

Thesaurus d'història de Catalunya: creació i anàlisi d'un llenguatge documental aplicat a la història

MARIBEL CUADRADO BENÍTEZ

CARMEN GARCIA LORÉS

ESTEVE PERPINYÀ I PERPINYÀ

MEI PERPINYÀ I MORERA

Resum: L'article analitza el procés de creació d'un thesaurus monolingüe en català sobre història contemporània de Catalunya; com s'ha estructurat definitivament i quina funció compleix com a instrument d'anàlisi d'una base de dades bibliogràfica d'història.

El *Thesaurus d'història de Catalunya* és un thesaurus monolingüe en català que es va crear per servir d'instrument d'anàlisi documental a la base de dades bibliogràfica d'història local contemporània de Catalunya, elaborada pel SDHL¹ de la Universitat Autònoma de Barcelona. És el fruit d'un llarg esforç que es va iniciar l'any 1987 i que en aquests moments presenta ja una edició definitiva amb 2.479 termes.

Es va intentar crear un llenguatge documental controlat que permetés una indexació homogènia dels documents que s'anaven incorporant a la base de dades, i que a la vegada en fes possible la recuperació de forma exhaustiva i fàcil. La decisió de crear un thesaurus específic d'història de Catalunya es va prendre davant la inexistència de llenguatges documentals adients a aquesta disciplina. Necessitàvem una eina de treball que permetés recuperar la informació a través de la terminologia i els punts de vista més familiars per als nostres usuaris:

1. Servei de Documentació d'Història Local de Catalunya. Departament d'Història Moderna i Contemporània. UAB. Edifici B. -Despatx B7-123- 08193 Bellaterra (Barcelona). Tel.:(93)581 25 37 Fax: (93) 581 20 01.

historiadors i estudiosos de la història de Catalunya.

Vàrem descartar, doncs, llenguatges documentals més generals (*Llista d'encapçalaments de matèria en català*) en favor d'un llenguatge més tècnic i especialitzat. D'altra banda, no ens servien els thesaurus existents perquè la seva aplicació a l'àmbit català no recollia molts aspectes importants de la nostra història. "La experiencia muestra que existen pocos thesaurus que contruidos para las necesidades de una documentación concreta, puedan ser utilizados tal cual para otra documentación"². L'adaptació de llenguatges externs a la documentació pròpia, com a receptes ja construïdes, difícilment esdevé satisfactòria per a les exigències de l'anàlisi documental.

La possibilitat de crear el nostre propi llenguatge documental ens permetria també realitzar una valoració crítica de la terminologia emprada en la historiografia catalana. La recopilació dels conceptes utilitzats i l'anàlisi posterior d'aquests (relacions d'equivalència, associació i jerarquia) permeten la reducció dels significats dels termes a un sol significant, eliminant ambigüitats en uns o donant contingut a d'altres. Aquests estudis terminològics tan desenvolupats en altres disciplines, com ara el dret, són encara molt precaris en història. Així, per exemple, es dona la paradoxa que diferents historiadors utilitzen un mateix terme a l'hora de conceptuar continguts diferents. És el cas del terme "Poder local", per exemple, que s'utilitza tant per designar les oligarquies polítiques locals, el poder de l'Administració local, o els grups de poder econòmic. Aquest és sols un exemple per posar de relleu la necessitat d'unificar la terminologia com a pas previ a l'intercanvi científic i crític.

Aquest article pretén mostrar la nostra experiència en la construcció del thesaurus i analitzar-ne les característiques principals.

En primer lloc, s'estudien els tipus d'unitats lèxiques utilitzades i l'estructura semàntica del thesaurus. L'anàlisi posa de manifest la complexitat que té l'elaboració d'un thesaurus d'una ciència multidisciplinària com és la història. La interdisciplinarietat es fa patent en les diferents àrees de coneixement que conformen l'estructura del thesaurus.

A continuació, apuntem el mètode emprat en el procés de construcció i manteniment, amb una atenció especial al programari utilitzat en la gestió del thesaurus. I per acabar, presentem una anàlisi de la consulta i utilització del thesaurus per part d'usuaris i documentalistes, mostrant les diferents presentacions del thesaurus i explicant la filosofia d'aquest com a instrument per a la indexació.

2 Van Slype, Georges. *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide: Fundación Germán Sánchez Ruipérez, 1991.

1. Les unitats lèxiques del thesaurus

Els components fonamentals dels thesaurus són els termes, que es defineixen com “la representació d'un concepte”³. Aquests termes “són unitats lingüístiques factibles de ser examinades lingüísticament i gramaticalment”⁴ i no tots compleixen la mateixa funció dins del thesaurus i del sistema documental.

Dins del *Thesaurus d'història de Catalunya* s'inclouen tres categories d'unitats lèxiques: les àrees temàtiques, els descriptors i els no-descriptors.

Les àrees temàtiques o termes capçalera són els títols que encapçalen les grans microdisciplines. Agrupen el conjunt de descriptors, i al seu interior s'organitzen els termes. Aquests títols no són utilitzats per indexar els documents ni les consultes; només es fan servir per subdividir temàticament els descriptors. S'identifiquen gràficament perquè van anotats entre claudàtors. Aquest thesaurus s'estructura a l'entorn de 15 grans àrees de coneixement, dividides a la vegada en 49 subàrees.

Els descriptors o termes preferents són paraules o expressions que s'han triat —a partir d'un conjunt de sinònims, de quasi-sinònims i de termes emparentats semànticament—, per representar de forma unívoca un concepte susceptible d'intervenir en els documents i en les consultes de la base de dades. Aquests elements s'utilitzen per a la indexació. El thesaurus inclou, en aquest moment, 2.071 descriptors.

Els no-descriptors, anomenats també termes equivalents o termes no preferents, són sinònims o quasi-sinònims dels descriptors, i designen en el llenguatge natural conceptes molt pròxims als descriptors. No poden ser utilitzats per indexar els documents ni les consultes, però són inclosos al llistat alfabètic del thesaurus perquè cadascun d'ells reenvia als descriptors que sí representen el concepte adequat. Com a no-descriptors tenim 408 termes, que representen un 20% del thesaurus.

Batles
USEU *Alcaldes*

Dictablanda
USEU *Govern Berenguer*

1. 1. Característiques formals dels descriptors i normalització⁵

Tal com apunta Luis Lizasoain⁶, l'objectiu que persegueix la normalització del

3. Definició de: “Directrices para el establecimiento y desarrollo de tesauros monolingües, Normas PNE- 50 106, Norma ISO 2788-1986 (E)”: *Revista española de documentación científica*, 1989.

4 Currás, Emilia. *Tesauros: lenguajes terminológicos*: Madrid, Paraninfo, 1991.

vocabulari recollit en el thesaurus és reduir el nombre de termes que expressen un mateix concepte, amb la finalitat que els termes emprats en la indexació coincideixin amb els usats en la recerca. Per a aquesta tasca s'han establert unes normes que fixen la forma escrita en què podem presentar els descriptors i els no-descriptors, i que regulen entre d'altres: les formes sintàctiques dels termes; la utilització del singular o plural; el tractament de les xifres numèriques; els termes homògrafs i polisèmics i el gènere dels descriptors.

Formes sintàctiques dels termes. Els descriptors poden ser simples (formats per una sola paraula), preferentment substantius (*Edificis, Comerç*), adjectius substantivats (*Afrancesats*), o verbs substantivats (*Excomunió, Pavimentació*), o compostos (contenen més d'una paraula) formant frases substantivades.

Hem acceptat els termes en la seva forma precoordinaada quan: el terme compost resulta familiar en el camp del thesaurus i la seva descomposició en podria dificultar la comprensió (*Aplicacions informàtiques, Frau electoral*); el fraccionament sintàctic podria conduir a ambigüitats o pèrdues de significat (*Nova cançó, Exposicions universals*); es tracta de noms propis (*Corts de Cadis, Decret de Nova Planta*); els termes compostos tenen un modificador que va perdre el significat original (*Economia submergida*) o relaciona els objectes o els esdeveniments sense connexió lògica (*Plantes depuradores, Claustre de professors*); són termes que no poden definir-se sense el recurs a un altre substantiu (*Grans superfícies, Capitols matrimonials*). Però, sobretot, hem mantingut els termes compostos quan hem considerat que un descriptor simple cobreix un camp d'estudi ampli o bé expressa un concepte molt comú a la nostra disciplina i admet subdivisions, amb la finalitat de matisar-ne el significat o descongestionar i agilitar les recerques d'informació, (*Indústria, Política industrial, Indústria cotonera. Ensenyament primari, Ensenyament secundari, ...*). Tanmateix, l'hem utilitzat per reduir el nombre dels termes assignats a un document, encara que això suposí comptar amb un major volum de descriptors en el thesaurus.

Com a termes compostos s'han de considerar, sintàcticament, les frases adjectivades (*Arxius municipals, Escoles rurals*), les preposicionals (*Impost sobre*

5. Per a l'elaboració d'aquestes directrius de normalització hem utilitzat: "Documentación - Directrices para el establecimiento y desarrollo de tesauros monolingües, Normas PNE- 50 106, Norma ISO 2788-1986 (E)": *Revista española de documentación científica, 1989*; "Documentación - Directrices para el establecimiento y desarrollos de tesauros monolingües (Parte II)": *Revista española de documentación científica, 1990*; "Documentación. Métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización": *Revista española de documentación científica, 1991*.

6. Lizasoain Hernández, Luis. *El sistema de documentación en Ciencias de la Educación: elaboración de un tesoro de investigación educativa*. Madrid: Universidad Complutense de Madrid, 1988. (Col. tesis doctorales; 43/88). Tesis doctoral dir. per Arturo de la Orden Hoz, Madrid, 1987.

la renda, Exposicions a les corts), les adverbials (*Treballadors no qualificats*), les mixtes (*Ferrocarrils de via estreta, Jutjats de primera instància*) i les copulatives (*Llegats i donatius, Seguretat i higiene laboral*).

L'ús de frases copulatives (expressions compostes per dues paraules substantives unides per la conjunció copulativa "i") demana un aclariment, ja que la seva utilització s'aparta de les normes ISO, on no hi són contemplades, encara que molts thesaurus, com ara el de la Unesco, les facin servir⁷. Les principals raons d'aquesta divergència van ser: poder utilitzar un mateix descriptor per a temes molt pròxims que no tenen prou entitat en la nostra disciplina per ser dividits en dos descriptors diferents (*Licors i aiguardents*); relacionar paraules que en el llenguatge col·loquial és difícil de distingir (*Autoserveis i supermercats*); mantenir junts termes i conceptes que sovint trobem units en els documents (*Mapes i plànols, Aliances i tractats*); respectar les nocions coetànies als documents juntament amb les accepcions més actuals (*Rifes i tòmbols*).

Utilització de la forma singular i plural. Com a norma general, sempre que hem pogut, s'ha preferit la forma de singular a la de plural. Però, de vegades, ha estat més pràctic escollir el plural, per tal de donar més sentit als termes i evitar ambigüitats i confusions. En l'elecció de la forma singular o plural ens vàrem trobar amb dues situacions: les formes singulars i plurals tenien el mateix significat o bé un de diferent. En el primer cas, hom distingeix entre substantius quantificables i substantius no quantificables. Dins dels substantius quantificables hi ha unes paraules que poden respondre a la pregunta "Quants-Quantes?" i aquestes s'utilitzen en plural (*Quants Partits Polítics hi ha?, Quantes Batalles?*), i altres paraules que poden respondre a la qüestió "Quant-Quanta?", i aquestes s'han d'usar en singular (*Quant Petroli hi ha?, Quanta Boira hi ha?*). Els substantius usats en sentit general, com un conjunt d'elements (tipus o generalitzacions), s'expressen en plural (*Fonts d'informació, Serveis municipals*). Quan un substantiu fa referència als fenòmens naturals, fem servir el singular (*Humitat atmosfèrica, Temperatura*).

Quant als substantius no quantificables s'empra el singular en els descriptors que representen activitats (*Auditoria, Inspecció*), procediments particulars (*Innovació tecnològica*), noms propis (*Govern Berenguer*), disciplines (*Història, Química*), i creences o ideologies (*Catolicisme, Socialisme*). El plural s'utilitza en els descriptors indicatius de les propietats i materials genèrics (*Càrrecs públics, Equipaments sanitaris*) i els objectes (*Objectes de culte*).

Finalment, quan hem trobat les formes singulars i plurals amb significats diferents, s'ha optat preferentment per afegir un modificador i desfer l'ambigüitat (*Mercat de capitals, Mercats municipals; Consum, Impost de comsums*). Si

⁷ Aitchison, Jean (ed.); Viet, Jean (trad.). *Tesaurus de la Unesco*. Paris: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 1984.

aquesta solució no és possible, utilitzem ambdues formes afegint-hi una nota d'aplicació amb el seu significat. És el cas dels termes *Memòria i Memòries*:

Memòria

- NA Com a tipologia documental s'usa en singular per no confondre'l amb memòries (gènere biogràfic)
- TC *[Materials d'informació]*
- TG *Documents primaris*
- TE *Memòria d'activitats*
Memòria de gestió
Memòria econòmica
Memòria mèdica
- TR *Actes*
Anuaris
Informes

Tractament de les xifres numèriques. S'entén fàcilment que en una base de dades d'història, les dades factuais siguin un element important. Així, els números ordinals es consideren descriptors, encara que siguin adjectius, ja que no són susceptibles de substantivació. Nosaltres els utilitzem per a la numeració dels fenòmens històrics i els posem en forma inversa per facilitar la indexació (*República I, Guerra carlina III*)

Quant als números cardinals, considerem només les dates de fets històrics, que s'escriuen darrera del descriptor (*Crema de convents 1835*)

La qüestió dels **termes homògrafs i polisèmics** s'ha resolt de dues formes: amb l'addició d'un qualificador (*Arquitectura (Ciència)*); o amb la transformació del terme en un sintagma (*Producció, Producció industrial*).

Pel que fa al **gènere dels descriptors**, hem fet un intent d'avançar cap a un llenguatge no androcèntric, encara que estiguem lluny d'haver-ho aconseguit. Així, hem buscat sempre que ha estat possible la paraula més neutra o genèrica per definir un concepte, sense incloure els qualificadors de sexe: (*Personal administratiu* i no *Administratius*). No obstant això, aquest sistema no es pot aplicar a tots els descriptors. La solució més adient hauria estat la indicació dels dos gèneres (*Dissenyadors, dissenyadores*) però no ens resultava útil ja que es multiplicaven el termes del thesaurus i aquest perdia operativitat. Hem adoptat, doncs, la forma masculina com a genèrica. Només hem trencat la norma i hem mantingut els dos gèneres en el cas de les persones vinculades a l'Església (*Clergues, Religioses ; Abats, Abadesses*), donat que compleixen funcions diferents.

Grafia i ortografia dels descriptors: accents, nombre de caràcters, majúscules i minúscules.

El programari utilitzat per a la gestió del thesaurus limita a un màxim de 30 el

nombre de caràcters que pot tenir un descriptor; és, doncs, un altre aspecte a tenir en compte en el moment de la normalització. Els descriptors tenen la primera lletra en majúscula i la resta en minúscules accentuables. Els signes de puntuació estan molt limitats i només s'admet el parèntesi per desfer la polisèmia: *Arquitectura (Ciència)*.

1. 2. Descriptors anacrònics i amb significat històric

Construir un llenguatge documental útil per a la indexació d'un gran volum de documents referents a la història contemporània de Catalunya ha resultat de gran complexitat ja que ha estat necessària la introducció d'un element addicional: el factor temps. Això ha afegit una nova dificultat en la fase de normalització dels descriptors perquè, en treballar amb fonts històriques, ens hem trobat anacronismes i conceptes que han sofert una evolució històrica.

L'anacronisme pot ser motivat per l'obsolescència del concepte o per polisèmia. Per tal de normalitzar aquests termes hem optat per seguir dos criteris; si el concepte no està en ús, però és representatiu d'una etapa històrica i pot ser un descriptor per la quantitat de documentació que l'empara, s'ha mantingut. Hom especifica, però, les relacions de complementarietat amb el descriptor que l'ha substituït. Per exemple, el descriptor *Establiments benèfics*, és un concepte anacrònic que avui ha estat substituït per *Equipaments assistencials*, però nosaltres hem admès ambdós termes per la seva importància en la documentació consultada i per la necessitat tècnica d'acotar les accepcions dels mots dins d'una cronologia concreta.

Quan l'anacronisme es genera per qüestions de polisèmia rep el mateix tractament que tindria un terme obsolet, és a dir, si la paraula té molta importància a la documentació analitzada es manté com a descriptor; si no és considerada així, només resta en el thesaurus alfabètic com a no-descriptor (*Apotecaris*, USEU *Farmacèutics*)

L'últim problema de normalització que varem haver de solucionar és el de l'evolució conceptual i els canvis de denominació que han experimentat alguns termes al llarg del temps. A tall d'exemple, podríem citar els casos dels canvis que durant aquests prop de 300 anys ha sofert la divisió territorial del nostre país, l'evolució de l'Administració de justícia des de les acaballes de l'antic règim fins a l'actualitat, les transformacions que han conegut les diverses administracions que han actuat sobre Catalunya, o les observades en l'àmbit de l'economia, el comerç i la indústria, amb sectors i activitats força importants en un altre temps i ara desapareguts.

Hem optat per mantenir totes les denominacions utilitzades en els documents contemporanis i en els estudis històrics, donat que aquestes transformacions són una de les principals característiques dels processos històrics.

.Indústria

- ..
- ..*Colònies industrials*
- ..
- ..*Protoindustrialització*
- ..
- ..*Indústries*
 - ...*Indústria del metall*
 - ...*Farga catalana*

.Administració local

- ..*Ajuntament*
- ..*Consell de Cent*
- ..*Consell Comarcal*
- ..*Mancomunitat de municipis*
- ..*Administració metropolitana*
- ..*Diputació Provincial*
- ..*Mancomunitat de Catalunya*

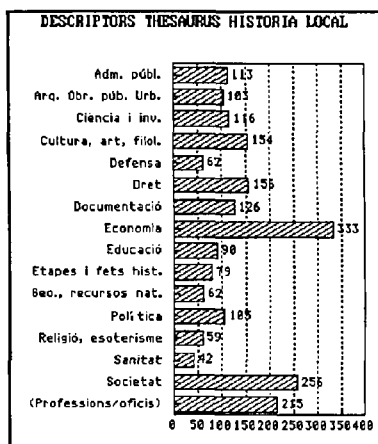
2. Estructura semàntica del thesaurus

En l'edició actual, el thesaurus està estructurat en 15 grans àrees de coneixement subdividides en 49 subàrees, en les quals s'ha intentat desglossar la disciplina de què ens ocupem, la història, i, més concretament, la referida a l'àmbit territorial de Catalunya i en la cronologia que va de l'any 1700 ençà. La complexitat de l'anàlisi documental ha provocat que els perfils definitius d'aquests grans conjunts de termes hagin anat variant al llarg del temps; els seus marges s'han anat precisant a posteriori sobre la base d'un disseny inicial i de successius esborranys del thesaurus. Heus aquí l'estructura principal del thesaurus:

- | | |
|---|---------------------------------------|
| [Administració pública] | [Cultura] |
| [Polítiques d'actuació] | [Filologia] |
| [Serveis públics] | [Arts] |
| [Hisenda pública] | |
| [Funció pública] | [Defensa i ordre públic] |
| [Organització administrativa] | [Documentació i centres d'informació] |
| [Organització territorial de l'estat] | [Materials d'informació] |
| [Divisió territorial] | [Centres d'informació] |
| [Arquitectura, obres públiques i urbanisme] | [Dret] |
| [Arquitectura] | [Ordre jurídic] |
| [Infraestructures i obres públiques] | [Administració de justícia] |
| [Urbanisme] | [Economia] |
| [Ciència i investigació] | [Economia general] |
| [Ciències] | [Economia agrària] |
| [Tecnologia] | [Economia pesquera] |
| [Estudis i investigacions] | [Economia industrial] |
| [Cultura, filologia i arts] | [Comerç] |
| | [Hoteleria i turisme] |
| | [Finances] |
| | [Serveis personals i |

domèstics]	[Medi geogràfic]
[Economia de l'empresa]	[Recursos naturals]
[Treball]	[Impacte i política ambiental]
[Transports i comunicacions]	[Política]
[Educació]	[Teoria política]
[Etapes i fets històrics]	[Acció política]
[Història de Catalunya, s. XVIII-1808]	[Religió i esoterisme]
[Història de Catalunya, 1808-1874]	[Religió]
[Història de Catalunya, 1875-1930]	[Pràctiques esotèriques]
[Història de Catalunya, 1931-1994]	[Sanitat]
[Història mundial]	[Societat]
[Geografia, recursos naturals i medi ambient]	[Demografia]
	[Sistema social]
	[Benestar social]
	[Lleure i activitats recreatives]
	[Comunicació social]
	[Vida pública i associativa]

Entre les diferents microdisciplines n'hi ha unes de més desenvolupades que d'altres. Si es té en compte que el vocabulari és recollit de la indexació lliure dels documents, i per tant s'incorpora al thesaurus en funció d'una necessitat conceptual, es fa evident que entre elles apareguin diferències notables. Aquestes disfuncions deixen patent el propi estat actual de la disciplina: temes més estudiats, enfocaments més utilitzats, dominis inèdits,...



Algunes de les àrees (Economia, Societat i Cultura) presenten un major nombre de descriptors donat que indexen una quantitat major de documentació en la base de dades; així mateix, han requerit un grau superior d'estructuració per eliminar el "soroll" que es produïa en algunes recerques.

En l'estructura del thesaurus, els camps semàntics on s'inclouen els descriptors es dinamitzen per una sèrie de relacions lingüístiques sistematitzades entre els termes, la qual cosa dota de grans possibilitats al diàleg documental. L'estructura semàntica del thesaurus és la que proporciona l'ajuda a l'usuari en aquest diàleg. Aquesta xarxa de relacions, en el moment de la consulta al sistema, permet estendre la recerca a altres conceptes, ja siguin més específics, ja siguin més genèrics o, simplement, als associats, enriquint-la de contingut: múltiples nocions, a priori independents, són susceptibles de ser entrelaçades; l'usuari que interroga el sistema a través del thesaurus s'orienta d'aquesta manera cap a tots els documents que puguin ser del seu interès.

En realitat, el thesaurus funciona com un sistema on cada element que el compon és interdependent amb els altres als quals està unit. La possibilitat de combinacions és enorme i permet realitzar tants creuaments terminològics com es vulgui, des de qualsevol angle conceptual.

Les principals relacions que s'estableixen en el thesaurus són les relacions jeràrquiques, les associatives i les d'equivalència.

La relació jeràrquica és la relació bàsica que marca la diferència entre un thesaurus sistemàtic i un llistat no estructurat de paraules claus, com ara un glossari o un diccionari. Es pot definir com una relació asimètrica entre dos descriptors en la qual un d'ells és superior a l'altre per un caràcter normatiu⁸. Els termes es relacionen en funció del nivell i grau d'especificitat-generalitat. Els descriptors, en el cas del nostre thesaurus, formen cadenes jeràrquiques que oscil·len fins a un màxim de 3 a 5 nivells per cadena. Els descriptors independents, no inclosos en cap cadena, són considerats caps de jerarquia.

.Indústria

..

..Indústries

...Indústria alimentària

....Indústria càrnica

....

....Indústria vinícola

.....Cava

.....Licors i aiguardents

.....Vi

El maneig eficient d'una gran quantitat de termes esdevindria una tasca molt complexa si no existís una estructura interna que anés subdividint i agrupant els descriptors en funció dels seus significats. La jerarquitització semàntica facilita la

⁸ Van Slype, *Op. cit.*

indexació dels documents alhora que suggereix, en estar en la mateixa escala conceptual, altres possibilitats o l'elecció dels descriptors més pertinents, de manera que s'elimina "soroll" en el moment de la recerca.

Les agrupacions internes de descriptors responen a diferents criteris. El més comú és el de reunir-los i estructurar-los per temes tot seguint la coherència lògica de cada microdisciplina. L'avantatge d'aquest sistema rau en el fet que l'usuari sol estar acostumat a raonar en clau de matèries i copsa fàcilment aquesta lògica interna.

En l'estructura del thesaurus també hi ha intervingut la jerarquització a partir de "facetes", per agrupar els descriptors segons el seu significat i funció o per la seva naturalesa dins d'un determinat domini. El seu ús dins del conjunt del thesaurus és limitat, en el sentit que sempre es procura mantenir per a cada matèria la classificació tradicionalment admesa; quan aquesta no existeix, és massa simple o és insatisfactòria per a la indexació de la documentació, hem recorregut a la utilització de "facetes";

.Equipaments educatius

..Centres d'ensenyament

...[Classificació per nivells]

....Escoles bressol

....Parvularis

....Escoles primàries

...

...[Class. per mode d'administració]

....Escola privada

....Escola pública

...[Segons ideari del centre]

....Escola laica

....Escola religiosa

...[Segons mètodes pedagògics]

....Escola nova

....Escola racionalista

...

Les relacions associatives es donen entre dos descriptors que designen conceptes susceptibles d'evocar-se per associació d'idees; els dos termes relacionats no poden ser equivalents i tampoc no poden ser a la mateixa cadena jeràrquica car la proximitat física ja facilita a l'usuari la correcta elecció. La importància de fer explícita aquesta connexió en el thesaurus rau en el fet que és de gran ajuda en el procés d'indexació o de recuperació, indicant l'existència o bé l'ús més pertinent d'altres descriptors. La relació associativa és recíproca.

Accidents de trànsit

TC *Transports i comunicac.]*
 TG *Seguretat viària*
 TR *Assegurances d'automòbils*

Assegurances d'automòbils

TC *[Finances]*
 TG *Assegurances*
 TR *Accidents de trànsit
 Vehicles*

En un tipus de thesaurus com el nostre, d'història entesa com a ciència multidisciplinària, les relacions associatives entre termes resulten de gran utilitat i n'afavoreixen la versatilitat i ductilitat, car actuen enviant enllaços entre descriptors d'àrees diverses, precisant els diferents plans a què pot remetre la informació. Es podria dir que és a través d'aquesta relació que les diferents grans àrees del thesaurus no són compartiments estancs sinó que esdevenen complementàries les unes de les altres, actuant com a vasos comunicants entre elles. Així, per exemple, en un estudi sobre un grup de masies d'una població, la indexació haurà de recollir els diversos aspectes que, probablement, inclourà el document: arquitectura de l'edifici, tipus de construcció, usos a què es va destinar, modificacions sofertes en el temps, etc. Per tant, una indexació correcta hauria d'incloure descriptors com "Masies" i "Explotacions agràries". Les relacions entre ambdós termes s'expressen en el thesaurus de la manera següent:

Masies

TC *[Arquitectura]*
 TG *Arquitectura civil*
 TR *Edificis
 Explotacions agràries*

Explotacions agràries

TC *[Economia de l'empresa]*
 TG *Empreses*
 TE *Granja experimental*
 TR *Agricultura
 Ramaderia
 Masies*

Finalment, **les relacions d'equivalència** entre descriptors són les relacions que s'estableixen entre termes diferents que tenen el mateix significat. S'utilitzen en el cas de sinonímies, quasi-sinonímies o equivalències semàntiques, enviant l'usuari des dels termes no-descriptors al descriptor, mitjançant el codi USEU ; recíprocament, el descriptor preferent, acceptat en el thesaurus, haurà d'anar acompanyat de la notació UP, (Utilitzat per). En general, es pot dir que s'han incorporat els termes més habituals en els documents, més estesos a les disciplines o bé més propers a l'usuari de la base de dades. Aquest tipus de relacions no consta en la presentació jeràrquica del thesaurus i per consultar-les cal adreçar-se a la presentació alfabètica, on apareix tota la notació completa dels descriptors.

D'entre les particularitats que presenta el nostre thesaurus respecte d'altres

caldría comentar la inclusió d'una àrea innovadora, la que consta com “[*Etales i fets històrics*]”, destinada a recollir, ordenar i sistematitzar aquells períodes i aquells fets que bé la documentació històrica o bé la historiografia s’han encarregat de singularitzar i identificar. Com a exemple, entre els usuaris de la base de dades hom identifica ràpidament i sense problemes conceptes i descriptors com *Guerra de Successió*, *Guerra del Francès*, *Sexenni democràtic*, o *República espanyola II*; la consulta al thesaurus alfabètic acaba de precisar, en cas de dubte, el significat del terme, el seu abast cronològic, etc. La creació d’aquesta àrea ha servit per recollir i identificar la documentació que es referia al període o al fet en qüestió, precisant, bé que indirectament, l’abast cronològic de la consulta. S’ha procurat donar-li idèntic tractament documental al que tenien altres àrees, dotant-la també d’una estructura seqüencial (en el temps històric) i jeràrquica (per nivells d’amplitud semàntica). Aquesta solució ens ha servit alhora per evitar el “soroll” que presentaven algunes consultes, donada la gran quantitat de registres associats a aquests descriptors; s’ha possibilitat així l’elecció del grau de precisió i especificació de la consulta.

.*Sexenni democràtic*
 ..*Revolució 1868*
 ..*Regnat d’Amadeu I d’Espanya*
 ..*República espanyola I*
 ..*Guerra de Cuba 1868-1875*

També presenta un tractament diferenciat l’àrea de “[*Professions i oficis*]”, entre d’altres raons per l’elevat nombre de documentació a què es referien (biografies de personatges als quals calia identificar professionalment) i per la necessitat d’augmentar el nivell de precisió de la consulta; per això se l’ha dotat d’una estructura paral·lela lligada a l’activitat realitzada, donant una estructura facetada a cadascuna de les àrees del thesaurus.

<i>Hoteleria</i>	<i>Construcció</i>	<i>Sanitat</i>
...
<i>Cambres</i>	<i>Ènguixadors</i>	<i>Forenses</i>
<i>Cuiners</i>	<i>Manobres</i>	<i>Llevadores</i>
	<i>Paletes</i>	<i>Metges</i>
	<i>Tapiadors</i>	

3. Procés de construcció, manteniment i tractament formal del thesaurus

3.1. Procés de construcció

El procés de construcció del thesaurus ha contemplat diferents tasques: recopilació del vocabulari, selecció i normalització dels descriptors, i establiment de les relacions d'equivalència, jerarquia i associació entre descriptors.

El primer pas era la recopilació del vocabulari o terminologia. Un primer estudi a priori de les fonts terminològiques més característiques de la disciplina ens va permetre establir una primera gran estructura amb les possibles grans àrees de coneixement o microdisciplines⁹. Era obligada la consulta de glossaris, llistats de termes, vocabularis o altres thesaurus afins, ja que, molts d'ells responen a estudis terminològics i d'anàlisi documental prou rellevants, i són fruit d'un treball que no podíem obviar.

Posteriorment, i a partir d'un primer mostreig fet a diverses biblioteques pels nostres especialistes amb un sistema d'indexació lliure, vàrem establir una relació de "descriptors candidats". Mitjançant un mètode deductiu reagrupàrem els descriptors en funció de la seva temàtica. Aquesta primera recopilació de vocabulari ens permetia ja establir els grans temes o àrees de coneixement específics del thesaurus, és a dir, una primera estructura dins la qual en podíem anar incorporant d'altres (de manera que el termes específics que pertanyessin a un mateix camp quedessin agrupats en "microdisciplines" i "facetes"). Aquestes àrees s'han anat revisant contínuament al llarg del procés de construcció, subdividint aquelles amb massa descriptors o reagrupant-ne d'altres amb pocs termes. Actualment, podem afirmar que els termes presents en el thesaurus són el producte d'un conjunt representatiu de documentació, els 35.000 registres de la base de dades d'història local de Catalunya.

En una segona fase es va normalitzar el vocabulari recollit i es va concretar en un llenguatge documental. S'examinà cadascuna de les microdisciplines amb la finalitat de seleccionar-ne i normalitzar-ne els descriptors. Aquest procés suposa una primera tria entre descriptors i no-descriptors i, per tant, l'elaboració d'una primera estructura d'equivalència semàntica entre termes sinònims.

Una vegada seleccionats i normalitzats els termes descriptors s'havia d'elaborar una estructura jeràrquica dintre de cadascuna de les microdisciplines. La creació de cadenes jeràrquiques permeté detectar i eliminar els dobles usos de descriptors i trobar possibles llacunes que s'ompliren afegint els conceptes que faltaven per mantenir-ne la coherència.

Finalment s'examinà cadascun dels descriptors del thesaurus per poder esta-

9. Per consultar les fonts terminològiques utilitzades, vegeu la bibliografia final.

blir les seves relacions associatives amb d'altres descriptors localitzats en altres àrees de coneixement.

3.2. Manteniment i actualització

El thesaurus és una eina viva i dinàmica que es va nodrint de nous descriptors com a conseqüència de la indexació de nous documents i dels canvis qualitatius que el mateix procés de construcció va generant. El manteniment i l'actualització periòdica d'aquest llenguatge documental és, doncs, una necessitat. Se n'ha de controlar l'ús i tenir en compte les dificultats que es presenten i les solucions a adoptar. Es fa necessari actualitzar-lo periòdicament tant per corregir errors i omissions com per adaptar-lo a l'evolució científica de les disciplines que aplega.

“El mantenimiento de los lenguajes documentales es una necesidad pues el lenguaje natural evoluciona a medida que los conocimientos progresan y las actividades se diversifican. Conservar los lenguajes documentales permite, asimismo tener en cuenta la experiencia que ha adquirido la unidad de información, tanto en la descripción como en la búsqueda de información; algunos descriptores resultan inútiles, otros demasiado precisos, otros muy generales y ambiguos...”¹⁰

El seguiment d'ús del thesaurus és una tasca dels responsables de la seva construcció, comptant sempre amb els suggeriments i problemes plantejats pels analistes encarregats de la indexació dels documents i del usuari que l'utilitzen per plantejar les seves recerques.

Es controla la freqüència d'ús dels descriptors (gran quantitat de descriptors que es fan servir molt poc o, per contra, descriptors que s'utilitzen amb massa freqüència); l'absència de conceptes que sí apareixen en els documents (descriptors candidats); les dificultats d'ús del thesaurus (descriptors amb un sentit molt similar que necessiten la inclusió d'una nota d'aplicació i ús); noves relacions semàntiques o associatives entre descriptors, ...

Aquest seguiment que fan conjuntament els responsables de l'administració del thesaurus i els analistes es materialitza en informes o propostes concretes que es porten a discutir en les reunions periòdiques organitzades amb aquest objectiu. Les resolucions adoptades es centralitzen en una persona, l'administradora del thesaurus, que és l'encarregada de portar a terme els canvis procurant mantenir-ne la coherència. Així, si la supressió o incorporació de descriptors pot repercutir en la indexació, aquesta persona estudia i reindexa els documents afectats.

Tenint en compte que la incorporació d'un nou terme ha d'estar precedida de l'estudi de diferents fonts terminològiques, vàrem trobar operatiu designar especialistes responsables per a cada àrea de coneixement del thesaurus. D'aquesta

10. Van Slype, Georges ; *Op. cit.*

manera s'agilita la consulta de bibliografia i s'evita la duplicitat de tasques. Així, cadascun dels descriptors candidats suggerits pels documentalistes s'assigna en principi i d'una manera lògica a la microdisciplina a la qual hauria de pertànyer. L'especialista d'aquesta àrea serà l'encarregat d'estudiar la viabilitat o no del descriptor i la ubicació més adient, així com les possibles relacions jeràrquiques, associatives i d'equivalència. L'informe de cada responsable es porta a reunió per tal de ser comentat. En els inicis de la construcció del thesaurus les reunions dels responsables foren molt freqüents (setmanals o mensuals). Una vegada provada la validesa d'utilització per part d'analistes i usuaris, i mentre l'estudi es troba en procés d'edició, el seguiment i l'actualització del thesaurus es poden fer anualment, coincidint amb les noves edicions.

3.3. Tractament del thesaurus: el sistema gestor de CDS-ISIS

El procés de creació i manteniment d'un thesaurus és dinàmic; mai no es pot parlar d'una edició definitiva i, per tant, el seu tractament formal demana unes facilitats que assegurin el control del llenguatge documental i que a la vegada agilitin les tasques de seguiment.

L'automatització del procés de gestió del thesaurus aportà una apreciable ajuda al possibilitar majors manipulacions del vocabulari: selecció, classificació alfabètica i edició. La utilització de la informàtica en el procés de tractament del thesaurus pot ser molt diversa.

En un primer moment optàrem per l'aprofitament de les prestacions d'un programa de tractament de textos. Aquests permeten portar a terme les principals tasques de creació i manteniment del thesaurus; això no obstant, el temps invertit no disminueix ja que gran part de les operacions no es poden fer de forma automàtica (per exemple afegir o esborrar relacions entre descriptors). Els principals avantatges de l'ús del processador de textos són les seves prestacions de cara a l'edició del thesaurus.

Posteriorment resolguérem agilitar les tasques i utilitzar un sistema de gestió i manteniment de thesaurus, de processament per lots, amb moltes més prestacions.

El desenvolupament d'una recent aplicació específica per al manteniment de thesaurus sobre el programari CDS/ISIS (sistema de gestió documental que administra la nostra base de dades) ens va oferir la possibilitat d'accedir i gaudir de les seves prestacions.

L'aplicació per a la gestió i manteniment de thesaurus desenvolupada sobre CDS/ISIS va ser concebuda com un programa pilot per resoldre els múltiples problemes que comportaven les tasques de creació i manteniment de thesaurus multilingües. La primera versió fou implementada per TERMNET (Xarxa Internacional de Terminologia) el 1990 i posteriorment millorada l'any 1991, amb la incorporació, entre d'altres facilitats, d'un subsistema KWIC per a la creació

d'índexs permutats.

```
***** Thesaurus maintenance System *****

C - Change database
G - Thesaurus system configuration
T - Thesaurus maintenance services
K - Thesaurus KWIC subsystem
E - ISISSTENT - Data entry services
S - ISISRET - Information retrieval services
P - ISISPRT - Sorting and printing services
I - ISISINV - Inverted file services
D - ISISDEF - Data base definition services
M - ISISXCH - Master file services
U - ISISUTL - System utility services
A - Programming facilities
X - Exit (to MSDOS)

Data base      :THES      Worksheet: Thes
Max MFN       : 2479      Format: ALL
```

El programari pot ser considerat com un prototip avançat i flexible per crear thesaurus amb l'avantatge que pot treballar amb nou llengües, de les quals una es pot definir com a llengua preferent, mentre que les altres queden com a secundàries.

L'aplicació compta amb les següents facilitats: possibilitat de produir relacions invertides automàticament; prestacions per a l'ordenació i impressió ; capacitat d'organitzar els termes per microdisciplines; creació d'índexs permutats ...

Una de les principals característiques d'aquest programari és la flexibilitat del sistema per adaptar-se a les necessitats de cada thesaurus. L'aplicació és una base de dades de terminologia en la qual cada descriptor o no-descriptor del thesaurus conforma un registre propi. No existeixen limitacions en el nombre de registres a gestionar, si bé hi ha una limitació en la longitud del descriptor de 30 caràcters.

El sistema genera automàticament la taula (FDT) amb els camps de la base de dades predefinits (Descriptor, Scope Note, UF, BT, NT, USE, RT, Ascriptor) i una altra taula (FST) amb els camps indexables que conformaran l'arxiu invers. Aquestes taules poden completar-se afegint-hi els camps que es considerin útils o específics de cada sistema documental.

L'usuari pot també configurar, en funció de les seves necessitats, els seus propis fulls d'entrada de dades i d'impressió, així com l'ordenació i la indexació dels camps que vulgui.

Una altra de les característiques bàsiques del programa és que permet conservar un fitxer històric (fitxer LOG), on consten les modificacions sofertes pels

descriptors: termes obsolets, relacions ja no vàlides, notes d'aplicació modificades... D'aquesta manera es pot fer un seguiment cronològic dels descriptors i a la vegada de les modificacions i evolucions que es van produint en l'estructura del llenguatge documental.

L'adopció d'aquest sistema en la gestió del nostre thesaurus ens ha reportat molts avantatges, sobretot quant a l'estalvi de temps i recursos i a un control més exhaustiu de l'estructura del thesaurus. D'altra banda, la possibilitat d'intercanviar registres des del sistema a un processador de textos ens permet gaudir de les prestacions d'aquests pel que fa a l'edició i manipulació de les dades.

4. Presentació del thesaurus

El *Thesaurus d'història de Catalunya* té diferents presentacions per a la seva utilització i consulta: la presentació alfabètica, la jeràrquica o sistemàtica, i l'alfabètica permutada o índex KWIC.

En la presentació alfabètica tots els descriptors i no-descriptors s'organitzen en una seqüència alfabètica única. Els no-descriptors, tipogràficament diferenciats, sols remetent als seus equivalents o descriptors acceptats mitjançant l'indicador USEU.

Eleccions locals (descriptor no admès)

USEU *Eleccions municipals* (descriptor vàlid)

La informació que cada descriptor porta incorporada en la presentació alfabètica és la següent:

TC Microdisciplina (Terme capçalera) a la qual pertany el descriptor.

TG Indica el terme genèric al qual està adscrit el descriptor.

TE Referència als termes específics que conté el descriptor.

TR Indica els termes relacionats; les relacions associatives del descriptor amb d'altres del thesaurus.

UP Remet d'un descriptor a d'altres no-descriptors semànticament equivalents. Presenta la relació inversa de l'identificador USEU.

NA Nota explicativa o d'aplicació que dóna informació suplementària sobre el concepte o la seva utilització.

NOTA Nota d'ús intern per als analistes. Conté aclariments i definicions per assegurar una correcta utilització del descriptor. (Aquesta nota interna queda exclosa en l'edició del thesaurus).

Guerra del Francès

NA [1808-1814]

- UP *Guerra de la Independència*
- TC *[Història de Catalunya 1808-1874]*
- TE *Batalles del Bruc*
 - Setge de Tarragona*
 - Setges de Girona*
- TR *Administració napoleònica*
 - Afrancesats*
 - Crisi de l'antic règim*
 - Guerra*
 - Juntisme*
 - Miquelets*
 - Epoca de Napoleó I*

La presentació jeràrquica o sistemàtica mostra les diferents microdisciplines o àrees de coneixement en què s'estructura el thesaurus i, dins de cadascuna, les seves facetes i descriptors, amb les seves relacions genèrico-específiques degudament jerarquizades. Els diferents nivells de jerarquia es visualitzen mitjançant un desplaçament dels termes a la dreta precedits per un punt. El nombre de punts indica el nivell jeràrquic que ocupa cada terme en qüestió. Els claudàtors s'utilitzen com a indicadors classificatoris que identifiquen els termes que no tenen funció de descriptor, ja sigui perquè són facetes o bé àrees de coneixement.

[Vida pública i associativa]

. *Associacions*

.. *[Recreativo-culturals]*

... *Associacions excursionistes*

... *Associacions filatèliques*

... *Associacions folklòriques*

.... *Associacions pessebristes*

.... *Colles castelleres*

.... *Colles sardanistes*

.... *Esbarts*

... *Associacions recreatives*

.... *Casinos*

... *Centres cívics*

... *Centres regionals*

.... *Cercles catalans*

La presentació alfabètica permutada o índex KWIC permet un accés més fàcil als descriptors compostos ja que agrupa els termes allunyats en la presentació

alfabètica no permutada. Facilita la recerca d'un descriptor del qual es desco-neix la forma exacta, partint d'una sola paraula.

La permutació es realitza a partir de cadascuna de les paraules significatives dels descriptors simples o compostos. Es genera una ordenació alfabètica de tots els descriptors comptant també amb les paraules significatives no ubicades en primer lloc. Els termes sempre apareixen complets, i la paraula alfabetitzada se situa en la columna central.

	Comerç
	Comerç colonial
	Comerç detallista
	Comerç exterior
	Comerç interior
	Comerç majorista
	Comerç marítim
Cambres de	comerç
Representants del	comerç
Distribució	comercial
Frau	comercial
Política	comercial
	Comercialització
Acords	comercials
Agents	comercials
Guies	comercials

5. El thesaurus com a eina de la indexació

Fins ara hem analitzat el thesaurus des del punt de vista de la seva construció com un llenguatge combinatori d'estructura jeràrquica. Però, bàsicament, un thesaurus es defineix per la funció que exerceix en el procés documental: la d'instrument d'indexació dels documents en la fase d'anàlisi documental i, posteriorment, la d'instrument de formulació de la recerca en la fase de recuperació de la informació. L'objectiu, doncs, d'aquest apartat és analitzar quin és el paper del thesaurus en el conjunt del sistema documental de la base de dades d'història local.

Segons el disseny del sistema documental la descripció del contingut dels documents es fa en base a les quatre variables que es consideren definitòries de la disciplina històrica: la cronològica, l'àmbit geogràfic, el tema i, finalment, els protagonistes de l'acció. Cadascun d'aquests elements resta reflectit en camps diferents a l'estructura de la base de dades. En aquest apartat ens ocuparem exclusivament del Temàtic, ja que és en aquest camp on s'hi recullen els descriptors que informen sobre la temàtica o temàtiques del document i que

s'extreuen del thesaurus elaborat amb aquesta finalitat. El thesaurus, doncs, és l'instrument lèxic de l'anàlisi temàtica: només els descriptors del thesaurus poden ser utilitzats per omplir aquest camp.

Aquest és de tots els camps d'anàlisi el més difícil d'omplir ja que, tant en el procediment d'extracció de conceptes com en la posterior assignació de descriptors, intervenen molts factors que poden afectar l'eficàcia del sistema documental. De fet, un bon sistema documental no depèn exclusivament de disposar d'un bon llenguatge documental, sinó del bon ús que se'n faci. En aquest sentit, és fonamental establir uns criteris homogenis de lectura dels documents per tal d'expressar-los temàticament, com també és molt important saber assignar-hi els descriptors més correctes.

5.1. Extracció de conceptes

El primer pas de l'anàlisi documental és la identificació de les matèries de què tracta el document a través del seu examen. Hom recomana als analistes la lectura de les parts més significatives: títol, sumari, introducció, conclusions i paràgrafs inicials dels capítols. La lectura d'aquestes parts ha de permetre identificar els conceptes més significatius del document. Aquests conceptes han de respondre a una sèrie de variables que la normativa d'anàlisi utilitzada imposa, a saber: el tema principal, els temes secundaris o col·laterals, els factors, les conseqüències, els agents, l'enfocament, el període històric i la tipologia documental.

Quan s'inicia l'anàlisi d'un document s'ha d'identificar el tema principal amb el descriptor més específic possible. Ara bé, quan un llibre o document tracta moltes de les matèries compreses dins d'un descriptor genèric, amb aquest sol n'hi ha prou per definir-lo i no cal desglossar-lo. Per exemple, seria el cas d'un estudi general sobre una comarca en el qual es parla d'aspectes molt diferents (econòmic, cultural, social,...) cada un d'ells desglossats en temàtiques més específiques (agricultura, comerç, indústria, turisme,...). En aquest cas amb el terme *Economia* definiríem l'aspecte econòmic i no faria falta utilitzar els descriptors més específics. De la mateixa forma que l'aspecte cultural (tradicions populars, manifestacions culturals, patrimoni artístic,...) quedaria definit amb el descriptor *Cultura*. També, l'analista pot considerar que el llibre aporta informació sobre d'altres aspectes secundaris o col·laterals al tema principal, o bé creu que cal destacar un aspecte concret del tema principal perquè té un tractament més ampli en el document.

Una vegada definit el tema o temes objecte d'estudi del document, l'analista s'ha de preguntar pels factors o causes possibles que l'han provocat. Per exemple: en un llibre sobre la crisi agrària de finals del segle XIX possiblement es parli detingudament de la *Plaga de la fil·loxera*. Així també tindrà en compte les conseqüències, que fan referència a les repercussions o efectes del tema principal

del document. Per exemple, en un estudi sobre *epidèmies*, probablement es parli de la *mortalitat* causada per la malaltia.

Un altre element que s'ha de fer constar són els protagonistes socials que intervenen en la temàtica del document: empresaris, obrers, partits polítics, associacions, etc. Per exemple, en una exposició al govern per demanar una política econòmica proteccionista, és molt important els agents que intervenen en l'exposició: burgesia comercial, empresaris, associacions econòmiques, etc. A la vegada, el documentalista analitzarà l'enfocament metodològic que l'autor ha donat a la temàtica del document; no és el mateix un estudi historiogràfic que un llibre de text sobre la *Setmana Tràgica*. De la mateixa manera, no és el mateix un estudi sociològic sobre la família que un estudi antropològic.

Quant al període històric, hem de dir que no l'utilitzem per assenyalar el període cronològic que abasta el contingut del llibre, sinó que s'utilitza per definir obres historiogràfiques que analitzen les repercussions polítiques i/o econòmiques del període com a conjunt. Per exemple, a la història d'una cooperativa obrera des del 1900 fins al 1923, no caldrà posar *Restauració*, mentre que a un llibre que tracti del Cooperativisme durant la Restauració serà necessari fer-ho.

Sota el criteri de tipologia documental englobem els elements que fan referència a les característiques formals del contingut del document. Ex.: *biografia*, *memòries*, *reglaments*, *legislació*, *exposicions al govern*, *sessions inaugurals*, *almanacs*, *dietaris*, *cròniques*, etc. Aquest tipus de descriptors aporten més informació sobre el document però per si sols no el poden definir i, per tant, han d'anar acompanyats sempre de descriptors conceptuals. Per exemple, la biografia d'un alcalde franquista (a més de posar el nom del personatge en el camp d'identificadors personals) podria tenir com a descriptors: *Biografia*; *Polítics*; *Alcaldes*; *Política municipal*; *Franquisme*.

Altres descriptors com *Reglaments*, *Legislació*, *Recull de premsa*, *Directori*, etc., també han d'anar acompanyats de termes referents a la temàtica del document ja sigui el tipus d'associació a què fan referència els reglaments, ja sigui la temàtica a la qual afecta la legislació, el directori o el recull de premsa.

Així doncs, l'analista quan examina un document ha de tenir en compte aquestes vuit variables per determinar els conceptes representatius del document, la quantitat i qualitat dels quals ha de satisfer els paràmetres d'exhaustivitat i especificitat definits pels teòrics de l'anàlisi documental.¹¹

L'exhaustivitat és el nombre de descriptors usats i està en funció de l'acomodació del document als criteris de selecció del sistema documental. En aquest sentit una monografia local tindrà un examen molt més detallat que no pas un llibre del qual només ens interessi una part, o que la temàtica de què parli sigui

11. Chaumier, Jacques. *Análisis y lenguajes documentales: el tratamiento lingüístico de la información documental*. Barcelona: Mitre, 1986.

tangencial a les àrees de coneixement definides en els criteris de selecció. I l'especificitat o pertinença és l'exactitud amb què un concepte particular d'un document està representat per un terme d'indexació. Tal i com ja hem dit, a un document, sempre se li ha d'assignar el descriptor més específic possible, de manera que una noció no estigui representada per un descriptor gaire general. Igual que en el cas anterior, aquesta normativa que se segueix estrictament en documents d'història de Catalunya es va flexibilitzant en els casos en què el document o una part sigui tangencial als interessos temàtics del sistema documental.

5.2. Traducció al llenguatge documental

Una vegada s'han extret els conceptes del document cal traduir-los al llenguatge documental. S'ha de consultar el thesaurus i trobar els descriptors que representin amb més exactitud cada concepte o noció extreta del document.

No sempre una noció té la seva equivalència semàntica en un sol descriptor del thesaurus. En virtut de la postcoordinació, fonament de la construcció d'un thesaurus, un concepte pot estar representat per la combinació de diferents descriptors. Per exemple, si un document tracta de la història d'una associació religiosa juvenil caldran dos descriptors per definir-lo: *Associacions religioses i Joventut*.

Si no se'n troba cap d'adient caldrà proposar, tal i com s'explica en l'apartat de manteniment del thesaurus, un descriptor candidat per tal que els responsables de la seva construcció n'estudïin la possible inclusió.

Pel que fa a la quantitat de descriptors a utilitzar en l'anàlisi d'un document cal dir que, seguint les recomanacions de la norma ISO 5963-1985, no hem imposat cap limitació. En conseqüència, el nombre de descriptors assignats per document depèn de diferents factors: en primer lloc, de la quantitat d'informació del propi document; en segon lloc, de la major o menor pertinença dels documents a l'àmbit temàtic de la base de dades i, en tercer lloc, de la informació ja continguda en els altres camps de l'anàlisi documental. En una mostra aleatòria de 100 registres la mitjana de descriptors assignats, si només tenim en compte el camp temàtic, és de 6 descriptors, mentre que si comptabilitzem també els descriptors dels camps geogràfic, cronològic i identificadors, és de 12.

6. Valoració del sistema documental

L'eficàcia d'un sistema documental es mesura bàsicament a través de la taxa de coherència. La taxa de coherència correspon al nombre de descriptors iguals assignats a un document pels diferents analistes, i pot variar des de 0 (cap descriptor comú) fins a 1 (tots els descriptors comuns). Una taxa de coherència d'u

implica evitar la distorsió en la recuperació de la informació. I això depèn, a part d'una bona formulació de l'equació de consulta, principalment de l'homogeneïtat de l'anàlisi dels documents, ja que només recuperem la informació a partir dels descriptors assignats als documents.

L'homogeneïtat de l'anàlisi i, en conseqüència, l'operativitat del sistema, s'aconsegueix tenint cura de tres factors: la validesa del llenguatge documental, una normativa d'anàlisi rigorosa i el nivell tècnic i la formació dels indexadors.

Pel que fa a la validesa del llenguatge documental, en aquest article ja s'ha parlat a bastament de la filosofia i la construcció del *Thesaurus d'història de Catalunya*: aquest ha de permetre un diàleg eficaç entre els documents i els usuaris a través de la utilització d'un lèxic familiar als historiadors; el lèxic usat ha de ser representatiu de la disciplina en qüestió i ha de ser prou ric com per permetre analitzar tots els documents susceptibles d'introduir-se a la base de dades. En segon lloc, s'ha redactat una normativa d'anàlisi que intenta uniformar tant els criteris de lectura dels documents com els de l'assignació de descriptors. Aquest element és molt important en especial en serveis de documentació en què, com en el nostre, hi participen molts documentalistes. I, finalment, un altre element influent en l'eficàcia del sistema és el nivell tècnic i la formació dels indexadors. Tots els indexadors de la base de dades d'història local són historiadors, amb coneixements amplis de la història contemporània de Catalunya, als quals es dona una sòlida formació en biblioteconomia i documentació.

Bibliografia:

Aitchison, J.; Allen, G. G. *Bibliografía de vocabularios "Thesauri", encabezamientos de materias y esquemas de clasificación de ciencias sociales (mono y plurilingües)*. Paris: Unesco, 1983.

Aitchison, Jean (ed.); Viet, Jean (trad.). *Tesaurus de la Unesco*. Paris : Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 1984.

Aitchison, J.; Gilchrist, A. *Thesaurus construction: a practical manual*. 2nd ed. London: Aslib, 1987.

Booth, Barbara; Blair, Michael. *Thesaurus of Sociological Indexing Terms*. San Diego, California: Sociological Abstracts, 1989.

Casòliva, Mercè; et al. *Llistat d'encapçalaments de matèries per a biblioteques*. Barcelona: Diputació de Barcelona, Biblioteca de Catalunya, 1985.

CEDEPOP. *Thesaurus de la formación profesional*. Madrid: INEM, 1990.

CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN CIENTÍFICA (CINDOC). *Tesaurus Isoc de economía (Versión provisional) 1992*. Madrid: CINDOC: CSIC, 1992.

Chaumier, Jacques. *Análisis y lenguajes documentales: el tratamiento lingüístico de la información documental*. Barcelona: Mitre, 1986.

CSIC; CINDOC; Instituto Vasco de Administración Pública. *Tesaurus ISOC de urbanismo*. Madrid: CSIC, 1992.

Currás, Emilia. *Tesaurus: lenguajes terminológicos*. Madrid: Paraninfo, 1991.

"Directrices para el establecimiento y desarrollo de tesauros monolingües". PNE- 50 106. ISO 2788 - 1986 (E). *Revista española de documentación científica*. Núm. 12,4 (1989) p. 463-483., núm. 13,1 (1990), núm. 14, 1 (1991).

Enciclopedia Universal Ilustrada Europeo-americana. Madrid; Bilbao; Barcelona: Espasa Calpe, 1920.

Fundamentos de información y documentación. López Yepes, José (comp.). Madrid: EUDEMA, 1989. 485 p.

Gabriel, Pere; et al. *Ictineu: diccionari de les ciències de la societat dels Països Catalans (segles XVIII-XX)*. Barcelona: Edicions 62, 1979. (Cultura Catalana Contemporània, 8).

García Gutiérrez, Antonio Luis; Lucas Fernández, Ricardo. *Documentación automatizada en los medios informativos*. Madrid: Paraninfo, 1987.

Gran enciclopèdia catalana. 2a. ed. Barcelona: Enciclopèdia Catalana, 1986 (reimp., 1988). 21 vol.

Gran enciclopedia Larousse. 3ª ed., Barcelona: Planeta, 1977. 10 vol.

Guinchat, Claire; Menou, Michel. *Introducción general a las ciencias y técnicas de la información y documentación*. Madrid: CINDOC: Unesco, 1992.

Institut d'Estadística de Catalunya. *Thesaurus de termes estadístics*. Barcelona: Generalitat de Catalunya, Institut d'Estadística de Catalunya, 1992.

Instituto de Información y Documentación en Ciencia y Tecnología (ICYT-CSIC); Unesco. *Tesaurus Spines: un vocabulario controlado y estructurado para el tratamiento de información sobre ciencia y tecnología para el desarrollo*. Ed. española de la versión revisada de 1984. Madrid: Unesco; ICYT, 1988.

Laureilhe, Marie-Thérèse. *Le Thesaurus: son role: sa structure: son elaboration*. 2e éd. Villeurbanne: Presses de l'ENSB, 1981.

Lizosoain Hernández, Luis. *El sistema de documentación en Ciencias de la Educación: elaboración de un tesaurus de investigación educativa*. Madrid: Universidad Complutense de Madrid, 1988. (Col. tesis doctorales, 43/88). Tesis doctoral dir. per Arturo de la Orden Hoz, Madrid, 1987.

Maniez, J. *Los lenguajes documentales y de clasificación: concepción, construcción y utili-*

- zación de los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez, 1992.
- Mestre i Campi, Jesús (dir.). *Diccionari d'història de Catalunya*. Barcelona: Edicions 62, 1992.
- Nomenclàtor Internacional Normalitzat en els camps de la Ciència i de la Tecnologia. (segons projecte UNESCO): anglès, català, castellà*. Barcelona: Generalitat de Catalunya, Departament d'Ensenyament, 1982.
- Palladó Comas, M^a Flores. *Tesaurus de Medio Ambiente*. Madrid: MOPU, Secretaria General de Medio Ambiente, 1990.
- Population Information Network (POPIN). *Tesaurus de POPIN: tesauro multilingüe sobre población*. 2^a ed. española. New York; Paris: Population Information Network: Comité Internacional de Cooperación de Investigaciones Nacionales sobre Demografía (CICRED): Fondo de las Naciones Unidas para las Actividades en Materia de Población (FNUAP), 1984.
- Sebastià i Salat, Montserrat. *Thesaurus d'història social de la dona*. Barcelona: Generalitat de Catalunya, Departament de la Presidència, Comissió Interdepartamental de Promoció de la Dona, 1988.
- Van Slype, Georges. *Les langages d'indexation: conception, construction et utilisation dans les systemes documentaires*. (Versió espanyola: Madrid: Piràmide, 1991.)
- Viet, Jean. *Thesaurus for information processing in Sociology: pour le traitement de l'information en Sociology/Sociologie*. Paris : The Hague, 1971.