

## TERMINOLOGÍA Y FRASEOLOGÍA

M. Teresa Cabré, Rosa Estopà, Mercè Lorente  
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 esta comunicación, 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 (1985), 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 puede 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.

## **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 (no el reconocimiento) de cada uno de los términos usados en un ámbito) a su ámbito de origen 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. 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.

Por consiguiente y desde el punto de vista de su estructura morfológica, las unidades terminológicas, como unidades léxicas que son, presentan estructuras diferentes. Así, existen términos con una estructura simple y términos con una estructura compleja. Los términos complejos se pueden agrupar en tres grandes grupos en función del proceso gramatical que los ha generado: un *proceso morfológico*, un *proceso fonológico* o un *proceso sintáctico*. A su vez, estos tres grupos se pueden desglosar más si se considera las piezas léxicas que intervienen en la combinación morfológica.

El siguiente esquema presenta una síntesis de la tipología de términos en la que ubicamos nuestro trabajo:

**1. términos simples:** *sol, carbón, tierra, rayo, ácido, culpa, junta, crimen, demanda, etc*

**2. términos complejos:**

**2.1 términos complejos formados por un proceso morfológico**

**2.1.1 prefijación:** *prejudicial, enjuiciar, autocontrato, desembargar, desnitrificación, desecho, desarenado, agraviar, alineación, etc.*

**2.1.2 sufijación:** *evaporación, refinado, olector, incineradora, cesionista, heredero, falsedad, tributario, etc.*

**2.1.3 conversión:** *arresto, atestado, juzgado, danyo, baja, etc.*

**2.1.4 composición propia:** *bancarota, salvaguarda, malversación, tragabasuras, etc.*

**2.1.5 composición culta:** *fotosíntesis, agricultura euritermo, orografía, toxicología, democracia, criminología, genocidio, criminología, etc.*

**2.2 términos complejos formados por un proceso fonológico**

**2.2.1 siglación:** *ADN, EEG, TAC, TV, FAO, etc.*

**2.2.2 abreviación:** *foto, mono, tele, etc.*

**2.2.3 acronimia:** *Termesp, Mercosur, informática, bit, etc.*

**2.3 términos complejos formados por un proceso sintáctico:** *energía nuclear, energía geotérmica, gas natural, contaminantes primarios, contenedores estáticos con mecanismos de compactación, estufa de secado, impacto ecológico, bienes muebles, sociedad limitada, sociedades de garantía recíproca, perdón del ofendido, matrimonio consumado, matrimonio rato, ley orgánica dle tribunal constitucional, cheque para abonar en cuenta, setencia con autoridad de cosa juzgada, juicio en rebeldía.*

Las unidades de estudio en la que centramos nuestro trabajo son los términos complejos formados por un proceso sintáctico, denominados diversamente *unidades polilexemáticas, compuestos, sinapsis, lexías, unidades sintácticas, compuestos impropios, locuciones, sintagmas travados...*<sup>1</sup>.

---

<sup>1</sup>Sobre la recopilación de los diferentes nombres que han recibido este tipo de unidades lingüísticas, Kocourek (1991) da una llista de 60 términos diferentes.

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

- 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 autónomas<sup>2</sup>;
- desde el punto de vista de su **uso**, forman parte de un ámbito temáticamente especializado.

Una *unidad terminológica polilexemática* será **una unidad léxica (unidad) propia de un dominio de especialidad (terminológica) que está formada por más de un lexema (polilexemática)**. 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.

Las *unidades terminológicas polilexemáticas sintácticas* pueden ser de dos tipos atendiendo al **grado de fijación** de sus elementos:

- unidades terminológicas polilexemáticas lexicalizadas (UTPL)*, cuyos componentes están totalmente fijados, y
- unidades terminológicas polilexemáticas fraseológicas (UTPF)*, cuyos componentes no están totalmente lexicalizados y, por tanto, pueden presentar variación<sup>3</sup>.

#### **4. La fraseología desde la óptica de la terminología**;Error! Marcador no definido.

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

---

<sup>2</sup> De esta manera se excluyen por un lado, todas las palabras salidas de un proceso de conversión o de afijación, i 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.

<sup>3</sup>Un elemento de diferenciación de las UTPL y las UTPF es que las primeras pueden ocupar en una frase una posición de constituyente sintáctico autónomo mínimo, mientras que las segundas no.

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, porqué la **contextualización** de éstos puede facilitar su precisión semántica.

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** i 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.

### **Problemas teóricos:**

1. La noción de fraseología es compleja y esconde 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.

2. 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.

3. Pero, tal vez, el problema más grave se refiera a la falta de delimitación conceptual de la noción "fraseología". No todos los autores que reflexionan sobre la noción 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 UTPS, 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 las UTPS que son términos y las UTPS que son construcciones fraseológicas, 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, 199 ); otros han acumulado criterios diferentes (Roberts, 1995). 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 léxico-semánticos (o terminológicos)

- sobre el grado de lexicalización

-Criterios léxico-sintácticos

- 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 visión de las unidades terminológicas polilexemáticas y distinguiremos entre las unidades terminológicas polilexemáticas lexicalizadas y las unidades terminológicas polilexemáticas fraseológicas. Y asumiremos que la diferenciación entre las UTPL y las UTPF, 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 variación de sus componentes

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

## **5. Aplicación**

### **5.1 Corpus**

El 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 alrededor 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 exponer la viabilidad de los criterios.

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 en dos niveles: en el 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;Error! Marcador no definido.**

(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 t&egrave;nic parlem de vaixell per referir-nos a qualsevol construcci&oacute; destinada a la navegaci&oacute; mar&iacute;tima o fluvial.</s><s>A aquesta idea atenen tamb&eacute;, en general, els ordenaments positius moderns en determinar, amb major o menor amplitud, la noci&oacute; jur&iacute;dica del vaixell.</s></p>

## 2) Documento preprocesado

```
<div1 n=6 complete=n>
<head type=main>DRET DE LA NAVEGACI&Oacute;</head>
<div2 n=72>
<head type=main>ESTATUT JUR&Iacute;DIC 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 jur&iacute;dica</head>
<p><s>En sentit t&egrave;nic parlem de vaixell per referir-nos a qualsevol construcci&oacute; destinada a la navegaci&oacute; mar&iacute;tima o fluvial.</s><s>A aquesta idea atenen tamb&eacute;, <loc pos='D'>en general</loc>, els ordenaments positius moderns en determinar, amb major o menor amplitud, la noci&oacute; jur&iacute;dica del vaixell.</s></p>
```

## 3) Documento lematizado

(información morfológica con ambigüedad '|')

```
## 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 desanbiguado 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--MS
## 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           e|\AMS
9   TOK   VAIXELL      vaixell\N5-MS
10  TOK   I            i\C
11  TOK   DE          de\P
12  TOK   L           e|\A6S
##- DLI   '           =\DELIM
13  PGR   AERONAU    EOS    aeronau\N5-FS
##   TAG   </head>
##   TAG   <div3 n=1>
##   TAG   <head type=main>
14  TOK   EL      BOS   e|\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>

```

## 5.2 Pruebas de reconocimiento

Superadas las cuatro fases anteriores, se ha aplicado al texto dos tipos de pruebas para reconocer las unidades terminológicas polilexemáticas sintácticas, separándolas de las unidades discursivas libres:

- A/pruebas estadísticas
- B/prueba de referencia

### A. Pruebas estadísticas

Para las pruebas estadísticas, hemos partido de los siguientes listados de frecuencias:

- a/ listado de frecuencias de formas ordenadas alfabéticamente<sup>4</sup>
- b/ listado de frecuencias de formas en orden descendente de frecuencias
- c/ listado de frecuencias de lemas ordenados alfabéticamente
- d/ listado de frecuencias de lemas en orden descendente de frecuencias

Paralelamente, se han establecido seis patrones de sintagmas perfectos y, mediante un programa informático, se han obtenido automáticamente un listado de expresiones de casa tipo de estructura:

- N (A) prep (det)N
- V prep (det) N
- N (A) prep infinitivo

---

<sup>4</sup>En todas estas listas de frecuencias no se ha tenido en cuenta las palabras gramaticales.

## V prep infinitivo

Hemos partido del supuesto que el núcleo del patrón sintáctico condiciona la naturaleza de la UTPS. Así, a modelos estructurales idénticos, la naturaleza del núcleo condiciona la naturaleza gramatical de la UTP resultante. De esta manera, asumimos que si el núcleo es un nombre la UTPS tenderá a ser una *unidad terminológica polilexemática sintáctica lexicalizada*, en cambio si el núcleo es un verbo tenderá a ser una *unidad terminológica polilexemática sintáctica fraseológica*.

### B. Prueba referencial

Por lo que se refiere al filtro referencial, se ha tenido en cuenta que si una unidad sintáctica determinada figuraba en el corpus lexicográfico de referencia con la indicación de que pertenecía a una área temática, era automáticamente considerada unidad terminológica. Para llevar a cabo esta selección se ha partido de un diccionario electrónico elaborado a partir de las acepciones marcadas con la etiqueta de área temática *dr* del *Hiperdiccionari d'Enciclopèdia Catalana*.

### 5.3 Resultados de las pruebas estadísticas

### 5.4 Resultados de la búsqueda en los diccionarios automatizados

### 5.5 Comparación y síntesis

Después de haber aplicado los dos tipos de filtros, hemos obtenido unidades simples y unidades complejas altamente frecuentes en el texto, pero documentadas en el corpus de referencia. En este estudio nos interesan las unidades complejas sintácticas, las cuales pueden ser de tres tipos diferentes si consideramos el grado de fijación de sus elementos:

1. unidades terminológicas polilexemáticas sintácticas **lexicalizadas** (UTPSL)
2. unidades terminológicas polilexemáticas sintácticas **fraseológicas** (UTPSF)
3. unidades sintácticas **discursivas** o unidades libres

### . Conclusiones

- Sobre la utilidad de los corpus para establecer criterios en terminología
- Sobre la posibilidad de reconocer automáticamente los términos
- Sobre las estrategias de reconocimiento de las unidades terminológicas
- Sobre los criterios estructurales de delimitación de las unidades terminológicas de estructura sintagmática

-- Sobre las diferencias entre terminología y fraseología

-- Sobre las posibilidades de diferenciar automáticamente las unidades terminológicas y las fraseológicas