

**Los sintagmas nominales extensos
especializados en inglés y en español:
descripción y clasificación
en un corpus de genoma**

Gabriel Ángel Quiroz Herrera

Papers de l'IULA. Sèrie Monografies, 10

Barcelona

Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada

2005

Quiroz Herrera, Gabriel Ángel

Los Sintagmas nominales extensos especializados en inglés y en español : descripción y clasificación en un corpus de genoma. - (Papers de l'IULA. Sèrie monografies ; 10)

Text en castellà, resum en castellà i anglès. - Bibliografia

I. Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada II. Títol

III. Col·lecció: Papers de l'IULA. Sèrie monografies ; 10

1. Gramàtica comparada i general - Sintagma nominal 2. Castellà - Sintagma nominal

3. Anglès - Sintagma nominal 4. Anglès - Gramàtica comparada - Castellà 5.

Castellà - Gramàtica comparada - Anglès 6. Llengües d'especialitat

806.0-316.4

802.0-316.4

Dirección de las Publicaciones del IULA: M. Teresa Turell

Coordinación de las Publicaciones del IULA: Lluís Codina, Gemma Martínez

Primera edición: octubre de 2005 (versión electrónica)

© el autor

© Institut Universitari de Lingüística Aplicada

La Rambla, 30-32

08002 Barcelona

Depósito legal: B-45.887-2005

Los sintagmas nominales extensos especializados en inglés y en español: descripción y clasificación en un corpus de genoma [^]

Gabriel Ángel Quiroz Herrera
gabriel.quiroz@upf.edu

Institut Universitari de Lingüística Aplicada
Universitat Pompeu Fabra
Barcelona

En el presente trabajo se presenta un estado de la cuestión desde diferentes disciplinas sobre los sintagmas nominales extensos especializados (SNEE) de más de tres *tokens* en inglés y en español en textos especializados del nivel experto-experto en el área del genoma. Se propone una metodología para describir y clasificar los SNEE a partir de 500.000 palabras en cada lengua de modo que se definan regularidades y se propongan soluciones para los diferentes profesionales del lenguaje.

In this working paper, a rationale of long specialized noun phrases of more than three tokens in English and Spanish is discussed. Furthermore, a methodology is proposed in order to describe and classify long specialized noun phrases using an expert to expert corpus of 500.000 words in the field of genomics. Thus, we will obtain regularities and define solutions for various language professionals.

[^] Este *working paper* es una versión modificada del proyecto de tesis doctoral dirigido por la Dra. Mercè Lorente y defendido en septiembre de 2004. El trabajo de investigación se enmarca dentro de los proyectos TEXTERM-2 (BFF2003-02111) y RICOTERM-2 (HUM2004-05658-C02-01), financiados ambos por el Ministerio de Educación y Ciencia. El autor quiere agradecer a la Dra. Monserrat González y el Dr. Jaume Llopis, investigadores del IULA y profesores del programa de doctorado en Ciencias del Lenguaje y Lingüística Aplicada, que han actuado como evaluadores de este *paper*, por todos sus comentarios y observaciones.

Índice

1.	Introducción	1
2.	Objetivos del trabajo	4
3.	Objeto de análisis	6
4.	Breve estado de la cuestión	9
4.1.	Introducción	9
4.2.	Un fenómeno, muchos nombres.....	9
4.3.	Gramática clásica	13
4.4.	Terminología	17
4.5.	Estudios de traducción	21
4.6.	Traducción automática	23
4.7.	Inglés para propósitos específicos (ESP)	24
4.8.	Lingüística computacional e ingeniería lingüística.....	24
4.9.	Escritura técnica	26
4.10.	Comparación de algunos estudios	27
5.	Hipótesis e ideas previas	32
6.	Metodología de trabajo	35
6.1.	Descripción del corpus	35
6.2.	Herramientas	36
6.3.	Metodología de análisis.....	36
7.	Bibliografía	48
8.	Programas de procesamiento y almacenamiento	57

1. Introducción

En un trabajo anterior (Quiroz 2003) se hizo una breve descripción y una clasificación de la premodificación de los sintagmas nominales (SN) de más de 4 tokens en un corpus de textos especializados (43.692 palabras) del nivel experto-experto en el área del genoma en inglés. Para esto se extrajeron 808 unidades y se obtuvo un total de 324 patrones de análisis diferentes para los 808 sintagmas nominales (SN) del corpus de análisis. El patrón N+N+N (sin determinantes ni números) es el más frecuente con 74 ocurrencias que equivalen a 9,15%. De los 324 patrones de análisis, 28 entre 5 y 74 ocurrencias tienen casi la mitad de los 808 SN del corpus de análisis de 423 SN que equivalen a 53,35%. Los patrones superficiales de nuestro corpus se contrastaron con los expuestos por Biber *et al.* (1999) y, preliminarmente, se encontró que: 1) El 2% de SN de entre 3-4 premodificadores dado en Biber *et al.* (1999) no concuerda con el 34,1% de SN de más de 5 tokens¹. Esto se debe quizá a que el corpus de Longman (LGWSE) tiene un rango más amplio de tipos de texto académico que nuestro corpus que tiene sólo un tipo. 2) La categoría léxica predominante como premodificador es el sustantivo (N) y no el adjetivo (A), al contrario de como lo prevén las gramáticas clásicas en general aun para el discurso académico. 3) En la mayoría de los patrones con frecuencia +5 predomina la secuencia de N. 4) Hay una clara diferencia entre el uso en inglés del participio pasado (-ed) y el participio presente (-ing) (29,62% vs. 8,02%). Se concluyó preliminarmente que: a) El discurso especializado de experto a experto es el hábitat natural de los sintagmas nominales con premodificación extensa; b) El problema de los sintagmas nominales con premodificación extensa se ubica fundamentalmente en la dificultad cognitiva de los profesionales intermediarios; c) La categoría premodificadora por excelencia en los sintagmas nominales con premodificación extensa es especialmente el sustantivo; d) Los patrones de premodificación en el discurso especializado varían de los estándares en el lenguaje común en cuanto a la frecuencia; y e) los sintagmas nominales con premodificación extensa presentan dificultades en el tratamiento automático.

En otro trabajo (Quiroz *et al* 2004), se presentó la metodología y los resultados de un ensayo de extracción y paralelización de SN y se analizaron los datos de los patrones contrastivamente con el fin de observar las regularidades de estos para la traducción. El ensayo se hizo sobre un minicorpus paralelo de 6.447 palabras tomado de la Revista *The Lancet* en inglés y su versión en español con 9.880 palabras de mismo nivel de especialidad que se empleó en Quiroz (2003), con la intención de afinar el formalismo diseñado para la extracción semiautomática de términos sintagmáticos extensos y así poder ampliar su aplicación en un corpus paralelo de mayores dimensiones en trabajos posteriores. Los textos se etiquetaron en la cadena de procesamiento del corpus del IULA. Al final de este proceso, se obtuvieron los 4 archivos etiquetados y desambiguados, aptos para ser procesados

¹ Puesto que Biber *et al* excluyen el núcleo del SN, un SN de 4 tokens de estos autores equivale a un SN de 5 tokens de nuestro corpus de análisis.

mediante los programas Repoker² y ALIPA. Se creó un conjunto de reglas 231 para entrenar el formalismo y luego se aplicaron las cuatro opciones del formalismo al corpus: extracción de SN exactos, extracción de SN aproximados por etiqueta del núcleo, extracción de SN aproximados a partir de las etiquetas de un SN muestra e introducción de los SN lemas alineados en los textos originales para ser usados luego en la alineación. Posteriormente se hizo un análisis contrastivo de los SN extensos para observar las regularidades que se presentaban tanto en cada lengua como en las soluciones presentadas para los SN en inglés. Se concluyó que:

- a) Los corpus etiquetados lingüísticamente ofrecen datos, al nivel de formas, lemas y patrones categoriales, que permiten establecer generalizaciones implementables en el diseño de herramientas de extracción de información.
- b) La combinación de estrategias (opciones de identificación en el caso de Repoker) se vuelve a demostrar como la vía más eficaz en el desarrollo de aplicaciones para la extracción de información, ofreciendo resultados con una reducida tasa de error en la identificación y en la paralelización de sintagmas nominales extensos, muy deseables para el tratamiento de corpus mucho más extensos de los utilizados en este experimento.
- c) Sin embargo, se precisan descripciones detalladas de estructuras formales y de relaciones semánticas para cada lengua y para cada ámbito de especialidad, con el propósito de mejorar los procesos de extracción y que esto, a su vez, nos permita resolver problemas traductológicos de manera eficaz y así evitar caer en el prescriptivismo hasta ahora encontrado en los manuales.
- d) Finalmente, el uso de estas herramientas de extracción y alineación en trabajos posteriores, nos permitirá agilizar la extracción y la paralelización y enriquecer el análisis contrastivo al mostrarnos las posibilidades que presenta un SN en inglés y sus soluciones en español y así, obtener el máximo de regularidades en un corpus de mayor tamaño. De este modo, podemos corroborar nuestras hipótesis de trabajo.

En trabajos posteriores se quieren corroborar estas conclusiones preliminares y extender el trabajo de manera contrastiva con el español para observar su comportamiento y así llegar a establecer unas regularidades que permitan a profesionales de las lenguas inglesa y española tener una herramienta que les permita solucionar adecuadamente los sintagmas nominales con premodificación extensa.

En resumen, los fenómenos que se analizarán en trabajos posteriores pueden resumirse básicamente en: a) el estudio de la premodificación compleja de los sintagmas nominales en inglés; b) la observación de la estabilización de dichas unidades tan extensas en el discurso especializado; c) el análisis de la correlación entre las estructuras y las categorías con las relaciones y otros aspectos

² Las herramientas Repoker y Alipa fueron creadas por el profesor Lluís de Yzaguirre del Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra.

semánticos; d) el análisis del crecimiento semántico gradual en este tipo de unidades; y finalmente e) la relación de la premodificación en inglés y la posmodificación en español en estas estructuras tan extensas que son del ámbito especializado con la creación y transmisión del conocimiento especializado en unos contextos delimitados.

2. Objetivos del trabajo

En trabajos posteriores se pretende resolver básicamente dos problemas: uno teórico y uno aplicado, como se planteo en la introducción. En el plano teórico se pretende

- Demostrar que la existencia de SN extensos (lexicalizados o no) refleja la complejidad semántica y pragmática del conocimiento especializado.
- Observar cómo se correlacionan la creación de conocimiento especializado y la transmisión de este conocimiento a través de estructuras que pueden ser formalmente secuenciales pero que semánticamente son jerárquicas.
- Contrastar los prejuicios existentes que están condicionando las descripciones y los trabajos teórico-descriptivos de los investigadores que trabajan en los estudios de traducción, en la enseñanza de lenguas para propósitos específicos, y en otras disciplinas mencionadas en el estado de la cuestión.

En el plano aplicado se pretende:

- Proponer recomendaciones para el tratamiento desde el punto de vista formal y semántico de estos sintagmas del inglés y sus correspondientes en español para que profesionales de la traducción, la lexicografía y la ingeniería lingüística, entre otros, puedan tener herramientas de trabajo.

Para cumplir estos dos objetivos generales, se propone una serie de objetivos específicos:

- Analizar cuantitativamente los sintagmas nominales especializados extensos (SNEE) en ambas lenguas en un corpus escrito de un ámbito especializado.
- Caracterizar formal y semánticamente los SNEE que sean unidades de significación especializada (USE), lexicalizadas o fraseológicas.
- Revisar la literatura en varias disciplinas relacionadas con la lingüística sobre la premodificación compleja, teniendo en cuenta aspectos morfológicos, sintácticos, semánticos, y cognitivos, implicados en la producción y comprensión de este tipo de estructuras.

- Probar que este fenómeno lingüístico tiene razones fundadas en la relación emisor-destinatario en un tipo de comunicación específico y natural y en la relación de estas estructuras con la producción de conocimiento y la interpretación de los diferentes destinatarios involucrados (expertos-diferentes profesionales de las lenguas).

- Diseñar un método de análisis de este tipo de sintagmas que permita interpretar estos sintagmas usando elementos lingüísticos resultado de las regularidades observadas.

- Estudiar si existen diferencias en el uso de estos sintagmas con respecto a lo que produce un experto en cada lengua y lo que reflejan los textos paralelos (traducidos) y observar si son convenciones retóricas inherentes al inglés y al español.

3. Objeto de análisis

Un sintagma nominal especializado extenso (SNEE) de la lengua inglesa es una frase nominal definida o indefinida de más de 3 *tokens* (2 premodificadores y un núcleo) que consta de un sustantivo nuclear (núcleo) precedido por diversos elementos (premodificación), en especial por determinantes, adjetivos, participios pasado y presente, sustantivos, en algunos casos, por adverbios, u otras categorías y otros elementos no verbales como abreviaturas, símbolos, signos, etc. Un sintagma nominal especializado extenso puede estar precedido por otras categorías gramaticales como una preposición. Estas unidades suelen tener términos o conjuntos de términos en su estructura combinados con unidades del lenguaje común.

Se establece un número mínimo de 3 tokens (2 premodificadores mínimo). Como se ha visto en Quiroz (2003) y en otros autores (Quirk *et al* 1985; Biber *et al* 1999) los SN extensos tienen más frecuencia en el registro formal. Es decir, los estudios realizados en córpora no presentan datos sobre sintagmas de más de 2 ó 3 elementos premodificadores como máximo. Estos córpora no incluyen un porcentaje significativo del tipo de texto que nosotros tratamos, sin embargo hemos analizado desde 3 tokens para observar la variabilidad de las unidades y la frecuencia de estos sintagmas.

Para la descripción del objeto de estudio, mantenemos la postura de que no existe una dicotomía entre lenguaje general y “lenguaje” especializado (o mejor dicho discurso especializado); no es un problema de “otro” lenguaje con características excluyentes. Sostenemos que los patrones que aquí describiremos y analizaremos se estudiarán desde la óptica de una gramática general y que sólo su uso y frecuencia en el discurso especializado se debe a razones pragmáticas, conceptuales o cognitivas.

Para ilustrar el fenómeno en cuestión se han coleccionado los ejemplos siguientes de las diferentes gramáticas (de referencia) y libros de sintaxis³:

- the beautiful big old neglected square red Jamaican stone plantation houses (11 palabras, 9 en posición premodificadora) (*Let's Write English*: 1980)
- an attractive tight-fitting brand-new pink Italian lycra women's swimsuit (11 palabras, 10 en posición premodificadora) (*The Cambridge Grammar of the English Language*: 2002: 543)
- Horseback riding school cafeteria breakfast menu substitution list (Levi 1978: 5)

³ Gran parte de estos ejemplos se presentaron en Quiroz (2004), pero nos parece oportuno presentarlos aquí.

- The hydrophobic polyethylene terephthalate (PET) and hydrophilic (water-soluble) polyethylene (PEO) blocks (Sager, Dungworth, and McDonald 1980: 221)

Los ejemplos siguientes se han coleccionado de corpórea textuales de diferentes áreas temáticas:

- 1-(-)2',3'-Dideoxythiacytidine (l(-)SddC, Lamivudine) resistant hepatitis B virus (HBV). (10 tokens, 7 en posición premodificadora)
- *Autographa californica* nuclear polyhedrosis virus (AcMNPV)-infected cells (8 palabras, 7 en posición premodificadora)
- a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model (13 palabras, 11 en posición premodificadora)
- an aperture emission mode scanning near-field optical microscope (SNOM)
- a 30-day-no-questions asked product return policy
- the Thermo-Sequenase fluorescent-labeled primer cycle sequencing kit
- a human acute lymphoblastic leukemia CCRF-CEM cDNA library
- *in vitro*-translated, radioactively labeled wild type and mutant Sox10 proteins

Los ejemplos siguientes se tomaron del corpus de genoma del IULA:

- A novel, red, low-potential, periplasmic copper protein
- a membrane bound proton-translocating pyrophosphatase
- a green fluorescent protein (GFP) fusion protein
- bright clear small dot-like fluorescent structures
- the trans-Golgi localized reversibly glycosylated polypeptide (RGP1)
- an efficient nuclear magnetic resonance (NMR)-based metabonomic approach
- *Drosophila melanogaster* maternal nuclear protein kinase Dm-nk
- The mitochondrial inner membrane AAA metalloprotease family

En español también podemos encontrar sintagmas nominales de esta extensión que se consideran términos, es decir, unidades especializadas. Por supuesto que no se tendrá una premodificación como en inglés dadas las características del español, pero puede verse que el concepto de sintagma nominal especializado extenso va más allá de los límites considerados “normales”. A guisa de ejemplo, se tomaron algunos sintagmas nominales especializados extensos del Diccionario Espasa de Medicina (1999: 320).

Como entrada principal:

- Déficit familiar de lecitín-colesterol-acil-transferasa

- Déficit neurológico isquémico reversible
- Déficit selectivo de subclases de IgG
- Déficit de alfa-1-antitripsina
- Años de vida ajustados según la calidad

Como subentrada:

- ácido graso Omega 3
- acidosis tubular renal distal hiperpotasémica
- rigidez muscular arteriosclerótica de Forester

En ámbitos técnicos también podemos encontrar SN extensos especializados como los ejemplos tomados de los diccionarios en CD-ROM de Routledge Spanish Dictionary of Business, Commerce and Finance - Diccionario Inglés de Negocios, Comercio y Finanzas (1998) y Spanish Technical Dictionary (1998):

- adaptive differential pulse coded modulation (modulación adaptable diferencial de impulsos en código)
- ammonium nitrate fuel oil (explosivo compuesto de nitrato amónico y fueloil)
- broadband-integrated services digital network (red digital de servicios integrados de banda ancha)
- compact disc programmable read-only memory (disco compacto de memoria de sólo lectura programable)
- containerized lighter aboard ship system (sistema de buques remolcadores para transporte de gabarras cargadas con contenedores)
- consolidated link-layer management message (mensaje de gestión consolidada de enlace entre capas)
- extended binary-coded decimal-interchange code (código ampliado de caracteres decimales codificados en binario)
- permanent income bearing share (acción productora de renta permanente)
- directly unproductive profit-seeking activities (actividades directamente improductivas con fines lucrativos)

En otras lenguas como el francés y el alemán podemos dar cuenta también de este fenómeno. Por ejemplo, Kocourek (1991: 140) cita el fenómeno, tomando ejemplos de otros autores:

- Maillot de bain féminin d'une seule pièce dégageant les côtes, les bas du dos et les hanches (Rey-Devobe 1973: 92)
- Donaudampfschiffahrtsgesellschaftskapitänwitwenrentenauszahlungstag (Heger 1971: 72)
- VDI Ultrakurzwellenüberreichweitenfernfunkverbindung (Heger 1971: 72)
- Valve spool dirt excluding rubber washer steel spacer (Horecký 1963: 275)

4. Breve estado de la cuestión

4.1. Introducción

Los sintagmas nominales extensos se han visto desde diferentes disciplinas relacionadas con el lenguaje: gramática clásica, lingüística teórica, lenguajes especializados (LSP), terminología, traducción, traducción automática, inglés para propósitos específicos (ESP), escritura técnica y recuperación de la información, entre otros. Sin embargo, no hay una descripción detallada de ellos hasta donde hemos explorado.

Muchas disciplinas ven este tipo de sintagma más bien como un obstáculo o una rareza de la lengua, y no como un fenómeno recurrente y “natural” en ciertos niveles de especialidad o géneros discursivos, como se ha ilustrado en el apartado anterior.

Es importante tener en cuenta que los fenómenos lingüísticos no deben considerarse desde una óptica prescriptivista como se verá más adelante, pues esto impide explicar por qué ocurre un fenómeno como éste. Hay que tener en cuenta que son los científicos los que usan los SNEE y no los lingüistas y que debe existir otro tipo de motivaciones para usar determinadas características de la lengua con más frecuencia y de algún modo específico.

4.2. Un fenómeno, muchos nombres

Se pueden distinguir básicamente dos tipos de denominación de este fenómeno dependiendo de si la unidad está lexicalizada o no. Para la mayoría de disciplinas, los nombres se confieren indistintamente si la unidad está lexicalizada; sólo los terminólogos hacen una distinción clara, por ejemplo, entre unidades términos o unidades libres (Cabré 1993) y, actualmente, entre términos, fraseología, colocaciones, locuciones y unidades libres (Lorente 2001).

Kocourek (1979) utiliza la denominación *lexical phrase* para este tipo de sintagma y, además, hace la diferencia entre estos y los sustantivos tipo *compounds* (que pueden ser parte de un *lexical phrase*). Igualmente, emplea *French noun-phrase terms*. Como puede verse en la siguiente lista, los nombres corresponden a unidades que pueden entrar en el lexicón de una lengua y no a unidades que pueden ser fruto de un encadenamiento accidental dentro del discurso.

lexie	(Rey 1977: 15; Pottier 1973: 251)
synapsie	(Benveniste 1966: 91)
paralexème	(Dubois <i>et al.</i> 1973: 354, Griemas 1966: 37-38)
mot complexe	(Auger 1975: 79, Picoche 1977: 23)
sytagne codé	(Rey 1977: 135)
groupe lexical	(Vachet 1964: 40)

lexie complexe	(Pottier 1963, 1968: 19, 1973: 251)
lexème complexe	(Goffin 1978)
groupe lexicalisé	(Rey 1975: 13)
locution composée	(Bally 1975: 250)
syntagme lexical	(Auger 1978)
syntagme autonome	(Guilbert 1967: 305)
syntagme lexicalisé	(Rey-Debove 1971: 113, Rey 1975: 11)
lexème syntagmatique	(Lyons 1978: 25-26)
composé syntagmatique	(Auger 1976: 66)
dénomination complexe	(Hollyman 1966: 97)
dénomination synaptique	(Benveniste 1966: 94)
groupe de mots	(Phal 1964: 47)
syntagme de lexique	(Auger 1976a: 67)
unité de signification	(Guilbert, Quemada 1972: 400)
unité lexicale complexe	(Phal 1969: 76)
unité sémantique complexe	(Dubois 1960: 62-63)
unité lexicale supérieure	(Marcellesi and Marcellesi 1969: 114)
unité syntagmatique de signification	(Guilbert 1967: 305)
unité lexicale syntagmatique complexe	(Dugas 1978)

Kocourek excluye la siguiente denominación:

unité lexicale à deux et plus de deux éléments

Esta primera clasificación de las denominaciones de estas unidades para el francés la realizó Kocourek (1979: 124), que luego publica en 1981 y 1991. Kocourek (1979) no tiene en cuenta, para este listado, las denominaciones en francés que también tienen sintagmas no lexicalizados.

Sin embargo, otros autores ponen en duda el carácter lexicalizado de algunas unidades y por esto, prefieren tomar un rango más amplio de estas unidades. Además, dependiendo del área de estudio como la traducción automática, los estudios de traducción y la enseñanza del inglés con propósitos específicos, el hecho de que una unidad esté o no lexicalizada pasa a un segundo plano, como lo plantea L'Homme (1994: 150):

“La distinction souvent évoquée, notamment en terminologie, entre un groupe lexicalisé et une combinaison libre joue un rôle secondaire de point de vue de la TA (traduction automatique). Les notions de lexicalisation ou de véhicule d'une signification ou d'une notion unique ne sont pas vraiment mises à contribution”.

Esto se debe a que en estas áreas un SN de estas características es un problema que se debe resolver en el acto: analizarlo y traducirlo o interpretarlo.

En los manuales de gramática y artículos de tipo más lingüístico no se habla de estos sintagmas con un nombre específico como se suele hacer en otras áreas; siempre son sintagmas nominales con premodificación compleja o un tipo de compuesto (de dos o más N o A):

noun phrases with multiple premodifiers	(Quirk <i>et al</i> 1985)
compound nouns	(Downing 1977: 810)
complex nominal	(Levi 1978)
noun strings, piled-up adjectives	(Georges 1996)
multiple premodification	(Biber <i>et al</i> 1999)

Esto se debe quizá a que se quiere potenciar más las partes que el todo. Por ejemplo, se dedican fragmentos de estas gramáticas a explicar las relaciones internas de los compuestos de dos elementos (*compounds*), los tipos de premodificación, el orden de los premodificadores, los tipos de núcleos, etc.

En terminología, como es de esperar, predominan las denominaciones relacionadas con las palabras término y terminología, es decir sintagma terminológico y término sintagmático; se considera que siempre están lexicalizadas. Para identificar la cantidad de palabras que contiene el sintagma se usan expresiones como multi-, poli-, y complejo. Sin embargo la palabra ‘complejo’ no siempre indica cantidad de palabras en algunos autores.

complex noun phrases	(Myking 1988)
syntagmes termes	(Portelance 1989: 400)
syntagmes nominaux complexes	(Jastrab 1987)
unités terminologique complexes	(Assal y Delavigne 1993)
sintagmas terminológicos	(Cabré 1993: 29)
terme complexe	(Bourigault 1993)
multiword compounds	(Maalej 1994: 142)
syntagme terminologique	(Collet 1997: 2003)
términos sintagmáticos	(Cartagena 1998: 282)
sintagmas terminológicos	(Aldestein 1998)
sintagmas terminológicos	(Cardero 2000)
unités terminologiques complexes	(Café 1999)
unités terminologiques polylexicales	(Estopa 2001: 219)
unidades terminologica complexa	(Faultisch 2003)
término sintagmático,	
término compuesto,	
complex term,	
grupo nominal complejo	(Méndez 2002: 191)
sintagma nominal compuesto	(Oster 2003: 138)
syntagme nominal pertinents	(Naulleau 1998: 33)

En traducción, no se usan, en muchos casos, nombres para designar el fenómeno como en el caso de López y Minett (1997), pero se intenta dar indicaciones para resolverlo o comentarlo. Por tanto, no se ve que predomine un nombre, pero podemos decir que autores como Newmark para el inglés y Alcaraz para el español, considerados autoridades, se pueden tomar como punto de referencia.

multiple noun compounds	(Newmark 1981: 115 y 1988: 41)
sintagma nominal largo	(Alcaraz 2000)
noun clusters	(Linder 2002)
multinoun compounds	(Quiroz <i>et al</i> 1996, 2000)
sintagma nominal extenso especializado	(Quiroz <i>et al</i> 2004)
sintagmas nominales complejos	(Zabala 1996)
compuestos multinominales	(Abril y Ortiz 1998: 291)
grupo nominal complejo	(Begoña 1995: 50)
yuxtaposición de adjetivos	(López y Minett 1999: 105)

En escritura técnica es más usual hablar de cadenas de sustantivos (o de adjetivos denominales), utilizando algún adjetivo de forma peyorativa para criticar el exceso de premodificación como podemos ver en los siguientes autores:

nounspeak	(Orwellian 1974)
noun strings	(Burnett 1992: 312; Huckin 1991)
sustantivos adjetivales en “caravana”	(Norman 1999)
string/excessive premodification, groups of premodifiers	(Kirkman 1992)
nouns as adjectives	(Blake and Bly 1993)
complex nominals	(Begoña 1996: 58)

En traducción automática (TA) e ingeniería lingüística es más importante la extensión del sintagma por sus posibilidades de interpretación que el hecho de que la unidad esté lexicalizada. Además no sería rentable para un sistema de TA incorporar en su lexicón estas unidades tan largas. En general se piensa que es mejor tener un analizador (*parser*) que permita hacer un análisis de los SN al ser estos recurrentes en los textos técnicos.

compound structures	(Chambers 1994)
noun sequences	(Lehrberger 1982: 29)
nominal compound	(Finin 1980, 1986)
compound nouns	(Bennett 1993: 43)
groupes nominaux	(L’Homme 1994: 148)
noun compounds	(Barker 1998)
compound nominal groups	(Woolley 1997)
two-noun compounds/nominal compounds	(Maalej 1994)
noun sequences, noun compounds, complex nominals	(Vanderwende 1995: 2)
noun + noun compound	(Downing 1977)
noun sequence	(Leonard 1984)

En ESP (*English for Specific Purposes*) tiende a predominar la palabra ‘compuesto’ no en el sentido estrictamente binario y lexicalizado que ven los lingüistas, para dar una idea de una unidad más larga de lo “normal”, sino que se agregan palabras como ‘*complex*’, ‘*phrase*’, ‘*in chain*’, ‘*group*’, *etc.*

noun strings	(Palmer 1968)
noun compounds (terms)	(Pugh 1984: 395)
long noun phrase	(Varantola 1985)
complex compounds, more complex compounds, very complex compounds, noun strings	(Trimble 1985)
compound nominal phrases	(Salager 1984)
nominal compounds	(Horsella y Pérez 1991: 125)
lessie complesse	(Casadei 1991: 58)
noms composes en chain	(Boughedaoui 1995)
nominal groups	(Thouvenin 1996)
complex nominals	(Begoña 1996)
nominal group	(Ventola 1996)
compound nominal groups	(Woolley 1997)
groupes nominaux	(Maniez 2001)
noms composés en chaîne, enchaînements syntagmatiques complexes	(Boughedaoui 2001: 139-137)
syntagmes nominaux complexes	
syntagme nominal fleuve	(Le Masle 2001: 65)
noms composés (lexicalisés), groupe nominal complexe	(Ormond 2001: 9)
noun clusters	(Limaye Pompian 1991; Salager 1985)

Puede verse en este corto recorrido por diversos autores que este fenómeno presenta diversas interpretaciones y denominaciones, dependiendo del área. Esto hace su abordaje más complejo, pero a la vez más enriquecedor.

4.3. Gramática clásica

Las gramáticas clásicas del inglés y los lingüistas se refieren poco a los SN largos; sólo algunos de ellos como Quirk *et al* (1985) y Biber *et al* (1999) mencionan el fenómeno y, en el último caso, hacen un recuento estadístico por niveles de lengua (registros) y explicar los casos internos de coocurrencias en la premodificación. Otros autores como Levi (1978) dan cuenta de este tipo de SN pero sin hacer un análisis de las estructuras largas; esta autora deja implícito que los SN con adjetivos no predicativos (denominales) de dos unidades (tipo compuesto) pueden servir para analizar los más extensos.

Los gramáticos clásicos (Quirk *et al* 1985; Huddleston y Pullum 2002) analizan estas unidades de diferentes modos. Quirk *et al* (1985: 1338) ponen de manifiesto que hay suficiente evidencia para decir que el uso de esta forma es más frecuente en el registro escrito que en el oral.

“There is indeed evidence of a higher proportion of three-or-more item sequences in written than in spoken English”

No obstante, cuando explican la premodificación múltiple Quirk *et al* (1985: 1.342) también ven este tipo de sintagmas como cosas pesadas e improbables:

“It should be noted, however, that if we introduce an adjective in this last noun phrase, already *clumsy* and improbable, the adjective has to come...” (itálicas nuestras).

Además plantean que los sintagmas nominales con premodificación compleja son un medio eficiente de comprimir la información, pero que implican las relaciones semánticas que hay entre los diferentes modificadores, lo que puede ocasionar problemas de interpretación para los hablantes.

Quirk *et al* (1985: 1.342-1.343) establecen que la premodificación múltiple presenta básicamente dos problemas: 1) orden relativo y 2) límite psicológico. Cuando hay más de un premodificador surgen problemas de orden. Analizan, por ejemplo, que el orden que hay en la siguiente premodificación lineal:

[Expensive [overseas [income [tax [office furniture]]]]]]

no es común y, por el contrario, pueden surgir varias interpretaciones como en:

A [new [giant size] [cardboard [detergent carton]]]

donde ‘size’ no premodifica a ‘cardboard’ y ‘cardboard’ no premodifica a ‘detergent’.

Quirk *et al* (1985) y Biber *et al* (1999) reconocen que este fenómeno no es algo oscuro, o excesivamente largo, sino que piensan que esto depende del tipo de destinatario al que va dirigido el texto. Su enfoque se centra más en términos de la implicación de las relaciones semánticas para quienes no son expertos:

“such an example is not, of course, obscure. Indeed, it is generally the cause that obscurity in premodification exists only for the hearer or reader who is unfamiliar with the subject concerned who is nor therefore equipped to tolerate the radical reduction in explicitness that premodification entails”. Quirk *et al* (1985: 1343).

Otros teóricos de la lingüística, en especial, las gramáticas clásicas del inglés, han detectado el fenómeno y han descrito unos patrones de coocurrencia más frecuentes (hasta 3 tokens) para la lengua inglesa y en varios géneros textuales. Sin embargo, no han estudiado a fondo este fenómeno en el tipo de texto del

presente estudio. (Quirk *et al* 1985: 1.338; Biber *et al* 1999: 595-597; Huddleston y Pullum 2002: 453).

De entre todos estos autores, Biber *et al* (1999) han hecho un análisis cuantitativo general de la premodificación compleja de hasta tres tokens y han explicado las relaciones semánticas de tipo compuesto (como ‘*compounds*’) que pueden tener en varios géneros discursivos: lengua general, noticias y prosa académica, en especial. Para estos autores (1999: 589), la premodificación múltiple es más común en las noticias que en la prosa académica, pero ambas en conjunto tienen una frecuencia más alta que en el resto de registros⁴. Además, proporcionan unos patrones de coocurrencia de la premodificación de los SN⁵ y afirman que la premodificación es más común que la posmodificación en este registro:

“In all registers, NP (noun phrases) with premodifiers are somewhat more common than those with postmodifiers (578). Proportionally, in academic prose, almost 60% of all NP have some modifier of which 25% have a premodifier and 20% have a postmodifier” (Biber *et al.* 1999: 579).

Como bien señalan Biber *et al* (1999: 579) se puede explicar que, en el género académico, haya más sintagmas nominales con premodificación que con posmodificación, debido a que la nueva información se empaqueta a modo de modificadores en el SN:

“Much of the new information presented in academic texts is packaged as modifiers in NP (noun phrases), resulting in a very high density of information”.

Sin embargo, Abberton (1977: 29-72), establece lo contrario con base en un corpus de ocho textos del corpus de *Survey of English Usage*, cuatro de novelas y cuatro de escritura científica no popular (*non-popular scientific writing*). De estos últimos, un texto pertenece al área de la biología, dos a la física y la matemática y uno a la química. Esta autora concluye que:

“both types of English examined are remarkably similar (subrayado del propio autor) in nominal group premodification structure: the vast majority of nominal groups are not premodified or are premodified by only one determiner and/or adjective (...) the main differences between the fiction and the science texts are that in the scientific writing postmodified nominal groups predominate; ...The spread over form class types is similar in both styles of writing and there is no form class (pattern) exclusively preferred by one or the other. In both the fiction and the science texts the ‘favourite’

4 En este trabajo se trata de usar la terminología empleada por cada autor. Registro se usa en el sentido de niveles de lengua y concuerda, en cierto sentido, con nivel de especialidad.

5 Ver Quiroz (2003), apartado 5.

patterns for premodified nominal groups are the same, although postmodified nominal groups are far more common in the science texts.” (62-63).

Esto se debe principalmente al tipo de corpus (registro de la lengua) y quizás al área del conocimiento en cuestión. Sin embargo, casi todos los corpóra del inglés tienen nuestras áreas en cuestión: biología, química, física, etc. Una revisión rápida de estos corpóra nos muestra que no hay, o hay muy pocos, textos del registro o nivel realmente (muy) especializado. Es decir, que hay pocos textos (y cantidad de palabras) que tengan una situación comunicativa del nivel experto-experto. Con lo cual, si tenemos en cuenta la cantidad de textos que se producen en el ámbito científico-técnico, las generalizaciones que se hacen en algunos casos no son representativas ni describen situaciones comunicativas reales. Además, como se ve en la cita, este estudio contradice los datos en cuanto a la cantidad de pre- y posmodificación en los textos especializados. Puede verse que aún en estudios con corpus hay generalizaciones contradictorias que, como muestra este estudio (1977), se deben al tipo de corpus y a la cantidad de palabras. Conviene señalar que para la fecha del estudio no existían muchos avances que permitieran tener grandes cantidades de corpus ni herramientas.

Por otro lado, Biber *et al* (1999: 589) establecen que los premodificadores son más comunes en las noticias que en la prosa académica. Sostienen que también los adjetivos comunes (i.e., adjetivos no participios) son la categoría gramatical que, como premodificador, es más común en todos los registros:

- Los adjetivos premodificadores son extremadamente comunes en la prosa académica.
- Los sustantivos dan cuenta del 40% de los premodificadores en las noticias y un 30% en la prosa académica.
- Los modificadores terminados en –ed son algo más comunes en la prosa académica que en otros registros.

Asimismo, establecen que “only about 2% of premodified noun phrases have 3 or 4 word premodification”, sumando todos los registros del corpus *Logman*. Además, argumentan que el género de noticias es un poco más común de premodificación compleja. Sin embargo, no explican las relaciones internas de la premodificación con su núcleo en los SNEE de más de tres tokens.

Puede verse que no hay un acuerdo en los datos, pues casi todos dependen del tipo de corpus, como se dijo antes. A pesar de esto, puede observarse que este tipo de unidades tan extensas no tiene una frecuencia alta en la lengua inglesa en conjunto pero sí en determinados registros o géneros.

Al referirse a la extensión de la premodificación Quirk *et al* (1995: 1.337-1.338) argumentan que:

“the problem becomes even more acute with longer strings of premodifiers. Although there is, theoretically, no grammatical upper limit to the number of premodifiers, *it is unusual to find more than three or four*” (las *italicas* son nuestras).

No obstante, matizan que “premodification is an area of English grammar where there is considerable variation among the varieties of the language”.

Algunos autores como Levi (1978) y Woollie (1997) establecen que los análisis bajo la Teoría de la Barra X presentan problemas para tratar los atributos y complementos de un SN, quizá por la cantidad de irregularidades que pueden presentar.

Finalmente, dada la capacidad de la Gramática Sistémico-Funcional de Halliday para tratar la premodificación de un SN como *deitic + numerative + epithet + classifier + head/thing*, algunos autores han analizado la premodificación de los SNEE bajo esta misma perspectiva (Thouvenin 1996; Boughedaoui 2001: 138; Ormod 2001: 12). De este modo, se puede distinguir entre los adjetivos que funcionan como atributos del sustantivo y los que funcionan como complemento de la posmodificación. Igualmente, Woollie (1997) concluye que se ha estudiado poco la estructura interna de los sintagmas nominales largos.

4.4. Terminología

Desde el punto de vista terminológico y de los lenguajes de especialidad, se ha descrito muy poco la estructura interna y el comportamiento de los SNEE en los textos de especialidad, no desde la perspectiva de la lexicalización sino como unidades (libres) de los discursos especializados que condensan una gran cantidad de información en poco espacio. Consideramos que el uso de este tipo de estructura puede, en un momento dado, establecer la densidad conceptual de un texto y nos permite, por tanto, decir que un texto es más especializado que otro.

Por otro lado, estas unidades presentan una serie de problemas como el comportamiento de elementos verbales y no verbales, e.g. siglas, acrónimos, formas cortas, inclusión de siglas en otras siglas, abreviación discursiva (en una revista sí y en otra no), cifras y códigos, entre otros fenómenos que deberían ser explicados. Otro aspecto importante sería observar el grado de lexicalización de dichas unidades y saber si tienden a ser unidades terminológicas especializadas o son simplemente un conjunto de ellas o una mezcla de lenguaje general y lenguaje especializado (discurso especializado) mediante la repetición de ellas en otros textos.

Además, estos SNEE están presentes en los diccionarios técnicos como vimos en los ejemplos de los tres diccionarios presentados en el apartado 3.

Al hablar de la lexicalización de los sintagmas terminológicos, Kocourek (1991: 140) plantea que dichas unidades tan extensas no pueden ser nombres sino definiciones y que, además, carecen de estatus léxico: “...une expression telle que *liqueur alcalique saturé de la matière colorante du bleu de Prusse* est ‘moins que un nom qu’une définition’...” En esa misma línea, Cartagena (1998: 281-296) concuerda con Kocourek en cuanto a que “(desde luego que) existe una relación directa entre longitud, el grado de especialización y la estabilidad sintáctica del término; a mayor longitud, mayor especialización e inestabilidad.” En este sentido, Cabré (1993: 304) dice que para los sintagmas muy largos se suelen utilizar formas reducidas:

“Comunicativamente, los sintagmas terminológicos excesivamente largos se suelen utilizar de forma abreviada en el discurso, y, a la larga, se suelen resolver fijando las formas reducidas respectivas como términos de uso corriente:

unidad central de proceso:	unidad central
hoja de cálculo electrónico:	hoja de cálculo
red de área local:	red local”

Cartagena (1998) hace una clasificación de los patrones más regulares en español de hasta 4 *tokens* en el ámbito de la anatomía y señala que no se han estudiado los términos sintagmáticos en el ámbito del LSP para el español. A continuación, se presentan para el español los patrones de tres y cuatro premodificadores de acuerdo con Cartagena (1998: 283-284):

N+A+A+A (el más frecuente)
 A+N+A+A (el más frecuente)
 N+Prep+D+N+A+A
 N+Prep+D+N+A+Prep+D+N
 N+Prep+N+Prep+N+Prep+N
 N+A+Prep+D+N+A
 N+A+Prep+D+N+Prep+N
 N+A+A+Prep+D+N

N+A+A+A+Prep+D+N
 N+A+A+Prep+D+A+N
 N+A+Prep+D+N+A+A

En este estudio, Cartagena presenta estos patrones para explicar la variabilidad de los sintagmas terminológicos. No presenta una estadística de su frecuencia respecto del corpus; simplemente se limita a decir las veces que aparece dicha estructura.

Cartagena (1998) está de acuerdo además con autores como Cabré (1993) y Hoffmann (1985) en cuanto a que la combinación de los componentes de los sintagmas terminológicos está regida por la gramática de la lengua común.

“La sintagmación, como recurso formal de obtención de unidades léxicas, se basa en la formación de una nueva unidad a partir de una combinación sintáctica jerarquizada de palabras. Las nuevas unidades así obtenidas *respetan las reglas combinatorias del sistema lingüístico al que pertenecen*, e incluyen muy frecuentemente conectores gramaticales...” -itálicas nuestras. (Cabré 1993: 177).

Como se indicó anteriormente, en este trabajo mantendremos la postura de que no existe una dicotomía entre lenguaje general y lenguaje especializado (o mejor dicho discurso especializado). No es un problema de lenguajes distintos con características excluyentes, sostenemos que la descripción y el análisis de los patrones se debe llevar a cabo desde la óptica de una gramática general, y que sólo su uso y frecuencia en el discurso especializado se debe a razones conceptuales o cognitivas.

En español, hay pocos estudios desde el punto de vista terminológico que incluyan unidades tan grandes. Cardero (2000 y 2004) hace una descripción de los SN terminológicos del español de México en un corpus de cinco áreas: cinematografía, el *Tratado de Libre Comercio*, redes de computación, telefonía celular y control de satélites. Esta autora no proporciona el número total de palabras del corpus ni el método de extracción. Además, no establece el tipo de corpus: textual o terminológico; parece ser mixto. Del total de términos, selecciona aleatoriamente 1.368 unidades (20% del total). De estos, el 5,19% corresponden a 66 términos de 12 estructuras que la autora denomina poco frecuentes. Estas estructuras se dan principalmente en el área de la telefonía celular. Su estudio no profundiza en patrones morfológicos ni relaciones semánticas.

Dentro del marco de la teoría comunicativa de la terminología (TCT), Estopà (1999 y 2001) hace una clasificación de las unidades terminológicas especializadas de acuerdo con las preferencias de los usuarios. Esta autora establece que las unidades que los traductores seleccionan son segmentos de las unidades terminológicas poliléxicas (principalmente nominales y adjetivales). Según ella, esto se debe a que son las que ocasionan problemas durante el proceso traductivo. A pesar de la clasificación de las unidades, la autora no analiza unidades tan grandes; las estructuras de más ocurrencias tienen un núcleo con dos modificadores (dos complementos) del tipo N+A+A con 30 ocurrencias. De su trabajo podemos comparar con nuestro estudio los datos de tipo morfológico y observar si las estructuras que la autora plantea (1999: 114) se presentan en nuestro corpus y si éstas simplemente se expanden o se forman otras estructuras no descritas. De entrada, por ejemplo, en Quiroz (2003) se compararon sintagmas desde dos premodificadores y coincidimos en que una de las estructuras más frecuentes en inglés es A+A+N y todas las soluciones en español están representadas en la estructura N+A+A. Sin embargo, en el mismo trabajo otras estructuras recurrentes como A+N+N = N+N+A, Adv+PP+A+A+N = N+A+A+A+Adv, N+Adv+N+N (*nucleoside reverse-transcriptase inhibitors*)=

A+N+Prep+D+A+Adv (inhibidores nucleósidos de la transcriptasa inversa), entre otras, no están descritas en Estopà (1999). Debe aclararse que en nuestros trabajos tomamos SN lexicalizados y no lexicalizados y, por tanto, no conviene precipitarse haciendo generalizaciones inadecuadas. A pesar de esto, todas las estructuras lexicalizadas ya descritas en otros trabajos (Cartagena 1998; Estopà 1999; Café 2000; Cardero 2000) las tendremos en cuenta puesto que para la traducción son igualmente relevantes.

Dentro del marco de la fraseología contrastiva en el ámbito de la radiología, Méndez (2000 y 2002) incluye los grupos nominales complejos como una estrategia discursiva de este tipo de discurso. Méndez (2002) se centra en una serie de estructuras muy diversas a partir de las concordancias sacadas con el programa WordSmith. Por tanto, no hay un análisis de estructuras (patrones), ni semántico ni estadístico de esta variedad de estructuras que incluyen, entre otros, los SN. El estudio tiene una orientación pragmática de la fraseología para traductores y redactores. Sin embargo, además de las estrategias fraseológicas que puede aplicar un traductor o un redactor, estos necesitan también las regularidades en el plano lingüístico para tener estrategias claras y, así, resolver los SN con independencia de los elementos que los acompañen en la predicación. Nos interesar más las relaciones al interior del sintagma sin dejar de reconocer la importancia de las otras estrategias que tienen que ver más con elementos de coherencia y cohesión como esta autora lo plantea.

Finalmente, Café (2000) presenta una clasificación de la expansión y las reglas de formación de las unidades terminológicas complejas para el portugués de Brasil en el ámbito de la biotecnología desde la perspectiva de la gramática funcional de Dik. Café parte también del principio de que una gramática de la lengua general debe abarcar los llamados lenguajes de especialidad y, por tanto, para crear las reglas de expansión, se basa en una teoría de la predicación. Las reglas están compuestas por una base (el núcleo), un argumento (el complemento) y los satélites (complementos externos). Cada base, argumento o satélite tiene una función semántica y una función sintáctica (sujeto, complemento nominal, complemento circunstancial, etc.). Para nuestros propósitos, algunos elementos del análisis de Café son interesantes, e.g., las funciones semánticas de un ámbito similar al nuestro. Sin embargo, consideramos que esta autora hace un análisis forzado al adaptar unas funciones de la predicación al ámbito de los SN, cuestión que pertenece al ámbito del verbo. Por otro lado, este estudio no contiene un análisis estadístico, un análisis de relaciones semánticas ni un análisis contrastivo entre dos lenguas como pretendemos en este proyecto. El análisis de los satélites de los SN que hace Café no tiene una función clara dentro de las reglas de formación pues no los incluye. Finalmente, la regla de formación más extensa da cuenta de SN con una extensión máxima de tres modificadores del núcleo (*molécula de DNA circular extracromossómica*).

4.5. Estudios de traducción

Desde la traducción, algunos autores como Vázquez-Ayora (1977), López/Minett (1997), Zabala (1998), Linder (2001 y 2002), entre otros, afirman que los sintagmas nominales extensos especializados son un problema de traducción dado el problema que plantea el orden de los premodificadores y la falta de relaciones explícitas de las relaciones semánticas entre los diferentes elementos. Además, su traducción en las lenguas romances como el español, el catalán o el francés presenta muchas posibilidades para explicitar las relaciones a través de la posmodificación.

Ya en 1959, Vinay y Dalbarnet dedican un apartado a explicar los problemas que los sintagmas nominales y los compuestos ocasionan en la traducción general y advierten de los peligros que ellos albergan (1959: 152-153).

Abril y Ortiz (1998: 291) centran el problema de los sintagmas nominales extensos como una de las características generales del lenguaje médico y, en especial, desde el punto de vista gramatical:

“Así, en inglés médico son habituales la nominalización, los verbos en pasado, el participio pasado, el participio presente, los compuestos multinominales (*mitogen-triggered lymphocyte DNA synthesis*)...”

De igual modo, Newmark (1988: 40) establece *a priori* que los compuestos con múltiples sustantivos (“*multinoun compounds*”) son una característica de los textos formales que se suelen traducir (según su tipología). Quiroz *et al* (1997) muestran en un corpus pequeño en el área de las enfermedades tropicales, que los SN extensos son una característica relevante de los textos especializados. Sin embargo, estos autores no proponen un análisis sintáctico-semántico o de otra índole para solucionar el problema. Las propuestas, cuando las hay, son muy intuitivas y no responden a una lógica sistemática de estos casos. Algunos de ellos, como Linder (2002), también los ven como un problema.

Uno de los pocos autores que tiene una perspectiva similar a la nuestra es Cartagena (1998: 282) dentro del marco de variabilidad de términos para el par alemán-español en el ámbito de la anatomía. Cartagena comenta que algunos autores como Coseriu (1973: 11) afirman que este tipo de unidades no presenta ninguna dificultad para la traducción ya que no se trata de transposición de significados sino de mero reemplazo de significantes en relación uno a uno. Calonge (1995: 184-185) también argumenta que parece evidente que el vocabulario científico no tiene nada que ver con la lengua general “[...] los sintagmas que representan a significados especializados [...] en general son fácilmente traducibles, debiendo evitarse solamente el calco”. Cartagena se apoya en su estudio y en el de Artz (1982: 114-117) para rebatir estos argumentos y

mostrar que los sintagmas terminológicos presentan considerables problemas de traducción, incluso aquellos SN relativamente sencillos. Además, en su estudio muestra que los sintagmas terminológicos y su variabilidad responden a la sintaxis de la lengua general.

Desde una perspectiva prescriptivista, Vázquez-Ayora (1977: 123) establece que la manera lógica de resolverlos es: “*to translate the premodifier closest to the nucleus and continue from there translating each successive adjectives to the left and so on*”. Este tipo de solución puede dar resultados en algunos casos, pero no es la única solución; sólo cubriría un tipo de patrón que quizá no sea el más frecuente. Sin establecer unas regularidades o una lógica, López y Minett (1997: 103-109) proponen que deben analizarse los elementos adjetivales y reordenarlos en grupos alrededor del núcleo del sintagma. Linder (2002: 266) establece una serie de parámetros y dice que se debe, en primer lugar, buscar el núcleo; en segundo lugar, determinar el orden e importancia relativa de los elementos y, finalmente, aplicar la siguiente estrategia (sin un orden predeterminado): maximizar el número de elementos en español, variar las preposiciones, omitir sustantivos o partes del SN si se repiten dentro del SN (variación denominativa) y usar tantas técnicas de transposición como se pueda (p. ej., de sustantivo a verbo).

La estrategia de Linder tiende a ser prescriptivista y no responde a una lógica de análisis lingüístico o traductivo. En primer lugar, no explica cómo se debe determinar el orden y la importancia relativa de los elementos del SN, al no analizar las tendencias en un corpus textual o terminológico. En segundo lugar, propone una estrategia para omitir sustantivos del SN, elemento peligroso puesto que un SN se puede diferenciar de otro sólo por un sustantivo que haga referencia a otro concepto o a un concepto que funcione como hipónimo o hiperónimo en la jerarquía. Además, en la traducción especializada esto podría considerarse como una falta de coherencia en el uso de un término, lo que podría ocasionar problemas de cohesión. Finalmente, no se puede maximizar el número de elementos sin razón alguna. Además, el uso de una preposición en una lengua no responde a una elección caprichosa del hablante sino que responde a una serie de restricciones gramaticales, pragmáticas y, en especial semánticas.

La extracción y la paralelización de sintagmas nominales extensos llevada a cabo en Quiroz *et al* (2004) permitió hacer un corto análisis sobre las regularidades que ellos albergan y rebatir, a pequeña escala, las propuestas de los autores antes mencionados. Como luego veremos, ninguno de estos autores ha trabajado observando las regularidades de un corpus.

Finalmente, pensamos que existen regularidades en los textos que nos pueden ayudar a abordar mejor la traducción de estos SN, que las relaciones semánticas se deben identificar y que los corpus nos ayudan a observar qué soluciones podemos proponer.

4.6. Traducción automática

Desde el ámbito de la traducción automática existen trabajos que han abordado el tema (Woolie 1997; Lehrberger 1982; L'Homme 1994; Maalej 1994). Tal es el caso de Woolie (1997) quien detectó que el principal problema de los traductores automáticos comerciales actuales es la traducción de sintagmas nominales complejos (de longitud variable), lexicalizados o no, que no están incluidos en el lexicón de excepciones del sistema. Para solucionar este problema hay dos alternativas. La primera es tener los SN en el lexicón del sistema, lo cual es muy inviable, poco productivo y poco predecible en términos de la variabilidad de los textos y áreas. La segunda es estudiar con profundidad los patrones sintácticos y las relaciones semánticas entre los diferentes componentes del SN para observar las regularidades y así introducir un conjunto de reglas que permita al sistema solucionar correctamente los SN.

El problema radica en que la formalización de estos SN es muy difícil por las supuestas irregularidades que se pueden encontrar. Por consiguiente, los errores causados por el orden de los premodificadores en un sintagma nominal complejo en las gramáticas formales hace que los analizadores morfosintácticos de los programas de traducción automática del inglés hacia una lengua romance produzcan resultados desastrosos (Woolie 1997: 3).

De esta dificultad planteada para formalizar los SNEE se derivan problemas para la traducción automática. Lehrberger (1982: 92-94) también estableció algo similar para el sistema de traducción automática TAUM que traduce informes meteorológicos del inglés al francés. Este mismo sistema estaba proyectado para traducir manuales de aviones en los cuales los SNEE jugarían un papel importante. Ya que este “sublenguaje” (como lo llama el propio autor) es muy restrictivo, Lehrberger estableció 50 relaciones sintáctico-semánticas suficientes para las combinaciones posibles de los manuales⁶. Sin embargo, no se hicieron predicciones para la lengua general (1982: 94)

Igualmente, Montero (1996) hace un pequeño estudio empleando un corpus terminológico de 4.235 términos de 2 a 5 tokens y concluye, igualmente, que la traducción de los SN extensos es una de las dificultades mayores y más visibles de los traductores automáticos.

⁶ El mismo Lehrberger (1982, 92-94) no es preciso indicando la publicación donde aparecen las 50 relaciones y por tanto no se ha podido ubicar. Sin embargo, la publicación debe ser de antes del año 1980.

4.7. Inglés para propósitos específicos (ESP)⁷

Incluso en la enseñanza del inglés con propósitos específicos (ESP), los SNEE son un verdadero problema para el profesor de lengua inglesa a la hora de preparar material didáctico. Las estrategias propuestas por diversos autores para enseñar a interpretarlos (Trimble 1985: 130-135, 163-165) se basan en esencia, en la intuición del hablante. Las relaciones semánticas internas implícitas de los SNEE se explicitan mediante paráfrasis que, en muchos casos, no resuelven las supuestas ambigüedades de los SNEE. El mismo Trimble (1985: 136) recomienda dejar los más extensos para el profesor del área de especialidad.

Quizá el estudio más importante cuantitativa y cualitativamente desde la óptica del inglés para propósitos específicos (ESP) es el llevado a cabo por Salager Mayer (1984: 135-146). Esta autora compara los SNEE del lenguaje general, el “lenguaje médico” y el “lenguaje de la técnica” llegando a la conclusión de que el promedio de la extensión entre los tres corpus es similar (2,06 en los tres). Sin embargo, el porcentaje de ocurrencia es más alto en medicina y en técnica (9,76% y 12,37%) que en el lenguaje general (0,87%)⁸. En cuanto a los SNEE de cuatro y cinco palabras, la frecuencia en medicina y técnica es muy superior que en el lenguaje general (20 y 7 veces más, respectivamente). Esto significa que, si bien el promedio en la extensión es similar respecto del número de palabras totales, la distribución de la cantidad de tokens en los discursos especializados es mayor. Este estudio no hace ningún análisis de patrones gramaticales, relaciones semánticas, clases semánticas o distribución o función de los SNEE en el discurso especializado.

4.8. Lingüística computacional e ingeniería lingüística

Desde la lingüística computacional y la ingeniería lingüística, algunos autores han trabajado las relaciones semánticas de los SNEE “cortos” teniendo en cuenta el concepto de ‘*compound*’⁹. Si un SNEE tiene, por ejemplo, seis elementos, se hacen tres pares de ‘*compounds*’ y se explicitan las relaciones semánticas de los tres pares. Las relaciones existentes entre los tres elementos con el núcleo no se explicitan generalmente. Además, se han creado algunos ‘*bracketers*’ para identificar semiautomáticamente las relaciones semánticas de dos elementos (Baker y Szpakotwicz 1998: 96-102)

Muchos estudios (Woollie 1995; Thouvenin 1996) extraen las ocurrencias estadísticamente empleando unos mínimos de frecuencia de 10 ó 25 apariciones y no usan etiquetaje automático o manual para observar las regularidades de los patrones. Thouvenin concluye que este método deja de lado muchos SNEE de

⁷ No se han consultado obras sobre enseñanza del español para fines específicos.

⁸ De cada 100 palabras, menos de 1 palabra formaría un compuesto.

⁹ Definido como la formación de una palabra a partir de una base más un afijo (prefijo o sufijo) que funciona como una unidad semántica.

“baja frecuencia” que podrían ser términos en un área de conocimiento. Incluso advierte, desde la perspectiva del ESP, que estas unidades deberían aparecer en los diccionarios del área a pesar de su relativa fijación o baja frecuencia.

Por otro lado, es importante tener en cuenta que los etiquetadores automáticos pueden generar gran cantidad de ruido y silencio. Por ejemplo:

[the trans-Golgi localized reversibly glycosylated polypeptide (RGP1)]

xxx antibodies raised against [the *trans-Golgi localized* reversibly glycosylated polypeptide xxx]
of

MO6 NN6P H6 P [A666 X V6A66 D6 X NN6S MO6] P

Puede verse que parte del SN se interpretó como una forma verbal finita. Esto obstaculiza las búsquedas de manera automática o semiautomática puesto que si se interroga un sistema con una búsqueda de patrones generará mucho silencio, como se ve en el ejemplo anterior.

Respecto del ruido, se realizó en el programa BwanaNet (herramienta de interrogación al Corpus Técnico del IULA) la siguiente consulta: N+N+N+N+N con mínimo 1-3 tokens obligatorios por cada N en todo el subcorpus de genoma:

Date: Thu Jun 19 11:49:13 2003
Corpus: Corpustecnicen
Subcorpus: CORPUSTECNICEN:Last
Number of Matches: 103
Left display context: 7 tokens
Right display context: 7 tokens

Query text: CORPUSTECNICEN; a: [pos="N.*"]{1,3} [pos="N.*"]{1,3} [pos="N.*"]{1,3} [pos="N.*"]{1,3} [pos="N.*"]{1,3} :

Number of concordances: 103

1. Because of Watson-Crick affinity , the *Probes capture DNA strands hat* contain Boston's name (TCGGACTG) .

NN6S Z</s> <s>P X NN6S Z A666 *NN6P NN6S NN6S NN6P NN6S* V6R6S NG6S NN6S Z X Z Z</s>

2. the latter approach , most often a *bait protein finds prey candidates*- sometimes many - and those might

A666 JA NN6S Z DS D6 A66S *NN6S NN6S NN6P NN6S NN6P* Z D6 R66A6666 Z C6 RD666666 V6666

3. and why we get old , " *remarks geneticist Richard K. Wilson* of Washington University , one partner in

C6 D6 RE666N61P V6R6S JA Z Z *NN6P NN6S NN6S NN6S NN6S* P NN6S NN66 Z MC6 NN6S P

4. Most single mutations that increase *Activation produce amino acid substitutions* that increase the negative charge of region

NN6S Z</s> <s>A666 JA NN6P AD6S V6R6S *NN6S NN6S NN6S NN6S NN6P* X V6R6S A666 JA NN6S P NN6S

Puede verse que, en los casos anteriores, la desambiguación es errónea, pues el etiquetador analiza los verbos en tercera persona del singular o plural como sustantivos en plural o en singular extrayendo SNEE erróneos. Por tanto, es difícil el reconocimiento y la extracción de unidades de tal extensión. Esto se debe quizá a que los etiquetadores se crearon para la lengua general y no tienen en cuenta sintagmas de este tipo.

4.9. Escritura técnica

Los autores de escritura técnica en inglés (*technical writing*) son, quizá, los principales detractores de los sintagmas con premodificación extensa y siempre advierten sobre el mal uso de ellos (Blake *et al* 1993; Burnett, 1992; Huckin/Olsen 1991; Kirkman 1992, entre otros). Dichos autores atacan contundentemente, por ejemplo, lo que llaman ‘*noun as adjective*’:

“Strictly speaking, it is poor writing to use a noun as an adjective”. Blake and Bly (1993: 88).

Agregan que la premodificación extensa obscurece el significado. Como solución dicen que se debe romper con frases preposicionales o del participio como en:

- A 15,000-lb steam/h pulverized-coal-fired power boiler → a pulverized-coal-fired power boiler generating 15,000 pounds of steam per hour

Otros autores ven los SN extensos como una característica que molesta al lector, debido a que su premodificación es excesiva:

“A particularly disturbing feature of technical writing is excessive ‘premodification’ – the piling up of adjectives, or words being used adjectivally, in front of a single noun: ...a mobile hopper fed compressed air operated grit blasting machine”. Kirkman (1992: 32-33).

Este mismo autor incluso compara la forma escrita con la oral. Por supuesto que en la forma oral no se usaría una premodificación tan compleja:

“To piled up ‘modifiers’ in this way is utterly unnatural language behaviour. We would not normally dream of telling someone we had been

to a store and bought a ‘new green leather suede-lapelled patch-pocketed tie-belted jacket’ ”. Kirkman (1992: 32).

Burnett (1992: 312) ve las cadenas de sustantivos como un problema estilístico:

“Imprecise diction also results from noun strings: a series of two or more nouns in which the first nouns modify the later ones”

Además, el autor sostiene que cuando una cadena de sustantivos alcanza cinco o más palabras, se requiere tiempo extra para calcular las relaciones entre las palabras y, como consecuencia, se puede interpretar de múltiples maneras y se vuelve indescifrable. Las preguntas que caben aquí son las siguientes: ¿para quién y en qué circunstancias una cadena es ambigua e indescifrable? y ¿qué toma más tiempo y esfuerzo cognitivo para un lector experto, un sintagma nominal con múltiple posmodificación como sucede en las lenguas romances en las cuales se puede perder la referencia, o un sintagma nominal con múltiple premodificación como sucede en inglés?

4.10. Comparación de algunos estudios

En la tabla 1 se han recogido los datos de varios estudios relacionados con los sintagmas nominales especializados extensos, terminológicos o libres (formados con unidades especializadas).

Autor	Corpus	Tokens	Método	2	3	4	5	6	7	8
Herzog (1971) ¹⁰	Computadores	ND ¹¹	ND	10%	36%	40%	12%	2%	-	-
Varantola (1984)	Ingeniería/ noticias-general	20.000/ 11.526	Lingüístico	23,1 /22,9	33,3/ 32,3	12,6/ 8,7	4,7/ 2,4	2,6/0, 5	-	-
Salager (1984)	Medicina/ técnica/ general	20.000/ 20.000/ 20.000	Estadístico	6,05/ 4,15/ 0,38	6,84/ 2,36/ 0,42	0,61/ 0,82/ 0,04	0,61/ 0,22/ 0,03	-	-	-
Goffin (1985) ¹²	ND	ND	ND	52	28	5+	-	-	-	-
Horsella <i>et al</i> (1991) ¹³	Química	17.500 ¹⁴	¿?	70,77 ¹⁵	22,66	5,09	1,48	-	-	-
Montero (1995) ¹⁶	Computadores (diccionario)	4.235	Estadístico? ND	82,9	14,89	1,9	0,1	-	-	-
Thouvenin (1996) ¹⁷	Electricidad	246.435	Estadístico	ND	ND	ND	ND	ND	-	-
Biber <i>et al</i> (1999) ¹⁸	Noticias/ prosa académica/ general	10.7 mill./ 5.3 mill./ 40 mill.	Lingüístico	80	20	2+		-	-	-
Oster (2003) ¹⁹	Cerámica	1.216	Lingüístico	57 ²⁰ 52,5	13,2 7,6	2,1 1,0	0,4 0,1	ND ND	0,1 0,0	- -
Quiroz (2003) ²¹	Medicina	44.000	Lingüístico	7,5	27	34	16,4	8,2	2,5	1,9

Tabla 1: Cuadro comparativo de estudios sobre los SNEE

Estos estudios se han realizado en varias áreas del conocimiento (química, medicina, técnica, etc.), varios niveles de especialidad (general vs. especializado) y con diferentes métodos de análisis y extracción (lingüístico, estadístico, manual, etc.). Puede observarse que:

1. Los métodos lingüísticos o híbridos dan porcentajes más altos, excepto Montero (1996) que toma un corpus de términos cerrado y no un corpus de textos.

¹⁰ Se desconoce si es corpus textual, terminológico o lexicográfico.

¹¹ ND: No disponible.

¹² Citado por Cabré (1993). Estudio hecho para el alemán.

¹³ Sólo trabaja con compuestos del tipo N+N+...N.

¹⁴ Hace una comparación de tres niveles de especialización diferentes. Aquí sólo hemos tomado los datos del nivel más especializado.

¹⁵ Este porcentaje sólo se refiere al total de unidades pero no es frente al total de palabras.

¹⁶ Este estudio no es sobre corpus de textos sino sobre un diccionario.

¹⁷ Se analizan las unidades de 2 a 4 tokens pero no hay estadísticas, ni patrones, ni frecuencias.

¹⁸ Se analiza la modificación múltiple, pero no hay estadísticas sobre tokens de más de 3. Además, la estadística de los tokens de +3 se combina con posmodificación.

¹⁹ El aspecto cuantitativo se realiza en un corpus terminológico alemán-español.

²⁰ Los datos de la primera línea se refieren al alemán y la segunda al español.

²¹ Incluye determinantes del sintagma.

2. Hay diferencias muy grandes entre un autor y otro por las razones siguientes. El tipo de corpus empleado por cada autor tiene diferentes niveles de especialidad (además del concepto de nivel de especialidad que cada autor entiende), utiliza un número diferente de tokens, emplea un método diferente de extracción y una combinación de subcorpus dentro de cada estudio (si los hay). Los corpórea exclusivamente especializados presentan porcentajes más altos de aparición de este tipo de sintagmas y premodificación más extensa. Por el contrario, los corpórea que combinan varios géneros o niveles de especialidad (desde el nivel de experto-lego hasta experto-experto) presentan porcentajes más bajos de sintagmas nominales especializados extensos y menos extensión de éstos.

3. En casi todos los estudios se comienza con los porcentajes más bajos en 2 premodificadores (sin el sustantivo nuclear), se llega a un pico en los porcentajes con 3 y 4 premodificadores y, por último, se desciende desde 5 hasta 9 premodificadores, de modo que, existen menos sintagmas nominales con premodificación compleja cuanto más extensión haya en la premodificación. El hecho de haber un pico porcentual entre 3 y 4 puede revertir en estabilización y posiblemente en la lexicalización de estas unidades.

4. La mayoría de estos estudios no miran la premodificación de los SNEE desde la lexicalización, sino como un fenómeno que, en apariencia, se sale de los cánones de la lengua general. Para algunos autores, la premodificación se enmarca dentro del sistema de la lengua en general.

Para interpretar la Tabla 1, es necesario saber que algunos autores observan:

1. Cuántas palabras son parte de la premodificación respecto del total de palabras del corpus.

2. Cuántos sintagmas hay en un corpus y, de éstos, cuántos corresponden a la premodificación múltiple.

3. Cuántos sintagmas nominales de premodificación compleja hay de cierto número de tokens (de más de 2, 3 ó 4). Se hace la estadística sobre ese parámetro de tokens y no sobre el total de sintagmas nominales de un corpus, es decir, desde un premodificador hasta el límite encontrado.

Algunos autores observan la opción 1 pero no la 2 ni 3, o las opciones 2 ó 3 pero no la 1. Sólo un caso observa las opciones 1 y 2 (Salager-Mayer 1985). Otros autores observan la opción 3 y luego contabilizan el número total de tokens para obtener la ratio de palabras de los SN seleccionados con el total de palabras del corpus. Es cierto también que algunos estudios previos a los años 90 no disponían de corpórea ni de herramientas para el procesamiento, el almacenamiento y la

manipulación de los datos. A pesar de los avances en la creación de córpora y de las herramientas para su procesamiento, los pocos estudios que hay en la actualidad tienen las mismas carencias que los anteriores.

Puede verse, entonces, que el asunto es mucho más complejo de lo que muestran las simples cifras. En primer lugar, no hay acuerdo sobre: a) los métodos de extracción de los SN, b) el tipo de corpus, c) el nivel de especialidad del corpus, d) el uso de lenguaje general versus lenguaje especializado, y e) los métodos estadísticos para tratar y analizar los datos.

Además, la mayoría de los estudios, excepto Biber *et al* (1999) en algunos análisis, no tienen en cuenta: a) análisis categorial, b) análisis de patrones, c) análisis semántico (clases, relaciones, etc.), d) análisis morfológico, e) análisis tipográfico (números, cifras, siglas, etc.) y f) análisis textual (cohesión, etc.). El único estudio que tiene realmente un análisis semántico es el de Oster (2003) para el par alemán-español.

Como consecuencia, no hay una visión amplia ni un conjunto amplio de análisis que permita hacer generalizaciones sobre este fenómeno. Una visión parcial de este fenómeno se ve favorecida por las siguientes causas:

1. La baja frecuencia de los SNEE en el conjunto de la lengua general.
2. Algunos estudios sólo se centran en el discurso especializado y en unos niveles de especialidad que en muchos casos pueden presentar una cantidad baja de SNEE; en ellos se encuentra muy poca cantidad del tipo de texto que aquí se estudia ('journal').
3. En general, los córpora de los diferentes estudios, excepto Biber *et al*, no podrían considerarse representativos ni del lenguaje general ni del especializado debido al tamaño de los córpora (muy pequeños), los tipos de texto, etc. Únicamente el corpus de base de Biber *et al* puede ser discutible en cuanto a que se hacen generalizaciones sobre el discurso especializado (llamado '*academic prose*').

Un aspecto que puede destacarse es que casi todos los estudios provienen de investigadores de ESP o de traducción, puesto que es en estas disciplinas donde este fenómeno se considera un verdadero problema.

Sin embargo, los mayores aportes los han hecho los gramáticos (Quirk *et al* 1985; Biber *et al* 1999; Huddleston y Pullum 2002) en aspectos como a) la categoría gramatical que predomina en la premodificación, b) el orden de los premodificadores, las restricciones categoriales, el orden "natural", etc. y c) las

relaciones y clases semánticas que pueden ocurrir en la premodificación. Estos aspectos no se especifican para los sintagmas con premodificación Extensa o muy extensa, pero pueden ser muy útiles para nuestro estudio, es decir, nos pueden ayudar a observar el orden de los premodificadores, las relaciones y clases semánticas que predominan en el discurso especializado y las funciones que cumplen.

5. Hipótesis e ideas previas

En este trabajo partimos de la idea de que la noción de sintagma nominal especializado extenso (SNEE) es el resultado de una serie de parámetros y rasgos que afectan:

- Las características de cada uno de los elementos del SNEE
- Las relaciones que se establecen entre los diferentes elementos que constituyen un SNEE y éste percibido como una unidad
- La relación entre el SNEE y el resto de la oración
- La función del SNEE dentro de la oración.

Además, la noción de SNEE presenta rasgos diferenciales respecto de la noción general de SN:

- Los SNEE presentan una diferencia estadística mayor en determinados registros y tipos de texto que lo propuesto por la bibliografía.
- Las condiciones de interacción de los interlocutores, reflejadas en el tipo de texto, hacen de los SNEE una característica muy relevante del discurso científico, como lo proponen Swales (1974: 29) y Halliday (1998: 193).
- Los SNEE son un elemento fundamental en la cohesión y compactación del discurso especializado.

Hasta donde se ha explorado en los trabajos anteriores (Quiroz 2003 y Quiroz *et al* 2004), los SNEE se manifiestan a través de una variedad de estructuras no descritas sistemáticamente para el par inglés-español.

Finalmente, los SNEE pueden detectarse y extraerse de los textos a partir de criterios léxico-sintácticos y semánticos. Hasta ahora sólo se han empleado los primeros.

De acuerdo con Cabré (2001), un texto especializado tiene unas condiciones tanto en la estructura textual como en la estructura informativa, que presentan a su vez características graduables como la precisión, la concisión, la sistematicidad, la objetividad y la impersonalidad. Todas ellas hacen que un texto sea más o menos especializado. Los SNEE pueden vehicular una gran densidad léxica y una alta precisión conceptual en un texto de especialidad mediante los componentes internos de un SN (acrónimos, símbolos, códigos, clasificaciones, compactación conceptual de la premodificación, entre otros).

Para esto se proponen las siguientes hipótesis o puntos de partida:

- Los SNEE no son un problema del discurso especializado, son un fenómeno de la lengua y, como tal, deben ser descritos y explicados.

Como se ha visto en el estado de la cuestión, muchos autores ven este fenómeno básicamente desde una perspectiva prescriptivista. Dentro del marco de la teoría comunicativa de la terminología (Cabré 1999: 34-35) y en consonancia con los principios metodológicos *f* y *g*²², este trabajo se enmarca dentro de una perspectiva descriptivista en la cual las unidades se extraen de textos reales que reflejan las condiciones comunicativas de un grupo de hablantes expertos o de un área en un nivel de especialidad determinado. Por tanto, se reconoce que estas unidades son un fenómeno natural dentro de la situación comunicativa en la que están situadas.

- La gramática de una lengua puede describir, clasificar, explicar y predecir todos los fenómenos lingüísticos de los discursos de los ámbitos de especialidad.

Partimos del planteamiento de Cabré (1993) en el cual los sintagmas terminológicos y libres son fenómenos lingüísticos de los discursos especializados que pueden explicarse mediante el uso de la gramática de la lengua general. La dicotomía, ya clásica, entre lenguaje general y lenguaje especializado o entre gramática de la lengua general y la gramática de los “lenguajes especializados” no se tendrá en cuenta en este trabajo de investigación puesto que la diferencia radica en que hay áreas de especialidad que usan determinados recursos lingüísticos, como los sintagmas nominales con premodificación extensa, con mayor frecuencia por razones pragmáticas debido a la situación comunicativa. Cabré (1999: 33), al explicar el carácter interdisciplinario de la terminología, afirma que la terminología recibe los aportes de “...una *teoría del lenguaje* que dé cuenta de las unidades de significación especializada (USE) dentro del lenguaje natural teniendo

²² f) El método es necesariamente descriptivo y consiste en la recopilación de las unidades reales usadas por los especialistas de un campo en distintas situaciones de comunicación. Esta diversidad de situaciones presupone que el corpus de extracción de los términos debe ser heterogéneo y representativo. Ello no impide que para un trabajo determinado pueda ser homogéneo tanto en su nivel de especialización y en el tipo de textos seleccionados, como también en la perspectiva de tratamiento del tema. Los términos seleccionados son unidades reales, no necesariamente satisfactorias ni normalizadas, simplemente reales. Solo en caso de que el trabajo pretenda ser una representación de los términos normalizados, se entrará en la actividad de fijación de una variante y/o reducción de la variación denominativa.

g) Las unidades retenidas en los textos como representativas del conocimiento especializado pueden ser términos (nominales, verbales, adjetivales) o unidades más amplias, combinaciones frecuentes en un determinado ámbito de especialidad. Estas combinaciones pueden ser terminológicas (unidades polilexemáticas denominativas de un concepto), fraseológicas (sintagmas no autónomos comunicativamente, que contienen por lo menos un término, habitualmente verbales frecuente y específicamente usados en una materia), o combinaciones aun más vastas que constituyen unidades oracionales propias de un área específica (órdenes informáticas, interjecciones deportivas de valor oracional, etc.).

en cuenta que participan de todas sus características, pero singularizando su carácter especializado y explicando cómo se activa este carácter en la comunicación...”

- Cuanto más especializado sea un texto, más SNEE debe tener.

Ya Swales (1974: 129) había planteado que el uso de sintagmas nominales extensos era ya una evidencia de especialización por la complejidad que ellos albergan. Sin embargo, Swales no proporciona evidencia empírica del hecho.

Dentro de la TCT, Cabré (2002) plantea que, desde el punto de vista de la densidad y la complejidad de un texto, a una secuencia más larga o a un patrón más largo desde el punto de vista estructural, el crecimiento de complejidad es exponencial. Esta complejidad es semántica y por tanto no es lineal; es jerárquica y por tanto la sintaxis y la morfología juegan con la semántica un papel fundamental. Los elementos semánticos de constructo jerárquico no son transparentes porque la sintaxis no permite ver claramente qué depende de qué, en muchos casos.

6. Metodología de trabajo

Para alcanzar los objetivos propuestos y demostrar las hipótesis planteadas, llevaremos a cabo la siguiente metodología.

6.1. Descripción del corpus

Para este trabajo se proponen tres tipos de corpóra de referencia para la extracción de los diferentes subcorpóra: un corpus en inglés, un corpus en español y un corpus paralelo inglés-español.

6.1.1. Descripción del corpus en inglés

Se ha recogido un corpus de aproximadamente 500.000 palabras en inglés tomado de varias revistas, entre ellas *The Lancet*, *Genomics* y *FEBS Letters*. Este corpus es parte del corpus técnico del IULA. Todos los textos son revistas con el formato IMMRAD (*Introduction, Materials and Methods, Results and Discussion*). Los criterios de selección para cada artículo son:

- Pertener al área del genoma
- Estar publicado en una revista de reconocido prestigio
- Estar disponible en versión electrónica
- Estar escrito por un hablante nativo del inglés: para esto se observa que el autor tenga apellidos de origen inglés; si esto es difícil de establecer, entonces se tiene en cuenta:
 - que estuviera escrito en un país de habla inglesa (Reino Unido, Estados Unidos, Canadá y Australia, principalmente),
 - que por lo menos un laboratorio o universidad de habla inglesa estuviera involucrado en la redacción.

6.1.2. Descripción del corpus en español

Se ha constituido un corpus de aproximadamente 500.000 palabras tomado del Corpus Tècnic del IULA. Los textos pertenecen a varias revistas con el formato IMMRAD (*Introduction, Materials and Methods, Results and Discussion*). Los criterios de selección para cada artículo son:

- Pertener al área del genoma
- Estar publicado en una revista de reconocido prestigio
- Estar disponible en versión electrónica
- Estar escrito por un hablante nativo del español: para esto se observa que el autor tenga apellidos de origen hispano, si esto es difícil de establecer, entonces se tiene en cuenta:

- que estuviera escrito en un país de habla española (Latinoamérica y España),
- que por lo menos un laboratorio o universidad de habla española estuviera involucrado en la redacción.

6.1.3. Descripción del corpus paralelo inglés-español

Por último, se ha recogido el corpus paralelo de unas 100.000 palabras aproximadamente tomado de la revista *The Lancet*, versión inglesa, y su respectiva versión española. Cada versión consta de 24 textos completos relacionados con el genoma. Para su elaboración, se emplearon palabras clave tales como: *gene*, *genome*, *mutation*, *genomics*, etc. Los artículos en inglés están en formato electrónico y la versión española sólo se encuentra en papel, con lo cual hubo que escanear los artículos. Todos los artículos están escritos entre los años 1995 y 1999, año en el cual se dejó de traducir al español la revista *The Lancet*. Se usaron los criterios de selección de los apartados 6.1.1 y 6.1.2. Sin embargo, dada la gran dificultad para conseguir corpórea paralelos de este nivel de especialidad, primó el criterio de disponibilidad y adquisición de la versión española.

6.2. Herramientas

A lo largo de todo el trabajo de investigación, se emplearán varios tipos de herramientas para procesar, almacenar y presentar los datos. Los corpus de referencia del inglés y el español ya están procesados. Para procesar los datos del corpus paralelo se usarán las herramientas del Corpus Tècnic del IULA.

Para extraer los SN de los corpórea de referencia emplearemos el Repoker o las herramientas del IULA, y para procesar los patrones y extraer los tipos de núcleos y otros datos aplicaremos el Simple Concordance Program (SCP)²³ y el System Quirk²⁴. Para comparar los SN del corpus paralelo, se usará el paralelizador Alipa del IULA.

Para almacenar los SN y los datos de análisis, usaremos la base de datos MS-Access, y de igual modo se almacenarán en Multiterm iX²⁵ para editarlos e imprimirlos dadas las prestaciones que tiene este programa frente a los otros.

6.3. Metodología de análisis

6.3.1. Criterios para la extracción de los SN

²³ SCP es un analizador de texto de libre acceso (<http://www.textworld.com>).

²⁴ System Quirk es una estación de trabajo terminológica de libre acceso creada en la Universidad de Surrey (<http://www.msc.surrey.ac.uk>).

²⁵ Multiterm iX es una base de datos terminológicos bajo licencia de la empresa Trados. (<http://www.trados.com>).

A partir de los resultados obtenidos en Quiroz (2003), y en especial en Quiroz *et al* (2004), además de la observación de trabajos que incluyen unidades similares, hemos decidido extraer todas las unidades que, en general, cumplan con el criterio de tener al menos dos premodificadores en el SN que sean categorías abiertas, es decir, sustantivos, adjetivos (-ing, -ed) y adverbios. Las otras categorías no se tendrán en cuenta cuantitativamente, pero sí para otros tipos de análisis. De igual modo, en español no se contarán categorías frecuentes en los SN, como los determinantes que están a la izquierda del núcleo. Además, se extraerán todos los SN que cumplan el requisito de extensión sin importar su grado de lexicalización. Por tanto, para el análisis cuantitativo tendremos en cuenta sintagmas endógenos y exógenos, compuestos (*compounds*) del tipo N+N+...N, etc.

Para extraer estas unidades de los textos, y teniendo en cuenta las dificultades expuestas en trabajos anteriores, se procederá a usar las reglas actuales, a crear nuevas reglas a partir del 20% de cada texto del corpus, a aplicar las tres opciones del Repoker y, posteriormente, a clasificar por patrones.

El Repoker nos permite extraer:

- los SN exactos obtenidos de las reglas
- los SN aproximados por etiqueta del núcleo de las reglas
- los SN aproximados a partir de las etiquetas de un SN muestra.

e introduce los SN paralelizados en los textos originales para observar posteriormente la alineación.

Para usar el programa Repoker, se deben dar los siguientes cuatro parámetros en una secuencia de comandos de Perl:

- opciones de la regla [1, 2, 3, 4]
- archivo de reglas en una de las lenguas [EN, ES]
- el nombre del archivo del corpus
- el nombre del archivo de salida

En el corpus paralelo, se extraerán los SN en inglés y luego se extraerán sus correspondientes SN en español mediante el mismo procedimiento descrito antes.

6.3.2. Criterios para el análisis estadístico

Para la depuración y el análisis estadístico de cada uno de los corpóra, usaremos una estrategia similar a la descrita en Quiroz (2004).

En primer lugar, se clasificarán y contarán los SN de cada lengua de menor a mayor extensión con su respectivo patrón. Se calcularán los porcentajes por número de palabras de categoría léxica abierta (4, 5, 6, 7, y 8 palabras) y se calcularán contra el número total de palabras (vocabulario). Así, podemos verificar si la tendencia observada en Quiroz (2004) continúa, es decir, corroboraremos si el porcentaje de SN extensos es mayor que en los otros estudios mencionados. Con ello reforzaremos nuestra hipótesis, desde un punto de vista cuantitativo, de que este fenómeno es “natural” en este tipo de texto y que su presencia es un rasgo que le caracteriza; se mostrará que lo es desde el punto de vista contrastivo con todas las implicaciones y dificultades que conlleva su comparación o traducción en el par inglés-español.

En segundo lugar, se extraerán los patrones en un fichero para ser analizados en el Simple Concordance Program (SCP), el cual nos permite ver las regularidades de los patrones de superficie y de las diferentes categorías si esto no se puede hacer con el Repoker. Se calcularán los porcentajes de las frecuencias de los patrones en general y se calcularán los porcentajes de las frecuencias de los patrones por predominio de categoría y por exclusión de categoría, es decir, se clasificarán los patrones que tienen N, A, Adv, PP, PPI en la premodificación y también, de forma excluyente, patrones que tienen N en la premodificación pero no A y patrones que tienen A en la premodificación pero no N. De este modo, se definirán los patrones de superficie más frecuentes, la categoría gramatical que predomina en la premodificación y el comportamiento del resto de categorías en los patrones.

Para el corpus paralelo, se seguirá el mismo procedimiento, pero siempre ordenando ambas lenguas al mismo tiempo para poder observar luego las regularidades en las soluciones respecto del inglés.

6.3.3. Análisis de las categorías

Con los datos estadísticos se procederá a analizar cada categoría y a confrontar algunas de ellas dentro de la premodificación, como se indica en los numerales siguientes.

6.3.3.1. Análisis morfológico por categoría gramatical

Se observarán los rasgos morfológicos más frecuentes de cada categoría gramatical en cuanto a prefijos, sufijos, etc. Para este análisis y dadas las características del discurso biomédico se usará el programa Tapafoc del IULA para el español, el cual trataremos de adaptar al inglés.

- Análisis por categoría gramatical N
- Análisis por categoría gramatical A
- Análisis por categoría gramatical PP
- Análisis por categoría gramatical PPI
- Análisis por categoría gramatical Adv

- Análisis por categoría gramatical Conj
- Análisis por categoría gramatical D
- Comparación de las categorías gramaticales N y A
- Comparación de las categorías gramaticales PP y PPI
- Análisis de otras categorías presentes

En español, incluiremos la preposición y los determinantes que no sean parte del núcleo. Este análisis no se realizará en el corpus paralelo.

6.3.3.2. Análisis de los patrones de superficie por coocurrencias

Los estudios llevados a cabo por Boughedaoui (1995, 1996, 1997, 1998, 2001), Maniez (2001) y Ormord (2001) sobre la coocurrencia de algunos patrones binarios dentro de la premodificación en inglés y las posibles traducciones al francés nos parecen útiles para el par inglés-español pues también nos ayudarían a crear posibles reglas para resolver la traducción o el tratamiento de estos SN.

Por tanto, se estudiarán los patrones de coocurrencia y otros patrones más frecuentes entre los siguientes tipos de estructuras para el inglés:

- Análisis de las secuencias de sustantivos N+N
- Análisis de la secuencia Adv-Iy-PP
- Análisis de la secuencia Adv-A
- Análisis de la secuencia A-N
- Análisis de la secuencia A-A

Para el español, tendremos en cuenta:

- Análisis de las secuencias de N+N
- Análisis de la secuencia Adv-mente-PP
- Análisis de la secuencia Adv-A
- Análisis de la secuencia A-N
- Análisis de la secuencia N A

6.3.4. Análisis de los tipos de modificadores respecto del orden dentro de la premodificación

Para el análisis de los tipos de modificadores y las restricciones de orden de los SN, seguiremos los criterios de Quirk *et al* (1985: 1341 y ss)

Quirk *et al* (1985) dividen el análisis del orden de la premodificación en cuatro zonas: *Precentral*, *Central*, *Postcentral* y *Pre-head*²⁶, y estas a su vez en otras zonas, como se explica a continuación y se resume en el gráfico 1 de la página 43.

La primera zona se denomina *Precentral*: Esta incluye ‘non-gradable adjectives’ (‘intensifying adjectives’):

- *emphasizers* (certain, definitive, plain, pure,...)
- *amplifiers* (absolute, entire, extreme, perfect, total)
- *downtoners* (feble, sligth)

En la segunda zona denominada *Central* se encuentran los ‘gradable adjectives’ (adjetivos de grado) (big, powerful, slow, thick,...), es decir, aquellos que se les puede aplicar el test para adjetivos (uso atributivo en el SN, uso predicativo en la oración y modificación por ‘very’). Estos se dividen en los ‘Non-derived adjectives’ (big, powerful, slow) y en los ‘derived adjectives’: deverbales (interesting, interested, hesitant) y los denominales (angry, rainy, peaceful).

Según estos autores, el orden en la segunda zona en caso de coocurrencia es ‘nonderived+deverbal+denominal’. Además, los adjetivos de tamaño, longitud y altura van antes que los adjetivos no derivativos (nonderived adjectives).

El grupo de adjetivos emotivos, evaluativos o subjetivos (lovely, nice, wonderful, terrible, horrible,...) suelen preceder a los otros adjetivos centrales.

En la tercera zona, la *postcentral*, se pueden encontrar los participios de presente y pasado y los adjetivos de color.

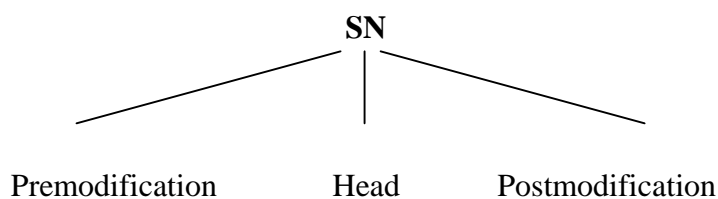
En la cuarta zona, la *prenuclear*, se localizan los premodificadores que menos se acercan al adjetivo pero que tienen más carácter nominal. Esta zona la podemos dividir en tres:

- adjetivos con propiedades de nombre propios que denotan nacionalidad, origen y estilo: American, Gothic...
- adjetivos con características morfológicas o semánticas relacionadas con sustantivos (con el significado de ‘consisting of’ o ‘relating to’: annual, economic, medical, social, political, rural,...)
- sustantivos

²⁶ Dejamos en muchos casos el nombre en inglés pues la traducción de algunos de ellos es problemática en español.

Si dos premodificadores de la misma clase coocurren, los adjetivos que denotan lugar/tiempo deben ir antes (*local economic interest, annual linguistic meeting*) y normalmente no pueden coordinarse.

El principio general sobre el cual Quirk *et al* (1985: 1341) se rigen para ordenar los modificadores es la polaridad subjetiva/objetiva: las propiedades inherentes al núcleo, visualmente observables y objetivamente reconocibles o accesibles, tienden a ir más cerca del núcleo. Si el adjetivo es una cuestión de opinión que no se puede observar visualmente tenderá a ir más lejos del núcleo.



Premodification

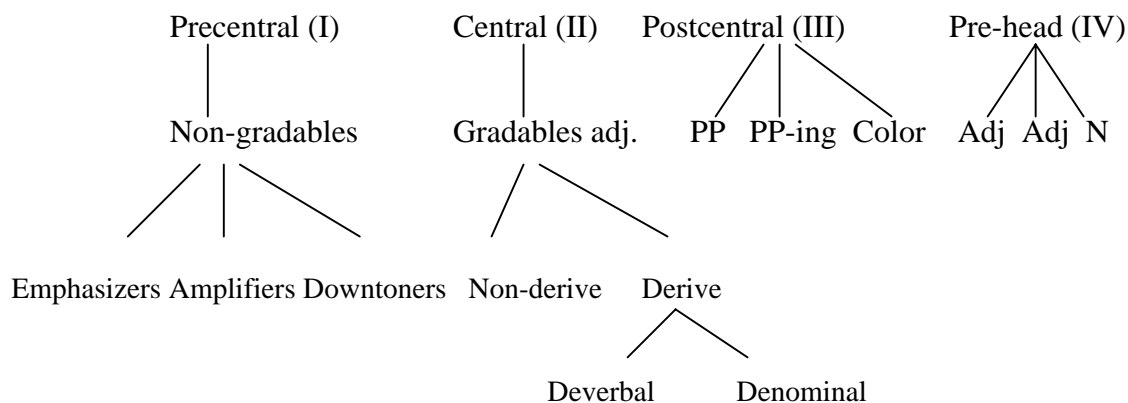


Gráfico 1. Orden de la premodificación en inglés (según Quirk et al 1984)

Se hará una plantilla en el programa MS-Access que refleje el gráfico anterior para observar las regularidades en la premodificación y qué orden y tipos de premodificadores predominan en el área del genoma.

Por ejemplo, el siguiente SN puede analizarse así:

[subsequent [placebo-controlled] [clinical [trials]]]]

A N PP A N

I III IV

También, nos ayudaremos de los criterios de Huddleston y Pullum (2002: 452-454) de los cuales podremos incluir:

Pre-head internal modifiers > Head

Dentro de los modificadores internos prenucleares, Huddleston y Pullum (2002: 452-454) distinguen dos grupos: ‘early pre-head internal modifiers’ y ‘residual pre-head internal modifiers’. Los ‘early pre-head internal modifiers’ son realizados por los determinantes, superlativos y adjetivos primarios.

Los ‘residual pre-head internal modifiers’ se han dividido en las siguientes clases:

Evaluative>General property>Age>Colour>Provenance>Manufacture>Type

an [attractive tight-fitting brand-new pink Italian lycra women’s] swimsuit

Cada clase de las anteriores puede incluir una subclasificación de modificadores del siguiente tipo:

- Evaluativos: modificadores que incluyen la evaluación del hablante y no una propiedad general objetiva (good, bad, etc)
- Propiedades generales: tamaño, dimensión, sonido, tacto, gusto, etc.
- Edad
- Color
- Procedencia
- Manufactura: modificadores de composición, modo, etc.
- Tipo

Para complementar los tipos de modificadores tendremos también en cuenta la clasificación clásica de Dixon (1977), modificada por Folgerà (2001) en un estudio sobre el adjetivo en ámbitos especializados:

- Adjetivos calificativos: color, dimensión, velocidad, forma, composición, edad, tacto, sonido, sabor, consistencia, aptitudes intelectuales, concreción, multiplicidad, evaluación, emoción, cantidad y comparación.
- Adjetivos relacionales: de tiempo, espacio, entidades, modales, etc.

De este modo, podemos observar las clases semánticas y el orden (y las restricciones) que predominan en los SN extensos.

6.3.5. Análisis de las relaciones semánticas dentro de la premodificación

Para el análisis semántico, se ha decidido utilizar el modelo planteado por Oster (2003) por las siguientes razones:

- Esta autora hace un análisis exhaustivo de las tendencias semánticas que pueden aportar al análisis semántico de los compuestos: semántica léxica, semántica oracional, semántica cognitiva, terminología, etc.
- Tiene en cuenta a muchos autores alemanes que no tenemos en general a nuestro alcance y que han hecho aportes relevantes para el análisis de sintagmas en ámbitos de especialidad.
- Hace un análisis contrastivo entre el alemán y el español de los sintagmas terminológicos del ámbito de la cerámica que será de mucha utilidad para nuestro trabajo en el ámbito del genoma para el par inglés-español.
- Su estudio se basa en un corpus.
- Su estudio aboga por un análisis multinivel en el cual se usan la noción de esquema como protocategoría semántica que abarca cada una de las funciones determinadas que establecen, a su vez, unas relaciones específicas a partir de los roles de los participantes del sintagma. Esta clasificación evita muchos solapamientos de relaciones y funciones que pueden ser interpretables en determinado momento.

En nuestro estudio se hará una selección del 20% de los SN y aplicaremos el análisis de esta autora. Si bien Oster (2003: 274 y ss) toma elementos para su análisis semántico de análisis hechos para la lengua general y desde las propuestas de autores clásicos de la terminología, no sabemos si muchos de estos esquemas semánticos, las funciones, las relaciones y los roles pueden hacer variar nuestra clasificación final. Por ello, tendremos en cuenta estudios más recientes del ámbito de la terminología como los de Feliu (2004), puesto que trabaja en el mismo ámbito de especialidad y marco teórico.

Según Oster (2003: 275), un esquema relacional es una estructura abstracta de conocimiento que sirve para reconocer e identificar el tipo de relación que une varios elementos. Oster (2003: 291) parte de la idea de los cuatro 'estados de

cosas' básicos de Dik (1989), i.e. acción, proceso, estado, posición, como lo hace Café (1999), para plantear una clasificación de esquemas relacionales, pero sólo se centra en las acciones y los procesos. La gran diferencia entre estas dos autoras radica en que Oster (2003: 289) no hace la distinción "forzosa" que hace Café (1999) entre predicado y argumento para analizar las relaciones semánticas de los constituyentes de los términos complejos. Dicha distinción es errónea pues el estado de cosas se refiere al predicado y para el caso de los términos no da a lugar. Podemos resumir el esquema de análisis de la siguiente manera:

1. Esquema de acción
2. Extensiones del esquema de acción
 - 2.1 Esquema factivo
 - 2.2 Esquema instrumental
 - 2.3 Esquema final
3. Esquema de proceso
4. Extensiones del esquema de acción o de proceso
 - 4.1 Esquema causal
5. Esquema de estado
 - 5.1 Esquema atributivo
 - 5.2 Esquema predicativo
 - 5.3 Esquema de oposición
 - 5.4 Esquema de localización espacial
 - 5.5 Esquema de localización temporal
6. Esquema de posición
7. Esquema de imagen
 - 7.1 Esquema contenido-contenedor
 - 7.2 Esquema partitivo
8. Esquema metafórico

Y las funciones necesarias como elementos de los esquemas relacionales anteriores son:

- Acción: actividad controlada que tiene una finalidad directa
- Agente: la entidad que controla una acción
- Meta: la entidad afectada por la acción de un agente o el control de un positor
- Proceso: actividad que se produce sin constituir una finalidad en sí
- Procesado: la entidad que sufre un proceso
- Instrumento: medio empleado por un agente para realizar una acción

- Finalidad: resultado que un agente persigue con su acción
- Causa: actividad que origina un proceso o la aparición de una entidad
- Efecto: resultado de una opción o proceso
- Positor: la entidad que controla una posición
- Cero: la entidad implicada en primer lugar en un estado
- Referencia: la entidad implicada en un segundo término en un estado o con referencia a la cual se entiende que se da la relación.

Además, para las relaciones entre secuencias N+N se tendrá en cuenta la propuesta de Barker y Szpakowics (1998: 97) y, para los otros tipos de funciones semánticas que no incluye Oster, las propuestas de intensidad y agrupamiento de Café (1999).

6.3.6. Análisis de los tipos de núcleos

Para asignar el tipo de núcleo de cada SN emplearemos las 25 categorías de WordNet 2.0²⁷. Para ello, contrastaremos todos los núcleos de los SN con las listas que tenemos de WordNet usándolas como palabras clave o de inclusión en un sistema como SCP o System Quirk. Por medio de estos programas, sabremos cuáles núcleos se refieren a las clases semánticas de procesos, sustancias, objetos, formas, estados, lugares, fenómenos, etc. Los resultados nos permitirán saber qué tipos de núcleos predominan en los SN del genoma. A continuación se presenta la lista de categorías del WordNet:

{act, action, activity}
{animal, fauna}
{artifact}
{attribute, property}
{body, corpus}
{cognition, knowledge}
{communication}
{event, happening}
{feeling, emotion}
{food}
{group, collection}
{location, place}
{motive}
{natural object}
{natural phenomenon}
{person, human being}
{plant, flora}
{possession}
{process}
{quantity, amount}
{relation}

²⁷ WordNet 2.0 es un programa de libre acceso creado en la Universidad de Princeton (www.WordNet.princeton.org).

{shape}
{state, condition}
{substance}
{time}

6.3.7. Análisis de la expansión de los SN

Para el subcorpus de expansión de los SN se harán los siguientes análisis:

- Expansión lineal del SN
- Expansión por coocurrencias de patrones de dos elementos
- Expansión por coocurrencias de patrones de tres elementos
- Expansión por círculos concéntricos
- Relaciones entre las expansiones y el núcleo
- Otros tipos de expansión.

6.3.8. Análisis comparativo del corpus paralelo

El análisis estadístico de patrones ya se ha realizado en el corpus paralelo. Asimismo, se ha hecho también el análisis semántico a un subcorpus de éste. Con estos datos podemos realizar los siguientes análisis:

- Comparar los patrones de las unidades seleccionadas en inglés y sus correspondientes en español
- Comparar los patrones del corpus paralelo en español con los patrones del corpus de referencia en español
- Analizar las regularidades de los patrones y sus soluciones de traducción
- Comparar las secuencias de las relaciones semánticas.

Esto nos permitirá definir los patrones más regulares en inglés, sistematizar las soluciones en español de los SN del inglés y observar las relaciones semánticas y los tipos de núcleos más frecuentes cuando las dos lenguas están en contacto, para así plantear recomendaciones para su tratamiento en la traducción.

6.3.9. Propuesta de ficha para almacenar la información

Para almacenar y manejar adecuadamente los datos concernientes a cada sintagma nominal, se propone el siguiente modelo de ficha:

Sintagma nominal:

Contexto:

Patrón superficial:

Tipo de núcleo:

Tipos de modificadores:

Relaciones semánticas internas:

Estructura jerárquica:

Para ejemplificar su uso hemos tomado dos SN extensos de Quirk *et al* (1985)

Sintagma nominal: a new giant size cardboard detergent carton
Contexto: ...A new giant size cardboard detergent carton...
Patrón superficial: D A A N N N N
Tipo de núcleo: {artifact}
Tipos de modificadores: age, size, property, material, substance, object1
Relaciones semánticas internas: property, kind of, kind of, content
Estructura jerárquica: A [new [giant size] [cardboard [detergent carton]]]]

Sintagma nominal: expensive overseas income tax office furniture
Contexto: ... expensive overseas income tax office furniture...
Patrón superficial: A N N N N N
Tipo de núcleo: {artifact}
Tipos de modificadores: property, place, object2, object2, place, object1
Relaciones semánticas internas: property, kind of, kind of.
Estructura jerárquica: [expensive [overseas [income [tax [office furniture]]]]]]

Finalmente, los datos se almacenarán en la base de datos terminológicos Multiterm iX para su edición e impresión pues, en este sentido, tiene más prestaciones que MS-Access.

7. Bibliografía

Abberton, Evelyn. (1977). "Nominal Group Premodification Structures". En: Bald, W-R. and Ilson, R. (Eds). (1977). *Studies in English Usage: The resources of a present-day English corpus for linguistic analysis*. Frankfurt am Main: Peter Verlag. 29-72.

Abril Martí, Isabel; Ortiz Urbano, Cocha. (1998) "Formación de intérpretes de conferencia en el ámbito biosanitario inglés/español: la experiencia de la Facultad de Traducción e interpretación de la Universidad de Granada". En: *Traducción e interpretación en el ámbito biosanitario*. Fernández Félix, Leandro; Ortega Arjónilla, Emilio (Coords.) Granada: Comares.

Alcaraz Varó, Enrique. (2000). *El inglés profesional y académico*. Madrid: Ed. Alianza. ISBN 84-206-5754-9.

Adelstain, Andreína. (1998). "Condiciones de la reductibilidad léxica de los sintagmas terminológicos". En: *Actas del VI Simposio Iberoamericano de Terminología RITerm*, La Habana. CD-ROM.

Alonso, Araceli; Cabré, Teresa; De Yzaguirre, Lluís; Tebé, Carles. (2002). "La utilización de corpus paralelos alineados en la docencia de la traducción y de los lenguajes de especialidad". En: Iglesias, L.; Doval, S. (ed.) *Studies in Contrastive Linguistics. Proceedings of the Second International Contrastive Linguistics Conference*, pp. 71-82. Santiago de Compostela: Publicación de la Universidade de Santiago de Compostela. ISBN: 84-9750-027-X.

Amaro de Melo, Bianca. (1998). *Las unidades terminológicas complejas en el área de telecomunicaciones*. Trabajo de investigación. Dir. Teresa Cabré y Enilde Faulstich. IULA-UPF, Barcelona.

Banks, David. (Ed.). (2001). *Le group nominal dans le texte spécialisé*. Paris: L'Harmattan. Pp XII+179. ISBN 2-7475-1584-2

Bark, Julia. (1980). *Let's Write English*. New York: Academic Press.

Barker, Ken. (1998). "A Trainable Bracketter for Noun Modifiers". En: *AI*, 196-210.

Barker, Ken; Szpakowicz, Stan. (1998). "Semi-automatic Recognition of Noun Modifier Relationships". *Coling-ACL '98, Proceedings 36th Annual Meeting of the Association for Computational Linguistics*. Montreal. 96-102.

Bauer, Laurie. (1982). *English Word-formation*. Cambridge: Cambridge University Press. ISBN 0-521-28492-9.

Bennett, Paul (1993) "A Multilingual Translation-Oriented Typology of Compound Nouns". En: *T.A.L.* vol. 34 (2) pp. 48-58.

Biber, Douglas; Johansson, Stig; Conrad, Susan; Finegan, Edward. (1999). *Logman Grammar of Spoken and Written English*. London: Logman.

Blake, Gari; Bly, W. Robert. (1993). *The Elements of Technical Writing*. New York: MacMillan. ISBN 0-20-013085-6.

Boughedaoui, Mourand. (1995). "Les séquences collocationnelles et la dynamique de la composition adjectivale". En: *Les Cahiers de l'Apliut*, 15 (1). 47-58.

Boughedaoui, Mourand. (1996). "Essai de categorisation sémantique des adjectives composés (1)". En: *Les Cahiers de l'Apliut*, 16 (1).

Boughedaoui, Mourand. (1996). "Essai de categorisation sémantique des adjectives composés (2)". En: *Les Cahiers de l'Apliut*, 16 (2). 37-54.

Boughedaoui, Mourand. (1997). "Contribution à l'amélioration de la compréhension et de la traduction des adjectifs composés en classe de langue de spécialité". En: *Asp*, 15-18. 225.

Boughedaoui, Mourand. (1998). "Comparative Study of the Distribution of Adverb-Adjective Combinations with a Special Concern in English for Statistics". En: *Les Cahiers de l'Apliut*, 17 (2). 37-54.

Boughedaoui, Mourand. (2001). "Contribution des associations syntagmatiques adjectivales à la complexification du groupe nominal dans le texte spécialisé". En: Banks, David (ed.). *Le group nominal dans le texte spécialisé*. Paris: L'Harmattan. 137-150.

Bourigault, Didier. (1993). "Analyse syntaxique locale pour le repereage de termes complexes". En: *T.A.L.*, Vol 34 (2). 105-117.

Burnett, Rebeca. (1992). *Technical Communication*. California: Wadsworth.

Cabré, María Teresa. (1993). *La terminología: teoría, metodología, aplicaciones*. Barcelona: Antártida. [traducido por Tebé, Carles].

Cabré, María Teresa. (1999). “Elementos para una teoría de la terminología: hacia un paradigma alternativo”. En: *La Terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra. pp. 69-92.

Cabré, María Teresa. (1999). “Hacia una teoría comunicativa de la terminología: aspectos metodológicos”. En: *La terminología: Representación y comunicación*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. 129-150.

Café, Ligia. (1999). *La Description et l'analyse des unités terminologiques complexes en langue portugaise (variété brésilienne)*. Tesis doctoral. Dir. Pierre Auger et Enilde Faultish. Université Laval.

Cambridge Dictionary of American English en CD-ROM. Cambridge: Cambridge University Press, 2000.

Cardero, Ana María. (2000). “En torno a la frecuencia de algunas estructuras sintácticas en terminología”. En: *Actas del VIII Simposio Internacional de la Red Iberoamericana de Terminología*. Lisboa: Editorial Colibrí.

Cartagena, Nelson. (1998). “Acerca de la variabilidad de los términos sintagmáticos en textos españoles especializados”. En: Wotjak, Gerd (ed.). *Estudios de fraseología del español actual*. Madrid y Frankfurt: Iberoamericana. 281-296.

Coates, Jennifer. (1977). “A Corpus Study of Modifiers in Sequence”. En: Bald, W-R. and Ilson, R. (Eds). (1977). *Studies in English Usage: The resources of a present-day English corpus for linguistic analysis*. Frankfurt am Main: Peter Verlag. 9-27.

Collet, Tanja. (1997). “La réduction des unités terminologiques complexes de type syntagmatique”. En: *Meta, journal des traducteurs*, 42 (1). Montréal: Presse de l'Université de Montréal. 193-206.

Collet, Tanja. (2003). "A Two-Level Grammar of the Reduction Process of French Complex Terms in Discourse" En: *Terminology* 9 (3). JB. 1-27.

David, Sophie. (1993). *Les unités nominales polylexicales: Éléments de description et reconnaissance automatique*. Tesis doctoral. Dir. F. Corblin. Université Denis Diderot, Paris.

Diccionario Espasa Medicina; Facultad de Medicina, Universidad de Navarra. Madrid: Espasa Calpe. 1999

Dixon, Robert. (1977). "Where Have All the Adjectives Gone?" En: *Studies in Language*, 1. 19-80.

Downing, Pamela. (1977). "On the Creation and Use of English Compound Nouns". En: *Language*, Vol 53, 4, pp 810-842.

Drouin, Patrick. (1997). "Une méthodologie d'identification automatique des syntagmes terminologiques: L'apport de la description du non-terme". En: *Meta, journal des traducteurs* 42 (1). Montréal: Presse de l'Université de Montréal. 45-54.

Estopà, Rosa. (1999). *Extracció de terminologia: elements per a la construcció d'un SEACUSE*. Tesis Doctoral. Dir. Teresa Cabré. Institut Universitari de Lingüística Aplicada: Barcelona.

Estopà, Rosa. (2001). "Les unités de signification spécialisées: élargissant l'objet du travail en terminologie" En: *Terminology* 7 (2), 217-237.

Faultisch, Enilde. (2003) "Formação de termos: do constructo e das regras às evidências empíricas". [S.D].

Feliu C., Judit. (2004). *Relacions conceptuais i terminologia: anàlisi i proposta de detecció semiautomàtica*. Tesis doctoral. Dir. Teresa Cabré. IULA- Univeristat Pompeu Fabra.

Finin, Timothy W. (1980). *The Semantic Interpretation of Compound Nominals*. University of Illinois at Urbana-Champaign. University Microfilms International.

Finin, Timothy W. (1986). "Constraining the Interpretation of Nominal Compounds in a Limited Context". En: Grishman, Ralph; Kittredge, Richard

(Ed.). *Analysing Language in Restricted Domains*. New Jersey: Lawrence Erlbaum Associates. 163-173. ISBN: 0-89859-620-3.

Folguerà, Rosana. (2002). *Adjectius en el discurs especialitzat: Una primera descripció dels adjectius en els textos del Genoma Humà*. Trabajo de investigación. Dir. Mercè Lorente i Casafont. Institut Universitari de Lingüística Aplicada: Barcelona.

Georges, T. M. (1996). *Analytical Writing for Science and Technology*. En: <http://tgeorges.home.attbi.com/write/les13.htm> [Consultado el 15 de julio de 2003].

Halliday, Michael M. K. (1998). "Things and Relations: Regrammaticising experience and technical knowledge". En: Martin, J.R. and Veal, Robert (eds.). *Reading science: critical and functional perspective on discourses of science*. London: Roudledge.

Halliday, Michael M. K.; Hasan, Ruqaiya. (1976). *Cohesion in English*. London: Logman.

Halliday, Michael. M. K. (1991). *An Introduction to Functional Grammar*. London: Edward Arnold.

Herzon, Robert. (1971). "Gegenwartige Tendenzen in der terminologischen Wortbildung". En: *Mitteilungsblatt für Dolmetscher und Übersetzer* 17, 9/10. 3-6. En: Gotti, Maurizio (1991). *I linguaggi specialistici*. Nuova Italia.

Hetzon, Robert. (1978). "On the Relative Order of Adjectives". In: *Language Universal*. Tübingen: Gunter Narr. 165-184.

Hoffmann, Lothar. (1987) [1975]. *Kommunikationsmittel Fachsprache: Eine Einführung*. Tübingen Gunter Narr: Tübinga.

Hoffmann, Lothar. (1998). *Llenguatges d'especialitat: selecció de textos*. Barcelona: IULA.

Horsella, María; Pérez, Fresia. (1991). "Nominal Compounds in Chemical English Literature: Towards an Approach to Text Typology" En: *English for Specific Purposes*, Vol. 10 pp 125-138.

Huddleston, Rodney; Pullum, Geoffrey. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hurford, Jim. (1998). "The Interaction between Numerals and Nouns". En: Plank, F., (ed.). *Noun Phrase Structure in the Languages of Europe*. European Science Foundation.

Jastrab de Saint-Robert, Marie-Josée. (1987). "Les syntagmes nominaux complexes en anglais et en française". En: *Meta, journal des traducteurs*, 32 (3). Montréal: Presse de l'Université de Montréal. 260-266.

Kirkman, John. (1992). *Good Style: Writing for Science and Technology*. London: Chapman & Hall.

Kocourek, Rostilav. (1991). *La langue française de la technique et de la science: vers une linguistique de la langue savante*. 2.º edición. Wiesbaden: Brandstetter Verlag.

Ladouceur, Jacques; Drouin, Patrick. (1997). "Une analyse terminométrique pour le repérage automatique des descripteurs complexes dans les textes de spécialité". En: *Meta, journal des traducteurs* 42 (1). Montréal: Presse de l'Université de Montréal. 207-218.

Lehrberger, John. (1982). "Automatic Translation and the Concept of Sublanguage". En: Kittredge, Richard ; Lehrberger, John (eds.). *Sublanguage: Studies of Language in Restricted Semantic Domains*. Berlin: Walter de Gruyter. 81-106.

Leonard, Rosemary. (1984). *The Interpretation of English Noun Sequences on the Computer*. North-Holland Linguistic Studies, Elsevier, the Netherlands.

Levi, Judith N. (1978). *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.

L'Homme, Marie-Claude. (1994). "Traitement des groupes nominaux en traduction automatique: opportunité d'un codage conceptuel". En: *Proceedings of the Workshop on Nominal Compounds: Multilingual Aspects of Nominal Composition*. Bouillon, Pierre; Estival, Dominique (eds). Université de Genève. 147-161.

L'Homme, Marie-Claude. (1997). "Méthode d'accès informatisé aux combinaisons lexicales en langue technique". En: *Meta, journal des traducteurs* 42 (1). Montréal: Presse de l'Université de Montréal. 15-23.

Linder, Daniel. (2002). "Translating Noun Clusters and 'Nounspak' in Specialized Computer Text". En: Chabas, José; Gaser, Rolf; Rey, Joëlle (eds). *Translating Science*. Barcelona: Universitat Pompeu Fabra. ISBN 84-477-0820-9.

López Guix, Juan Gabriel; Minett Wilkinson, Jacqueline. (1997). *Manual de traducción español-inglés*. Barcelona: Gedisa.

Lorente, Mercè. (2001). "Altres elements lèxics". Solà, Joan (Dir.) *Gramàtica del català contemporani (Gcc)*. Barcelona: Empúries, Cap. 8.

Maalej, Zouhair. (1994). "English-Arabic Machine Translation of Nominal Compounds". En: *Proceedings of the Workshop on Nominal Compounds: Multilingual Aspects of Nominal Composition*. Bouillon, Pierre; Estival, Dominique (eds). Université de Genève. 135-146.

Malgorzata, Tryuk. (2000). "La phraséologie en terminologie: Quelques problèmes de traduction". En: *Babel* 46 (1). 66-76.

Maniez, Françoise. (2001). "Extraction d'une phraséologie bilingue en langue de spécialité: corpus parallèles et corpus comparable". En: *Meta, journal des traducteurs* 46 (3). Montréal: Presse de l'Université de Montréal. 553-563.

Méndez Cedón, Beatriz. (2002). *Estrategias fraseológicas en el género discursivo de los artículos científicos médicos en lengua inglesa*. Tesis doctoral, Dir. Purificación Nistal F., Universidad de Valladolid.

Meunier-Crespo, Mariette. (1997). "Les locutions nominales dans les dictionnaires de spécialités". En: *Meta, journal des traducteurs* 42 (1). Montréal: Presse de l'Université de Montréal. 69-71.

Meyer, Ingrid; Mackintosh, Kristen. (1996). "Refining the terminographer's concept-analysis methods: How can phraseology help?" En: *Terminology* 3 (1). 1-26.

Montero, Begoña. (1995). "Noun Premodifications vs. Postmodification in Scientific English". En: *Unesco-Alsed LSP Newsletter* 18, 2 (40). 14- 27.

Montero, Begoña. (1995). *La estructura del grupo nominal complejo en el inglés científico escrito. Sus componentes premodificadores y sus correspondencias en español*. Microfilmed Doctorate Thesis. Valencia: Universitat de Valencia.

Montero, Begoña. (1996). "Technical Communication: Complex Nominals Used to Express New Concepts in Scientific English - Causes and Ambiguity in Meaning". En: *The ESP* 17 (1). 57-72.

Myking, Johan. (1989). "Complex Noun Phrase as a Problem of Terminological Practice". En: Laurén, Christer; Nordman, Marianne (Eds). *Special Language: From Humans Thinking to Thinking Machines*. Clevedon: Multilingual Matters Ltd. 265-274. ISBN: 1-85359-033-9.

Naulleau, Eli. (1998). *Apprentissage et filtrage syntaxico-sémantique de syntagmes nominaux pertinents pour la recherche documentaire*. Tesis doctoral. Dir. Daniel Kaiser, Université Paris XIII.

Newmark, Peter. (1981). *Approaches to Translation*. London: Oxford Pergamon Press.

Newmark, Peter. (1988). *A Textbook of Translation*. London: Prentice Hall International.

Norman, Guy. (1999). *Cómo escribir un artículo científico en inglés*. Madrid: Astrazeneca. ISBN 84-921124-2-5.

Ormod, Janet. (2001). "Construction discursive de noms composés dans des textes scientifiques anglais". En: Banks, David (ed.). *Le group nominal dans le texte spécialisé*. Paris: L'Harmattan. 9-24.

Oster, Ulrike (2003). *Los términos de la cerámica en alemán y español: Análisis semántico orientado a la traducción de los compuestos nominales en español*. Tesis doctoral. Dir. Amparo Alcina Caudet y Pilar Elena García. Univesitat Jaume I, Castellón.

Pugh, A. K.; Ulijn, J. M. (1984). "Reading for Professional Purposes". London: Hienemann.

Pugh, Jeanette. (1984). "Contrastive Analysis of Noun Compound Terms in English, French, and Spanish within a Restricted, Specialized Domain". En: Hartmann, R. R. K. *Proceedings LeXeter '93*. Tübingen: Niemeyer. 395-400.

Quirk, Randoldh; Greenbaum, Sidney; Leech, Geoffrey; Svartik, Jan. (1985). *A Comprehensive Grammar of the English Language*. London: Logman.

Quiroz, Gabriel. (2003). *Los sintagmas nominales especializados extensos en inglés: Primera descripción en un corpus de genoma*. Trabajo de investigación de primera línea de doctorado. Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona.

Quiroz, Gabriel; De Yzaguirre, Lluís; Lorente, Mercè. (2004). “El uso de corpus paralelos para la identificación de sintagmas terminológicos extensos: ingeniería lingüística al servicio de problemas de la traducción”. En: *Actas del 3.º Congreso Internacional de Traducción Especializada*. Universitat Pompeu Fabra, Barcelona.

Quiroz, Gabriel; Muñoz, Carlos. (1997). *La traducción hacia lengua extranjera mediante rastreos terminológicos en área de la malaria*. Trabajo de investigación de especialización. Universidad de Antioquia, Medellín.

Sager, Juan Carlos; Dungworth, D.; McDonald, P. F. (1980). *English Special Languages. Principles and Practice in Science and Technology*. Wiesbaden: Brandstteter.

Salager-Mayer, Françoise. (1984). “Compound Nominal Phrases in Scientific-Technical Literature: Proportion and Rationale”. En: Pugh, A. K.; Ulijn, J. M. (eds). *Reading for Professional Purposes*. London: Heinemann.

Stockwell, Robert; Bowen, J. Donald; Martin, John W. (1965). *The Grammatical Structures of English and Spanish*. Chicago: University of Chicago Press.

Swales, John. (1974). *Writing Scientific English*. [S.L]: Nelson.

Thouvenin, Susan P. (1996). *The Identification and Exemplification of Multi-Word Units within a Technical Corpus of English, Including an Investigation of Nominal Groups*. M.Sc. Thesis, University of Aston, England. Tomada de <http://www.les.aston.ac.uk/lisu/diss/>.

Trimble, Louis. (1985). *English for Science and Technology*. Cambridge: Cambridge University Press.

Vanderwenden, Lucretia. (1995). *The Analysis of Noun Sequences Using Semantic Information Extracted from On-Line Dictionaries*. Tesis doctoral, Dir. Donald Lortiz. Georgetown University.

Varantola, Krista. (1984). *On Noun Phrase Structures in Engineering English*. Turku: Turun Yliopisto.

Vázquez-Ayora, Gerardo. (1977). *Introducción a la traductología*. Washington: Georgetown University Press.

Velásquez, Gonzalo. (1994). *Proceso, Métodos y Técnicas de la Traducción*. Escuela de Idiomas, Universidad de Antioquia, Colombia.

Vinay, Jean Paul; Dalbènet, Jean. (1958). *Stylistique comparée du français et de l'anglais*. Paris: Didier. ISBN 2-278-00894-3.

Woolley, Reuben. (1997). *Compound Nominal Groups in the Machine Translation of Medical English: Lexical Units or Analysable Sequences?* M.Sc Thesis, University of Aston, England. Tomada de <http://www.les.aston.ac.uk/lsu/diss/>.

Zabala, Igone. (1998). “La traducción al vasco de los sintagmas nominales complejos del lenguaje técnico”. En: *Actes del III Congrès Internacional sobre Traducció, UAB*. 589-603.

8. Programas de procesamiento y almacenamiento

BwanaNet y Corpus Tècnic de l'IULA de la Universitat Pompeu Fabra (CT-IULA).

Constraint Grammar en línea. En: www.Conexor.fi

Multiterm 95+ Database Management System de Trados. 1995.

Multiterm iX Database Management System de Trados. 2003.

Repoker. Programa para la extracción de datos lingüísticos etiquetados. Desarrollado por el Dr. Lluís de Yzaguirre. 2004

Simple Concordance Programe V. 4.0.6 (1997-2002) by Alan Reed.
<http://www.textworld.com/scp/index.html>

System Quirk V. 2.1 Beta. University of Surrey. AI Group, SEEITM.

The Webster's New World Dictionary & Thesaurus en CD-ROM. Versión 2.0
Macmillan Publishers, 1998.

Thesarus.com. Lexico Publishing Group, LLC. <http://www.thesaurus.com>, 2003.
[Consulta el 22.01.2003 y 28.07.2003].

WordNet 2.0. Universidad de Princeton, 1995-2004.