

## **Criterios para la elaboración de un corpus de fitonimia catalana: vaciado terminológico, informatización de los datos y aspectos metodológicos**

Cristina Bofill TERMCAT, [1], M. Àngels Bonet TERMCAT, [2], M. Antònia Julià [3]  
y Joan Vallès [4]

### **1. Introducción**

En esta comunicación presentamos un proyecto en curso de elaboración en el que se ha abordado la creación de un corpus lexical de fitonimia catalana, es decir, una recopilación de nombres catalanes de plantas. Dicho proyecto, que lleva por título Recull de noms catalans de plantes, es fruto de una iniciativa conjunta de la Universidad de Barcelona (concretamente, el Laboratorio de Botánica de la Facultad de Farmacia) y del Centro de Terminología TERMCAT.

Empezaremos por explicar brevemente las características del proyecto, para, a continuación, exponer con mayor detalle las decisiones metodológicas que ha sido necesario tomar para tratar adecuadamente, desde el punto de vista terminológico, la ejecución de una obra que reúne un buen número de singularidades.

Concretamente, nos referiremos al establecimiento del corpus de vaciado, a la extracción e informatización de las unidades terminológicas, y al tratamiento de los datos obtenidos. Incluiremos, además, una sucinta descripción de las herramientas informáticas que se han empleado para ejecutar este trabajo terminográfico de una manera correcta y eficaz.

En el marco de esta jornada dedicada al papel de los corpóra en el trabajo terminológico, nos proponemos transmitirles nuestras reflexiones metodológicas aplicadas a un caso práctico en el que el tratamiento del corpus ha sido fundamental.

### **2. Características del proyecto**

Al tratarse de un proyecto en curso de realización, debemos hablar de previsiones. Según nuestra estimación, la recopilación resultante de nuestro trabajo incluirá finalmente alrededor de 25.000 denominaciones catalanas referidas a unos 4.800 táxones de flora vascular (es decir, helechos y plantas con flores) presentes en el dominio lingüístico catalán, ya sea en forma silvestre, cultivada o comercial. Para

todos ellos se recogerá el nombre científico y la adscripción a la familia botánica correspondiente.

Adicionalmente, con el fin de facilitar la consulta de la obra a sus usuarios y dado su carácter divulgativo, la publicación incluirá unas 300 ilustraciones de plantas, así como diversos anexos: un índice de nombres científicos, un índice sistemático por familias botánicas y un índice de las obras constituyentes del corpus de vaciado.

El proyecto se ha abordado en distintas etapas y, siguiendo su desarrollo previsible, la publicación de la obra llegará el 2008.

### **3. Establecimiento del corpus de vaciado**

En la primera fase del proyecto, empezado en 1993, se definieron las características que éste había de tener, partiendo de un objetivo inicial: el propósito de poner al día la obra más completa editada sobre fitonimia catalana, *Els Noms de les plantes als Països Catalans*, del botánico Francesc Masclans. Dicha obra, publicada en 1981 a partir de una primera edición de 1954 —que apareció bajo un título distinto—, está agotada editorialmente desde hace años. En el transcurso del trabajo se hizo patente, como ya intuíamos desde el inicio, que era imprescindible que la información que contenía estuviese de nuevo a disposición de los usuarios, accesible en un formato y un soporte adaptados a las necesidades actuales.

El libro de Masclans posee un valor simbólico y referencial indiscutible en el ámbito de la botánica y la lengua catalanas: contiene toda la nomenclatura catalana referida a las denominaciones de plantas disponible hasta el momento de su publicación. Recoge unas 9.500 denominaciones catalanas de plantas con su correspondiente nombre científico. Se trata de términos que el autor documentó mediante el vaciado de 119 publicaciones que cubren todo el territorio lingüístico catalán: la *Flora de Catalunya* y otras publicaciones botánicas relevantes; el herbario del Institut Botànic de Barcelona; diversos diccionarios de la lengua catalana, entre los cuales el diccionario normativo de Pompeu Fabra, y grandes obras de la literatura catalana. Por otro lado, Masclans reunió también en su libro la información resultante de su propio trabajo de campo, realizado durante años en distintas áreas de la geografía catalana.

Nuestro corpus de vaciado se completó gracias a una búsqueda documental que permitió confeccionar una lista de más de 350 obras pertinentes, en formatos y soportes varios, publicadas desde 1981 hasta el presente.

#### **4. Extracción e informatización de las unidades terminológicas**

Una vez establecido el corpus de vaciado, comenzó el trabajo de extracción e informatización de las unidades terminológicas. En la primera etapa del proyecto, el libro de Francesc Masclans se consideró como corpus lexical de exclusión. De esta manera, se procedió al vaciado de cerca de 200 obras botánicas, labor de extracción cuyo resultado se tradujo en la obtención de unas 20.000 denominaciones referidas a 4.500 táxones.

En una segunda etapa del trabajo de vaciado, iniciada el año 2005, se fijó como requisito indispensable la validación del vaciado de nombres científicos obtenido en la primera etapa, siguiendo las propuestas nomenclaturales de la tercera edición de la *Flora manual dels Països Catalans*. Una vez concluida esta validación, deberá incorporarse el contenido de la obra de Masclans a nuestra base de datos. Paralelamente se continuará con el vaciado de más de 100 obras publicadas desde 1994, de material inédito cedido por botánicos y de más de 25 fuentes documentales botánicas en línea. Asimismo, se prevé la inclusión en nuestro corpus lexical del conjunto de nombres de plantas que aparecen en el diccionario normativo del Institut d'Estudis Catalans de 1995.

#### **5. Tratamiento metodológico de los datos del corpus de vaciado**

La ingente cantidad de datos obtenida mediante el vaciado ha comportado la necesidad de tomar decisiones metodológicas ajustadas a la realidad del proyecto y fijar criterios de selección adecuados, tanto de tipo general como específico.

##### *5.1 Criterios generales*

Los criterios generales de selección de términos reflejan el espíritu que ha guiado la concepción de este trabajo y que es el alma de su proceso de elaboración. A continuación los enumeraremos sintéticamente.

En primer lugar, la recopilación de denominaciones se rige por una voluntad descriptiva. Por ello, como planteamiento de partida no se ha considerado la

necesidad de priorizar una forma en detrimento de otras, sino que se ha recogido sin distinción alguna el conjunto de denominaciones catalanas aplicables a un determinado taxon botánico.

En segundo lugar, éste es un trabajo concebido con un propósito de exhaustividad. La composición del corpus de vaciado, constituido a partir de la ya citada obra de Masclans y del resultado de la búsqueda documental, cumple con los requisitos de adecuación, fiabilidad e integridad que nos habíamos propuesto.

En tercer lugar, el pilar fundamental de este trabajo terminográfico es el rigor metodológico, basado en la labor conjunta de terminólogos y especialistas, y sustentado en el asesoramiento experto de botánicos y lingüistas.

Finalmente, cabe destacar que al plantearnos los objetivos del proyecto se consideró imprescindible facilitar a los usuarios la consulta y la recuperación de la información. Así, por ejemplo, este requisito ha conllevado la necesidad de estructurar el diccionario a partir del nombre científico y la conveniencia de añadir anexos e índices complementarios que contribuyan a la manejabilidad de la obra.

## *5.2 Criterios singulares*

Desde el punto de vista del trabajo terminológico sistemático que realizamos en TERMCAT, el *Recull de noms catalans de plantes* representa un caso muy particular. Es una obra que requiere un tratamiento metodológico específico, puesto que supone operar simultáneamente con nomenclatura científica botánica (que exige un tratamiento singular del vaciado, el procesamiento y la ordenación de los términos) y con terminología popular referida a nombres de plantas (que incluye un elevado número de variantes geolectales y funcionales).

Los táxones botánicos tomados en cuenta en este proyecto incluyen mayoritariamente el género y la especie, pero también algunos táxones infraespecíficos hasta el nivel de variedad. La categoría más significativa desde el punto de vista de las denominaciones populares es, sin lugar a dudas, la especie, por lo cual la mayor parte de nombres catalanes recogidos en nuestra base de datos se refiere a táxones específicos.

### *5.2.1 Tratamiento de la nomenclatura científica*

De acuerdo con los principios de la nomenclatura científica establecida por Linneo y regida por el Código Internacional de Nomenclatura Botánica –que se actualiza cada seis años–, los táxones botánicos, tanto a nivel de especie como infraespecífico, se designan con un binomen latinizado. Estos nombres científicos se componen de un nombre de género más un epíteto específico, asociados al nombre abreviado del autor de la combinación; si se trata de un taxon infraespecífico, se debe añadir el epíteto subespecífico y, de nuevo, el autor de la combinación, y así sucesivamente. Su notación tipográfica es además bastante compleja: una combinación de cursivas y redondas, mayúsculas y minúsculas, acompañada en ocasiones de paréntesis u otros signos gráficos. Tomemos como ejemplo la genciana, cuyo nombre científico es *Gentiana acaulis* L. subsp. *alpina* (Vill.) O. Bolòs et Vigo.

### *5.2.2 Tratamiento de las variantes geolectales*

La exhaustividad y fidelidad asumidas en el vaciado de las fuentes documentales ha comportado la recopilación de variantes geolectales sin priorización previa de una forma sobre las restantes. De este modo, como ya hemos dicho, no se ha otorgado preferencia a ninguna de las formas documentadas y no se han establecido jerarquías entre un término principal y sus posibles sinónimos complementarios. Sin embargo, nuestra próxima tarea, en la cual actualmente trabajamos, consiste ineludiblemente en establecer criterios priorizadores, asumibles por igual por botánicos, lingüistas y terminólogos, con el objetivo de depurar el corpus de vaciado.

Veamos ahora, como muestra representativa, algunos de los fenómenos concretos de variación geolectal que nos obligan a fijar criterios de selección con el fin de depurar el corpus de vaciado.

En primer lugar, nos encontramos con la sinonimia, situación en la que un mismo taxon recibe más de un nombre común; es decir, tenemos diversas denominaciones catalanas para un único nombre científico. En muchas ocasiones esta diversidad fitonímica mantiene una buena correlación con la articulación dialectal del catalán, que se vertebra en dos grandes bloques que marcan una

división más o menos vertical del territorio: el catalán oriental y el catalán occidental. Entre los muchos ejemplos que podríamos citar, se encuentra el caso del romero, de nombre científico *Rosmarinus officinalis*, una especie aromática muy representativa de la vegetación mediterránea, presente en todo el territorio de habla catalana. Esta planta, usada frecuentemente como condimento en la cocina mediterránea, recibe los nombres de *romer* en la zona occidental y *romaní* en la oriental.

Pero la sinonimia ofrece casos mucho más complejos, en los que el eje de variación dialectal puede no tener valor distintivo: la especie *Agrimonia eupatoria*, la *agrimonia*, perteneciente a la familia de las rosáceas, es designada en el macizo del Montseny, un área montañosa reducida de unos 800 km<sup>2</sup> de superficie a tan solo 50 km de la ciudad de Barcelona, por lo menos con cinco nombres populares distintos. Esta planta, de reconocidas virtudes medicinales, recibe los nombres de *agrimònia* (con distintas variantes, tales como *agramònica* i *gramònica*), *herba de sant Antoni*, *herba del mal estrany*, *set sagnies* y *herba tossera*. Y si consideráramos todo el territorio catalán, deberíamos añadir otros nombres populares, como *serverola* y *herba de la sang*. Se usa tradicionalmente en la zona con finalidad antihipertensiva (para “rebajar la sangre”, en la terminología popular), aplicación a la que parece aludir el nombre catalán *set sagnies*, si bien éste no es el único uso medicinal al que se destina.

En segundo lugar, se da el fenómeno inverso a la sinonimia, la *polisemia*, con una ocurrencia significativa entre las denominaciones recogidas. Un ejemplo muy ilustrativo es el de *árnica*. Con este nombre popular se designan, en la misma área del Montseny citada anteriormente, hasta cinco especies diferentes: *Arnica montana* (la auténtica *árnica*, ausente del Montseny, pero cuyos habitantes la adquieren en herboristerías), *Pallenis spinosa*, *Inula montana*, *Inula viscosa* e *Inula salicina* (las cuatro autóctonas de la zona). Todas estas especies tienen en común el hecho de pertenecer a la misma familia botánica (la familia de las asteráceas o compuestas), poseer flores (o, mejor dicho, capítulos florales) de color amarillo y ser usadas en medicina popular por vía externa como antiinflamatorio (para mitigar

golpes y dolores óseos o musculares), previa maceración en alcohol.

### *5.3 Herramientas informáticas*

Llegados a este punto resulta evidente que el tratamiento adecuado de toda esta información sólo es posible mediante el empleo de herramientas informáticas específicas. En nuestro caso, hemos contado con la ayuda del Gestor de Terminología de TERMCAT (GdT), herramienta basada en un planteamiento conceptual y multilingüe de las fichas terminológicas, que posibilita el tratamiento igualitario de las distintas lenguas de trabajo y admite la atribución de jerarquías en cualquiera de ellas, dos premisas necesarias para abordar con garantías la elaboración de la obra que nos ocupa.

La estructura del Gestor de Terminología permite organizar el diccionario a partir de cualquiera de las lenguas de trabajo: en este caso, el catalán y el nombre científico latinizado. Con el fin de tratar los datos con el máximo rigor metodológico, tanto desde el punto de vista botánico como terminológico, se ha partido de la denominación científica como referente conceptual unívoco. Ello significa que el nombre científico es el elemento a través del cual se organiza toda la información en las fichas terminológicas.

A efectos prácticos, pues, el nombre científico recibe la consideración de lengua principal del diccionario, con el consiguiente resultado sobre la presentación de informaciones en la obra, especialmente sobre la ordenación alfabética de los términos en el diccionario propiamente dicho y en los índices alfabéticos. Sin menoscabo de este planteamiento, aprovechando las prestaciones técnicas de nuestro gestor, la obra ofrecerá adicionalmente una clasificación sistemática de las denominaciones catalanas.

## **6. Conclusiones**

Para finalizar, reseñaremos a modo de conclusiones las ideas que hemos expuesto a lo largo de este artículo sobre la definición del proyecto y sus características y singularidades más relevantes.

Este proyecto, denominado *Recull de noms catalans de plantes*, consiste en la elaboración de un corpus lexical de fitonimia catalana, que se estima que

contendrá alrededor de 25.000 denominaciones catalanas relativas a unos 4.800 táxones con sus correspondientes nombres científicos.

Se ha pretendido que el corpus de vaciado fuese exhaustivo para poder recoger el conjunto de denominaciones documentadas tanto en obras de referencia botánicas y lexicográficas como en bases de datos terminológicas, y las obtenidas a partir de material inédito cedido por botánicos.

Como elementos esenciales que definen el trabajo terminológico llevado a cabo destacamos la voluntad descriptiva de la obra, así como el rigor metodológico asumido en el establecimiento de unos criterios de selección adecuados, basado en el asesoramiento y revisión por parte de botánicos, lingüistas y terminólogos. En esta misma línea, se ha atendido a las singularidades propias de este ámbito de especialidad que se dan en esta obra en concreto: notablemente, la complejidad de la nomenclatura taxonómica y la elevada ocurrencia de variantes geolectales. El tratamiento adecuado de esta información ha sido posible mediante el empleo de herramientas informáticas específicas; en nuestro caso, el Gestor de Terminología de TERMCAT.

Finalmente, sólo cabe decir que el principio de proximidad a las necesidades del usuario y el intento de adecuación a sus expectativas han sido los principios rectores que han impulsado desde el principio esta singular experiencia terminológica.

## **7. Referencias**

BOLÒS, O. de [et al.]. *Flora manual dels Països Catalans*. 3a ed. rev. i ampl. Barcelona: Pòrtic, 2005. (Pòrtic Natura; 9)

GREUTER, W. [et al.]. [eds.]. *International code of botanical nomenclature* (St. Louis Code) [En línia]. Königstein: International Association for Plant Taxonomy, 2003. <http://www.bgbm.fu-berlin.de/iapt/nomenclature/code/SaintLouis/0000St.Luistitle.htm>

INSTITUT D'ESTUDIS CATALANS. *Diccionari de la llengua catalana*. Barcelona; Palma de Mallorca; València: Edicions 3 i 4: Edicions 62: Moll: Enciclopèdia Catalana: Publicacions de l'Abadia de Montserrat, 1995.



MASCLANS, F. *Els Noms de les plantes als Països Catalans*. Granollers; Barcelona: Montblanc-Martín: Centre Excursionista de Catalunya, 1981. (Monografies locals. Botànica; 24).

MASCLANS, F. *Els Noms vulgars de les plantes a les terres catalanes*. Barcelona: Institut d'Estudis Catalans, 1954. (Arxius de la Secció de Ciències; 23)

**Note**

[1]

Centre de Terminologia. c/ Diputació 119, 5<sup>a</sup> 08015 Barcelona - [cbofill@termcat.cat](mailto:cbofill@termcat.cat)

[2]

Centre de Terminologia. c/ Diputació 119, 5<sup>a</sup> 08015 Barcelona - [mabonet@termcat.cat](mailto:mabonet@termcat.cat)

[3]

TERMCAT, Centre de Terminologia. c/ Diputació 119, 5<sup>a</sup> 08015 Barcelona - [majulia@termcat.cat](mailto:majulia@termcat.cat)

[4]

Laboratori de Botànica, Facultat de Farmàcia. Av. Joan XXIII, s/n 08028 Barcelona - [joanvalles@ub.edu](mailto:joanvalles@ub.edu)