lexicographer Files

The names of the lexicographer files and their corresponding file numbers are listed below along with a brief description each file's contents.

| File Number | Name | Contents |
|-------------|--------------------|--|
| 00 | adj.all | all adjective clusters |
| 01 | adj.pert | relational adjectives (pertainyms) |
| 02 | adv.all | all adverbs |
| 03 | noun.Tops | unique beginners for nouns |
| 04 | noun.act | nouns denoting acts or actions |
| 05 | noun.animal | nouns denoting animals |
| 06 | noun.artifact | nouns denoting man-made objects |
| 07 | noun.attribute | nouns denoting attributes of people and objects |
| 08 | noun.body | nouns denoting body parts |
| 09 | noun.cognition | nouns denoting cognitive processes and contents |
| 10 | noun.communication | nouns denoting communicative processes and contents |
| 11 | noun.event | nouns denoting natural events |
| 12 | noun.feeling | nouns denoting feelings and emotions |
| 13 | noun.food | nouns denoting foods and drinks |
| 14 | noun.group | nouns denoting groupings of people or objects |
| 15 | noun.location | nouns denoting spatial position |
| 16 | noun.motive | nouns denoting goals |
| 17 | noun.object | nouns denoting natural objects (not man-made) |
| 18 | noun.person | nouns denoting people |
| 19 | noun.phenomenon | nouns denoting natural phenomena |
| 20 | noun.plant | nouns denoting plants |
| 21 | noun.possession | nouns denoting possession and transfer of possession |
| 22 | noun.process | nouns denoting natural processes |
| 23 | noun.quantity | nouns denoting quantities and units of measure |

Web WordNet 2.0 - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

← Atrás → Búsqueda Favoritos Multimedia

Dirección http://www.cogsci.princeton.edu/cgi-bin/webwn2.0?stage=1&word=cell

WordNet 2.0 Search

Search word: Find senses

Overview for "cell"

The noun "cell" has 7 senses in WordNet.

1. **cell** -- (any small compartment, "the cells of a honeycomb")
2. **cell** -- ((biology) the basic structural and functional unit of all organisms; cells may exist as independent units of life (as in monads) or may form colonies or tissues as in higher plants and animals)
3. **cell**, electric cell -- (a device that delivers an electric current as the result of a chemical reaction)
4. **cell**, cadre -- (a small unit serving as part of or as the nucleus of a larger political movement)
5. cellular telephone, cellular phone, cellphone, **cell**, mobile phone -- (a hand-held mobile radiotelephone for use in an area divided into small sections (cells), each with its own short-range transmitter/receiver)
6. **cell**, cubicle -- (small room in which a monk or nun lives)
7. **cell**, jail cell, prison cell -- (a room where a prisoner is kept)

Search for: Domain of senses 2

Show Coordinate Terms

Show Hypernyms (cell is a kind of...)

Show Hyponyms (...is a kind of cell), brief

Show Hyponyms (...is a kind of cell), full

Show Holonyms (cell is a part of...), regular

Show Holonyms (cell is a part of...), inherited

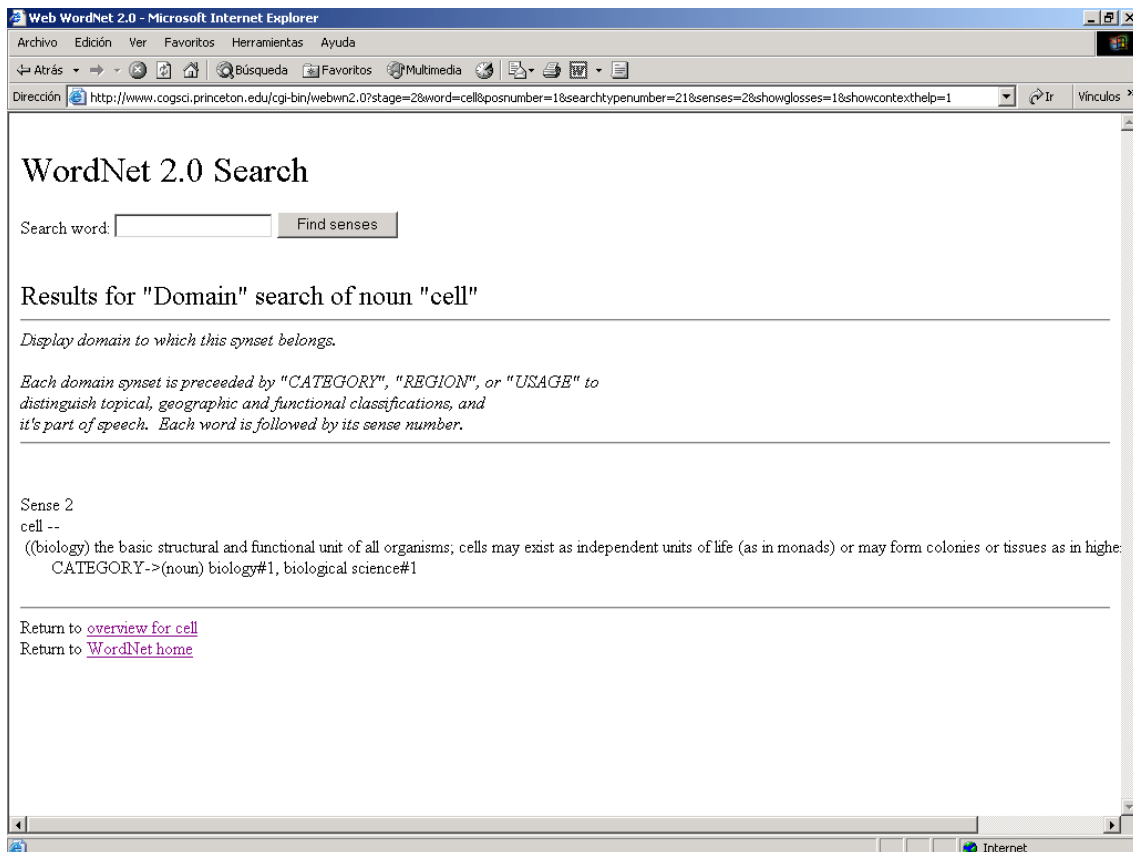
Show Meronyms (parts of cell), regular

Show Meronyms (parts of cell), inherited

Return to: Domain

Domain Terms

Familiarity



c) El proyecto SIMPLE propone una lista de catorce papeles temáticos y rasgos semánticos para la descripción semántica del léxico de las lenguas (<http://www.ub.es/gilcub/SIMPLE/liege/spanish/Spanish.html>). Reproducimos la lista de los papeles temáticos utilizados en este proyecto y un ejemplo de los rasgos que SIMPLE utiliza para describir el contenido de las unidades léxicas a través de la unidad:

| |
|------------------------------------|
| SemanticRole list |
| RoleKinship |
| RoleIconic |
| RoleHeadQuantified |
| RoleHeadCollective |
| RoleProtoAgent |
| RoleProtoPatient |
| Role2Participant |
| RoleOblique |
| RoleLocation |
| RoleDirection |
| RoleOrigin |
| RoleSOAArg |
| RoleAdjunct |
| RoleUnderspecified |

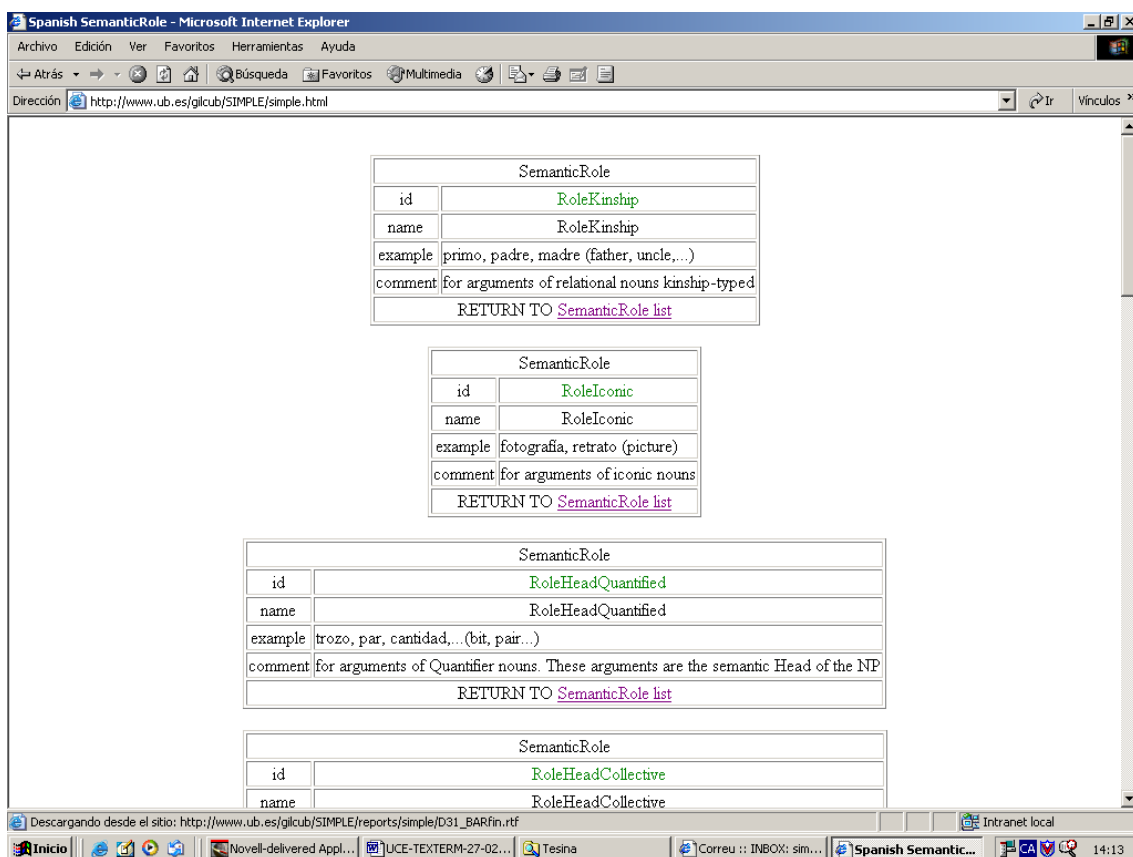
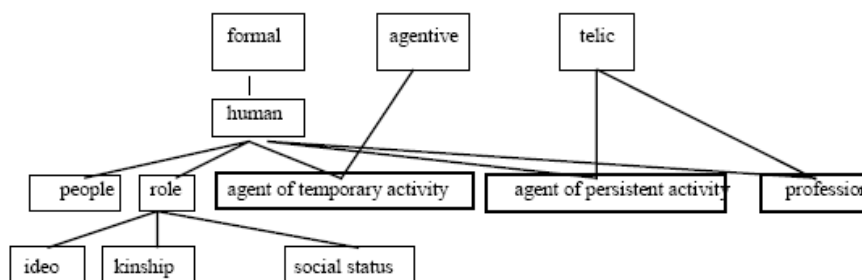


Figure 2: Subset of the SIMPLE Core Ontology representing human beings



Todas las clasificaciones aquí expuestas se han valorado para establecer la ontología del prototipo de banco de conocimiento que se ha desarrollado en el presente proyecto. Este banco, denominado GENOMA, puede consultarse en <http://www.iula.upf.edu>. A continuación, como muestra, se reproduce el resultado de la búsqueda en este banco del concepto *enfermedad genética* a partir del término *enfermedad genética*:

3.2.5 Clasificación por el papel lingüístico-conceptual que ejercen en el discurso especializado

Aparte de la clase semántica, pero relacionada con ella, las UCE representan en la estructura conceptual de un ámbito papeles distintos según constituyan un nodo o se integren en un nodo, o bien representen relaciones entre nodos, o ambas cosas a la vez, como ya ha sido descrito en Cabré (2002b).

De acuerdo con este criterio, las UCE que forman parte de las estructuras conceptuales pueden describirse en forma de conjuntos de dos rasgos: el rasgo *nodo* y el rasgo *relación*. Así tendremos las siguientes combinaciones pertinentes para la representación cognitiva del conocimiento especializado:

| | |
|------------------------------------|-------------------------|
| [+ <i>nodo</i> – <i>relación</i>] | <i>gen, clon,</i> |
| [+ <i>nodo</i> + <i>relación</i>] | <i>clonación, cruce</i> |
| [– <i>nodo</i> + <i>relación</i>] | <i>clonar, cortar</i> |

Los nodos poseen “pesos” conceptuales distintos según su mayor proximidad o alejamiento del nodo central. Cuanto más cercanos al nodo central, mayor peso conceptual; cuanto más distantes, menor peso conceptual.

Otro de los elementos que condiciona el papel de las UCE en una estructura conceptual es su condición terminológica o fraseológica: son o bien representaciones de nodos, o bien forman parte de predicaciones que especifican su función, su utilidad, sus

características, etc. Desde este punto de vista semántico, son estructuras complejas normalmente descriptivas de una acción que ejerce la unidad que ocupa el nodo mínimo, o una expansión o restricción o simplemente especificación de las propiedades de este nodo. De acuerdo con este criterio, en el que intervienen al mismo tiempo condiciones sobre la estructura de las unidades y condiciones sobre la relación y alcance de los nodos, diferenciamos entre, por un lado, las unidades terminológicas, de estructura léxica o sintáctica, pero en este último caso lexicalizadas, fuertemente cohesionadas y representativas de nodos mínimos, y, por otro lado, las unidades fraseológicas, de estructura sintáctica, de grado menor de cohesión interna, y representativas de expansiones de los nodos mínimos. En este contexto, las oraciones especializadas representarían combinaciones libres que contendrían unidades terminológicas y unidades fraseológicas. Existe un solapamiento entre la estructura y la condición de nodo mínimo de las estructuras conceptuales que ha provocado una gran confusión entre los distintos tipos de unidades que hay que retener, así como una gran variación denominativa.

3.2.6 Clasificación por la cohesión interna

En este trabajo, como ya hemos avanzado, consideramos *unidades terminológicas* las unidades de carácter léxico, con entrada en el lexicón del hablante en tanto que especialista del ámbito, de estructura morfológica simple, derivada, compuesta o sintagmática (entre las que se incluyen las siglas y las formas lexicalizadas abreviadas), semánticamente específicas, que constituyen un nodo mínimo de un ámbito (no de un texto).

Y consideramos *fraseología*, y hablaremos por lo tanto de unidades fraseológicas a aquellas unidades pertinentes en un ámbito especializado, de estructura sintáctica (nominal, verbal, adjetival o adverbial), de combinación restringida, que representan expansiones de los nodos mínimos. Entre ellas, las más estudiadas son las estructuras correspondientes a sintagmas verbales, en las que las UT ejercen la función de complemento interno, en la mayoría de los casos, aunque pueden ejercer también más raramente la de complemento externo.

Si estas unidades fraseológicas verbales son muy representativas de un ámbito, suelen alternar con estructuras sintagmáticas nominales formadas por el verbo de base, realizado como nombre deverbal seguido de un sintagma preposicional que introduce el complemento, la mayor parte de las veces determinado. Con menor frecuencia encontramos también otras estructuras que se caracterizan porque el léxico es el mismo aunque bajo estructuras superficiales distintas (*aislar, aislamiento, aislamiento de proteínas, aislamiento de las proteínas, aislamiento del virus, AC aislado; agente alquilante, alquilar, alquilación*).

Como resumen de todo lo expuesto, recalcamos una vez más que sólo consideramos UT las UCE que cumplen los requisitos de estar lexicalizadas, es decir, insertas en la gramática como piezas léxicas, usarse en un dominio temático con un sentido definido y ser necesarias en la representación conceptual de dicho dominio. Estas unidades lexicalizadas se han convertido, de hecho, en unidades léxicas a todo los efectos, aunque

tal vez no tengan todas las posibilidades gramaticales que tienen las unidades originariamente léxicas.

Las unidades de sistemas no lingüísticos que, aparte de su pertenencia a otro tipo de sistema, cumplen todas las demás características de las UT y pueden sustituirlas en un texto, las trataremos como *UT subsidiarias*. Desde el punto de vista funcional, actúan como unidades terminológicas (es el caso de los símbolos o de las nomenclaturas); desde el punto de vista de sus posibilidades discursivas, presentan diferencias en relación a las unidades lingüísticas, ya que, aparte de haber sido acuñadas siempre *in vitro*, en el exterior del discurso y para finalidades esencialmente normalizadoras, son inmunes a propiedades del lenguaje natural como la variación (*adenina o agrobacterium*) y al cambio evolutivo. Estas unidades funcionan prácticamente siempre como nombres, porque son los nombres las unidades más prototípicas tanto por la forma de expresar el conocimiento como por la evidencia psicológica que de ellas tienen los especialistas, como por su número y frecuencia en los discursos especializados, son las unidades de carácter nominal.

4. Análisis descriptivo de las UCE léxicas (UT) del ámbito de Genómica

El proyecto global de descripción de las unidades de conocimiento especializado propias del ámbito de la Genómica se ha dividido en dos fases:

- a) En una primera fase –cuyos resultados se presentan en este informe- analizamos las UCE correspondientes a unidades terminológicas (UT) (unidades léxicas cuya estructura corresponde a una unidad léxica de origen o lexicalizada) procedentes de un corpus lexicográfico especializado de Genómica. Este análisis tiene el objetivo de disponer de las estructuras más representativas de las unidades de Genómica, en tanto que han sido reconocidas explícitamente por los expertos como unidades terminológicas que merecen ser una entrada del diccionario de Genómica que ellos han elaborado. Daremos pues por supuesto que toda entrada –o subentrada- de un diccionario de Genómica elaborado por un experto del ámbito en cooperación con un terminólogo (en nuestro caso el de Kaufmann, U. y Bergenholtz, H. (1998) *Diccionario enciclopédico de Ingeniería genética*. 2 vols. Toronto, Lugus Libros LatinAmerica Inc.) proporciona unidades terminológicas reconocidas explícitamente como tales por los expertos. Se presupone, por lo tanto, que estas unidades forman parte del acervo léxico del conocimiento experto en Genómica

- b) En una segunda fase -que corresponde a la continuación de este trabajo en un nuevo proyecto-, analizaremos las mismas unidades en su contexto discursivo, y por lo tanto dentro de texto y teniendo en cuenta las características enunciativas de dicho texto como discurso. Ello nos dará información sobre la representatividad real en el uso de las unidades codificadas en el diccionario y nos proporcionará la confirmación o refutación de los patrones que se habrían considerado más representativos de las unidades del ámbito de la Genómica¹¹.

En el marco de la primera fase del trabajo, en este informe analizamos las UCE de Genómica procedentes de un corpus lexicográfico extraído del *Diccionario Enciclopédico de Ingeniería Genética*. De todo el corpus, se han analizado sistemáticamente las entradas correspondientes a las letras A y C, en total 1301 entradas que corresponden a un 24% de las entradas de este diccionario.

Hemos analizado solamente las unidades que el diccionario incluía como lema (entrada principal). No obstante cabe señalar que esta obra incluye una cantidad de fraseología nominal y verbal, y de UCE oracionales en muchos de los artículos del diccionario. Estas unidades se reconocen gráficamente por estar introducidas por un pequeño triángulo al final de las entradas. La gran cantidad de combinaciones concurrentes que incorpora este diccionario, que pueden corresponder a fraseología o a simples colocaciones, no se suelen recoger en muchas obras terminográficas, y, en el caso que nos ocupa, pensamos que esta peculiaridad responda al perfil profesional de los autores de la obra:

aminoácido m *amino acid*

Los aminoácidos son compuestos orgánicos que contienen un grupo amino, -NH₂ y un grupo de ácido carboxílico, -COOH. Hay 20 aminoácidos diferentes que constituyen las unidades estructurales de los péptidos y las proteínas. Una proteína de tamaño promedio se compone de unos 300 aminoácidos conectados por enlaces peptídicos. El posible número de secuencias de aminoácidos, en una proteína de 300 aminoácidos, es aproximadamente 10⁹⁰. Por lo tanto, la cantidad de proteínas que puede lograrse es prácticamente ilimitada. Muchas biomoléculas importantes son péptidos, es decir, consisten de unos pocos aminoácidos.

- Además vimos que en cada una de las proteínas, el zipper y la región de enlace del ADN, estaban separadas exactamente por 6 aminoácidos.
▲ cada ~ está codificado por *each ~ is encoded by*; los ~s están incorporados a ~s *are incorporated into*; ser traducido a un ~ *be translated into an ~*

ARNm m mRNA

▲ ~ de célula animal *animal cell ~*; degradar el ~ *degrade ~*; el ~ incorpora las instrucciones genéticas *embodies genetic instructions*; el ~se pliega *the ~ folds*; ~ de la hormona de crecimiento *growth-hormone ~*; el precursor del ~ *the precursor of ~*; ~ purificado *purified ~*; reconocer el sitio de enlace del ~ *recognize the ~ binding site*; una subfracción del ~ *a subfraction of the ~*; ~ truncado *truncated ~*

Las unidades analizadas —las entradas de las letras A y C— se distribuyen de la manera siguiente, si tenemos en cuenta su categoría gramatical:

¹¹ Poseemos sin embargo datos sobre el análisis realizado en el marco de este proyecto sobre los verbos en Genómica (Lorente 2003).

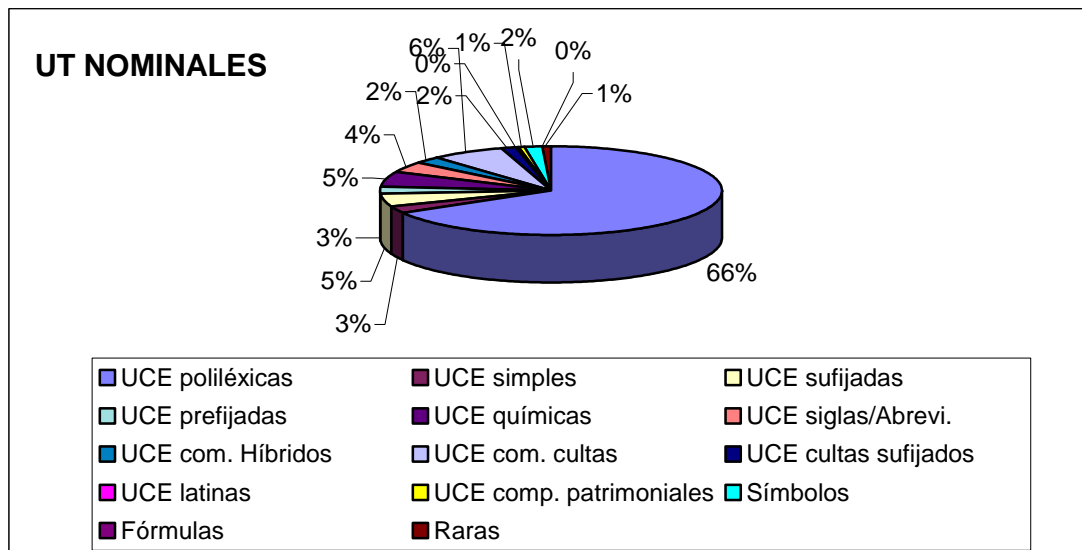
| Categoría gramatical | nominal | Verbal | Adjetival |
|----------------------|---------|--------|-----------|
| Número de UCE | 1189 | 21 | 85 |
| % | 91,39% | 1,61% | 6,53% |

Estos datos confirman la idea de que los expertos asocian la terminología principalmente a los nombres; pero no solo eso, sino que además conciben la estructura básica de una disciplina a través de las unidades lingüísticas nominales, dando pie a la afirmación general de que “conocer una disciplina es conocer su terminología”.

5. Estructuras más relevantes de las Unidades terminológicas del corpus de Genómica

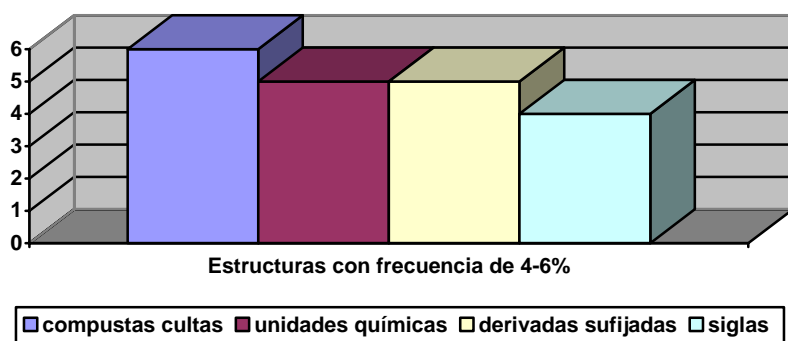
5.1 Las Unidades terminológicas nominales

Desde el punto de vista de su estructura, las 1189 UT nominales del vocabulario de Genómica que se han analizado se distribuyen de la manera siguiente:

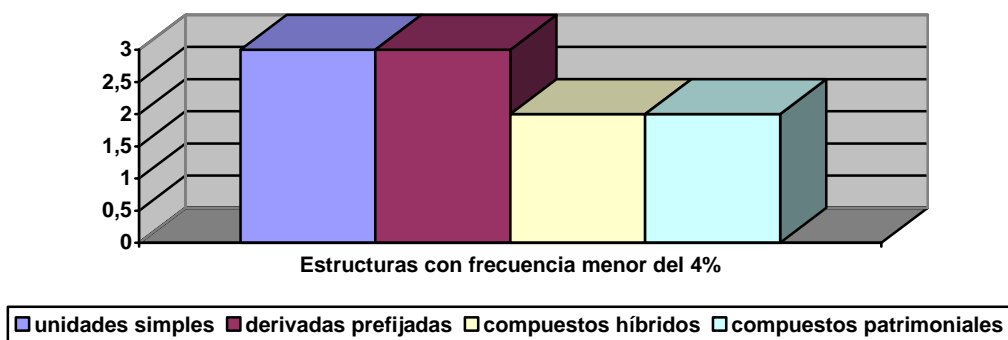


Como puede apreciarse, las UT poliléxicas son estructuralmente las únicas representativas, ya que corresponden al 66% de todas las estructuras. El 34% de los casos se distribuye en nueve tipos de estructura, que en función de su mayor o menor representatividad (representatividad siempre débil) se reparten en dos grupos:

- a) las estructuras que presentan una frecuencia de 4-6% (entre 83 y 33 casos en el corpus analizado, que cubren el 20% del total): estructuras compuestas cultas (6%), unidades químicas (5%), derivadas sufijadas (5%) y siglas (4%):



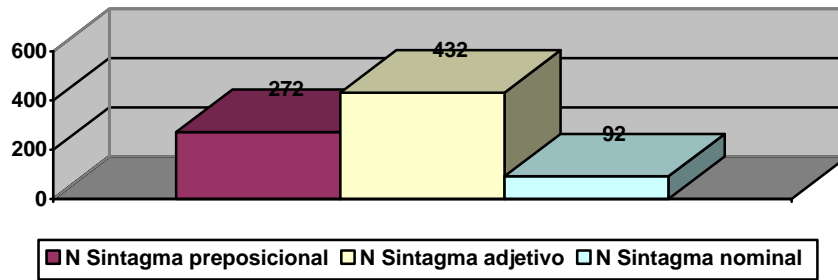
- b) las estructuras que presentan una frecuencia por debajo del 4% (que corresponden a menos de 33 casos y cubren el 13% del total): unidades simples (3%), estructuras prefijadas (3%), compuestos híbridos (2%), compuestos patrimoniales (2%).



Dentro de las 796 unidades poliléxicas, se dan tres estructuras básicas:

- [[N] Sintagma preposicional]: 272 unidades
- [[N] Sintagma adjetivo]: 432 unidades
- [[N] Sintagma nominal]: 92 unidades

Unidades poliléxicas



Cabe notar el número elevado de unidades de estructura [[N] SA] que es la estructura sintagmática más productiva también en la mayoría de terminologías, como han mostrado diferentes estudios. Otro tema es la variación que presenta cada una de estas tres estructuras básicas.

La estructura con más tipos variantes es [[N] Sintagma preposicional]: 272 unidades

- N de N: 99
- N de art N: 24
- N de Npl: 14
- N de N A: 12
- N en N: 11
- N de art sigla: 9
- N de sigla: 7
- N de art N A: 5
- N de A N: 5
- N de N de N: 5
- N de Npl Apl: 3
- N de Npropi: 3
- N en N A: 3
- N por N: 3
- sigla de N: 3
- N en N de N: 2
- A N de N: 2
- N A de A N: 2
- N A de N: 2
- N de art Npl: 2
- N de N de A Npl: 2
- N sin N: 2
- sigla de A N: 2
- N A de N A: 1
- N en N: 1
- N por A N: 1
- N a art N: 1
- N A de A sigla: 1
- N A de art N: 1
- N A de Npl: 1
- N A de sigla: 1
- N A por N A: 1

N A por N de N: 1
 N con N: 1
 N de A N A: 1
 N de art N de art N: 1
 N de art N de N: 1
 N de art N sigla: 1
 N de art símbolo: 1
 N de N A a N: 1
 N de N a art N: 1
 N de N de art Npl: 1
 N de N sigla: 1
 N de N símbolo: 1
 N de Nllatí: 1
 N de N-N: 1
 N de Npl A: 1
 N de Npl Apl: 1
 N de Npl de N A: 1
 N de un N A: 1
 N de una N para Npl: 1
 N en A N: 1
 N en Npl: 1
 N entre Npl: 1
 N in Nllatí: 1
 N por art N: 1
 N por N en N: 1
 N por Npl: 1
 N por Npl de art Npl: 1
 N por Npl de art símbolo: 1
 sigla de art N: 1
 sigla de N A: 1
 sigla N A de art sigla: 1

Algunos ejemplos de estas unidades serían (se puede consultar la lista completa en el CD anexo): *cromatografía en columna, ADN de doble hebra, cromatografía en capa delgada, código de una letra para aminoácidos, código del ADN, código de nucleótidos, condición de fermentación, concentración al vacío, aberración de los cromasomas, control de la contaminación, complejo proteico de ADN, adaptación del regulador del crecimiento, cultivo en agar, cultivo en agar blando, ciclo sin sentido, almacenamiento de las enzimas, cepa obtenida por ingeniería genética, cruzamiento de prueba de tres puntos, ADN polimerasa dependiente del ARN, ADN polimerasa dependiente del ADN, etc.*

Un análisis más detallado de las unidades de estructura [[N] SP] permite observar que en muchas de ellas aparece un nombre deverbal como núcleo del SN y el complemento está determinado por un artículo definido (*aislamiento de las proteínas, aislamiento del virus, almacenamiento de las enzimas, alteración del ADN, amplificación del cloranfenicol, análisis de los comosomas, célula de la membrana. Aberración de los cromosomas*). Estos casos corresponden *grosso modo* a unidades fraseológicas, y no a unidades terminológicas, según nuestros criterios.

La variación es substancialmente menor en la estructura [[N] Sintagma adjetivo]: 426 unidades

N A: 346
Sigla A: 44
NAA: 28
Sigla AA: 3
Sigla Adv A: 2
N sigla A: 2
N símbolo A: 2
NAAA: 2
NA Adv A: 1
Sigla A A adv: 1
Sigla no A: 1

Algunos ejemplos de estas unidades son (se puede consultar la lista completa en el CD anexo): *aborto bacteriano, contaminación microbiana, cadena no ramificada, célula no diferenciada, cromosoma doble diminuto, actividad fermentativa residual, célula bacteriana donante competente, cortes escalonados, ARN autoempalmante, ARNm enmascarado, ARN nuclear pequeño, ADN no codificador, ADN moderadamente repetitivo, etc.*

Finalmente, la estructura [[N] Sintagma nominal] presenta un grado de variantes similar a la anterior estructura pero corresponde a un número de unidades muy inferior: 90 unidades

Sigla N: 26
NN: 25
N símbolo: 16
N sigla: 6
NNN: 5
Sigla Símbolo: 5
NAN: 2
N sigla A: 1
Sigla N N: 1
N A símbolo: 1
N abreviatura: 1
N fórmula: 1

Ejemplos de estas unidades son (la lista completa se encuentra en el CD anexo): *análisis Cot, célula T, ARN satélite, célula madre, cepa Hfr, AP endonucleasa, cromatidio hijo, complejo antígeno-anticuerpo, célula CHO, configuración cis, ADN Z, etc.*

De la misma manera que ocurría con un corpus terminográfico de derecho y de medicina estudiado anteriormente (Estopa, 1996), la mayoría de UT poliléxicas corresponden a un número muy bajo de patrones estructurales y, contrariamente, existe una gran dispersión de estructuras morfológicas que equivalen a muy pocas UTP, de hecho, como también pasaba en los estudios mencionados, la mayoría de patrones

presentan una sola ocurrencia. Así, las seis estructuras siguientes corresponden al 75% de las UT poliléxicas del corpus analizado:

NA (346)
NdeN (99)
Sigla A (44)
NAA (28)
Sigla N (26)
NN (25)
N de art N (24).

Dentro de la lista de UT nominales, llama la atención la baja presencia de unidades léxicas no construidas. Si este dato fuera significativo nos llevaría a suponer que la Genómica posee un vocabulario básico propio extraordinariamente reducido, y que la mayoría de sus conceptos se han denominado a través de unidades formadas por combinación morfológica o léxica. En nuestra muestra (sobre el 25% de las entradas del Diccionario) sólo aparecen 30 unidades no construidas: *aborto, ácido, agar, agua, ala, alcali, alelo, alga, ambiente, angstrong, antera, azúcar, calor, cortes, cáncer, caldo, callo, cap, carbono, célula, cresta, cepa, cigoto, clon, código, codón, colonia, column, copia, cultivo*.

Cabe resaltar también el equilibrio frecuencial entre la composición en su conjunto (10%) y la derivación también en su conjunto (13%), y su contraste en relación a los procesos de truncación, entre los que sólo se da la siglación (4%).

Es interesante notar el grado de productividad de algunas siglas, especialmente de ADN y ARN, que constituyen el núcleo o el complemento de numerosas unidades de carácter sintagmático. Por citar un ejemplo, la unidad ADN, que es quizás la más representativa de este léxico, constituye el núcleo de 62 unidades poliléxicas: *ADN contrasentido, ADN copia, ADN copia simple, ADN cormosómico, ADN de célula animal, ADN de doble hebra, ADN de plastidio, ADN del cloroplasto, ADN del organelo, ADN desnaturalizado, ADN donador, ADN extraño, ADN glicosilasa, ADN circular cerrado covalentemente* y en 15 unidades poliléxicas forma parte de su complemento (*amplificación del ADN, ARN polimerasa dependiente del ADN, cadena del ADN, cantidad de ADN, código del ADN, composición de ADN, cebador del ADN*).

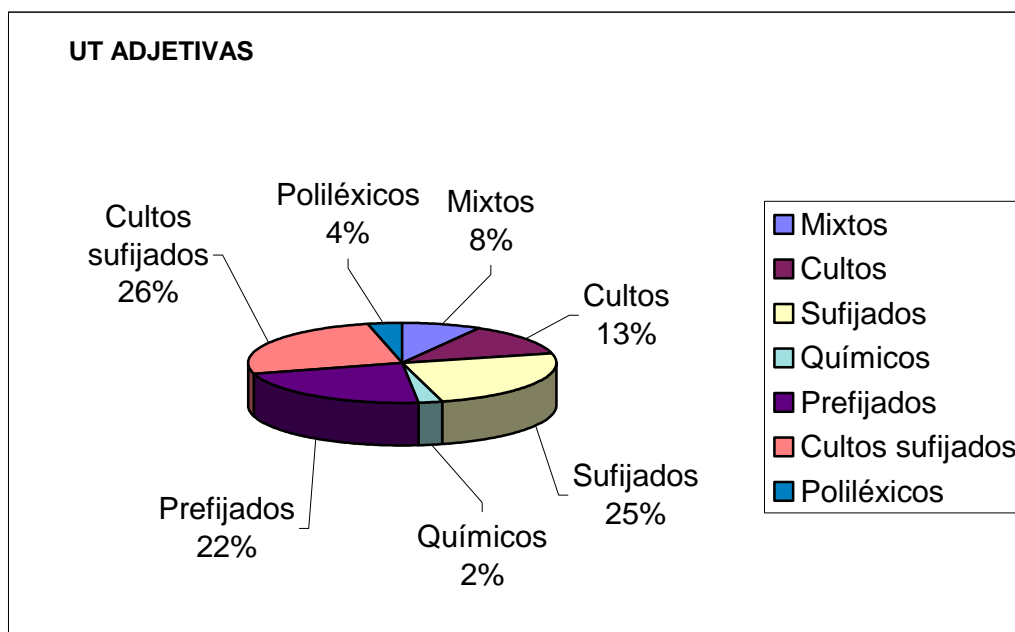
Una mención aparte merecen los casos denominados “UT químicas”, que corresponden a sustancias de la química formadas por sufijos especializados creados por consenso internacional. El sufijo más productivo en nuestro corpus es *-ina* (*abzima, acetilcolina ciclina amilopectina alanina albúmina adenosina*, etc.), seguido de *-asa* (*1-amilasa, acetiltransferasa, amiloglucosidasa, B-amilasa, celulasa*), *-osa* (*agarosa, amilosa, celobiosa, celulosa*) *-ón* (*almidón, aplicón, catión*), y *-ol* (alcohol).

En relación a las clases semánticas de las unidades terminológicas nominales, podemos observar que la mayoría se describen como entidades, aunque dentro de ellas destaca un grupo de unidades nominales procedentes de unidades verbales, que, al heredar el rasgo de eventividad del verbo, permiten etiquetar la unidad nominal también como acciones o procesos. Observamos así cómo algunas unidades, en función del punto de vista o de los rasgos activados, pueden considerarse entidades o procesos: *ciclo ovárico, cinética enzimática, cinética del crecimiento, cinética de la reacción, cinética de la muerte*

térmica, ciclo de infección lítica, ciclo sin sentido, ciclo lítico. O bien unidades relacionadas con adjetivos que pueden considerarse entidades o propiedades: *constante de espiralización, constante de fijación, constante de fijación.*

5.2 Las Unidades terminológicas adjetivas

La frecuencia de las estructuras correspondientes a las UT adjetivas analizadas (86 unidades) es la siguiente:



De entrada observamos que entre los adjetivos no aparece ni un adjetivo simple¹². Esta constatación puede obedecer a dos razones: o bien en Genómica no existen adjetivos propios acuñados en el origen conceptual de la disciplina, y la adjetivación en el ámbito de la Genómica sólo se realiza a través de unidades derivadas o compuestas a partir de los nombres o verbos propios; o bien los adjetivos de estructura simple usados en Genómica, procedentes del fondo común del léxico general o de otra materia aunque asociados a un sentido específico en el ámbito, no se reconocen como terminológicos (en el sentido que no funcionan aisladamente sino siempre se integran en una UT poliléxica formando parte de su complemento), lo que daría una pista sobre cómo los especialistas atribuyen estatus de término a una unidad, es este caso adjetiva¹³.

En los ejemplos analizados observamos que la derivación cubre el 75% de los casos; la composición, reducida a la composición culta o mixta, el 21% de los casos, lo que cubre casi el total de la adjetivación (96%).

¹² Esta misma situación se ha descrito en el caso de los nombres.

¹³ Esta cuestión ya se ha tratado en Estopà (1999) en relación al ámbito de la Medicina. Dejamos para estudios posteriores el análisis de este tema en Genómica.

Entre las estructuras sufijadas no cultas (25%), el total se reparte casi equitativamente entre las que proceden de un verbo (12 casos) y las que proceden de un nombre (9 casos, más dos si incluimos aquí los parasintéticos). Los sufijos que aparecen son los siguientes (con exclusión de los que solo aparecen una vez):

Combinados con bases verbales aparecen los siguientes sufijos adjetivales:

- (a/e)nte
- ivo
- (a)do (correspondiente a la forma de participio pasado de verbo)
- ble

Combinados con bases nominales se presentan las siguientes formas de sufijo:

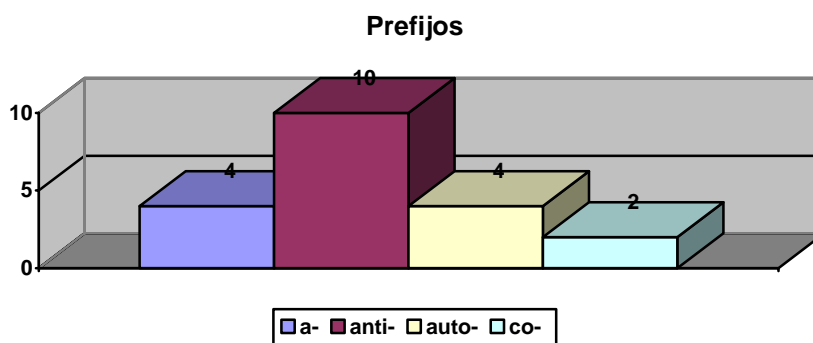
- ar
- ico
- al

Las formaciones adjetivas sufijadas cultas presentan los sufijos siguientes:

- ico (que aparece en el 77% de los casos)
- ivo (dos casos)
- al (un solo caso)
- ista (un caso)

Si a las formaciones patrimoniales sufijadas, les sumamos las estructuras cultas sufijadas y las estructuras con sufijos propios sólo de la Química (dos casos con el sufijo -ino) se cubre el 55% del total de los adjetivos.

Todas las estructuras prefijadas proceden de un adjetivo real en la lengua castellana. A este adjetivo se le antepone uno de los prefijos siguientes: a- (*avirulento, acéntrico, ameiótico*), -anti (*antiviral, antimicótico, antiséptico*), auto- (*autooxidable, autotrófico*), y co(n)- (*consanguíneo, codominante*). Los prefijos tienen distinto grado de productividad, como muestra el cuadro siguiente:

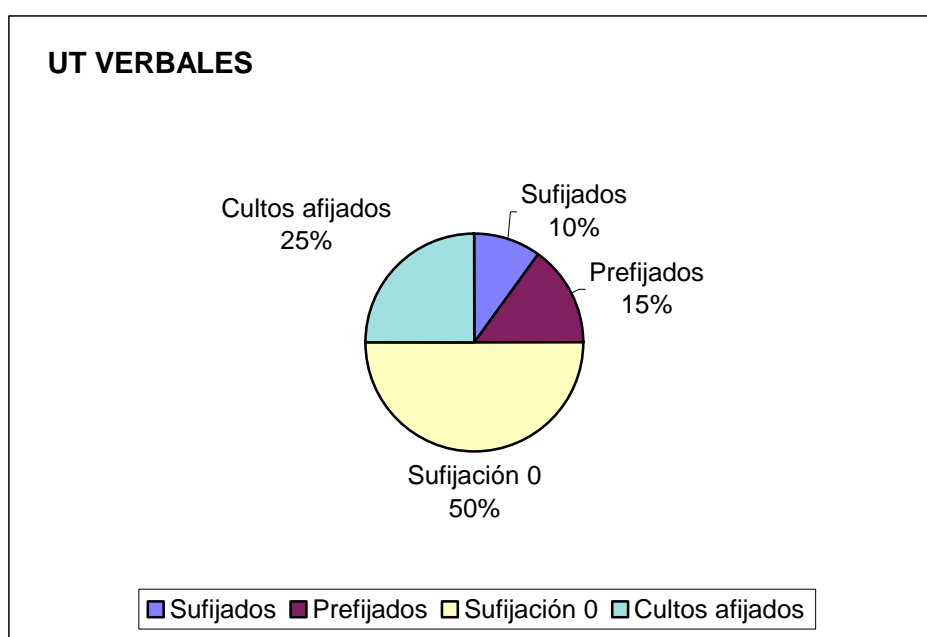


Los prefijos más productivos (a-, anti- y auto-) constituyen adjetivos relacionales (como también ocurre en otros ámbitos, como en medicina), y no calificativos, y participan en la constitución de una taxonomía nocional en el ámbito de la Genómica.

En cuanto a otros tipos de formación destacamos como dato curioso en este vocabulario la presencia de estructuras sintagmáticas adjetivas con carácter terminológico, especialmente raras en los ámbitos de especialidad, integradas por un núcleo adjetivo (*calórico, molecular, sensible*) con modificación cuantitativa antepuesta, cuantificación que se actualiza lingüísticamente siempre en el adverbio *altamente*¹⁴.

5.3 Las Unidades terminológicas verbales

Finalmente, las UT verbales analizadas (20 unidades) responden a las estructuras siguientes:



En relación a los verbos que aparecen en el corpus procedente del diccionario especializado, uno de los datos más relevantes es la importante presencia (importante en relación al número total de casos) de verbos formados por conversión sintáctica a partir de nombres, que podrían considerarse también como verbos simples que han producido nombres por conversión sintáctica (*cortar, cruzar, clonar, coagular, cultivar, acetilar, controlar, centrifugar, alquilar*). De todos ellos sólo la unidad *clonar* es propia y original de Genómica; *coagular, acetilar* y *centrifugar* proceden de unidades propias de otras materias especializadas; el resto forma parte del acervo léxico general, y en el campo de la Genómica, se utilizan con un sentido específico¹⁵.

La agrupación de las unidades verbales simples y las derivadas regresivas permite observar que el conjunto de los verbos morfológicamente simples corresponde al 52%

¹⁴ Esta estructura es el calco de la unidad del inglés *highly*.

¹⁵ Sobre la identificación de los verbos de Genómica que adquieren un sentido especializado y cómo se representa el cambio de sentido a través de la estructura argumental y de cualia (Pustejovski 1995), tenemos un trabajo en curso de realización.

de los casos. Sólo cinco verbos presentan una estructura derivada prefijada, dos totalmente patrimonial: *aislar*, *asimilar* y *aparear* (el primero procedente de base nominal, y los dos restantes de base adjetiva), dos de estructura culta (*adaptar* y *absorber*) y uno de estructura híbrida (*congelar*). Únicamente dos verbos del corpus corresponden a una estructura derivada sufijada (*amplificar* y *codificar*).

Aparte del hecho de que las unidades verbales de nuestro corpus posean una u otra estructura, es relevante la correspondencia entre verbos, nombres y adjetivos, simples, derivados o compuestos, de la misma familia derivativa, como puede observarse en la lista siguiente. La presencia constante de un nombre cuando existe el verbo es un dato a tener en cuenta para formular generalizaciones.

Verbos del corpus

Lexemas de la misma familia que el verbo, presentes en el diccionario

| | |
|-------------|---|
| absorber | absorción, absorbente, absorbancia |
| acetilar | acetilación, acetil, acetilcolina, acetiltransferasa, acetil-CoA |
| activar | activador, activación, actividad |
| adaptar | adaptación, adaptador, adaptabilidad |
| aislar | aislamiento, aislado |
| alquilar | alquilación (<i>sic</i>), aquilante |
| almacenar | almacenamiento |
| amplificar | amplificación |
| aparear | par, apareamiento, apareado |
| asimilar | asimilable, asimilación |
| clonar | clon, clonación, clonaje |
| centrifugar | centrifugación, centrífuga |
| codificar | código, codificación, codificado, codificador |
| coagular | coagulación, coágulo, coagulado, coagulante |
| congelar | congelación |
| controlar | control |
| construir | construcción |
| cortar | corte |
| cruzar | cruzamiento, cruzado, cruce |
| cultivar | cultivo |

6. Recapitulación

En este trabajo se han presentado en primer lugar el concepto y la clasificación de las denominadas Unidades de Conocimiento especializado (UCE). Para delimitar específicamente el concepto de CE nos hemos situado en el marco de la comunicación especializada y en él en el discurso especializado que se materializa en los textos denominados también especializados.

La comunicación especializada se distingue de la comunicación general por los factores que intervienen en ella, los más representativos de los cuales son las condiciones del emisor, las situaciones en que se produce este tipo de discurso y su función prioritariamente informativa. En estas situaciones se producen textos especializados que se diferencian de los generales también por varios factores, aunque comparten con ellos el esquema general de descripción. Discursivamente las UCE aparecen en el discurso producido por especialistas o por mediadores entre los especialistas y el público.

Los textos especializados, como los generales, pueden describirse a partir de sus condiciones discursivas y sus características textuales. Estas últimas se materializan en tres estructuras interrelacionadas: una estructura formal, una estructura cognitiva y una estructura lingüística.

Desde el punto de vista formal, los textos especializados no se distinguen de los generales más que por el hecho de seleccionar géneros textuales específicos. Así, por ejemplo, el género tesis doctoral o el de *paper* científico solo se da en situaciones especializadas de comunicación.

Cognitivamente los textos especializados se distinguen de los generales por la precisión de sus conceptos, los cuales componen una estructura bien delimitada, explícitamente establecida y consensuada por los expertos de cada ámbito de especialidad, de forma que los conceptos se usan en el discurso sólo en relación a su ubicación a dicha estructura y adquieren en ella un valor preciso.

Lingüísticamente los textos especializados se caracterizan por encima de otras características por las unidades en las que basa la transmisión del conocimiento especializado.

Estas unidades, denominadas Unidades de Conocimiento Especializado (UCE), se han clasificado teniendo en cuenta diferentes aspectos que actúan de criterio de tipologización: el sistema al que pertenecen, la estructura interna, la categoría gramatical, la clase conceptual, el grado de cohesión interna y el papel que ejercen en el discurso.

De acuerdo con estas características se han definido las unidades terminológicas (UT), que son las más relevantes en la representación de los conceptos de toda especialidad, como unidades léxicas, de estructura léxica o sintáctica, necesarias en la estructuración conceptual de un ámbito, que corresponden a las unidades mínimas con autonomía conceptual. Además de las UT se han caracterizado las Unidades Fraseológicas Especializadas (UFE), relevantes también en la representación del conocimiento de un ámbito especializado, pues son expansiones de las UT y piezas clave en las relaciones que se establecen en un campo de conocimiento.

Hecha esta clasificación, en este trabajo se han analizado las UT de Genómica procedentes de un corpus lexicográfico (el diccionario de Kaufmann, U. y Bergenholtz, H. (1998) *Diccionario enciclopédico de Ingeniería genética*. 2 vols. Toronto, Lugus Libros LatinAmerica Inc.) con la finalidad de disponer de las estructuras más representativas de las unidades de Genómica reconocidas explícitamente por los expertos como unidades terminológicas.

En una segunda fase -que corresponderá a la continuación de este trabajo en un nuevo proyecto-, analizaremos las mismas unidades en su contexto discursivo, y por lo tanto dentro de texto, teniendo en cuenta las características enunciativas de dicho texto como discurso. Ello nos dará información sobre la representatividad real en el uso de las unidades codificadas en el diccionario y nos proporcionará la confirmación o refutación de los patrones que se habrían considerado más representativos de las unidades del ámbito de la Genómica.

Este trabajo descriptivo, tanto el correspondiente a la primera fase como al de la segunda fase, se encamina a la detección automática de unidades terminológicas, a su representación en el texto y a la generación (en forma de asistente de generación) automática o semiautomática de estructuras conceptuales a partir del tratamiento automático de texto.

7. Bibliografía

Esta bibliografía incluye también las publicaciones relacionadas con este Proyecto de investigación realizadas por las autoras de este trabajo

Bevilacqua, C. R. (1999) Unidades fraseológicas especializadas: estado de la cuestión y perspectivas. Barcelona: IULA-UPF. [Trabajo de investigación, inédito]

Bevilacqua, C. R. (2004) Unidades fraseológicas especializadas eventivas: descripción y reglas de formación en el ámbito de la energía solar. Barcelona: IULA-UPF. [Tesis doctoral, inédito]

Cabré, M. T. (1999a) *Terminology. Theory, Methods and Applications*. Edited by J. C. Sager. Translated by J. DeCesaris. Amsterdam/Philadelphia: John Benjamins Publishing Company. 252 p. HB: 90-272-1633-9. PB: 90-272-1634-7.

Cabré, M. T. (1999b) *La terminología. Representación y comunicación. Una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. (Sèrie Monografies, 3).

Cabré, M. T. (2000a) «Elements for a theory of terminology: Towards an alternative paradigm». En: *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 6, 1, p. 1-23.

Cabré, M. T. (2000b) «Sur la représentation mentale des concepts: bases pour une tentative de modélisation». En: BEJOINT, H.; THOIRON, P. (dir.) *Le sens en terminologie*. Lyon: Presses Universitaires de Lyon, p. 20-39. ISBN: 2-7297-0641-0 .

Cabré, M. T. (2002a) «Textos especializados y unidades de conocimiento: metodología y tipologización». En: García Palacios, Joaquín; Fuentes, M. Teresa (eds). *Texto, terminología y traducción*. Salamanca: Ediciones Almar, p. 15-36. ISBN: 84-7455-079-3.

Cabré, M. T. (2002b) «Análisis textual y terminología, factores de activación de la competencia cognitiva en la traducción». En: Alcina Caudet, A. y S. Gamero Pérez (eds.) *La traducción científico-técnica y la terminología en la sociedad de la información*. Castellón: Publicacions de la Universitat Jaume I, p. 87-105.

Cabré, M. T. (2003a) «Theories of terminology. Their description, prescription and explanation». A: *Terminology*, 9, 2, p. 163-200.

Cabré, M. T. (2003b) «El lenguaje científico desde la terminología». En: Gutiérrez, Bertha (2003) *Aproximaciones al lenguaje de la ciencia*. Burgos: Fundación Instituto Castellano y Leonés de la Lengua, p. 19-52. ISBN: 84-607-7286-1.

Cabré, M. T.; Estopà, R. “On the units specialised meaning uses in professional communication”. *Terminology Science and Research*, 2003, 1-2 (en premsa).

Cabré, M. T.; Estopà, R. (2002) “El conocimiento especializado y sus unidades de representación: diversidad cognitiva”. *Sendébar*, 13, 141-153 .

Cabré, M. T.; Feliu, J. (ed.) *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica (DGES PB96-0293)*. Barcelona, Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra, 2001. ISBN: 84-477-0744-X.

Cabré, M. T.; Lorente, M. (1999) «L'estructura argumental dels predicats derivats». A: *Estudi General*, Revista de la Facultat de Lletres de la Universitat de Girona. Vol. 17-18, p. 67-78.

Chierchia, G.; McConell-Ginet, S. (1990) *Meaning and Grammar. An Introduction to Semantics*. Cambridge, The MIT Press.

Corbin, D. (1992) "Hypothèses sur les frontières de la composition nominale". *Cahiers de grammaire*, 17, 26-55.

Corbin, D. (1997) "Locutions, composés, unités polylexématiques: lexicalisation et mode de construction". *Actes du colloque de 1994 "La Locution, entre langue et usage"*. Textes réunis par Michel Martins-Baltar/ENS Éditions Fontenay/Saint-Cloud, 1-29.

DAVID, S. (1993) *Les unités nominales polylexicales. Éléments de description et reconnaissance automatique*. Tesis doctoral. Paris: Université Denis Diderot.

Estopà, R. (2001) “Les unités de signification spécialisées: élargissant l’objet du travail en terminologie”. *Terminology*, 7, 2, 217-237.

Estopà, 2001 “Los adjetivos en las unidades terminológicas poliléxicas: un análisis morfosemántico”. *Organon*, 14, 28/29, 233-246.

Estopà, R. (2003 [1999]) *Extracció de la terminologia: elements per a la construcció d'un SEACUSE*. CD-ROM. 84-89782-02-4.

Estopà, R. (1996). Les unitats terminològiques polilexemàtiques en els lèxics especialitzats: dret i medicina. Universitat Pompeu Fabra. IULA. Treball de recerca de doctorat [texto inédito].

Feliu, J. (2000) *Relacions conceptuals i variació funcional: elements per a un sistema de detecció automàtica*. [Trabajo de investigación, inédito]

Feliu, J.; Solé, E.; Tebé, C. Cabré, M. T. (2002) “Las relaciones conceptuales: un elemento esencial en la estructuración del conocimiento especializado”. En: *Actas del VIII Simposio Iberoamericano de Terminología*. Cartagena de Indias, 28 a 31 de octubre de 2002. ISBN: 958-33-4022-7.

Feliu, J.; Vivaldi, J.; Cabré, M. T. (2002) *Ontologies: a review*. Working Paper, 34. Barcelona: Institut Universitari de Lingüística Aplicada. DL: 23.735-2002.

Feliu, J.; Vivaldi, J.; Cabré, M. T. (2002) «Towards an Ontology for a Human Genome Knowledge Base». *LREC2002. Third International Conference on Language Resources and Evaluation. Proceedings*. Las Palmas de Gran Canaria, mayo 2002, p. 1885-1890. ISBN: 295-1740-808.

Fellbaum, C. (ed.) (1998) *WordNet. An Electronic Database*. Cambridge, The MIT Press.

Junyent, C. (2003) “La transmisión del conocimiento especializado. Conocimiento y conceptos”. Cabré, M. T.; Freixa, J.; Tebé, C. *Terminología y conocimiento especializado*. Barcelona: IULA-UPF.

Kaufmann, U.; Bergenholtz, H. (1998) *Diccionario enciclopédico de Ingeniería genética*. 2 vols. Toronto, Lugus Libros LatinAmerica Inc.

Lorente, M. (2002) Verbos y discurso especializado. En *Estudios de Lingüística Española (ELIES)*, 16. Publicación electrónica <http://elies.rediris.es>

Lorente, M. (2003) Procesos de creación del léxico con significado especializado: unidades verbales en el discurso del Genoma Humano. En Bisetto, A.; Iacobini, C.; Thorton, A.M. (2003) *Seriti di Morfologia*. In onore di Sergio Scalise. Roma, Caissa Italia Editore. págs 129-148.

Pustejovsky, J. (1995) *The Generative Lexicon*. Cambridge/Massachussets: The MIT Press.

Sager, J. C.; Kageura, K. (1994/1995) “Concept Classes and Conceptual Structures: Their Role and Necessity”. En: *Alfa*, 7/8, p. 191-216.

8. Anexo

Adjuntamos en un CD anexo una base de datos en ACCESS que incluye las UT analizadas. Esta base está formada por los campos siguientes:

- número de identificación de la ficha
- la unidad
- el tipo de unidad según su estructura (simple, sufijada, prefijada, compuesta culta, compuesta patrimonial, compuesta híbrida, compuesta culta afijado, sufijada química, sigla, fórmula, poliléxica, símbolo, nombre latino)
- la categoría gramatical de la unidad
- en el caso de las unidades afijadas y de las compuestas con algún formante greco-latino, los formantes o afijos de la unidad
- en el caso de las UT poliléxicas, la estructura morfosintáctica básica
- en el caso de las UT poliléxicas, la estructura morfosintáctica detallada

- en el caso de las UT poliléxicas, el número de elementos pertenecientes a una categoría gramatical fuerte
- clase semántica (acción, proceso, entidad, propiedad, relación)
- comentario, para poder hacer cualquier tipo de anotación útil para análisis futuros.

También se incluye un archivo EXCEL que cuenta con todas las unidades clasificadas por clases gramaticales (en tres hojas diferentes: nombres, verbos, adjetivos) y cada clase gramatical ordenada por tipos de unidades en función de su estructura.