

TERMINOLOGÍA Y FRASEOLOGÍA

M. Teresa Cabré, Rosa Estopà, Mercè Lorente
Grupo IULATERM
Institut Universitari de Lingüística Aplicada
Universitat Pompeu Fabra (Barcelona)

1.- Introducción

La selección de terminología para la elaboración de un glosario especializado se encuentra en la etapa de extracción o vaciado terminológicos de la documentación ante cuatro problemas cruciales. El primer problema lo constituyen los criterios que deben usarse para determinar si una unidad es o no terminológica (**criterio de especificidad temática**). El segundo problema es saber si las unidades consideradas terminológicas son propias o no de un campo especializado (**criterio de atribución de ámbito**). La tercera cuestión consiste en establecer los criterios que permitan determinar si las unidades que aparecen en la documentación de trabajo son o no pertinentes para un determinado vocabulario (**criterio de pertinencia**). El cuarto problema lo constituye saber delimitar el segmento formal que corresponde a una unidad terminológica, o dicho de otro modo, establecer cuál es su forma lingüística; esta no ofrece dudas en el caso de unidades simples o sintagmáticas altamente específicas, pero provoca una gran confusión ante unidades semiespecíficas de forma sintagmática (**criterio de delimitación**).

Si bien estos problemas han sido planteados constantemente por la terminología, o mejor por la terminografía, la automatización del trabajo terminológico los ha puesto nuevamente sobre la mesa por una motivación exclusivamente práctica: para realizar extracción automática de terminología es necesario que el ordenador posea unos criterios absolutamente explícitos que le permitan realizar cuatro operaciones sobre los términos: reconocerlos, seleccionarlos, delimitarlos y representarlos.

En nuestro trabajo, basándonos en materiales obtenidos de un corpus textual del ámbito jurídico-administrativo, nos proponemos, a título experimental, varios objetivos. En primer lugar, queremos presentar una tipología estructural de las unidades terminológicas centrándonos en la definición de unidad terminológica de carácter sintagmático, que, siguiendo a Corbin (1994), consideraremos que es una unidad polilexemática. En segundo lugar, nos proponemos discutir brevemente la problemática frontera entre unidad terminológica polilexemática lexicalizada, sintagma de discurso y unidad fraseológica especializada, para proponer algunos criterios que un sistema automático pueda utilizar para distinguir entre unidades sintagmáticas libres, de un lado, y unidades sintagmáticas fijadas, de otro; y, dentro de estas últimas, llegar a establecer características que nos permitan diferenciar las unidades sintagmáticas totalmente fosilizadas (en todos los casos, correspondientes a términos), de aquellas concurrencias de alto grado de

frecuencia, que suelen corresponder a fraseologismos. Estos criterios se establecen a partir del análisis de materiales extraídos del corpus de lenguajes especializados que está desarrollando nuestro Instituto y se inscribe en la línea de investigación sobre Terminología.

La estrategia de trabajo que seguimos se basa, dentro del ámbito de la lingüística aplicada, en la relación entre descripción y aplicación; en esta línea, partiendo de la observación directa de los datos lingüísticos aportados por el corpus textual seleccionado, hemos establecido y aplicado criterios de detección y análisis, sirviéndonos de herramientas informatizadas que se proponen explorar los textos. Nuestra aportación pretende abordar el problema del reconocimiento automático de unidades terminológicas a partir de criterios de naturaleza distinta: cuantitativos y estructurales; en concreto nos proponemos explicitar estos criterios que servirán de base a nuestra futura experimentación en el campo de la extracción automática de terminología.

2. Criterios terminológicos y competencia

En la elaboración de terminología, como se ha dicho en el punto anterior, los terminólogos se encuentran ante cuatro problemas a resolver:

- el reconocimiento de los términos *versus* las palabras de uso común
- la atribución de cada término a un ámbito temático
- la consideración de su pertinencia para un determinado vocabulario
- la delimitación de las unidades complejas

La resolución de todas estas cuestiones, sin embargo, no puede llevarse a cabo de manera global, ya que la competencia requerida para abordarlas es diversa para cada una, y el conjunto de competencias necesarias no siempre coincide en el mismo tipo de especialista.

Ciertamente, si analizamos los dos primeros problemas observamos que mientras que el reconocimiento de los términos de un ámbito de especialidad es una tarea que sólo pueden resolver de manera directa los especialistas en la materia, la atribución a su ámbito de origen (noción distinta al reconocimiento de cada uno de los términos usados en un ámbito determinado) no puede ser resuelta por ellos, por cuanto no poseen conocimientos suficientes de todas las áreas de especialidad. Pueden identificar las unidades especializadas de su dominio, pero no las unidades de valor terminológico que pertenecen a dominios ajenos.

Paralelamente, una situación parecida se produce cuando un especialista intenta resolver las dos últimas cuestiones: considerar la pertinencia de un término para un vocabulario y delimitar las unidades complejas. Para resolver el primer caso, sólo alguien formado en metodología terminológica que sepa aplicar adecuadamente los principios de caracterización de un vocabulario (tema, destinatarios, funciones y funcionalidades) podrá definir si una unidad (independientemente del hecho de que sea originaria de uno u otro ámbito) debe formar parte de la nomenclatura de un diccionario especializado, ya que sólo los objetivos y características generales de un producto terminológico deben condicionar la selección de las unidades que este debe contener.

En lo que se refiere a la cuarta cuestión, para establecer los límites del segmento que corresponde a un término de estructura sintagmática, la competencia en el tema es una condición necesaria para el reconocimiento de su carácter de término, pero no es una condición suficiente. En efecto,

en los ámbitos especializados, más técnicos que científicos, observamos que aparecen muy frecuentemente combinaciones lexemáticas que corresponden a descripciones de conceptos; estas combinaciones a veces forman una unidad terminológica y otras son paráfrasis de dichas unidades. En estos casos, sólo un buen conocimiento lingüístico de las estructuras posibles y probables de cada ámbito especializado podrá ayudar a diferenciar las unidades terminológicas de las simples combinaciones de discurso que las describen, y que representan, *grosso modo*, su definición. Serán también los criterios gramaticales los que decidirán más explícitamente el segmento exacto que corresponde a una unidad terminológica. Finalmente, en la difícil frontera entre un término de estructura sintagmática y una combinación con un alto valor de aparición, sólo los criterios cuantitativos, los formales, aducidos por el análisis lingüístico, y el reconocimiento por la competencia permitirán establecer diferenciaciones con un cierto rigor y precisión.

3. Las unidades polilexemáticas

Las unidades lingüísticas formadas por más de un lexema són las unidades que realmente suponen un problema para los diseñadores de programas automáticos o semiautomáticos de extracción de terminología. En efecto, estos programas no tienen suficientes criterios para seleccionar por una lado, las unidades polilexemáticas de los fragmentos de discurso, y del otro las unidades terminológicas polilexemáticas de la fraseología especializada.

En este trabajo, partimos del supuesto que las cadenas sintácticas se pueden dividir en dos grandes grupos:

- unidades polilexemáticas
- fragmentos de discurso

A su vez, las unidades polilexemáticas se pueden subdividir en:

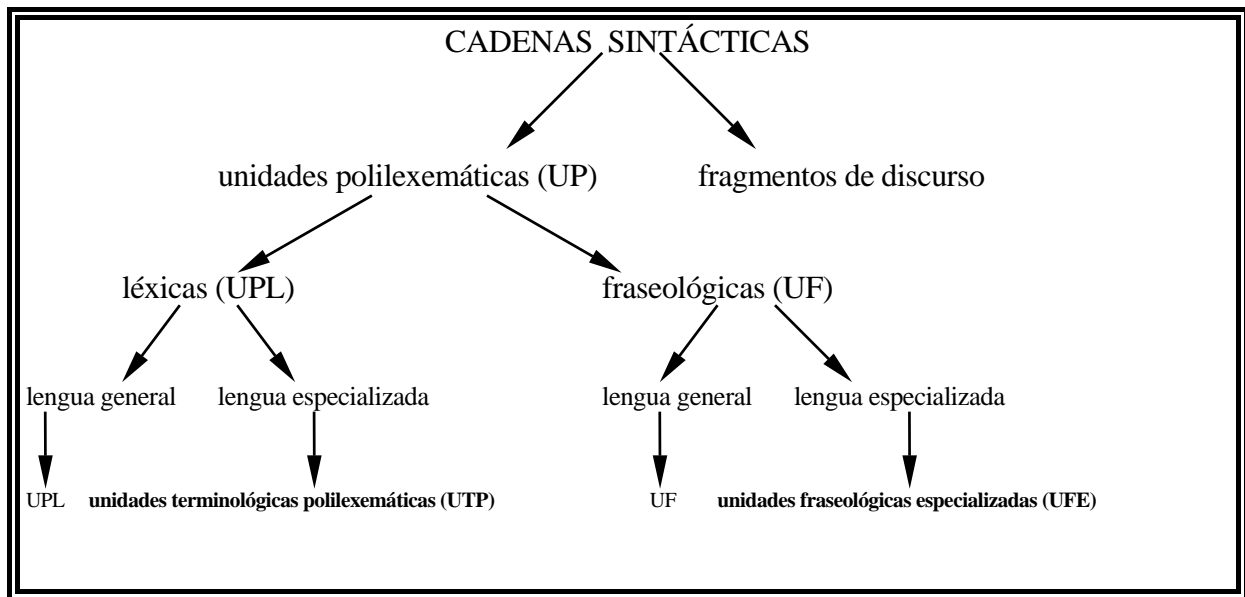
- unidades léxicas polilexemáticas (ULP)¹, es decir, unidades que pueden ocupar en una frase una posición de constituyente mínimo, y cuyos constituyentes están totalmente fijados.
- unidades fraseológicas polilexemáticas (UF), es decir, cadenas sintácticas altamente frecuentes que no pueden ocupar en una frase una posición de constituyente mínimo y cuyos componentes no están totalmente fijados, sino que pueden presentar variación.

Tanto las ULP como las UF pueden pertenecer a la lengua general o ser propias de un ámbito temáticamente especializado.

El siguiente cuadro refleja esta clasificación que hacemos de las cadenas sintácticas:



¹Este término es usado por D. Corbin (1994).



Las unidades de estudio en las que centramos nuestro trabajo son las **unidades terminológicas polilexemáticas**² en relación con las unidades fraseológicas especializadas y con los fragmentos discursivos.

4. La unidad terminológica polilexemática

Considerando la terminología como una materia de intersección y adoptando una **orientación lingüística**, consideraremos que la **unidad terminológica** es una *unidad léxica usada en un ámbito especializado concreto*. Partiremos del supuesto que la unidad básica de la terminología, y en el fondo de los lenguajes de especialidad, es la unidad terminológica concebida como una unidad lingüística que forma parte del lenguaje natural, y dejaremos de lado otros signos que, aunque presentes en el discurso especializado, no son de carácter lingüístico.

Así desde un punto de vista lingüístico, los términos, como unidades gramaticales de un código establecido, son básicamente sistemáticos, tanto en su forma como en su contenido. Formalmente, su estructura morfológica y sintáctica respeta las leyes de formación de palabras y las reglas de flexión de cada lengua. Sénticamente, los términos mantienen relaciones de significado con los demás términos que forman parte del sistema conceptual de un ámbito especializado. Y a nivel sintáctico, los términos son unidades distribucionales que se combinan con otros términos y otras palabras para formar sintagmas y oraciones, que a su vez se combinan entre sí para formar textos.

En este trabajo, siguiendo a Corbin (1994), utilizaremos la denominación **unidad terminológica polilexemática (UTP)** para referirnos a las unidades que cumplen, como mínimo, las cuatro premisas siguientes:

²Estas unidades han recibido muy diversas denominaciones: sinapsis, lexías, unidades sintácticas, compuestos, compuestos improprios, locuciones, sintagmas trabados, etc. Kocourek (1991) presenta una lista de más de 60 términos diferentes.

- desde el punto de vista **semiótico**, pueden ser asociadas de manera estable y codificada a elementos referenciales;
- desde el punto de vista **morfológico**, por lo menos dos de sus componentes son unidades léxicas autónomas³;
- desde el punto de vista de su **uso**, forman parte de un ámbito temáticamente especializado.
- desde el punto de vista **sintáctico**, pueden ocupar en una frase la posición de constituyente autónomo mínimo.

Una *unidad terminológica polilexemática* será una **unidad léxica** propia de un dominio de especialidad formada por más de un lexema. Así, las UTP pueden considerarse un subconjunto de las unidades terminológicas de los lenguajes de especialidad, que a la vez son un subgrupo de las unidades léxicas complejas de una lengua, que funcionan como una sola entidad, y que están formadas por más de una unidad de significación.

En contraposición, una *unidad fraseológica especializada* será una **unidad sintáctica**, pero no léxica⁴, propia de un dominio de especialidad formada por más de un lexema altamente frecuente.

5. La fraseología desde la óptica de la terminología

En los ámbitos científicos y profesionales de la terminología se ha venido detectando, en los últimos quince años, un interés creciente por la fraseología, demostrado feacientemente en diversos foros, como el congreso EUROPHRAS 88 (Strasbourg 1988), el *Symposium on Terminology and Phraseology* (Viena 1989), el coloquio *Phraséologie et terminologie en traduction et en interprétation* (Ginebra 1991), la mesa redonda *Terminologie, discours et textes spécialisés* del Congreso de Lingüistas (Québec 1992) o el seminario internacional sobre *Phraséologie*, celebrado en Hull (Canadá) en 1993, y otros coloquios sobre fraseología más recientes, como el de Ottawa 1994. No obstante, este interés por introducir aspectos discursivos en los estudios terminológicos y en los proyectos terminográficos no se ha limitado al debate teórico, sino que ha conducido a la reproducción de contextos textuales y de fraseología específica en las entradas de diversos vocabularios o bases de datos especializados.

Concretamente, los terminólogos han incorporado la fraseología a su objeto de estudio por diversas razones, de carácter teórico y aplicado.

1) Primero, se ha convenido que el estudio de la fraseología propia de los lenguajes de especialidad permite incrementar la identificación y la delimitación conceptual de los términos, porque la **contextualización** de éstos puede facilitar su precisión semántica.

³ De esta manera se excluyen por un lado, todas las palabras salidas de un proceso de conversión o de afijación, y del otro, todas las unidades léxicas complejas que uno de sus componentes no puede ocupar el lugar de una unidad léxica, es decir lo que se ha denominado tradicionalmente compuestos cultos.

⁴No es una unidad léxica porque no puede ocupar en una frase la posición de constituyente sintáctico autónomo mínimo.

2) Segundo, la descripción de la fraseología propia de cada lenguaje de especialidad y su representación en materiales terminográficos puede favorecer la **implantación** de la terminología en los usos lingüísticos de los hablantes, rechazando el estudio de la terminología descontextualizada por cuanto queda alejada de los actos comunicativos de los hablantes.

3) Tercero, la recuperación de la fraseología desde los textos especializados, para ser incorporada en los diccionarios de especialidad o en los bancos de datos terminológicos, se ha visto favorecida por el incremento de **técnicas informáticas** y por la constitución de corpus textuales, más o menos amplios, que recogen directamente los usos lingüísticos de los especialistas, sin la intervención de mediadores lingüísticos, como traductores, lexicógrafos o terminógrafos.

4) Cuarto, la voluntad de consignar información fraseológica en diccionarios y vocabularios ha hecho necesario que se establezcan **criterios** para la selección y la representación lexicográfica de la fraseología.

Aunque los motivos por los cuales la terminología se ha interesado por la fraseología son claros, las aportaciones que se han hecho en este campo se han encontrado con obstáculos de tipo teórico, heredados ya de la lingüística general. Los problemas teóricos más relevantes podrían resumirse en afirmaciones como las siguientes:

1. El conjunto de estructuras fraseológicas se corresponde con una gran diversidad de expresiones diferentes, por su categoría gramatical, su estructura interna, su significado, su frecuencia, su grado de lexicalización o su grado de fijación. Para poder establecer distinciones entre esta variación de estructuras fraseológicas tenemos que referirnos a criterios de naturaleza distinta: criterios semánticos, sintácticos, lexicológicos, terminológicos, estadísticos, etc.

2. Esta complejidad y la diversidad de criterios adoptados han producido una abundante proliferación de términos, como *coocurrencias*, *concurrentes*, *colocaciones*, *frasemas*, *fraseologismos*, *frases hechas*, *locuciones*, *idiomatismos*, *expresiones fijas*, etc. Diversidad de denominaciones que no se relacionan, casi nunca unívocamente, con conceptos bien delimitados⁵.

3. Pero, tal vez, el problema más grave se refiera a la falta de delimitación conceptual de la noción general de "fraseología". No todos los autores que reflexionan sobre la delimitación del concepto de fraseología se rigen por los mismos parámetros: mientras que algunos tienen una visión restringida de los fenómenos propiamente fraseológicos, otros observan una actitud más laxa que permite consignar bajo la etiqueta de fraseología cualquier estructura superior a la palabra.

4. Esta falta de precisión conceptual, si bien se detecta también en las aportaciones realizadas desde la lingüística teórica y descriptiva, se hace más compleja en el caso de la terminología, porque, si se acepta la existencia de términos de construcción sintagmática (estructuras superiores a una palabra que hemos denominado anteriormente UTP, unidades terminológicas polilexemáticas sintácticas), la teoría tiene que delimitar los ámbitos de la terminología y de la fraseología. En síntesis, es preciso poder distinguir entre los términos sintagmáticos (UTP) y las

⁵Además no siempre son pertinentes para la terminología y los lenguajes de especialidad.

construcciones fraseológicas especializadas (UFE), y además hay que establecer la relación que mantienen entre ambas.

Varias han sido las posiciones que los lingüistas han adoptado para centrar el tema. Algunos autores han optado por seleccionar uno o un reducido número de criterios de identificación (Sinclair, 1980); otros han acumulado criterios diferentes (Roberts, 1993). En resumen, podemos observar que se han aplicado los siguientes criterios para la definición de la fraseología:

-Criterios gramaticales

- relacionados con las estructuras sintagmáticas implicadas
- relacionados con restricciones sobre las categorías gramaticales

-Criterios semánticos

- sobre el significado exocéntrico o endocéntrico de las expresiones
- sobre la metaforización (sentidos interpretables)
- sobre el sentido figurado (opaco en el caso de idiomatismos muy fijados)

-Criterios lexicosemánticos (o terminológicos)

- sobre el grado de lexicalización

-Criterios lexicosintácticos

- sobre el grado de fijación de los componentes
- sobre las posibilidades de variación en las combinaciones
- sobre los paradigmas léxicos implicados y sus limitaciones

-Criterios cuantitativos

- sobre la frecuencia en el lenguaje de especialidad
- sobre la frecuencia en los textos

En este trabajo partiremos de la noción de unidad polilexemática y distinguiremos entre las unidades propiamente terminológicas (UTP) y las unidades fraseológicas especializadas (UFE). Y asumiremos que la diferenciación entre las UTP y las UFE, coincidentes aparentemente en su forma y estructura, es el resultado de la aplicación de criterios restrictivos de diversa índole, que, en nuestro trabajo, serán los siguientes:

- la categoría gramatical
- la estructura interna
- la frecuencia
- el grado de fijación y la variación de sus componentes

Estos criterios no son excluyentes entre sí, ni existe un claro predominio de unos sobre otros, sino que van interactuando de manera ordenada y así los aplicamos en nuestro trabajo.

6. Corpus

Nuestro trabajo, de carácter experimental, se basa en los materiales obtenidos a partir de un corpus textual especializado en catalán. Más concretamente, el corpus está formado por un fragmento de 10.000 palabras extraído del capítulo sobre derecho de navegación del libro *Dret*

mercantil de Rodrigo Uría publicado en 1995 . Esta obra es un manual dirigido a alumnos de los últimos años de la facultad de derecho. Se trata, pues, de un texto bastante especializado, y aunque la muestra sea pequeña, nos parece suficiente para explorar la viabilidad de los criterios que proponemos.

Des del punto de vista informático, hemos tratado el texto en cuatro fases sucesivas:

- en la primera fase, se ha realizado el etiquetado estructural del documento
- en la segunda fase, el preprocesamiento, de forma que se han rechazado los segmentos considerados no analizables
- en la tercera fase, la lematización, con lo que se han agrupado las diferentes formas flexionadas bajo el mismo lema
- finalmente, en la cuarta fase, la desambiguación, que se ha aplicado a dos niveles: en un primer nivel, sobre la base de reglas lingüísticas, y, en el segundo, sobre la base de reglas estadísticas.

Las muestras siguientes ejemplifican estas cuatro fases de tratamiento automático del texto:

0) Documento sin procesar

(fragmento de d0037 muestra 14 -*Dret Mercantil*-)

DRET DE LA NAVEGACIÓ

ESTATUT JURÍDIC DEL VAIXELL I DE L'AERONAU EL VAIXELL

Concepte i naturalesa jurídica

En sentit tècnic parlem de vaixell per referir-nos a qualsevol construcció destinada a la navegació marítima o fluvial. A aquesta idea atenen també, en general, els ordenaments positius moderns en determinar, amb major o menor amplitud, la noció jurídica del vaixell.”

1) Documento con marcas estructurales

```
<div1 n=6 complete=n>
```

```
<head type=main>DRET DE LA NAVEGACIÓ</head>
```

```
<div2 n=72>
```

```
<head type=main>ESTATUT JURÍDIC DEL VAIXELL I DE L'AERONAU</head>
```

```
<div3 n=1>
```

```
<head type=main>EL VAIXELL</head>
```

```
<div4 n=1.107>
```

```
<head type=main>Concepte i naturalesa jurídica</head>
```

```
<p><s>En sentit tècnic parlem de vaixell per referir-nos a qualsevol construcció destinada a la navegació marítima o fluvial.</s><s>A aquesta idea atenen també, en general, els ordenaments positius moderns en determinar, amb major o menor amplitud, la noció jurídica del vaixell.</s></p>
```

2) Documento preprocesado

```
<div1 n=6 complete=n>
```

```
<head type=main>DRET DE LA NAVEGACIÓ</head>
```

```
<div2 n=72>
```

```
<head type=main>ESTATUT JURÍDIC DEL VAIXELL I DE L'AERONAU</head>
```

```
<div3 n=1>
```

```
<head type=main>EL VAIXELL</head>
```

```
<div4 n=1.107>
```

```
<head type=main>Concepte i naturalesa jurídica</head>
```

<p><s>En sentit tènic parlem de vaixell per referir-nos a qualsevol construcció destinada a la navegació marítima o fluvial.</s><s>A aquesta idea atenen també., <loc pos='D'>en general</loc>, els ordenaments positius moderns en determinar, amb major o menor amplitud, la noció jurídica del vaixell.</s></p>

3) Documento lematizado

(información morfológica con ambiguedad '|')

```
## TAG <div1 n=6 complete=n>
## TAG <head type=main>
1 TOK DRET BOS dret\JQ--MS|dret\N5-MS
2 TOK de de\P|de\N5-FS
3 TOK la el\AFS|la\REEC3FS|la\N5-FS
4 TOK NAVEGACI&Oacute;; EOS navegaci&oacute;|\N5-FS
## TAG </head>
## TAG <div2 n=72>
## TAG <head type=main>
5 TOK ESTATUT BOS estatut\N5-MS
6 TOK JUR&Iacute;DIC jur&iacute;dic\JQ--MS
7 TOK de de\P
8 PGR l el\AMS
9 TOK VAIXELL vaixell\N5-MS
10 TOK i i\C|i\N5-FS
11 TOK de de\P|de\N5-FS
12 TOK L ?\N5-66|?\JQ--66
##- DLI ' =\DELIM
13 TOK AERONAU EOS aeronau\N5-FS
## TAG </head>
## TAG <div3 n=1>
## TAG <head type=main>
14 TOK el BOS el\AMS|el\REEC3MS
15 TOK VAIXELL EOS vaixell\N5-MS
## TAG </head>
## TAG <div4 n=1.107>
## TAG <head type=main>
16 TOK Concepte BOS concepte\N5-MS
17 TOK i i\C|i\N5-FS
18 TOK naturalesa naturalesa\N5-FS
19 TOK jur&iacute;dica EOS jur&iacute;dic\JQ--FS
## TAG </head>
## TAG <p>
## TAG <s>
```

4.1) Documento lematizado y desambiguado lingüísticamente

```
## TAG <div1 n=6 complete=n>
## TAG <head type=main>
1 TOK DRET BOS dret\N5-MS|dret\JQ--MS
2 TOK DE de\P
3 TOK LA el\AFS
4 TOK NAVEGACI&Oacute;; EOS navegaci&oacute;|\N5-FS
## TAG </head>
## TAG <div2 n=72>
## TAG <head type=main>
5 TOK ESTATUT BOS estatut\N5-MS
6 TOK JUR&Iacute;DIC jur&iacute;dic\JQ--MS
```

```

7   TOK   D           de\P
8   PGR   EL          el\AMS
9   TOK   VAIXELL     vaixell\N5-MS
10  TOK   I           i\C
11  TOK   DE          de\P
12  TOK   L           el\A6S
##-  DLI   '           =\DELIM
13  PGR   AERONAU     EOS   aeronau\N5-FS
##   TAG   </head>
##   TAG   <div3 n=1>
##   TAG   <head type=main>
14  TOK   EL      BOS   el\AMS
15  TOK   VAIXELL   EOS   vaixell\N5-MS
##   TAG   </head>
##   TAG   <div4 n=1.107>
##   TAG   <head type=main>
16  TOK   Concepte   BOS   concepte\N5-MS
17  TOK   i          i\C
18  TOK   naturalesa naturalesa\N5-FS
19  TOK   jur&iacute;dica EOS   jur&iacute;dic\JQ--FS
##   TAG   </head>
##   TAG   <p>
##   TAG   <s>

```

4.2) Documento desambiguado estadísticamente

```
## TAG <div1 n=6 complete=n>
## TAG <head type=main>
1 TOK DRET BOS dret\N5-MS
2 TOK DE de\P
3 TOK LA el\AFS
4 TOK NAVEGACI&Oacute; EOS navegaci&oacute;\N5-FS
## TAG </head>
## TAG <div2 n=72>
## TAG <head type=main>
5 TOK ESTATUT BOS estatut\N5-MS
6 TOK JUR&iacute;DIC jur&iacute;dic\JQ--MS
7 TOK D de\P
8 PGR EL el\AMS
9 TOK VAIXELL vaixell\N5-MS
10 TOK I i\C
11 TOK DE de\P
12 TOK L el\A6S
##- DLI ' =\DELIM
13 PGR AERONAU EOS aeronau\N5-FS
## TAG </head>
## TAG <div3 n=1>
## TAG <head type=main>
14 TOK EL BOS el\AMS
15 TOK VAIXELL EOS vaixell\N5-MS
## TAG </head>
## TAG <div4 n=1.107>
## TAG <head type=main>
16 TOK Concepte BOS concepte\N5-MS
17 TOK i i\C
18 TOK naturalesa naturalesa\N5-FS
19 TOK jur&iacute;dica EOS jur&iacute;dic\JQ--MS
## TAG </head>
## TAG <p>
## TAG <s>
```

7. Criterios para el reconocimiento de unidades

Una vez obtenida la versión lematizada y desambiguada del texto, se ha procedido a diseñar una estrategia adecuada para el reconocimiento automático, construida a partir de una serie de criterios que permiten ir diferenciando paulatinamente las siguientes parejas de variables:

- Unidades sintagmáticas vs. simples combinaciones
- Unidades suficientemente frecuentes vs. unidades poco frecuentes
- Unidades terminológicas y fraseológicas vs. combinaciones sintácticas libres
- Unidades terminológicas vs. unidades fraseológicas

Los criterios que proponemos deben aplicarse en el orden establecido para asegurar la extracción de todos y cada uno de los grupos diferenciales anteriores. La estrategia ordenada prevé que los criterios más generales (aplicables a textos de distintos ámbitos de especialidad) sean los

primeros, mientras que los criterios que afecten particularmente a las estructuras más productivas de cada lenguaje de especialidad se apliquen a continuación.

Dentro de la serie de criterios establecidos, distinguiremos entre supuestos previos, condiciones, filtros, acciones y pruebas, porque entendemos que se corresponden a elementos de diversa naturaleza y que ejercen funciones específicas en la aplicación.

Un **supuesto previo** consiste en una hipótesis de trabajo, que nos permite prever tendencias de correspondencia entre estructuras y clases de unidades. Una **condición** se caracteriza por un *input* informativo, basado en un conjunto de supuestos previos, que permite restringir el ámbito de aplicación de los filtros. Consideramos que un **filtro** es una orden informática que rechaza información no pertinente y crea un archivo con la información que cumple las condiciones establecidas previamente. Una **acción** es una orden informática que analiza y clasifica la información, a partir de las restricciones impuestas por las condiciones, y la reserva en archivos diferenciados, para su tratamiento posterior. Una **prueba** es una técnica de procesamiento, de carácter cualitativo o cuantitativo, que puede aplicarse indistintamente a todo el corpus o a los archivos producidos por el filtraje o por las acciones, para reforzar los resultados o para corregir errores. La aplicación de una acción o de un filtro siempre implica el establecimiento, con anterioridad, de supuestos previos o de condiciones restrictivas.

8. Criterios generales para la detección de UT simples (1ª fase del proceso)

Dado que el objetivo final debe distinguir entre unidades terminológicas sintagmáticas y unidades fraseológicas, y que por tanto trabajaremos con estructuras complejas, la primera parte de la estrategia de procesamiento terminológico del texto consistirá en la extracción de los términos de estructura simple. Partiendo del supuesto que una unidad terminológica simple es una estructura léxica, situada en el texto escrito entre blancos y frecuente en un texto especializado, esta primera parte del proceso se limita a aplicar una acción y dos filtros informáticos.

Sobre los materiales del corpus (textos marcados estructuralmente, preprocesados, lematizados y desambiguados como en la última muestra del apartado 5), se ejerce una **acción de segmentación de unidades simples**, consistente en aislar todos los lemas simples del texto. De manera que sobre las casi 10.000 palabras del corpus, que corresponden a 1840 lemas distintos, la segmentación nos ofrece un listado de unidades simples muy diverso, como las siguientes.

(1)	de	si
	ser	demanar
	contracte	orientació
	condomini	formulació
	marítim	típica

Sobre el listado de unidades simples, se aplica un **filtro de frecuencia para los lemas simples** (cuyo resultado aparece bajo el formato de índice alfabético de frecuencias de lemas simples) para resolver, mediante el análisis estadístico de frecuencias relativas a los textos analizados, qué lemas simples son frecuentes y cuáles no lo son. Este filtro rechaza los lemas poco frecuentes y

guarda en un archivo los lemas simples frecuentes, como en el ejemplo de (2) que muestra el número de frecuencias en el texto de cada lema.

(2) <i>Lemas frecuentes</i>	<i>Lemas rechazados</i>
de (614 ocurrencias)	demanar (1)
ser (60)	orientació (1)
contracte (34)	formulació (1)
condomini (8)	típica (1)
marítim (14)	
si (29)	

A continuación sobre el archivo de lemas simples frecuentes, se aplica el *filtro léxico*, que rechaza las unidades gramaticales (artículos, preposiciones, conjunciones, adverbios, verbos copulativos, etc), y guarda en un nuevo archivo las **unidades simples presumiblemente terminológicas (UT)**:

(3) contracte
condomini
marítim

9. Criterios generales para la detección de UTP i UF (2ª fase del proceso)

La segmentación y los filtros de la fase anterior, no reducen el corpus textual a analizar para detectar términos complejos y fraseologismos, porque tanto algunas de las piezas rechazadas como algunos de los candidatos a términos simples pudieran estar incluidos en una estructura sintagmática. Así, en esta segunda parte de la estrategia se procesará de nuevo todo el corpus lematizado y desambiguado para detectar estructuras sintagmáticas que cumplan el **requisito de unidad**, ya sea léxica o terminológica, ya sea fraseológica.

Sobre el corpus textual, se aplica la *acción de localización de coocurrencias léxicas*, para establecer un listado ordenado de combinaciones léxicas que tengan un grado de aparición significativo (índices alfabéticos de frecuencias de combinaciones de lemas. A continuación, sobre el listado de formas coocurrentes resultante, se aplica un criterio de naturaleza estadística, el *filtro de frecuencia para las coocurrencias*, para rechazar aquellas combinaciones poco frecuentes y reservar en un archivo aquellas otras combinaciones que tengan un grado de frecuencia significativo, en relación al texto, como las de (4).

(4) és a dir	fer referència
però destinades amb	com a dret a
adquirir la propietat	constituir la hipotèca
prèvia taxació judicial	caràcter permanent al servei
espècie de reconeixement oficial	determinació del valor dels vaixells
vida humana al mar	navegació de cabotatge
només més endavant	així succeeix amb
privilegi marítim en interès	creditor marítim al dret
contraure per preparar	vaixell per referir
pertànyer pro indiviso	procurar forma

En este punto de la estrategia, consideramos que el análisis de frecuencias anterior permite distinguir entre combinaciones libres y combinaciones recurrentes. Así suponemos que las coocurrencias poco frecuentes rechazadas serán presumiblemente combinaciones de discurso libres. En cambio, partimos del *supuesto sobre las coocurrencias frecuentes*, que establece que las coocurrencias léxicas frecuentes, como las de (4), podrán ser identificadas más adelante como:

- unidades terminológicas (UTP)
- unidades fraseológicas (UL), de la lengua general o de la lengua especializada
- segmentos repetidos

De manera que el *output* del filtro de frecuencia anterior sirve de *input* para el siguiente *filtro de unidad estructural*. Este cuarto criterio se aplica exclusivamente al archivo de combinaciones frecuentes, puesto que el objetivo de nuestro trabajo es delimitar unidades polilexemáticas sintagmáticas. El filtro de unidad estructural valida la condición de estructura que presente la combinación analizada. Así sólo reservará aquellas combinaciones de lemas que respondan a una unidad del tipo SV, SN, SA o SP (con todas sus variantes), y rechazará cualquier coincidencia de lemas de origen discursivo que no respondan a un esquema de estructura sintagmática completa. El resultado de la aplicación de este filtro se concreta en un archivo de unidades polilexemáticas sintagmáticas frecuentes, como las de (5), quedando fuera cualquier combinación no estructural.

(5)	<p>és a dir adquirir la propietat prèvia taxació judicial espècie de reconeixement oficial vida humana al mar privilegi marítim en interès de la hipoteca contraure per preparar pertànyer pro indiviso</p>	<p>fer referència constituir la hipotèca caràcter permanent al servei determinació del valor dels vaixells navegació de cabotatge creditor marítim al dret de preferència vaixell per referir procurar forma</p>
-----	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

El siguiente *filtro lexicográfico* permite comprobar si los sintagmas frecuentes (*output* del criterio anterior) están documentados en diccionarios de referencia. Los diccionarios de referencia pueden ser diccionarios de la lengua general, diccionarios bilingües, vocabularios terminológicos o simples glosarios, integrados en el proceso de automatización. Los diccionarios de referencia nos pueden aportar información sobre unidades sintagmáticas de diversa naturaleza, localizadas en entradas o en subentradas, precedidas por etiquetas o con indicaciones en la definición, que pueden desvelarnos su carácter especializado o fraseológico.

Cuando una unidad sintáctica determinada figure en el corpus lexicográfico de referencia con la indicación de que pertenece a una área temática, será automáticamente considerada una unidad terminológica (UTP). En nuestra experimentación, para llevar a cabo esta selección, hemos partido de un diccionario electrónico elaborado a partir de las acepciones marcadas con las etiquetas de área temática *dr* i *com/mar* del *Hiperdiccionari* de la editorial Enciclopèdia Catalana (Barcelona, 1993).

Si el diccionario de referencia incluye información fraseológica, generalmente corresponde a casos idiomáticos de alto grado de fijación y muy frecuentes en los discursos de la lengua

general. En este trabajo, el diccionario general mencionado ha permitido detectar algunas secuencias equivalentes a unidades fraseológicas de la lengua general (UF)⁶.

Así, consideramos que los sintagmas localizados en el diccionario de referencia, como entradas o subentradas, que coinciden con SN y que van precedidos por etiquetas de área temática son UTP reconocidas, mientras que los sintagmas localizados, que coinciden generalmente con SV y que llevan alguna indicación al respecto, son UF reconocidas. Por otro lado, asumimos que los sintagmas no documentados lexicográficamente podrán ser UTP nuevas o no documentadas, UF o combinaciones libres.

El resultado de la aplicación de esta prueba son tres archivos: dos compuestos respectivamente por las UTP y por las UF documentadas, que se reservarán, y otro formado por los candidatos a ser considerados UTP o UF, que no se excluirán, sino que serán objeto del proceso de reconocimiento que sigue a continuación.

(6) <i>No documentadas</i>	<i>Documentadas</i>
adquirir la propietat	navegació de cabotatge (UTP)
vida humana al mar	és a dir (UF)
constituir la hipotèca	fer referència (UF)
prèvia taxació judicial	
caràcter permanent al servei	
espècie de reconeixement oficial	
determinació del valor dels vaixells	
privilegi marítim en interès de la hipoteca	
creditor marítim al dret de preferència	
contraure per preparar	
vaixell per referir	
pertànyer pro indiviso	
procurar forma	

Dado que el resultado del filtro lexicográfico no distingue entre SN, SV, SA y SP, como podemos observar en los ejemplos no documentados de (6), establecemos el ***supuesto de tendencia por categoría gramatical de la estructura***, que se apoya en la observación de los datos aportados por las unidades documentadas lexicográficamente. Este supuesto nos permitirá aplicar filtros de distinción por la categoría del sintagma, para poder avanzar en la aplicación de criterios formales de reconocimiento, de manera que:

(a) Un SN filtrado, con un índice de frecuencia medio-alto respecto de un texto especializado, presentará una tendencia a ser terminológico, aunque puede no serlo.

(b) Un SV filtrado, con un índice de frecuencia medio-alto respecto de un texto especializado, no será terminológico, aunque puede ser fraseológico o de construcción libre.

De acuerdo con las hipótesis anteriores, ejerceremos una ***acción de clasificación de sintagmas por la categoría*** que, primero, localice los SA y los SP para considerarlos integrantes de otros sintagmas (SN o SV) y que, después, separe, en dos archivos, los SN y los SV. En este sentido,

⁶Difícilmente encontraremos unidades fraseológicas del discurso especializado en diccionarios generales, como el utilizado en nuestro trabajo.

consideraremos que las condiciones de “ser un SN” y la de “ser un SV” son fundamentales en el establecimiento sistemático de candidatos a ser presumibles UTP o UF, respectivamente.

(7) SN, candidatos a UTP	SV, candidatos a UF
vida humana al mar	adquirir la propietat
prèvia taxació judicial	constituir la hipotèca
caràcter permanent al servei	contraure per preparar
vaixell per referir	pertànyer pro indiviso
reconeixement oficial	procurar forma
determinació del valor dels vaixells	
privilegi marítim	
creditor marítim	
vaixell per referir	

Atendiendo al hecho que las categorías de SN y de SV pueden presentar la variación estructural que presentamos a continuación, la *acción de despliegue de estructuras* permite clasificar los conjuntos de SN y de SV resultantes de la acción de clasificación anterior en tantos subarchivos como estructuras distintas existan.

Variantes de SN: NA, AN, N Quant A, Quant N A, N prep (det) N, N prep V_{inf}

Variantes de SV: V (det) N, V prep (det) N, V prep V_{inf}

10. Criterios del ámbito especializado (lenguaje jurídico) (3ª fase del proceso)

A partir de aquí, la estrategia prevé que los criterios de esta tercera fase se apliquen exclusivamente sobre las estructuras filtradas y que puedan ser distintos para cada ámbito de especialidad. Así, los corpus textuales especializados de distintas áreas temáticas pueden compartir los criterios de las dos primeras fases, pero se procesaran, a partir de aquí, con criterios distintivos que darán cuenta de las diversas tendencias de construcción de secuencias sintagmáticas y oracionales de que disponen los discursos especializados. Estos criterios especializados se constituyen a partir de la toma en consideración de modelos de referencia, establecidos fuera del ámbito estricto de la experimentación, que pueden provenir de fuentes lexicográficas especializadas, de estudios descriptivos de un determinado lenguaje de especialidad o de convenciones generales sobre la terminología y la fraseología.

En nuestro caso, y para el corpus de lenguaje jurídico analizado, propondremos criterios adicionales aplicables sobre el archivo de SN desplegados y criterios adicionales aplicables tanto a los SN como a los SV.

Para el procesamiento del archivo de SN, partimos de la consideración del *supuesto previo de patrones productivos*, derivado del análisis de los esquemas sintagmáticos de las entradas y subentradas de un diccionario especializado en el ámbito determinado, que nos permite establecer qué estructuras sintagmáticas son más productivas en los discursos propios de ese ámbito⁷. Así, para los SN y para el lenguaje jurídico, consideramos que son estructuras productivas las siguientes: NA, N prep N, N prep det N. Y que, en cambio, son poco productivas

⁷En este trabajo hemos tomado como referència para los patrones productivos en el discurso jurídico, el análisis lexicográfico aportado en Estopà (1996).

otras estructuras que también aparecían en el despliegue de SN (AN, N Quant A, Quant NA, N prep V_{inf}).

La estrategia de detección nos permite establecer las siguientes hipótesis en relación con los productividad de patrones:

(a) Si un SN responde a un patrón productivo, esta condición refuerza la tendencia del SN a ser terminológico (SN presumiblemente terminológicos).

(b) Si un SN no responde a un patrón productivo, esta condición refuerza la tendencia del sintagma a ser libre, sin que esto signifique que no pueda ser fraseológico (SN presumiblemente NO terminológicos).

De acuerdo con lo anterior, aplicamos a continuación el *filtro de patrones productivos* sobre el archivo de SN, resultando

- un archivo de NA, candidatos a términos
- un archivo de N prep N , candidatos a términos
- un archivo de N prep det N, candidatos a términos
- i un archivo de rechazos, que inicialmente reservamos para un análisis posterior, constituido por los SN que no corresponden a patrones productivos y que serán presumiblemente libres o fraseológicos.

(8) *Ejemplos de candidatos a términos*

vida humana
taxació judicial
caràcter permanent
reconeixement oficial
determinació del valor dels vaixells
privilegi marítim
creditor marítim
vaixell per referir

Para continuar con el procesamiento terminológico del texto, se aplica un paquete de control, compuesto por diversas condiciones adicionales, que deben cumplir tanto los SN reservados como candidatos, como los SN que no cumplieran el requisito de estructura productiva, como los SV restantes. Este conjunto de condiciones, con un filtro final de cumplimiento de las condiciones, refuerza la estrategia de detección en la vertiente de criterios de naturaleza estructural, que deben interactuar con criterios de frecuencia o de productividad.

La primera de estas condiciones es la *condición del complemento V_{inf}* , que se refiere a la estructura interna de los SP integrados, a su vez, dentro de un SN o dentro de un SV. Partiendo de la hipótesis general lanzada en el *supuesto de tendencia por categoría gramatical de la estructura*, la *condición del complemento V_{inf}* , establece que:

Si el complemento de un SP es una forma verbal (infinitivo), el sintagma superior que lo incluye (SN o SV) será una combinación libre.

La segunda, la **condición del especificador**, se refiere, en este caso, a la estructura interna de los SN integrados dentro de otros sintagmas (SN, SP o SV). Siguiendo criterios más amplios sobre los procesos de lexicalización de unidades polilexemáticas sintácticas, esta condición establece que:

Si un SN, autónomo o integrado dentro de otro sintagma (SN o SV), está determinado por un especificador (det), será presumiblemente una combinación libre (UL).

Si un SN, autónomo o integrado dentro de otro SN, **no** está determinado por un especificador (det), el sintagma nominal superior será presumiblemente terminológico (UTP).

Si un SN, autónomo o integrado dentro de un SV, **no** está determinado por un especificador (det), el sintagma verbal será presumiblemente fraseológico (UF).

El tercer criterio adicional lo constituye la **condición de núcleo terminológico** que se refiere a la naturaleza terminológica de cada uno de los formantes léxicos de una unidad polilexemática. El carácter terminológico de una forma simple se comprueba mediante la consulta del fichero de unidades simples terminológicas (UT), que resultaba de la aplicación de los criterios de la primera fase del proceso (ver apartado 8). La **condición de núcleo terminológico** establece que:

Si el núcleo de un sintagma es un término del listado de frecuencias de lemas simples (UT), el sintagma será una unidad terminológica polilexemática (UTP).

Si el núcleo de un sintagma no forma parte del listado de frecuencias de lemas simples (UT), pero el complemento del sintagma sí, entonces el sintagma será una unidad fraseológica (UF).

Si ninguno de los formantes de una unidad polilexemática aparece en el listado de frecuencias de lemas simples (UT), el sintagma será una combinación libre (UL).

Aplicamos, a continuación, un **filtro de cumplimiento de las condiciones adicionales** para cada una de las condiciones propuestas en los criterios adicionales anteriores, para poder afinar la candidatura de los diversos sintagmas como unidades terminológicas (UTP), unidades fraseológicas (UF) o combinaciones libres (UL).

El ámbito de aplicación del filtro está restringido para cada una de las condiciones, de manera que *el filtro para la condición del complemento V_{inf}* se aplicará a las estructuras que contengan un infinitivo, (N prep V_{inf}) y (V prep V_{inf}); *el filtro para la condición del especificador* se aplicará a las estructuras que contengan un SN con especificador, (N prep det N) y (V prep det N); i finalmente *el filtro para la condición de núcleo terminológico* se aplicará a todas las estructuras.

Los resultados de la aplicación de este filtro serán los siguientes, como ejemplificamos en (9-11):

El filtro de cumplimiento de la condición del complemento V_{inf} rechazará, como candidatas a UTP o a UF, todas aquellas estructuras que contengan infinitivo, y las considerará combinaciones libres (UL).

- (9) *Candidatos rechazados*
vaixell per referir
contraure per preparar

El *filtro de cumplimiento de la condición del especificador* rechazará, como candidatas a UTP o a UF, todas aquellas estructuras que contengan determinante, y las considerará combinaciones libres (UL).

- (10) *Candidatos rechazados*
determinació del valor dels vaixells
adquirir la propietat
constituir la hipotèca

El *filtro de cumplimiento de la condición de núcleo terminológico* rechazará, como candidatas a UTP o a UF, todas aquellas estructuras en que ninguno de los formantes de la unidad polilexemática cumpla la condición de pertenecer al conjunto de términos simples, y las considerará combinaciones libres (UL).

- (11) *Candidatos rechazados*
vida humana
caràcter permanent
reconeixement oficial

Por otra parte las unidades polilexemáticas que cumplan las condiciones adicionales pasarán a formar parte de sendos archivos de candidatos a UTP y a UF, como en los ejemplos de (12-13). Así,

Los sintagmas nominales (SN) que cumplan las tres condiciones (no llevar infinitivo, ni determinante y que el núcleo pertenezca al conjunto de UT simples) serán **UTP**.

- (12) *Candidatos a UTP*
taxació judicial
privilegi marítim
creditor marítim

Los sintagmas nominales (SN) que cumplan las tres condiciones (no llevar infinitivo, ni determinante y que el núcleo no pertenezca al conjunto de UT simples, pero el complemento sí) serán **UF**.

Los sintagmas verbales (SV) que cumplan las tres condiciones (no llevar infinitivo, ni determinante y que el núcleo del SN que integra pertenezca al conjunto de UT simples) serán **UF**.

- (13) *Candidatos a UF*
pertànyer pro indiviso
procurar forma

11. Criterios finales de reconocimiento (4ª y última fase del proceso)

Sobre la totalidad de los sintagmas analizados (filtrados como UTP y UF o rechazados como UL), se aplicará una *prueba final de frecuencia* de lemas en coocurrencia, para detectar el grado de aparición en el texto de determinadas combinaciones de lemas. De manera que,

Si un candidato a UTP aparece con frecuencia, este hecho refuerza su naturaleza terminológica.

Si un candidato a UF aparece con frecuencia, este hecho también confirma su carácter de fraseología propia del lenguaje de especialidad.

Si una combinación ha sido considerada UL, pero aparece con un grado de frecuencia muy alto, hay que considerar cada caso para detectar posibles errores.

Y, en último lugar, sólo el *criterio de competencia especializada* sobre el área temática, que se relaciona con la identificación y delimitación de conceptos y referentes, permite concretar si una combinación que ha estado seleccionada como candidata a unidad terminológica (UTP), lo es realmente. De alguna manera, los especialistas como emisores del discurso especializado son los responsables de aplicar una prueba final que verifique la fiabilidad de la estrategia que acabamos de presentar y detecte qué grado de satisfacción ofrece cada fase del proceso.

12. Conclusiones

A la vista de la experimentación realizada, hemos llegado a algunas conclusiones, que presentamos muy sintéticamente organizadas en tres bloques:

(1) Conclusiones sobre el proceso de reconocimiento de las unidades terminológicas.

Las conclusiones a que hemos llegado en esta experimentación para el reconocimiento automático de unidades terminológicas polilexemáticas y de unidades fraseológicas propias de los lenguajes de especialidad, nos conducen a considerar como muy positivos los puntos siguientes:

(a) la utilidad de los corpus para establecer criterios en terminología;

(b) la utilidad del criterio de frecuencia como factor determinante en la detección automática de unidades terminológica simples gráficamente; y relevante para las unidades terminológicas sintagmáticas;

(c) la insuficiencia de un solo criterio, como el de la frecuencia, en la consideración de un proceso complejo de detección automática de términos y fraseología; y por consiguiente la necesidad de combinar diversos criterios estructurales y estadísticos;

(d) la importancia de la diferenciación establecida en el trabajo entre diversos tipos de criterios, según su función en el proceso:

criterios informativos (supuestos previos y condiciones)
ordenes informáticas (acciones y filtros)

(2) Conclusiones sobre las diferencias entre unidades sintagmáticas fijadas, en algún grado, versus combinaciones libres de discurso.

A la vista de la experimentación de este trabajo, podemos concluir, para diferenciar las UTP y las UF (combinaciones sintagmáticas con algún grado de fijación) de las combinaciones libres (UL):

(a) que no son suficientes los criterios estructurales, sino que conviene matizarlos con la presencia o la ausencia de determinados componentes y con la interacción con otro tipo de criterios (como los estadísticos);

(b) que es importante el uso de criterios de frecuencia y de las indexaciones de coocurrencias para detectar el grado de variación de ciertas combinaciones de lemas.

(c) y que la utilidad de un corpus lexicográfico especializado de referencia, morfológicamente etiquetado, agiliza la identificación de UTP reconocidas y permite establecer patrones de productividad en un ámbito especializado.

(3) Conclusiones sobre las diferencias entre UTP y UF

Podemos concluir que, para la distinción entre UTP y UF, son fundamentales los siguientes aspectos:

(a) La categoría de las unidades sintagmáticas polilexemáticas, por cuanto en el ámbito especializado que hemos analizado (derecho) la identificación entre UTP y los SN filtrados y la correspondencia entre UF y los SV filtrados es muy regular.

(b) La consideración del carácter terminológico de los componentes de las unidades polilexemáticas, con la distinción consiguiente entre la condición de término del núcleo del sintagma (para las UTP) y la condición de término del complemento del sintagma (para las UF).

En definitiva, deseamos finalizar esta ponencia con dos ideas, reforzadas por la consideración de la experimentación que hemos presentado.

La primera es que tanto las UTP como las UF en general, siempre lo son en el marco de un ámbito de especialidad; y que, por tanto, no consideramos que haya términos ni fraseología especializada en abstracto.

La segunda es que, si bien, para la fraseología, la suma de criterios estructurales y estadísticos es suficiente para la detección automática; en el caso de la terminología siempre nos veremos obligados a recurrir a la competencia del especialista, que incluye el conocimiento sobre la estructuración conceptual del área especializada y sobre las relaciones entre referente, concepto y denominación, para estar seguros de que las unidades sintagmáticas seleccionadas son efectivamente unidades terminológicas del ámbito especializado pertinente.

BIBLIOGRAFÍA

Cabré, M. Teresa (1992): *La terminología. Teoría, metodología y aplicaciones*. Barcelona: Antártida, 1993.

Estopà, Rosa (1996): *Las unitats terminològiques polilexemàtiques en els lèxics d'especialitat: dret i medicina*. Tesi de llicenciatura, Universitat Pompeu Fabra de Barcelona.

Pavel, Silvia (1992): *Bibliographie de la phraséologie 1905-1992*. Bureau de la traduction (Canada) & RINT (Réseau International de Néologie et Terminologie).

Roberts, Roda (1993): "La phraséologie: état des recherches". En *Terminologies nouvelles* (1993: 10), pp. 36-42.

Sinclair, John (1980): "Some implications of discourse analysis for ESP methodology". En *Applied Linguistics* (1980:1-3), pp. 253-261.

Phraséologie. Actes du séminaire international (Hull, mai 1993). *Terminologies nouvelles* (1993: 10)

Phraséologie et terminologie en traduction et interpretation (Actes du Colloque de Genève 1991). *Terminologie et traduction* (1992: 2-3)