

Experiències

L'observatori de terminologia Talaia: un exemple d'innovació per mitjà de la cooperació

Autors

Dolors Montes
TERMCAT

Mercè Vázquez
Universitat Oberta de Catalunya

En aquest article es fa una presentació detallada de l'observatori de terminologia de la societat del coneixement Talaia, fruit de la col·laboració entre la Universitat Oberta de Catalunya i el Centre de Terminologia TERMCAT. S'hi expliquen els objectius del projecte i s'hi descriu detalladament el procés de treball semiautomatitzat que ha donat lloc a l'observatori i que permetrà d'actualitzar-lo periòdicament.

Descripció general i objectius de l'observatori Talaia

Recentment, després de mesos de treball, ha vist la llum el projecte Talaia, un observatori de terminologia de la societat del coneixement fruit d'un acord de col·laboració entre la Universitat Oberta de Catalunya (UOC) i el Centre de Terminologia TERMCAT. L'observatori, que és consultable al web del TERMCAT (<http://www.termcat.cat>), recull la terminologia en català, castellà i anglès utilitzada en els articles de les revistes acadèmiques de la UOC, totes publicades en format electrònic. Aquestes revistes estudien, en general, els efectes i la influència de l'ús de les tecnologies de la informació i la comunicació (TIC) en les persones, les organitzacions i la societat i, com a publicacions d'especialitat, són una font importantíssima de terminologia i neologia.

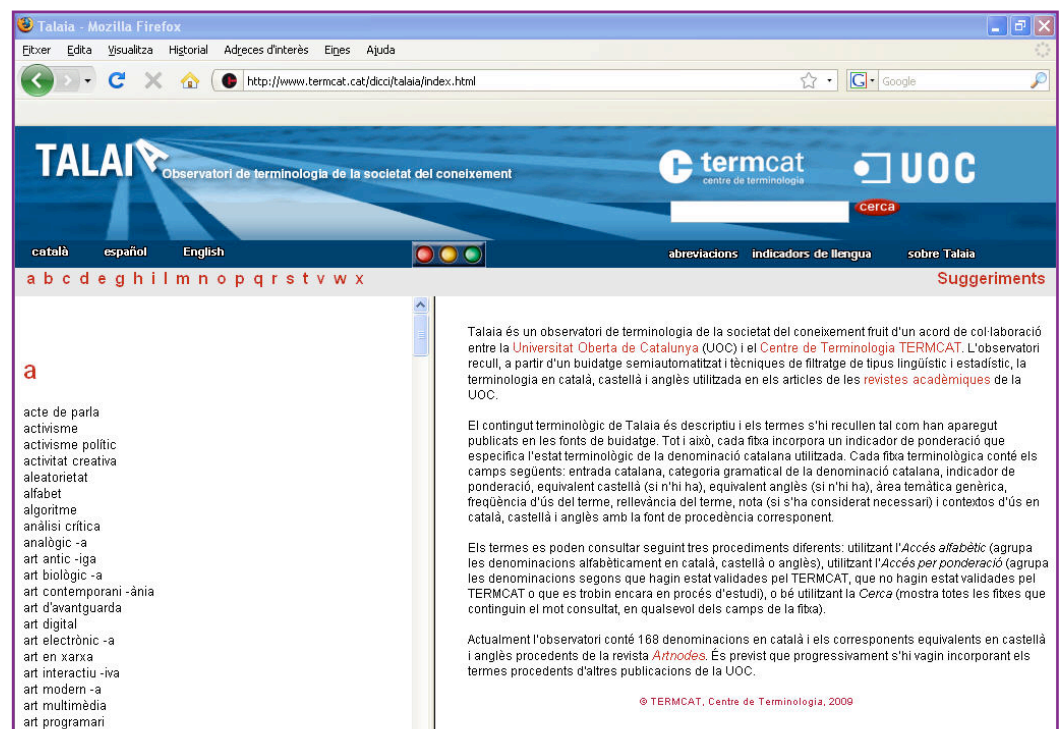
Una de les característiques més destacables i innovadores de Talaia, a banda estrictament de la terminologia que conté, és que és el resultat d'un treball semiautomatitzat, en què el buidatge de la terminologia es duu a terme per mitjà d'eines automàtiques d'extracció de dades textuais de codi lliure desenvolupades a la Unitat de Tecnologies Lingüístiques del Servei Lingüístic¹ de la mateixa UOC.

L'objectiu de Talaia és, en línies generals, posar a disposició pública terminologia neològica de qualitat en català, castellà i anglès, a fi de facilitar les comunicacions especialitzades i afavorir el treball multilingüe i l'elaboració d'estudis sobre neologia emergent o sobre formes que tenen un ús especialment rellevant en la comunitat d'especialistes. El fet que sigui un producte sorgit de la voluntat i la il·lusió de dues entitats diferents, amb missions distintes, persones amb capacitats i coneixements distints i públic objectiu només parcialment coincident, ha contribuït notablement —a partir d'un exercici de complementarietat i convergència d'interessos— a eixamplar els objectius per als quals ha estat concebut. Així, Talaia vol, entre altres coses:

1. <http://www.uoc.edu/serveilinguistic/home/index.html>

- Afavorir el treball multilingüe en l'espai europeu d'ensenyament superior.
- Aplegar un fons terminològic actual procedent de fonts especialitzades reglades (les revistes acadèmiques de la Universitat Oberta de Catalunya).
- Reaprofitar els esforços invertits en els processos editorials que hi ha al darrere de les revistes acadèmiques de la UOC i, d'aquesta manera, millorar la qualitat i l'eficiència d'aquesta feina.
- Difondre terminologia que pugui ser útil als professionals de la llengua en un sentit ampli.
- Facilitar la detecció precoç de neologismes dels sectors més innovadors en el seu context d'ús real.
- Facilitar l'observació de possibles canvis d'ús i de freqüència d'aparició dels termes.
- Afavorir els estudis sobre neologia.
- Explorar i valorar nous sistemes de buidatge terminològic basats en l'extracció automàtica de terminologia.
- Ajudar en l'anàlisi de la implantació dels termes normalitzats.

Imatge 1: Pàgina inicial de l'observatori Talaia



Fonts de procedència de la terminologia

La terminologia que es difon a través de l'observatori Talaia prové de les revistes acadèmiques digitals² de la Universitat Oberta de Catalunya: *Artnodes*, *IDP Revista d'Internet, Dret i Política*, *Digithum*, *UOC Papers*, *RUSC* i *Mosaic*. Aquestes sis revistes, com ja s'ha apuntat més amunt, estudien els efectes i les influències de l'ús de les TIC en les persones i la societat des d'àmbits de coneixement diferents: l'art, la ciència i la tecnologia, en el cas d'*Artnodes*; el dret i la ciència política, a *IDP*; les humanitats i les ciències socials, a *Digithum*; la societat del coneixement, a *UOC Papers*; l'ensenyament superior, a *RUSC*, i la comunicació multimèdia, a *Mosaic*.

Imatge 2: Pàgina inicial de les revistes acadèmiques de la UOC

The screenshot shows the 'Revistes digitals' page on the UOC website. At the top, there's a navigation bar with the UOC logo and several tabs: 'La Universitat', 'Estudia a la UOC', 'Recerca i innovació', and 'Difusió i publicacions'. Below this, the page title 'Revistes digitals' is displayed, along with 'Imprimir' and 'Recomanar' options. A sidebar on the right lists the journals: Artnodes, Digithum, IDP, Mosaic, RUSC, and UOC Papers. The main content area is divided into two columns, each featuring a journal's logo, title, and a short description of its scope. For example, 'Artnodes' is described as 'Espai coordinat pels Estudis d'Informàtica, Multimèdia i Telecomunicació...', while 'IDP' is 'Publicació dels Estudis de Dret i Ciència Política...'. Each entry also includes an RSS feed icon.

2. Les revistes acadèmiques digitals que publica la Universitat Oberta de Catalunya es poden consultar a l'adreça següent: http://www.uoc.edu/portal/catala/difusio_i_publicacions/revistes_digitals/index.html.

Les fonts de Talaia són un corpus reglat i multilingüe d'un àmbit de coneixement capdavanter, el de la societat del coneixement, nou, divers, canviant i que genera un gran nombre de denominacions especialitzades. Tots els articles són d'autor, passen per un procés de revisió d'experts i segueixen un procés editorial estàndard. De tots els articles de les revistes, el títol, el resum i les paraules clau es publiquen en català, castellà i anglès i el cos de l'article es publica en dos d'aquests idiomes. A més, aquestes sis revistes segueixen els processos i els formats estandarditzats de les revistes científiques: estan seriades (disposen d'ISSN), es publiquen periòdicament, tenen una línia editorial i unes normes de publicació, disposen d'un director, un consell de redacció i un consell assessor, i estan indexades en bases de dades de referència. Totes aquestes característiques fan que siguin un corpus òptim per al buidatge de terminologia.

En aquests moments (març del 2009), la terminologia recollida a l'observatori (prop de 200 termes) prové íntegrament de la revista *Artnodes*, de la qual s'han buidat els sis primers números. Progressivament, es preveu d'anar-hi incorporant els termes procedents del buidatge de la resta de publicacions. La pròxima revista que es constituirà com a corpus de buidatge és la *Revista d'Internet, Dret i Política*, a partir de la qual es recuperaran termes relacionats amb les tecnologies de la informació i la comunicació i els diferents camps del dret, la política i l'Administració pública.

Procés d'extracció terminològica

Fem	fer	VMIP1S0	
un	un	DIOMSO	
repàs	repàs	NCMS000	
del	del	SPCMS	
passat	passat	AQOMSP	
		Fc	
él	él	DAOMSO	
present	present	AQOCS0	
i	i	CC	
el	el	DAOMSO	
futur	futur	NCMS000	
de	de	SPS00	
l'	el	DAOCS0	
art	art	NCCS000	
digital	digital	AQOCS0	
		Fc	
i	i	CC	
més	més	RG	
específicament	específicament		RG
del	del	SPCMS	
programari	programari		NCMS000
i	i	CC	
el	el	DAOMSO	
codi	codi	NCMS000	
artístics	artístic		AQOMPO
		Fc	

El procés de buidatge de les revistes acadèmiques es duu a terme de manera automatitzada i se centra en les tasques de compilació del corpus, alineació i anàlisi lingüística dels continguts i extracció automàtica dels candidats a terme. Al final del procés es fa una revisió manual dels resultats.

Primerament, el contingut de les sis publicacions és compilat separatament per revista i classificat per llengua per a poder disposar del corpus monolingüe en català, espanyol i anglès de cada revista. Aquest corpus monolingüe serveix de base per a poder preparar corpus bilingües català-anglès i català-espanyol a partir d'un procés automàtic d'alineació de frases. Aquests corpus bilingües tenen com a objectiu localitzar els equivalents de traducció i els contextos d'ús de cada terme. Cal dir que els corpus bilingües es preparen tenint en compte que el català és la llengua de treball a partir de la qual es localitzen els equivalents de traducció i els contextos en les altres llengües.

Paral·lelament a la constitució dels corpus bilingües es prepara l'anàlisi lingüística dels corpus monolingües per a po-

der identificar de cada candidat a terme el lema, la categoria gramatical que correspon a aquest lema i les formes que hi estan associades, és a dir, es treballa amb uns triplets que contenen la informació següent: <forma d'una paraula, lema de la paraula, categoria gramatical de la paraula>. La tasca de crear aquests triplets es fa amb l'etiquetador de codi lliure i de lliure distribució FreeLing,³ el qual permet treballar amb corpus de gran volum i crear una llista de resultats de

Imatge 3: Mostra de resultats del FreeLing

3. <http://www.lsi.upc.edu/~nlp/freeling/>

manera automàtica. Amb tot, cal tenir en compte que aquesta eina pot introduir errors en els resultats finals fent una assignació incorrecta de lema (taxa d'error inferior al 5 %) i una assignació incorrecta de categoria gramatical (taxa d'error inferior al 2 %).

El fet de treballar amb lemes i no amb formes fa possible tenir una llista més breu de candidats a terme, ja que les diferents variants d'una paraula o combinacions de paraules no apareixen per separat, sinó que queden agrupades sota un mateix lema, cosa que permet fer una revisió homogènia dels resultats, i, alhora, disposar d'una dada única de freqüència d'aparició de cada lema i les formes corresponents en el corpus, amb la suma de les freqüències de les formes variants. Treballar a partir de lemes també té l'avantatge que es recullen per separat les freqüències de formes homògrafes que tenen categories gramaticals i sentits diferents i que, per tant, corresponen a termes diferents. Seria el cas, per exemple, d'*art net* i de *net art*: *net* és un adjectiu a *art net*, i probablement no seria escollit com a terme. En canvi, a *net art* és un substantiu i sí que s'acceptaria.

El resultat que s'obté en haver fet l'anàlisi lingüística dels corpus monolingües serveix per a preparar l'extracció automàtica de candidats a terme. Com que és molt important extreure uns candidats que siguin del tot rellevants en l'àmbit de la societat del coneixement, en el procés d'extracció automàtica s'apliquen diferents filtres a les dades:

- Filtre per longitud, per a recuperar solament els candidats a terme que tinguin una determinada longitud (per exemple, candidats de tres elements com ara *processador de textos*).
- Filtre per llongitud de freqüència, per a triar sols els candidats a terme que apareguin en el corpus un determinat nombre de vegades.
- Filtre de paraules buides en posicions extremes, per a evitar que la paraula inicial o final d'un candidat a terme sigui una paraula buida de contingut (article, pronom, preposició, conjunció, etc.). Així, aquest filtre ha permès recuperar com a termes candidats del tipus el *suport material* o la *realitat virtual*.

En aquest sentit, el fet de treballar amb la informació lingüística que hi ha associada a cada candidat a terme ha permès completar la llista d'aquests tres filtres inicials amb uns filtres basats en criteris més acurats:

- Filtre de candidats que tenen una categoria gramatical que no pot ser terme, per a no haver de revisar una llista de candidats que, des d'un punt de vista morfosintàctic, no poden ser termes (per exemple, es desestimen candidats com ara *programant pots crear*).
- Filtre de candidats que depenen d'un candidat més llarg que apareix al corpus amb la mateixa freqüència, tenint en compte que la freqüència idèntica és deguda al fet que el candidat més curt depèn del candidat més llarg (així, si *activisme* i *activisme polític* apareguessin amb la mateixa freqüència, es desestimaria *activisme* i es recolliria només *activisme polític*).
- Filtre de candidats del vocabulari general, que es basa en el càlcul estadístic de la rellevància de cada candidat del corpus d'especialitat en contrast amb un corpus de llengua general de diferents temàtiques i considera com a candidats a terme els candidats que són rellevants en el corpus d'especialitat però que no ho són en el de llengua general (per exemple, a partir d'aquest filtre s'han desestimat formes com ara *informàtica*, que té un ús molt generalitzat, i, en canvi, se n'han acceptat d'altres, com ara *intel·ligència artificial* o *tecnologia digital*, més especialitzades).

- Filtre de candidats amb noms propis, amb l'objectiu d'eliminar de la llista de candidats noms amb cognoms o sense, que són molt freqüents en el corpus d'especialitat de què es parteix (així, aquest filtre ha permès eliminar noms propis com ara *Pau Waelder*, *Gerfried Stocker* o *Andreas Broeckmann*).

La llista de resultats obtinguda al final del procés de filtratge de les dades és revisada per un terminòleg amb l'objectiu de tenir una llista final de termes en català a partir de la qual poder fer una cerca dels equivalents de traducció i dels contextos d'ús en els corpus paral·lels.

Així, doncs, la llista de termes propis de la societat del coneixement extreta del corpus de les revistes acadèmiques, els equivalents de traducció i els contextos d'ús passen a formar part d'un fitxer XML que és completat de manera automàtica amb la informació de la categoria gramatical del lema, les dades de freqüència i rellevància del lema, les formes que corresponen a cada lema i la font bibliogràfica on es pot documentar cada lema. La informació que conté aquest fitxer és la base de cada fitxa de l'observatori Talaia.

Procés de revisió terminològica

Abans de difondre's públicament a través de l'observatori en línia, les dades dels fitxers en XML, procedents del buidatge automatitzat i de la posterior revisió manual, es migren al Gestor de Terminologia del TERMCAT (GdT), a fi que puguin seguir encara un procés de depuració i de revisió terminològica.

La revisió de les dades, que duen a terme terminòlegs del Centre de Terminologia, es fa tenint en compte els contextos d'aparició de cada terme i se centra en diferents punts. En primer lloc, s'analitza i es valida la proposta de lema català, tant des del punt de vista de la segmentació (per exemple, que no s'hagi extret erròniament com a candidat a terme *art tradicional en comptes d'obra d'art tradicional*) com de la pertinència dels termes (noms propis no filtrats, per exemple), l'homonímia (homònims no separats com a termes diferents: per exemple, *multimèdia substantiu i multimèdia adjectiu*) o l'assignació exacta de lema i de la marca de flexió corresponent (que el lema proposat no sigui, per exemple, *efectes especials* en comptes de la forma en singular, *efecte especial*, o *art antic* en comptes d'*art antic -iga*). En segon lloc, s'assigna la categoria gramatical als lemes catalans i una àrea temàtica genèrica per a cada terme (l'assignació d'àrea es fa segons la revista de procedència del terme, tenint en compte els àmbits d'especialitat a què s'adscriu cadascuna de les publicacions). A continuació, es revisen els equivalents i s'estableixen els lemes castellans i anglesos a partir de les ocurrències extretes automàticament dels contextos. Finalment, s'analitza l'adequació de les denominacions catalanes des del punt de vista lingüístic i terminològic i, tenint en compte això, s'assigna una marca de ponderació específica a cadascuna de les entrades. Les marques de ponderació es complementen, quan es considera oportú o necessari, amb la redacció de notes explicatives sobre els termes, ja siguin de difusió pública o bé de treball.

Cal tenir present, de tota manera, que Talaia té una voluntat bàsicament descriptiva, no prescriptiva, i és per aquesta raó que els terminòlegs no introdueixen modificacions en la informació extreta del corpus de buidatge, sinó que bàsicament es limiten a completar-la des del punt de vista terminològic i, en tot cas, a valorar-ne l'adequació per mitjà de l'addició de notes o d'altres marques. És per això mateix que les fitxes terminològiques finals que apareixen a l'observatori mostren els termes en context, tal com han estat utilitzats a les fonts de què provenen. Una intervenció més aprofundida sobre els termes falsejaria les dades, dificultaria el procés periòdic d'actualització automatitzada i, sobretot, deixaria sense efecte alguns dels objectius fonamentals del projecte, especialment la voluntat de donar accés a dades reals que facilitin els estudis sobre neologia.

Pel que fa a la ponderació dels termes, que només es fa sobre les formes catalanes, s'estableixen diferents marques, que es visualitzen públicament en forma de tres grans grups: un grup de termes adequats o validats pel TERMCAT, que inclou termes normatius, normalitzats o sense cap problema lingüístic o terminològic destacable; un segon grup de termes conflictius, no validats pel TERMCAT, on s'agrupen casos terminològicament inconsistents o no plenament concordants amb les formes normatives o normalitzades, i un tercer grup de termes en estudi o pendents de ponderació definitiva.

La revisió terminològica, a banda de completar i millorar l'estat de les fitxes per a la seva difusió pública, també actua com a control de qualitat dels processos de treball precedents, i especialment dels que s'han dut a terme de manera automatitzada, i contribueix a perfeccionar-los. Com a millores futures, es preveu d'optimitzar el procés d'assignació automatitzada de categoria gramatical dels termes catalans, que actualment no funciona de manera adequada en els casos de denominacions sintagmàtiques (la màquina assigna una categoria a cadascun dels components del sintagma i no al sintagma globalment), i també introduir la lematització automàtica per als equivalents castellans i anglesos i la categoria gramatical d'aquestes denominacions.

Informació pública de Talaia

Talaia aplega terminologia de l'àmbit de la societat del coneixement en català, castellà i anglès. Cada fitxa terminològica conté els camps següents: entrada catalana, categoria gramatical de la denominació catalana, indicador de ponderació de la forma catalana, equivalent castellà (si n'hi ha), equivalent anglès (si n'hi ha), àrea temàtica genèrica, freqüència d'ús del terme, rellevància del terme, nota (si s'ha considerat necessari) i contextos d'ús en català, castellà i anglès amb la font de procedència corresponent.

Imatge 4: Exemple de fitxa de l'observatori Talaia

The screenshot shows the Talaia website interface in Mozilla Firefox. The page displays a list of terms under the letter 'C'. The term 'ciberespai' is highlighted, and a detailed entry is shown for it. The entry includes the following fields and annotations:

- denominació catalana:** ciberespai m
- equivalent castellà:** es ciberespacio
- equivalent anglès:** en cyberspace
- categoria gramatical:** -Art, ciència i tecnologia-
- àrea temàtica genèrica:** -Art, ciència i tecnologia-
- indicador de ponderació:** A traffic light icon showing red, yellow, and green lights.
- altres informacions:**
 - freqüència: 11
 - rellevància: 1.42
 - Notes: Terme normatiu.
 - Contextos: [Font bibliogràfica]

The entry also includes a list of related terms on the left: cibercepció, ciberespai, cibemètica, cibrí, ciències socials, cinema digital, cinemàtic -a, codi informàtic, codificar, collage, comunicació de masses, comunicació visual, comunitat virtual, connectivitat, corpografia, corporalitat, creació artística, creació col·lectiva.

Els indicadors de ponderació tenen una relació directa amb les marques de ponderació de què s'ha parlat en l'apartat anterior i tenen com a objectiu oferir als consultants una valoració general de l'adequació de les denominacions catalanes recollides, des del punt de vista lingüístic i terminològic. S'hi empren tres indicadors diferents: un indicador de color verd, que s'aplica als termes considerats adequats; un indicador de color vermell, que s'aplica als termes no validats pel TERMCAT, i un indicador de color taronja, que s'aplica als termes en procés d'estudi.

Les notes de difusió pública consignades a les fitxes dels termes poden ser de tipus lingüístic (sobre la formació del terme, sobre la sinonímia amb un altre terme, sobre la remissió al corresponent terme normatiu o normalitzat, sobre els equivalents, etc.) o bé de tipus conceptual.

Pel que fa als paràmetres de freqüència i rellevància, la freqüència mesura els cops que un terme apareix al corpus d'especialitat buidat i permet veure, doncs, quins són els termes més i menys usats. S'extreuen com a candidats els termes que tenen una freqüència mínima d'aparició al corpus de dues vegades. La rellevància, d'altra banda, intenta mesurar, a partir d'un corpus de la llengua general d'exclusió, i per mitjà d'unes fórmules matemàtiques establertes, fins a quin punt una forma lingüística determinada pot ser considerada un terme. Actualment, el contrast es fa amb un corpus actualitzat del diari *Avui*, per seccions temàtiques del diari. La rellevància segueix una gradació de 0 a 10. Com menys es documenta una determinada forma al corpus de contrast, més alta és la rellevància, és a dir, més possibilitats hi ha que efectivament es tracti d'una denominació d'especialitat.

És previst que a mesura que el corpus es vagi ampliant amb el buidatge de noves revistes s'incorporin a l'observatori uns gràfics amb l'historial de freqüència i de rellevància de cada terme. Aquests gràfics haurien de permetre veure l'evolució de l'ús i del grau d'especialització de cada forma.

Els contextos s'encapçalen sempre amb la font bibliogràfica de procedència, la qual remet a l'autor, l'article i la revista d'origen (amb l'URL corresponent). A més, s'agrupen en blocs de traducció, de manera que a sota d'un determinat context català es fa constar el mateix context traduït al castellà i a l'anglès (sempre que hi siguin a les fonts de buidatge), cadascun amb la seva font bibliogràfica. Això és possible gràcies als corpus bilingües (català-castellà i català-anglès) que es constitueixen en les fases inicials del procés per a poder localitzar els equivalents de traducció. Cal tenir present, de tota manera, que no sempre hi ha paral·lelisme en castellà i anglès dels contextos catalans, ja que, com s'ha dit més amunt, si bé els títols dels articles, els resums i les paraules clau sí que es publiquen en les tres llengües de treball, el cos de l'article pot ser que es publiqui només en dues d'aquestes llengües.

També en relació amb les fitxes terminològiques que es presenten a Talaia, cal tenir en compte que el lematitzador utilitzat en la confecció de la llista dels candidats a terme no incorpora desambiguador semàntic, per la qual cosa els homògrafs que comparteixen categoria gramatical no es distingeixen. Excepcionalment, doncs, aquestes formes consten a l'observatori en una mateixa fitxa terminològica, per bé que són termes diferents i des del punt de vista terminogràfic haurien de tenir fitxes diferenciades (per exemple, el substantiu *seqüenciació* en genètica i en audiovisuals). Els homògrafs de categoria gramatical diferent, en canvi, consten en fitxes diferents, si bé també hi pot haver un petit marge d'error a l'hora de fer-se efectiva aquesta distinció.

També excepcionalment i per motius d'automatització, els sinònims consten en fitxes terminològiques diferents (per exemple, *disseny gràfic* i *grafisme*). La referència a la remissió s'especifica sempre, però, al camp de nota.

L'observatori permet llistar totes les denominacions catalanes, castellanes i angleses, fer cerques per denominació en qualsevol de les tres llengües de treball (amb uns resultats que comprenen qualsevol camp de la fitxa) i llistar els termes per qualsevol dels tres paràmetres de ponderació. Properament, quan s'hi incorporin termes d'altres revistes, també es podran llistar els termes (en català, castellà o anglès) segons la revista de procedència.

Futures actualitzacions i primeres anàlisis de resultats

Talaia no és un producte estàtic. La idea, com ja s'ha apuntat, és que l'observatori es renovi periòdicament amb les dades terminològiques obtingudes d'altres revistes acadèmiques de la UOC. Els nous buidatges serviran per a augmentar el cabal terminològic de l'observatori, però també per a actualitzar la terminologia ja existent, fonamentalment en termes de freqüència, de rellevància i de contextos d'ús.

Dels continguts dels sis números de la revista *Artnodes* ja buidats i incorporats a Talaia, que s'han traduït en 168 denominacions catalanes, es poden començar a extreure algunes conclusions interessants, que caldrà anar perfilant i ampliant a mesura que el recull terminològic vagi creixent:

- En primer lloc —i tal com era de preveure, per tal com les revistes de la UOC segueixen un rigorós procés d'edició—, destaca l'adequació lingüística de la majoria de denominacions utilitzades: només un 2 % dels casos entren en contradicció amb algun terme ja normatiu o normalitzat, i la majoria dels casos que hi entren poden ser motiu, en qualsevol cas, de reflexió i anàlisi (per exemple, *mitjà de comunicació de masses*, en comptes de la forma normativa *mitjà de comunicació de massa*, o *hipervincle* en comptes d'*hiperenllaç*).
- També la presència majoritària de termes sintagmàtics (56 % del total), molts dels quals són neològics (*art electrònic -a*, *art en xarxa*, *art interactiu -iva*, *artista programador -a*, *comunitat virtual*, *estètica digital*, *ordinador digital*, *teoria de sistemes*, etc.).
- La presència també de nombrosos termes d'autor (si més no, aparentment), sovint formats per acronímia (*cibercepció*, *hacktivisme*, *smart mob*, *corpografia*, etc.).
- L'existència d'un percentatge interessant de termes (aproximadament un 8 % del total) susceptibles de seguir un estudi lingüístic i terminològic més aprofundit o, fins i tot, un procés de normalització, si els futurs buidatges en confirmen una freqüència d'ús i una rellevància prou elevades (entre d'altres, *media*, *software art*, *teoria queer* o *wiki*).

Conclusions

Talaia és un projecte de cooperació institucional creat amb la voluntat de millorar la qualitat de les comunicacions especialitzades i la gestió lingüística per mitjà de la terminologia i l'ús d'eines automàtiques de tractament de la informació. Els resultats es posen a disposició lliure a través d'Internet perquè la comunitat acadèmica i d'especialistes, i la societat en general, se'n puguin beneficiar.

Bibliografia

COLOMER, R.; VÁZQUEZ, M. (2008). «Assurer la qualité par la terminologie: le projet Talaia». Comunicació. A: *La Terminologie dans les Applications Avancées de Gestion (TAMA)*. Ottawa-Gatineau: Universitat del Quebec a Outaouais.

MORÉ, J.; RIUS, L.; VÁZQUEZ, M.; VILLAREJO, L. (2008). «L'observatori de terminologia Talaia: mètode i processos». A: *Revista Tradumàtica. Terminologia i Traducció*. Núm. 6.

<<http://www.fti.uab.es/tradumatica/revista/num6/articles/06/06art.htm>>

ISSN: 1578-7559

RIUS, L.; SÁNCHEZ, I.; VÁZQUEZ, M. (2007). «Projecte Talaia: cap a un observatori de terminologia de la societat del coneixement». A: *Nous reptes dels professionals en la comunicació especialitzada: actes de la II Jornada de Terminologia i Serveis Lingüístics*. Palma: Universitat de les Illes Balears. Servei Lingüístic. P. 69-77.

ISBN 978-84-8384-026-9