

## Annex 6: DESCRIPCIÓ DE LES DADES

### Índex

<b>LLISTA DE TAULES</b> .....	<b>3</b>
<b>1. DESCRIPCIÓ DELS DICCIONARIS NORMATIUS</b> .....	<b>5</b>
1.1. DICCIONARIS NORMATIUS DE LLENGUA CATALANA.....	6
1.1.1. DIEC1.....	6
1.1.1.1. Recursos d'actualització lèxica.....	6
1.1.1.2. Manlleus .....	7
1.1.1.3. Marca d'ús .....	8
1.1.1.4. Marca temàtica .....	9
1.1.1.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic.....	11
1.1.1.6. Contrast dels lemes dels diccionaris amb els corpus textuals .....	13
1.1.2. DIEC2.....	17
1.1.2.1. Recursos d'actualització lèxica.....	18
1.1.2.2. Manlleus .....	18
1.1.2.3. Marca d'ús .....	19
1.1.2.4. Marca temàtica .....	20
1.1.2.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic.....	21
1.1.2.6. Contrast dels lemes dels diccionaris amb els corpus textuals .....	23
1.2. DICCIONARIS NORMATIUS DE LLENGUA CASTELLANA .....	26
1.2.1. DRAE21 .....	26
1.2.1.1. Recursos d'actualització lèxica.....	27
1.2.1.2. Manlleus .....	27
1.2.1.3. Marca d'ús .....	28
1.2.1.4. Marca temàtica .....	29
1.2.1.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic.....	29
1.2.1.6. Contrast dels lemes dels diccionaris amb els corpus textuals .....	31
1.2.2. DRAE22 .....	35

1.2.2.1. Recursos d'actualització lèxica.....	36
1.2.2.2. Manlleus .....	36
1.2.2.3. Marca d'ús .....	37
1.2.2.4. Marca temàtica .....	38
1.2.2.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic.....	38
1.2.2.6. Contrast dels lemes dels diccionaris amb els corpus textuals .....	40
<b>2. DESCRIPCIÓ DELS DICCIONARIS D'ÚS .....</b>	<b>44</b>
2.1. GDLC .....	44
2.1.1. Recursos d'actualització lèxica .....	44
2.1.2. Manlleus .....	45
2.1.3. Marca d'ús.....	46
2.1.4. Marca temàtica.....	46
2.1.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic.....	47
2.1.6. Contrast dels lemes dels diccionaris amb els corpus textuals .....	49
2.2. DEA.....	54
2.2.1. Recursos d'actualització lèxica .....	54
2.2.2. Manlleus .....	54
2.2.3. Marca d'ús.....	55
2.2.4. Marca temàtica.....	55
2.2.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic.....	56
2.2.6. Contrast dels lemes dels diccionaris amb els corpus textuals .....	57

## Llista de taules

Taula 1: Contrast dels lemes dels diccionaris amb el corpus lexicogràfic del creuament DIEC1 .....	11
Taula 2: Representativitat dels lemes coincidents i no coincidents en el contrast amb corpus lexicogràfic (DIEC1) .....	11
Taula 3: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans (DIEC1) .....	13
Taula 4: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DIEC1) .....	15
Taula 5: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DIEC1) .....	17
Taula 6: Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DIEC2).....	21
Taula 7: Representativitat dels lemes coincidents i no coincidents en el contrast amb corpus lexicogràfic (DIEC2) .....	22
Taula 8: Representativitat dels lemes coincidents i no coincidents respecte el total de lemes en el contrast amb el Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans (DIEC2).....	23
Taula 9: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DIEC2) .....	24
Taula 10: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DIEC2).....	25
Taula 11: Contrast entre lemes amb corpus lexicogràfic (DRAE21)...	30
Taula 12: Representativitat dels lemes coincidents i no coincidents respecte el total de lemes (DRAE21) .....	30
Taula 13: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus de Referència del Español Actual de la Real Academia Española (DRAE21).....	32
Taula 14: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DRAE21) .....	33
Taula 15: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DRAE21) .....	34
Taula 16: Representativitat de les marques d'ús (DRAE22).....	37
Taula 17: Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DRAE22).....	38

Taula 18: Representativitat dels lemes coincidents i no coincidents en el contrast amb corpus lexicogràfic (DRAE22) .....	39
Taula 19: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus de Referencia del Español Actual de la Real Academia Española (DRAE22) .....	40
Taula 20: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DRAE22) .....	41
Taula 21: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DRAE22) .....	43
Taula 22: Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (GDLC) .....	47
Taula 23: Representativitat dels lemes coincidents i no coincidents respecte el total de lemes (GDLC) .....	48
Taula 24: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans (GDLC) .....	50
Taula 25: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (GDLC) .....	51
Taula 26: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (GDLC) .....	52
Taula 27: Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DEA) .....	56
Taula 28: Representativitat dels lemes coincidents i no coincidents respecte el total de lemes (DEA) .....	56
Taula 29: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus de Referencia del Español Actual de la Real Academia Española (DEA) .....	58
Taula 30: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DEA) .....	59
Taula 31: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DEA) .....	61

## 1. Descripció dels diccionaris normatius

Com expliquem en el capítol III sobre la metodologia de treball, comparem un conjunt de lemaris de diferents diccionaris que ens han proporcionat uns lemes i uns articles lexicogràfics.<sup>1</sup> A partir d'aquests lemes i dels seus articles, fem una primera descripció, segons diversos criteris: el de la innovació lèxica, el de la permeabilitat a la interferència, el de l'orientació dels usos i el de la coincidència en relació amb l'ús.

En la descripció de les dades segons cada criteri hem tingut en compte diverses variables:

- Innovació lèxica: nombre de lemes nous.
- Permeabilitat a la interferència: recurs d'actualització lèxica i manlleus.
- Orientació dels usos: marca temàtica i marca d'ús.
- Coincidència en relació amb l'ús: contrast dels lemes dels diccionaris amb corpus lexicogràfics i contrast dels lemes dels diccionaris amb corpus textuais.

Organitzem la presentació de les dades en dos grans apartats. En el primer, descrivim els resultats dels diccionaris normatius de llengua catalana i, en el segon, els resultats dels diccionaris normatius de llengua castellana.

Els creuaments que descriurem en llengua catalana són, en primer lloc, el DGLC-DIEC1 i, en segon lloc, el DIEC1-DIEC2. En llengua castellana, primer ens ocuparem del creuament DRAE16-DRAE21 i, posteriorment, del DRAE21-DRAE22.

Per fer més fàcil la interpretació de les dades, en els casos en què un mateix lema tingui més d'una etiqueta del mateix tipus en els apartats de recursos d'actualització lèxica, de marques d'ús i de marques temàtiques, comptabilitzem per separat aquestes marques. Per exemple, si un lema té com a recurs d'actualització lèxica les etiquetes de *manlleu* i *sufixació*, comptabilitzem el lema dues vegades, una per a cada categoria. Això significa que, en el recompte de cada una d'aquestes variables d'estudi, el nombre de dades recollides pot ser superior al nombre de lemes resultants de cada creuament.

En les anàlisis de contrast dels lemes amb corpus, tant lexicogràfics com textuais i, en canvi, no comptabilitzem per separat les etiquetes repetides d'un mateix lema ja que, com que no recuperem els

---

<sup>1</sup> Consulteu la *Base de dades de creuaments* (en CD-ROM).

contextos, i per tant no els analitzem, no podem atribuir una marca o una altra als lemes que en presenten més d'una. Així, en ocasions presentem les diferents marques com una seqüència de diverses marques.

## 1.1. Dictionaris normatius de llengua catalana

Duem a terme dos creuaments diferents amb els dictionaris de llengua catalana. En el primer, comparem els lemaris del DGLC i del DIEC1. En el segon, treballem amb el DIEC1 i el DIEC2.

### 1.1.1. DIEC1

En el creuament dels dictionaris DGLC i DIEC1 comparem les lletres *b* i *l* dels lemaris dels dos dictionaris per observar quins lemes nous hi ha al DIEC1 respecte del DGLC. Així, detectem 1.276 lemes, és a dir, lemes nous que figuren en el lemar del DIEC1 i no en el del DGLC. Més detalladament, detectem 781 lemes procedents de la lletra *b* i 495 de la lletra *l*. A continuació descrivim les dades d'aquest creuament segons els criteris exposats anteriorment.

#### 1.1.1.1. Recursos d'actualització lèxica

A partir de la classificació proposada en el marc del projecte *Fonaments, estratègies i eines per al processament, extracció i representació d'informació especialitzada (TEXTERM3)*<sup>2</sup>, classifiquem els lemes segons el seu recurs d'actualització lèxica.<sup>3</sup>

Detectem 22 recursos d'actualització lèxica diferents entre els 1.276 lemes nous del DIEC1.<sup>4</sup> Entre aquests recursos, els més freqüents són la sufixació (41,18 %), els lemes simples (10,91 %), la composició culta (10,60 %), la composició híbrida (8,9 %) i els manlleus (6,35 %). Com veurem més endavant, la sufixació, els manlleus i els lemes simples són tres recursos que acostumen a ser les més freqüents en la classificació segons el recurs d'actualització lèxica. Els recursos d'actualització lèxica que tenen una freqüència mitjana són els lemes simples construïts (4,9 %), la prefixació (4,48 %), la composició

---

<sup>2</sup> Vegeu <http://www.iula.upf.edu/rec/texterm/3/> per a més informació sobre el projecte.

<sup>3</sup> Vegeu l'apartat 2.4.5 del capítol III i l'Annex 2 per a més informació sobre els criteris de classificació dels recursos d'actualització lèxica que adoptem.

<sup>4</sup> Vegeu la taula 9 de l'Annex 4.

patrimonial (2,63 %), la sufixació taxonòmica (2,32 %), la derivació zero (2,17 %) i l'escurçament (1 %).

Per últim, els recursos menys representatius són el grup més nombrós: l'acronímia (0,77), els lemes simples formats a partir de noms propis (0,77 %), la lexicalització adjectiva (0,7 %), les formes prefixades (0,54 %), la lexicalització nominal (0,46 %), els compostos formats a la manera culta (0,39 %), els derivats amb prefixoides (0,31 %), els manlleus històrics (0,31 %), les variants ortogràfiques (0,15 %), els lemes simples construïts formats amb un sufix (0,15 %) i la parasíntesi (0,08 %). Tots aquests recursos tenen una freqüència igual o inferior a 10 lemes.

### **1.1.1.2. Manlleus**

Entre els 1.276 lemes nous del DIEC1 detectem només 82 manlleus. Aquests 82 manlleus procedeixen de 18 llengües diferents, entre les quals n'hi ha tres que predominen per sobre de la resta. Aquestes tres llengües són el francès (25,61 %), l'anglès (23,17 %) i el castellà (14,63 %).<sup>5</sup> Hi ha un grup de llengües de menor freqüència format per l'italià (6,10 %), l'alemany (3,66 %), l'arauac (2,44 %), el malai (2,44 %) i el rus (2,44 %). Aquestes llengües tenen una freqüència d'importància mitjana. Per últim, detectem un conjunt de llengües que tan sols presenten una freqüència relativa de l'1,22 %. Aquestes llengües són l'albanès, l'àrab, el bantu, el filipí, el japonès, el neerlandès, el romanès, el sànscrit i el serbocroat.

Segons la classificació anterior, observem que el nombre de manlleus adaptats (55) és superior al nombre de manlleus no adaptats (27). Les llengües que presenten manlleus adaptats i manlleus no adaptats (francès, l'anglès, el castellà, l'italià i l'alemany) tenen, en tots els casos, un nombre superior de manlleus adaptats, excepte en l'italià, que té 1 manlleu adaptat i 4 no adaptats.

També classifiquem els manlleus segons la llengua de procedència i l'àrea temàtica a la qual pertanyen, a més d'especificar si són adaptats o no.<sup>6</sup> Així, observem la diversitat de marques temàtiques que presenten els manlleus combinades segons la llengua i segons si el manlleu és adaptat o no. En cap cas detectem freqüències elevades, sinó que més aviat trobem una diversitat considerable en la combinació de marca temàtica i llengua de procedència. Si, a més, tenim en compte si el manlleu és adaptat, aquesta diversitat és encara més acusada. Per exemple, la marca temàtica més freqüent és

---

<sup>5</sup> Vegeu la taula 15 de l'Annex 4.

<sup>6</sup> Vegeu la taula 21 de l'Annex 4.

*genealogia i heràld.*, amb quatre casos procedents de manlleus adaptats del francès (4,88 %). La resta de casos que considerem rellevants són els que tenen una freqüència superior a 1 (a partir de l'1,22 %). Entre aquests casos en destaca la marca *d'esports*, on trobem tres manlleus adaptats del castellà (3,66 %) i dos de l'anglès (2,44 %). Altres exemples de marques temàtiques amb una freqüència relativa del 2,44 % són *informàtica* (anglès no adaptat), *agricultura* (francès adaptat), *hoteleria* (francès adaptat), *indústria de la fusta* (francès adaptat), *lèxic comú* (francès no adaptat), *història* (francès/serbocroat no adaptat) i *música* (italià no adaptat). Així doncs, el francès és la llengua que predomina entre aquests manlleus.

### 1.1.1.3. Marca d'ús

Les marques d'ús proporcionen informació sobre el context en què es poden utilitzar els lemes. En ocasions, els articles lexicogràfics dels lemes inclouen aquest tipus de marques, i en d'altres ocasions aquestes marques no apareixen explícitament, però sí que estan incorporades en el cos de la definició. La documentació del DIEC1 del web del diccionari les agrupa segons el punt de vista històric, social i estilístic (Institut d'Estudis Catalans s. d.:13):

Aquestes marques fan referència a l'ús tant des del punt de vista històric (*antigament, arcaic, obsolet*), com des del punt de vista social (*popularment, en el llenguatge infantil, vulgarment*) i des del punt de vista estilístic (*poèticament*).

En els articles lexicogràfics del nostre corpus trobem cinc marques d'ús diferents (4,22 %).<sup>7</sup> Així doncs, cal destacar el baix nombre de marques que detectem respecte el total d'articles lexicogràfics analitzats. Més concretament, predomina la marca *ANT.* ([emprat] *antigament*) per sobre de la resta (60,61 %). Aquesta marca pertany al vessant històric citat pel diccionari. La resta de marques tenen una representació anecdòtica: *POP.* ([emprat] *popularment*) (21,21 %), *OBS.* ([mot / sentit] *obsolet*) (12,12 %), *ABUS.* ([emprat] *abusivament*)<sup>8</sup> (3,03 %) i *VULG.* (*mot vulgar/ [emprat] vulgarment*) (3,03 %). S'ha de tenir en compte, doncs, que cap de les cinc marques té una

---

<sup>7</sup> Vegeu la taula 33 de l'Annex 4.

<sup>8</sup> Detectem també la marca *ABUS.* que no es recull en la classificació per contextos presentada en el diccionari.



representació important respecte el total de lemes resultants del creuament.<sup>9</sup>

L'escàs nombre de marques es deu al fet que el diccionari té una voluntat explícita de no incloure aquest tipus d'informació en les seves accepcions. Basem aquesta afirmació en la informació següent (Institut d'Estudis Catalans s. d.:12):

Des del punt de vista històric, el DIEC ha afegit la marca *en rec.* ('en recessió') a la llista de marques d'ús que heretava del diccionari de Fabra.

Quant a l'ús social, el DIEC ha prescindit d'algunes marques del DGLC com ara *h.* ('habitualment'), *ord.* ('ordinàriament'); ha prescindit també de la marca *fam.* ('propri del llenguatge familiar'), que ha estat assimilada a la marca *pop.* ('[emprat] popularment'); ha restringit la marca *vulg.* ('[emprat] vulgarment') a usos d'un grau de formalitat mínim.

Pel que fa als valoratius estilístics, el DIEC ha prescindit de la marca *pej.* ('pejorativament') del DGLC, els usos de la qual han estat assimilats a la marca *desp.* ('[emprat] despectivament'), i ha mantingut la resta de marques estilístiques del diccionari de Fabra.

Cal remarcar que, seguint el criteri de la Secció Filològica, el DIEC ha prescindit de les marques que en el DGLC fan referència a l'ús territorial de les paraules, és a dir, s'han eliminat les marques «*d.*» ('dialectal') i «*reg.*» ('regional') i només en alguns casos s'ha incorporat a la definició l'especificació «en algunes contrades».

#### 1.1.1.4. Marca temàtica

Obtenim les marques temàtiques dels lemes nous del DIEC1 en l'apartat de cerca avançada del web del mateix diccionari<sup>10</sup>, ja que és l'únic recurs que inclou aquesta informació.<sup>11</sup> Cal tenir present que en

---

<sup>9</sup> Vegeu la taula 27 de l'Annex 4.

<sup>10</sup> Vegeu <http://institucional.iec.cat/diec1/entrada/diec.asp> per a més informació.

<sup>11</sup> En la versió en paper del DIEC1 no s'inclouen explícitament les marques temàtiques en els articles lexicogràfics.

les instruccions del web s'explica que les marques temàtiques no són ni sistemàtiques ni cobreixen tot el leuari, tot i que són una bona orientació (DIEC1web):

El *Diccionari de la llengua catalana* (DIEC) editat el 1995 no inclou àrees temàtiques de manera explícita. La classificació temàtica proposada es va utilitzar per als treballs de redacció del DIEC. No es tracta d'una classificació completa ni exhaustiva, però pot aportar dades interessants per restringir la cerca. Amb la restricció per àrea temàtica obtindreu els articles del diccionari que inclouen almenys una accepció classificada temàticament a la base de dades de les Oficines Lexicogràfiques. Recordeu que aquesta accepció no apareixerà marcada.

Així, detectem 1.109 marques temàtiques entre els 1.276 articles lexicogràfics dels lemes nous del DIEC1 obtinguts.<sup>12</sup> En concret, detectem 82 marques diferents en 993 articles lexicogràfics. En aquest recompte no tenim en compte, però, la categoria *sense marca*, que en aquesta taula indica que alguns dels lemes, a banda de tenir alguna accepció amb una marca temàtica, també en tenen alguna altra sense marca. Això significa que el percentatge d'articles amb marques temàtiques és del 77,82 %. El nombre d'articles lexicogràfics sense cap tipus de marca és 283 (22,18 % dels lemes).

Les marques temàtiques més freqüents, amb una freqüència igual o superior a 50 lemes, són *medicina-farmàcia* (8,39 %), *química* (7,03 %), *història* (6,04 %), *plantes superiors* (5,95 %) i *ciència forestal* (4,51 %). Amb una freqüència mitjana, d'entre 49 i 10 lemes, trobem marques com *geologia* (3,97 %), *biologia* (3,43 %), *peixos* (2,98 %), *esports* (1,62 %), *agricultura* (1,53 %) o *art* (1,08 %). Ens crida l'atenció la marca *lèxic comú* amb 11 ocurrències (0,99 %). Creiem que la representació del lèxic comú en un diccionari normatiu que inclou aquesta marca hauria de ser més elevada. Més endavant comprovarem quin és el percentatge de lèxic comú en la resta de diccionaris. A partir dels resultats podrem valorar aquesta dada, que creiem que variarà en cada obra. Per últim, el grup més nombrós correspon a les marques temàtiques presenten una freqüència més baixa, inferior a 10 lemes. Alguns exemples són *psicologia* (0,72 %), *matemàtiques* (0,72 %), *informàtica* (0,54 %), *religió* (0,45 %), *dret* (0,36 %), *pedagogia* (0,27 %), etc.

---

<sup>12</sup> Vegeu la taula 39 de l'Annex 4.

### 1.1.1.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic

En aquest apartat contrastem els lemes nous del DIEC1 amb un corpus lexicogràfic, en aquest cas, el leuari del *Diccionari català-valencià-balear* (DCVB) d'Alcover Moll. En la taula següent es poden observar el nombre de lemes nous del DIEC1 documentats en el DCVB.<sup>13</sup>

<b>Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DIEC1)</b>			
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>	<b>Exemples</b>
0 entrades	814	63,79 %	<i>barbitúric, laïcat</i>
1 entrada	370	29,00 %	<i>biòpsia, llarguerut</i>
2 entrades	66	5,17 %	<i>barbeta, lletrada</i>
3 entrades	17	1,33 %	<i>batedor, llima</i>
4 entrades	7	0,55 %	<i>bull, lucà</i>
5 entrades	2	0,16 %	<i>bus, lluçà</i>
<b>Total</b>	<b>1.276</b>	<b>100 %</b>	

Taula 1: Contrast dels lemes dels diccionaris amb el corpus lexicogràfic del creuament DIEC1

El fenomen més habitual és la no-coincidència entre els lemes del corpus i el leuari del DCVB (63,79 %). En cas que els lemes siguin coincidents, són majoritaris els casos en què el lema només correspon a una entrada.

Vegem tot seguit una taula amb el total de lemes coincidents amb el DCVB i el total dels lemes que no coincideixen:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb corpus lexicogràfic (DIEC1)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	814	63,79 %
Coincidents	462	36,21 %
<b>Total</b>	<b>1.276</b>	<b>100 %</b>

Taula 2: Representativitat dels lemes coincidents i no coincidents en el contrast amb corpus lexicogràfic (DIEC1)

El nombre de lemes no coincidents és major que el nombre de lemes coincidents. El motiu pot ser que els dos diccionaris han estat

<sup>13</sup> En la taula es mostra el nombre de coincidències per entrada diferent. En el DCVB trobem que un mateix lema té una o més entrades diferents. Optem per mantenir aquesta informació desglossada en la taula.

elaborats en períodes de temps diferents i, a més, no es regeixen pels mateixos criteris d'elaboració. El DIEC1 té una voluntat estandarditzadora i el DCVB recull lèxic de totes les varietats geolingüístiques del català com bé s'explica al pròleg del diccionari (DCVB):

Segons el llarg subtítol que duu aquest Diccionari, és una obra comprensiva de la llengua que es parla "al Principat de Catalunya, al Regne de València, a les illes Balears, al departament francès dels Pirineus Orientals, a les Valls d'Andorra, al marge oriental d'Aragó i a la ciutat d'Alguer de Sardenya". El nom que designa més pròpiament aquesta llengua és l'adjectiu *català*: en dir *llengua catalana* queda denominat el conjunt de formes de parlar pròpies de totes les comarques indicades en el subtítol.

En el contrast dels lemes amb el DCVB comprovem quins són els recursos d'actualització lèxica més freqüents dels lemes.<sup>14</sup> La tendència més rellevant és la formació per sufixació (51,95 %), seguida dels lemes simples (14,72 %). Aquests dos recursos agrupen la majoria dels casos que detectem (66,47 %). També observem un grup de cinc recursos d'actualització lèxica que mostren una freqüència destacada per sobre de la mitjana. Aquests recursos són els lemes simples construïts (4,55 %), els manlleus (4,33 %), la composició culta (4,33 %), la derivació zero (3,68 %) i la prefixació (2,81 %). Per últim, hi ha un conjunt de recursos menys rellevants com l'acronímia, la composició híbrida o la composició que registren una freqüència relativa inferior a l'1,52 %. Aquests recursos són, per tant, els menys representatius del diccionari.

També comprovem quines són les marques temàtiques o les seqüències de marques més freqüents en els articles dels lemes nous documentats en el DCVB.<sup>15</sup> En total detectem 51 marques temàtiques soles en l'article i 62 seqüències de marques temàtiques. En cas que el lema presenti alguna marca, les més freqüents són *plantes superiors* (4,98 %), *història* (4,33 %), *medicina-farmàcia* (4,33 %), *transport (aigua)* (3,46 %), *indústria tèxtil* (2,81 %) i *ciència forestal* (2,38 %). La resta de casos (154) tenen una freqüència igual o inferior a vuit lemes (1,73 %). Entre aquests 154 casos detectem 68 marques

---

<sup>14</sup> Vegeu la taula 57 de l'Annex 4.

<sup>15</sup> Vegeu la taula 63 de l'Annex 4.

temàtiques i 86 seqüències de marques. Les marques més freqüents són *agricultura, ocells, química, filologia, biologia, peixos i indústries (general)*. I algunes de les seqüències són *sense marca/transports/informàtica/sense marca, arts gràfiques/ciència forestal, adoberia/plantes superiors*, etc.

Detectem 23 marques i 19 seqüències que no apareixen en cap dels lemes nous documentats en el DCVB. És el cas, per exemple, de les marques *comunicació, filosofia o física general* i de les seqüències *ciència forestal/biologia o antropologia/filologia*.

#### 1.1.1.6. Contrast dels lemes dels diccionaris amb els corpus textuais

Descrivim, tot seguit, els resultats de la consulta de tots els lemes nous del DIEC1 en els tres corpus textuais presentats en el capítol III sobre la metodologia de treball.

#### ***Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans***

Hi ha 590 lemes nous en el DIEC1 documentats en el CTILC de l'Institut d'Estudis Catalans (12.553 ocurrences). En la taula següent observem la representativitat dels lemes coincidents i dels no coincidents:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans (DIEC1)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	686	53,77 %
Coincidents	590	46,23 %
<b>Total</b>	<b>1.276</b>	<b>100 %</b>

Taula 3: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans (DIEC1)

El nombre de lemes no coincidents és lleugerament superior al nombre de lemes coincidents (53,77 %). Segons el recurs d'actualització lèxica<sup>16</sup>, classifiquem els 590 lemes en 19 recursos d'actualització lèxica i en 13 seqüències de recursos diferents. La sufixació és el recurs més freqüent (44,58 %). Els altres quatre

<sup>16</sup> Vegeu la taula 72 de l'Annex 4.

recursos d'actualització lèxica que també tenen un pes rellevant són els lemes simples (11,19 %), la composició culta (9,66 %), la composició híbrida (7,63 %) i els manlleus (6,78 %). Aquests cinc recursos representen una freqüència total de 471 lemes (79,84 %). En segon lloc, trobem un grup de recursos amb una freqüència inferior a 40 lemes. Alguns exemples d'aquests recursos són la prefixació (3,9 %), els lemes simples construïts (3,73 %), la composició patrimonial (2,03 %), la sufixació taxonòmica (1,86 %) i la derivació zero (1,86 %). En darrer lloc, amb una freqüència igual o inferior a sis lemes detectem recursos com l'escurçament (1,02 %), els noms propis (1,02 %), la lexicalització adjectiva (0,68 %), la composició a la manera culta (0,51 %), els manlleus històrics (0,51 %), l'acronímia (0,34 %) i la parasíntesi (0,17 %).

Les seqüències de recursos tenen, en tots els casos una rellevància molt escassa ja que tan sols les detectem en un sol lema (0,17 %). La diferència, però, la trobem en el nombre d'ocurrències, ja que en ocasions hi ha un nombre elevat d'ocurrències per a un sol lema, com a *bat*, *bar*, *bus* i *bola*. Per tant, els percentatges d'un mateix recurs o seqüència poden variar si tenim en compte els lemes (sense repeticions) o les ocurrències (amb repeticions).

Detectem 61 marques temàtiques diferents i 71 seqüències de marques.<sup>17</sup> El percentatge de lemes sense marcar és del 24,75 %. Hi ha un conjunt de marques força rellevants que tenen una freqüència relativa superior al 2 %, com *medicina-farmàcia* (7,97 %), *història* (4,58 %), *biologia* (3,73 %), *química* (3,39 %), *filologia* (2,88 %), *geologia* (2,71 %) i *ciència forestal* (2,2 %). Amb una freqüència de dos lemes (0,34 %), i per tant amb un pes menys rellevant, trobem altres marques temàtiques com *explotació animal*, *dret*, *defensa*, *art*, *enginyeria industrial*, *informàtica*, *jocs i espectacles*, *llenguatge administratiu*, *indústries (general)*, *meteorologia*, *peixos*, *obres públiques*, *numismàtica*, *pedagogia* i *indústria de la fusta*. També detectem alguna seqüència com *biologia/medicina-farmàcia* o *meteorologia/sense marca*.

També trobem lemes que presenten un sol registre, ja sigui amb una marca temàtica individual o amb una seqüència de marques (0,17 %). En la majoria dels casos aquests lemes presenten una seqüència de marques, com *agricultura/sense marca*, *agricultura/plantes superiors*, *plantes superiors/medicina-farmàcia/transports (aigua)* o *antropologia/filologia*. En cas que els lemes presentin alguna marca individual, algunes d'aquestes marques són *plantes inferiors*, *telecomunicació*, *física (general)*, *mamífers*, *enginyeria elèctrica*, *enginyeria electrònica*, *enginyeria industrial*, *invertebrats* o *oficines*

---

<sup>17</sup> Vegeu la taula 75 de l'Annex 4.

*comunicació*. Per últim, també detectem un conjunt de marques temàtiques i seqüències que no queden representades entre els lemes coincidents amb el CTILC. Algunes d'aquestes marques són *adoberia, col·lectius vegetals, economia domèstica o pesca*.

### **Corpus Tècnic de l'Institut Universitari de Lingüística Aplicada**

Detectem 227 lemes nous del DIEC1 en el Corpus Tècnic (17.224 ocurrències). Així doncs, el nombre de lemes no coincidents (82,22 %) està clarament per sobre del nombre de lemes coincidents (17,78 %). Mostrem a continuació la representativitat dels lemes coincidents respecte el total de lemes nous del DIEC1:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DIEC1)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	1,049	82,22 %
Coincidents	227	17,78 %
<b>Total</b>	<b>1.276</b>	<b>100 %</b>

Taula 4: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DIEC1)

S'han detectat 22 recursos d'actualització lèxica individuals i 14 seqüències de recursos.<sup>18</sup> Un cop més, segueixen predominant els lemes nous formats per sufixació (44,04 %), seguits de lluny pels manlleus (11,91 %). Detectem també una freqüència notable dels recursos de composició híbrida (7,58 %), de lemes simples (7,22 %), de composició culta (6,86 %), de lemes simples construïts (3,97 %) i de composició patrimonial (3,61 %). En tots aquests recursos la freqüència de lemes és igual o superior a deu ocurrències. Amb una freqüència mitjana, inferior a deu lemes, trobem recursos com la prefixació (3,25 %), la derivació zero (2,53 %), els noms propis (1,81 %), la sufixació taxonòmica (0,72 %), la lexicalització adjectiva (0,72 %) i els manlleus construïts (0,72 %). Els recursos que tenen una representativitat més baixa, amb una freqüència d'un sol lema (0,36 %), són els escurçaments i la derivació per prefixoide, a més d'un conjunt de seqüències de recursos d'actualització lèxica com nom propi/derivació zero, manlleu/derivació zero o manlleu/escurçament.

No detectem en cap cas els lemes formats a partir de recursos com l'acronímia, la lexicalització nominal, la variació ortogràfica, la parasíntesis o que siguin formes prefixades.

<sup>18</sup> Vegeu la taula 105 de l'Annex 4.

Detectem 53 marques temàtiques i 47 seqüències de marques entre els 277 lemes coincidents.<sup>19</sup> Tot i que els lemes sense marques formen un grup nombrós (28,52 %), detectem una freqüència elevada de marques com la *medicina-farmàcia* (5,42 %), *biologia* (3,97 %), *història* (3,61 %), *política* (3,25 %), *teoria econòmica* (2,17 %), *esports* (2,17 %), *sociologia* (1,81 %), *ciència forestal* (1,81 %), *ecologia* (1,81 %). Les marques que tenen una freqüència mitjana són aquelles que es repeteixen a entre quatre i dos lemes. Algunes d'aquestes marques són *defensa*, *psicologia* i *filologia*, amb un 1,44 %; *religió* i *química*, amb un 1,08 %; o *zoologia*, *agricultura*, *dret* o *hoteleria*, amb un 0,72 %.

Les seqüències de marques que només tenen un lema (i representen el 0,36 % respecte el total). Per exemple, detectem *economia domèstica*, *enginyeria industrial*, *arts gràfiques*, *enginyeria elèctrica*, *invertebrats*, *jocs i espectacles*, *llenguatge administratiu* o *matemàtiques*. Tot i això, si tenim en compte les ocurrencies d'aquests lemes en el corpus, el pes d'aquestes marques o seqüències varia. En ocasions detectem una freqüència elevada d'ocurrencies com, per exemple *llenguatge administratiu* (en el lema *laboral*, 1.834 ocurrencies, 10,65 %) o numismàtica (en el lema *lluís*, 497 ocurrencies, 2,89 %). Per últim, volem remarcar que hi ha un conjunt elevat d'àrees que no presenten cap ocurrencia en el corpus consultat. Algunes d'aquestes àrees són *adoberia*, *art*, *enginy. elèctrica*, *metal·lúrgia* o *pesca*.

### ***Banc BOBNEO de l'Observatori de Neologia de l'Institut Universitari de Lingüística Aplicada***

L'últim corpus on consultem els lemes nous del DIEC1 és el Banc BOBNEO de l'Observatori de Neologia. En la consulta només detectem 69 lemes coincidents (838 ocurrencies) (5,4 %). Així doncs, gairebé no hi ha lemes nous del DIEC1 en el Banc BOBNEO. Això significa que l'anàlisi dels pocs lemes coincidents s'han d'interpretar amb precaució, tot i que poden indicar tendències. En la taula següent mostrem la coincidència entre els lemes nous del DIEC1 i el Banc BOBNEO:

---

<sup>19</sup> Vegeu la taula 99 de l'Annex 4.



Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DIEC1)		
Coincidència i nombre d'entrades per lema coincident	Lemes	Freqüència lemes
No coincidents	1,207	94,6 %
Coincidents	69	5,40 %
<b>Total</b>	<b>1.276</b>	<b>100 %</b>

Taula 5: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DIEC1)

Detectem 11 recursos d'actualització lèxica diferents en els lemes coincidents.<sup>20</sup> Entre aquests recursos, n'hi ha quatre de molt freqüents (82,6 %). Aquests recursos són, per ordre de rellevància, la sufixació (37,68 %), els manlleus (21,74 %), la composició híbrida (11,59 %) i la patrimonial (11,59 %). La resta de recursos són set: lemes simples (7,25 %), composició culta (2,9 %), noms propis (1,45 %), derivació zero (1,45 %) i prefixació (1,45 %).

En dues ocasions detectem un lema que es pot haver format a partir de dos recursos d'actualització lèxica. És el cas de *bola* (lema simple construït o sufixació, 0,24 %), i *bat* (manlleu o escurçament, 1,45 %)

Alguns dels recursos que no detectem en els lemes nous documentats en el BOBNEO són, per exemple, les acronímies, els compostos a la manera culta, els escurçaments, les lexicalitzacions adjectives, les parasíntesis, etc.

Des del punt de vista de l'àrea temàtica<sup>21</sup>, detectem 27 marques individuals i 8 seqüències de marques diferents. Hi ha 20 lemes coincidents que no presenten marca temàtica (28,99 %). Les marques més habituals, amb una freqüència de quatre i tres lemes, són *esports* (5,8 %), *música* (4,35 %) i *plantes superiors* (4,35 %). Entre les que tenen una freqüència igual o inferior a dos lemes trobem, per exemple, les marques de *transports (aigua)*, *defensa* o de *medicina-farmàcia*, cadascuna d'elles amb un 2,9 %, o d'*arts gràfiques*, d'*antropologia*, d'*ecologia* o d'*agricultura* amb un 1,45 % respectivament.

### 1.1.2. DIEC2

En el creuament dels lemaris dels diccionaris DIEC1 i DIEC2 detectem 243 lemes nous en el DIEC2 que no trobem en el DIEC1. En concret, detectem 140 lemes nous de la lletra *b* i 103 de la lletra *l*.

<sup>20</sup> Vegeu la taula 111 de l'Annex 4.

<sup>21</sup> Vegeu la taula 117 de l'Annex 4.

### 1.1.2.1. Recursos d'actualització lèxica

En aquest apartat classifiquem els lemes del DIEC2 segons el seu recurs d'actualització lèxica.<sup>22</sup> Detectem 17 recursos d'actualització lèxica diferents entre els 243 lemes analitzats.

Els recursos d'actualització lèxica més freqüents són la sufixació (44,03 %), les formes prefixades (13,99 %) i els manlleus (11,11 %). Volem destacar l'elevat nombre de formes prefixades, que no trobem en l'edició anterior del mateix diccionari. Amb una freqüència mitjana, detectem la composició culta (7 %) i la composició híbrida (6,58 %). Els recursos d'actualització lèxica que tenen una freqüència més baixa, és a dir, inferior al 4 % (i menys de deu ocurrences), són els lemes simples (3,7 %), els lemes simples construïts (2,47 %), la composició patrimonial (1,65 %), la prefixació (1,65 %), la derivació zero (1,65 %), les formes sufixades (1,23 %), les variants ortogràfiques (1,23 %), la sufixació taxonòmica (0,82 %), els escurçaments (0,82 %), la lexicalització adjectiva (0,82 %), els prefixos (0,82 %) i els manlleus històrics (0,41 %).

### 1.1.2.2. Manlleus

Entre els 243 lemes nous del DIEC2 detectem 27 manlleus procedents de 12 llengües entre les quals en trobem una d'origen incert i una altra que no distingim si es tracta de l'hebreu o de l'arameu.<sup>23</sup> Entre aquestes dotze llengües en destaquen dues, l'anglès (40,74 %) i el francès (18,52 %). L'anglès, a més, és l'única llengua que presenta manlleus adaptats i no adaptats. La resta de llengües tan sols presenten un o dos manlleus, en ocasions adaptats i en ocasions no adaptats. Entre aquestes llengües destaquen l'italià i el bantu, amb un 7,41 % de manlleus respectivament.

A continuació classifiquem els manlleus segons si són adaptats o no adaptats i tenim en compte amb la seva àrea temàtica.<sup>24</sup> Per a cada llengua trobem diverses marques temàtiques. Això significa que els manlleus que detectem en el DIEC2 procedeixen de dotze llengües diferents i, a més, de diverses àrees temàtiques (32). La llengua que té una major diversitat d'àrees temàtiques és l'anglès, com és previsible, ja que és la llengua amb més presència entre els manlleus nous del DIEC2. En concret, els manlleus anglesos estan classificats en 11 temàtiques diferents. Entre aquestes 11 marques, els manlleus no adaptats s'agrupen en les àrees *EI* (enginyeria industrial, 1,92 %),

---

<sup>22</sup> Vegeu la taula 10 l'Annex 4.

<sup>23</sup> Vegeu la taula 16 de l'Annex 4.

<sup>24</sup> Vegeu la taula 22 de l'Annex 4.

GG (geografia, 3,85 %), GL (geologia, 1,92 %), *HIG* (genealogia i heràldica, 1,92 %), LC (lèxic comú, 5,77 %) i MD (medicina, 5,77 %). I els manlleus adaptats en ED (economia domèstica, 1,92 %), HO (hoteleria, 3,85 %), IMI (indústries en general, 1,92 %), LC (lèxic comú, 5,77 %) i PR (professions, 1,92 %). Només en un cas detectem manlleus adaptats i no adaptats en una mateixa àrea (LC). A més, aquesta àrea, juntament amb MD és la que presenta més casos, 6 per a LC (3 adaptats i 3 no adaptats) i 3 per a MD.

A banda de l'anglès també destaquen altres llengües com el francès, l'italià o el malai. En el cas del francès detectem que hi ha quatre àrees temàtiques diferents en manlleus no adaptats (*HIG* (genealogia i heràldica, 3,85 %), HO (hoteleria, 1,92 %), LC (lèxic comú, 1,92 %), MU (música, 1,92 %) i PR (professions, 1,92 %)). L'àrea *HIG* és la més nombrosa amb dues ocurrences. Quant a l'italià, trobem tres àrees temàtiques diferents en cinc manlleus no adaptats (*BB* (biblioteconomia, 1,92 %), JE (jocs i espectacles, 5,77 %), LC (lèxic comú, 5,77 %)). Detectem tres àrees temàtiques diferents (*AN* (antropologia, 3,85 %), AR (art, 1,92 %), IT (indústria tèxtil, 3,85 %)) en cinc manlleus del malai, igual que de l'italià. Les dues àrees més freqüents, amb dues ocurrences cadascuna són *AN* i *IT*. En la resta de llengües (bantu, castellà, hebreu/arameu, indi, llatí i una llengua d'origen incert) tan sols registrem una o dues àrees temàtiques. Algunes d'aquestes llengües, però, tenen una freqüència alta, com és el cas del bantu, amb 2 casos d'*AN* (antropologia, 3,85 %) i 4 de *FL* (filologia, 7,69 %).

### 1.1.2.3. Marca d'ús

El DIEC2 especifica en el seu pròleg que, de forma voluntària, s'han reduït les marques d'ús (o marques valoratives) en els articles lexicogràfics del diccionari. Així doncs, les marques d'ús en aquest creuament són pràcticament inexistent (DIEC2: X).<sup>25</sup>

“La Secció Filològica ha decidit reduir considerablement les marques valoratives. Així, no hi ha cap etiqueta específica que indiqui els usos absoluts de les entrades (*ABS.*) [...] Tampoc no s'indica amb cap senyal l'ús abusiu d'un mot (*ABUS.*). [...] No es destaca l'ús despectiu d'una paraula (*DESP.*). [...] També desapareix la marca abreviada *ESP./esp.* (especialment); continua utilitzant-se la referència a aquest concepte,

<sup>25</sup> Vegeu les taules 28 i 34 de l'Annex 4.

però amb la forma desenvolupada: *especialment*. No es marquen els sentits figurats (*FIG.*); [...] els usos hiperbòlics (*HIPERB.*): [...] Se suprimeix l'etiqueta *PER ANAL.* (per analogia) [...] En canvi, hom ha cregut convenient de mantenir les marques valoratives *ANT.* ([emprat] antigament), *EN REC.* ([sentit] en recessió), *OBS.* ([mot/sentit] obsolet), *PER EXT.* ([sentit] per extensió), *POP.* ([emprat] popularment) i *VULG.* ([emprat] vulgarment; [mot] vulgar); això no obstant, hi ha hagut modificacions respecte a la primera edició, en el sentit que algunes accepcions o subaccepcions que hi duen alguna d'aquestes marques, o no la mantenen o ha estat modificada.

En la nostra recerca detectem dos casos en què s'ha utilitzat la marca *[emprat] popularment (POP.)*, en els lemes *betza* i *bimba*. En els dos casos es tracta de noms comuns. En el primer lema, *betza* és una variant ortogràfica de *butza* (d'acord amb la definició del GDLC).<sup>26</sup> La marca popular reforça el fet que *betza* és una variant ortogràfica. En el lema *bimba*, es tracta d'un substantiu utilitzat especialment en el futbol per anomenar la pilota. Podríem dir que forma part de l'argot del futbol i, possiblement, per aquesta raó se li atribueix un ús popular. Així doncs, hi ha un gran desequilibri entre el nombre de mots que no presenten cap marca d'ús en el seu article lexicogràfic i els que sí que en porten. Queda comprovada la voluntat del DIEC2 de no incloure aquest tipus de marques en els seus articles.

#### 1.1.2.4. Marca temàtica

Obtenim les marques temàtiques dels lemes del DIEC2 en els articles lexicogràfics del mateix diccionari. A diferència del DIEC1, en què els articles no presenten les marques, i només es poden recuperar a partir d'una cerca avançada al web<sup>27</sup>, les marques temàtiques del DIEC2 sí que són explícites en el cos del diccionari, tant de l'edició en paper com de l'electrònica.<sup>28</sup>

---

<sup>26</sup> Definició de *betza*: [variant de *butza* (v. *esbotzar*)] f Panxa voluminosa. (GDLCweb)

<sup>27</sup> Vegeu el formulari de la cerca avançada al web del DIEC1: [http://institucional.iec.cat/diec1/entrada/diec\\_avan%C3%A7at.asp](http://institucional.iec.cat/diec1/entrada/diec_avan%C3%A7at.asp).

<sup>28</sup> Vegeu la taula 40 de l'Annex 4.

Trobem 378 marques temàtiques en els 243 articles lexicogràfics analitzats. En concret, detectem 51 marques temàtiques diferents. Entre aquestes marques la més freqüent és la de *LC (llengua comú)*, amb 128 ocurrences (33,86 %). Les marques *BOS (plantes superiors)* (5,82 %), *FL (filologia)* (5,03 %), *BOB (fongs i líquens)* (4,5 %), *MD (medicina)* (3,97 %), *GLM (minerals)* (3,44 %), *BI (biologia)* (3,17 %), *AN (antropologia)* (2,65 %), *HIG (genealogia i heràldica)* (2,65 %), *GLP (paleontologia)* (2,38 %) i *ZOP (peixos)* (2,12 %) tenen una freqüència entre un 6 % i un 2 % (entre 22 i 8 lemes) i considerem que formen un grup de representativitat mitjana. Per últim, el grup més nombrós de marques està format per aquelles que representen menys de 2 % dels casos. Trobem 40 marques diferents que compleixen aquesta característica, algunes de les quals són *RE (religió)* (1,85 %), *AGA (agricultura)* (1,59 %), *QU (química)* (1,59 %), *AGP (pesca)* (0,79 %), *DE (defensa)* (0,53 %), *FS (filosofia)* (0,53 %) o *IT (indústria tèxtil)* (0,53 %).

#### 1.1.2.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic

En aquest apartat comprovem quins lemes nous del DIEC2 es troben recollits en el leuari del DCVB. En la taula següent es poden observar el nombre de lemes del DIEC2 presents en el DCVB:

<b>Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DIEC2)</b>			
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Recòmpte</b>	<b>Freqüència relativa</b>	<b>Exemples</b>
0 entrades	170	69,96 %	<i>bàrman, líquènic</i>
1 entrada	60	24,69 %	<i>badaloní, llumeta</i>
2 entrades	8	3,29 %	<i>bitllar, llimonera</i>
3 entrades	4	1,65 %	<i>bevedor, lliterà</i>
6 entrades	1	0,41 %	<i>bet</i>
<b>Total</b>	<b>243</b>	<b>100 %</b>	

Taula 6: Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DIEC2)

Les dades mostren que la no-coincidència és la tendència més habitual en aquest contrast, amb 170 lemes no coincidents (69,96 %). En segon lloc, en cas que hi hagi coincidència entre el diccionari i els lemes del nostre corpus, el més habitual és trobar una única entrada per lema. Aquest fet es produeix en 60 lemes del DIEC2 documentats també en el DCVB (24,69 %). En canvi, detectar dues o més entrades en el DCVB per a cada lema del DIEC2 és menys habitual. Tan sols detectem 13 lemes que presenten aquesta característica (5,35 %).

A continuació incloem una taula que mostra la representativitat dels lemes que coincideixen, i els que no, respecte el total de lemes del DIEC2.

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb corpus lexicogràfic (DIEC2)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Recompte</b>	<b>Freqüència relativa</b>
No coincidents	170	69,96 %
Coincidents	73	30,04 %
<b>Total</b>	<b>243</b>	<b>100 %</b>

Taula 7: Representativitat dels lemes coincidents i no coincidents en el contrast amb corpus lexicogràfic (DIEC2)

Classifiquem els 73 lemes coincidents en deu recursos d'actualització lèxica diferents. Un cop més, la sufixació és el recurs més freqüent entre aquests deu recursos (73,97 %). La resta de recursos tenen una representativitat baixa: la derivació zero, les variants ortogràfiques i els lemes simples (tots tres amb un 4,11 % respectivament); la composició patrimonial, la lexicalització adjectiva, els manlleus i els lemes simples construïts (tots ells amb un 2,7 % respectivament); i, per últim, la composició culta i la composició híbrida (amb un 1,37 % cadascun).<sup>29</sup> Per últim, hi ha un conjunt de recursos que, un cop més, no tenen cap representació entre els lemes coincidents amb el DCVB. Aquests recursos són la prefixació, la sufixació taxonòmica, l'escurçament, les formes prefixades, les formes sufixades, els prefixos i els manlleus històrics. La no-detecció de formes prefixades, formes sufixades i prefixos és esperable, ja que el DCVB no inclou aquest tipus d'elements en el seu leuari.

En l'anàlisi de les marques temàtiques detectem 14 marques úniques i 16 seqüències de marques.<sup>30</sup> Entre aquestes 31 marques o seqüències incloem la marca *LC* de lèxic comú que, a més, és la que presenta una freqüència més elevada (31,51 %). La segona freqüència més elevada correspon a la seqüència *LC/LC*, amb un 15,07 %. Així doncs, la representativitat global de la marca *LC* és del 46,58 %. La resta de marques o seqüències tenen, com a màxim, una freqüència de tres lemes (4,11 %), com *ZOP* (peixos), *AGP* (pesca), *AGR* (explotació animal) o *BOS* (plantes superiors). Algunes de les marques menys representatives són *HO* (hoteleria, 2,74 %) i *BI* (biologia), *BOB* (fongs i líquens), *HIH* (història en general), *HIG* (genealogia i heràldica), *ZOR* (amfibis i rèptils) o *IMI* (indústries en general), cadascun amb un 1,37 % de representativitat. Hi ha un

<sup>29</sup> Vegeu la taula 58 de l'Annex 4.

<sup>30</sup> Vegeu la taula 64 de l'Annex 4.

conjunt de marques, com en totes les consultes, que no tenen representació entre els lemes coincidents. Algunes d'aquests recursos són *AF* (arts gràfiques), *BOI* (plantes inferiors), *CO* (comunicació) o *ZOM* (mamífers).

### 1.1.2.6. Contrast dels lemes dels diccionaris amb els corpus textuais

Descrivim en aquest apartat els resultats dels contrastos dels lemes del DIEC2 en els tres corpus textuais de contrast habituals.

#### ***Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans***

En la consulta dels lemes del DIEC2 en el CTILC, obtenim 65 lemes coincidents (540 ocurrences). Un cop més, el fenomen de la no-coincidència (73,26 %) és més habitual que la coincidència (26,74 %). En la taula següent mostrem el nombre de lemes coincidents entre el DIEC2 i el CTILC, juntament amb les freqüències relatives:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans (DIEC2)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	178	73,26 %
Coincidents	65	26,74 %
<b>Total</b>	<b>243</b>	<b>100 %</b>

Taula 8: Representativitat dels lemes coincidents i no coincidents respecte el total de lemes en el contrast amb el Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans (DIEC2)

Classifiquem els 65 lemes coincidents en deu recursos d'actualització lèxica diferents.<sup>31</sup> Entre aquests recursos el més habitual és la sufixació (64,62 %). Amb una freqüència força inferior trobem els manlleus (12,31 %) seguits de la composició híbrida (7,69 %). Els recursos de formació menys rellevants són els lemes simples, la composició culta i la composició patrimonial (amb un 3,08 % respectivament); i els manlleus històrics, els lemes simples construïts, els escurçaments i la derivació zero (amb un 1,54 % respectivament). Entre lemes coincidents no trobem recursos com la lexicalització adjectiva o substantiva, ni formes prefixades o sufixades, ni tampoc

<sup>31</sup> Vegeu la taula 73 de l'Annex 4.

l'acronímia o l'escurçament. També és rellevant la no-presència de lemes formats per prefixació.

Des del punt de vista de la variable de marca temàtica, detectem 15 marques i 19 seqüències de marques diferents.<sup>32</sup> La marca temàtica més freqüent és, un cop més, *LC* (*lèxic comú*). Aquesta marca representa el 26,15 % dels casos quan apareix de forma única en els articles lexicogràfics, i el 12,31 % quan apareix repetida dues vegades. Això significa que aquesta marca està present en el 38,46 % sense estar acompanyada de cap altra marca. La resta de marques o seqüències tenen una representativitat molt escassa respecte del total. Les úniques que destaquen entre les marques menys representatives són les que tenen una freqüència superior a un lema i que són *HO* (hoteleria, 4,62 %), *AGR* (explotació animal, 3,08 %), *FS* (filosofia, 3,08 %) i *FL* (filologia, 3,08 %). Les marques o seqüències que tenen una freqüència equivalent a un lema són molt diverses i nombroses (25). Algunes d'aquestes marques són *BOB* (fongs i líquens) *BO* (botànica), *BI* (biologia), *AGP* (pesca), *GLG* (mineralogia general), *SO* (sociologia), *QU* (química) o *RE* (religió). Un cop més hi ha un conjunt de marques que no tenen cap tipus de representació, ni com a marques úniques ni com a marques que formen part d'una seqüència. Algunes d'aquestes marques són *AF* (arts gràfiques), *BB* (biblioteconomia), *ME* (meteorologia) o *IT* (indústries tèxtil).

### **Corpus Tècnic de l'Institut Universitari de Lingüística Aplicada**

En la consulta dels lemes del DIEC2 en el Corpus Tècnic de l'IULA obtenim 38 lemes coincidents (437 ocurrències). Els lemes coincidents estan en minoria (15,63 %), respecte els lemes no coincidents (84,37 %). La taula següent mostra el nombre de lemes nous del DIEC2 presents en el Corpus Tècnic i les freqüències relatives:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DIEC2)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	205	84,37 %
Coincidents	38	15,63 %
<b>Total</b>	<b>243</b>	<b>100 %</b>

Taula 9: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DIEC2)

Classifiquem els 38 lemes coincidents en set recursos d'actualització lèxica diferents. Entre aquests set recursos en destaquen dos: la

<sup>32</sup> Vegeu la taula 76 de l'Annex 4.



sufixació (50 %) i els manlleus (26,32 %).<sup>33</sup> Els cinc recursos restants tenen una representativitat més baixa. D'entre aquests recursos, el més representatiu és la composició híbrida (10,53 %). La resta de recursos són la derivació zero (5,26 %) i els manlleus històrics, els escurçaments i la composició patrimonial, amb un 2,63 % cadascun. Alguns dels recursos que no han quedat recollits entre els lemes coincidents són la prefixació, la composició culta, l'acronímia, la siglació, les formes prefixades i sufixades, etc.

Detectem set marques temàtiques individuals i 19 seqüències de marques diferents.<sup>34</sup> Un cop més, la taula ens mostra que la marca temàtica més freqüent és *LC* (lèxic comú), amb un 31,58 % (suma del 21,05 % corresponent a *LC* com a marca única de l'article lexicogràfic i d'un 10,53 % corresponent a la seqüència *LC/LC*). Amb una freqüència molt inferior trobem dues marques, *HO* (hoteleria) i *IMI* (indústries en general), amb una freqüència del 5,26 % cadascuna. La resta de marques (o seqüències de marques) són 22 i només es troben en l'article lexicogràfic d'un sol lema (2,63 %). En la majoria de casos es tracta de seqüències de marques, com *BI/EG* (biologia/ecologia) o *LC/DE* (lèxic comú/defensa). Tot i això, detectem quatre marques simples, que són *BI* (biologia), *ED* (economia domèstica), *SO* (sociologia) o *SP* (esports). En cap cas detectem marques com, per exemple, *AF* (arts gràfiques), *BOC* (col·lectius vegetals), *CO* (comunicació), *GG* (geografia), o *GLM* (minerals).

### ***Banc BOBNEO de l'Observatori de Neologia de l'Institut Universitari de Lingüística Aplicada***

El nombre de lemes coincidents entre els nous del DIEC2 i els del Banc BOBNEO és 37 (349 ocurrences). Així doncs, el nombre de lemes coincidents és molt escàs (15,22 %). La taula següent mostra el nombre de lemes coincidents entre els dos corpus, juntament amb les freqüències relatives:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DIEC2)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	206	84,77 %
Coincidents	37	15,22 %
<b>Total</b>	<b>243</b>	<b>100 %</b>

Taula 10: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DIEC2)

<sup>33</sup> Vegeu la taula 94 de l'Annex 4.

<sup>34</sup> Vegeu la taula 100 de l'Annex 4.

Classifiquem els 37 lemes coincidents en 7 recursos d'actualització lèxica diferents.<sup>35</sup> Un cop més, la sufixació és el recurs més freqüent (56,76 %) seguit, novament, pels manlleus (27,03 %). La resta de recursos tenen una representativitat molt baixa. Aquests recursos són els lemes simples (5,41 %), la composició culta (2,7 %), la composició híbrida (2,7 %), els escurçaments (2,7 %) i els lemes simples construïts (2,7 %). Hi ha un conjunt de recursos que no apareixen entre els lemes coincidents amb el BOBNEO i que sí que tenen representació entre els lemes del DIEC2. Alguns d'aquests recursos són la prefixació, l'acronímia, la siglació, les formes prefixades i les sufixades, la lexicalització adjectiva i la nominal, la composició patrimonial o la derivació zero.

Detectem 14 marques temàtiques diferents i 12 seqüències de marques.<sup>36</sup> La marca *LC* (lèxic comú), segueix sent la més nombrosa (24,32 %). No obstant, en aquesta consulta, la marca *HO* (hoteleria) té més rellevància que en les consultes anteriors (8,11 %). La resta de marques o seqüències (24) presenten, un cop més, unes freqüències pràcticament irrelevantes (entre el 5 % i el 2 %). Alguns exemples són *LC/LC* ( 5,41 %) i la resta, com *ED* (economia domèstica), *FL* (filologia), *FLL* (filologia) o *FS* (filosofia), amb un 2,7 % de representativitat cadascuna. Algunes de les marques que no queden representades en els lemes nous del DIEC2 documentats en el Banc BOBNEO són *AF* (arts gràfiques), *AR* (art), *MU* (música), *PS* (psicologia) o *ZOA* (zoologia en general).

## 1.2. Dictionaris normatius de llengua castellana

Realitzem dos creuaments amb els dictionaris normatius de llengua castellana. En el primer creuament comparem els lemaris del DRAE16 i el DRAE21 (per obtenir els lemes nous del DRAE21) i, en el segon, els del DRAE21 i el DRAE22 (per obtenir els lemes nous del DRAE22).

### 1.2.1. DRAE21

En el DRAE21 detectem 1.152 lemes presents en la nomenclatura del DRAE21 i que no es recullen en la del DRAE16. En concret, detectem 854 lemes procedents de la lletra *b* i 298 de la lletra *l*. A continuació presentem una descripció de les dades obtingudes.

---

<sup>35</sup> Vegeu la taula 112 de l'Annex 4.

<sup>36</sup> Vegeu la taula 118 de l'Annex 4.

### 1.2.1.1. Recursos d'actualització lèxica

Els recursos d'actualització lèxica que detectem en els lemes del DRAR21 són 24. Entre aquests, els que presenten una freqüència més elevada són la sufixació (46,79 %), els manlleus (10,42 %) i els lemes simples (10,24 %).<sup>37</sup>

Els recursos que tenen una freqüència mitjana (entre 100 i deu ocurrences) són els lemes simples construïts (6,16 %), els compostos patrimonials (5,21 %), els compostos híbrids (4,17 %), els compostos cultes (3,73 %), la prefixació (3,21 %), la lexicalització adjectiva (2,52 %) i nominal (1,56 %), la derivació zero (1,13 %) i els noms propis (1,04 %). Per últim, els recursos d'actualització lèxica que tenen una freqüència més baixa (inferior a deu casos) són els escurçaments (0,69 %), la derivació amb sufixos taxonòmics (0,69 %), les formes prefixades (0,61 %), les formes sufixades (0,61 %), les acronímies (0,26 %), els prefixos (0,26 %), la derivació amb prefixoides (0,17 %), els lemes simples construïts en origen per sufixació (0,17 %), els sufixos (0,09 %), els manlleus amb forma de sigla (0,09 %), les variants ortogràfiques (0,09 %) i els llatinismes sintagmàtics (0,09 %).

Un cop més, observem que en el grup dels recursos d'actualització lèxica menys freqüents, la diversitat de recursos és més gran, mentre que en el grup dels recursos més freqüents, la diversitat és més escassa.

### 1.2.1.2. Manlleus

En el DRAE21 detectem 121 manlleus procedents de 20 llengües diferents, una de les quals és d'origen incert. També detectem dos manlleus que no distingim si provenen, el primer, del català o provençal i, el segon, del francès o l'italià. Entre aquests 121 manlleus la majoria són adaptats (95 manlleus) i només una minoria no han estat adaptats (26 manlleus).<sup>38</sup>

Les tres llengües més representatives són el francès (35,43 %), l'anglès (27,27 %) i l'àrab (9,09 %). En concret, la llengua que té un pes més significatiu és el francès, ja que detectem 43 manlleus, dels quals 38 són adaptats i 5 no adaptats. L'anglès és la segona llengua més representativa, amb 33 manlleus (20 d'adaptats i 13 de no adaptats). La tercera, i darrera llengua, és l'àrab, amb 11 manlleus, tots ells adaptats. L'àrab és l'única d'aquestes tres llengües que no inclou manlleus no adaptats. Aquest fet es deu a què el seu alfabet és

---

<sup>37</sup> Vegeu la taula 12 de l'Annex 4.

<sup>38</sup> Vegeu la taula 18 de l'Annex 4.

diferent a l'occidental i considerem que en tots els casos hi ha hagut una adaptació.

La resta de llengües tenen uns percentatges inferiors al 4 % i, en global, representen el 28,13 % dels manlleus. Aquestes llengües són nombroses: l'alemany (3,31 %), el català (3,31 %) i el neerlandès (2,48 %), amb una freqüència superior al 2 %; i el celta, l'afrikaans, el filipí, l'italià, el portuguès, el tahino bihao, el japonès, el caló, el bubi, el bantu, el quítxua, el rus i el taino, cadascun amb una representativitat de l'1,65 %. Observem que la tendència més habitual és incloure un nombre considerable de manlleus procedents de les llengües més predominants actualment. Per tant, s'inclouen pocs manlleus procedents de llengües minoritàries.

Classifiquem també els manlleus d'acord amb la marca temàtica, a més de seguir tenint en compte la variable d'adaptació del manlleu.<sup>39</sup> Els resultats indiquen que només hi ha deu marques temàtiques diferents. L'única marca temàtica que té més d'una repetició, en concret dues, és *informàtica*. En aquest cas, es tracta d'una repetició de la mateixa marca dins l'article lexicogràfic del mateix lema, *bit*. La resta de marques, com ja apuntem, no presenten repeticions i, per tant, cadascuna d'elles representa el 9,09 % respecte el total de marques detectades en els articles dels manlleus. Aquestes marques són *blasó*, *biologia*, *milícia*, *psicologia*, *química*, *tecnologia*, *esports*, *arquitectura* i *literatura*.

### 1.2.1.3. Marca d'ús

Detectem vuit marques d'ús diferents en 183 lemes entre els 1.152 articles lexicogràfics revisats.<sup>40</sup> Les marques d'ús més freqüents són tres: *ant.* (antic) (32,79 %), *fam.* (familiar) (24,04 %) i *fig.* (figurat) (24,04 %). Les marques menys freqüents són *p. us.* (poc utilitzat) (7,65 %), *vulg.* (vulgar) (5,46 %), *desus.* (en desús) (3,28), *rur.* (rural) (1,64 %) i *coloq.* (col·loquial) (1,09 %).

Les marques d'ús són molt escasses, de la mateixa manera que en els diccionaris de llengua catalana. Així, la tendència més habitual és no incloure en els lemes indicacions sobre l'ús (84,11 % dels casos). Aquesta tendència es manté constant en tots els diccionaris descrits.

---

<sup>39</sup> Vegeu la taula 24 de l'Annex 4.

<sup>40</sup> Vegeu les taules 30 i 36 de l'Annex 4.

#### 1.2.1.4. Marca temàtica

Detectem 206 marques temàtiques en els 1.152 articles lexicogràfics analitzats del DRAE21.<sup>41</sup> Més concretament, detectem 40 marques diferents en 159 articles lexicogràfics. Així doncs, el percentatge d'articles amb marques temàtiques és el 13,8 %.

Les marques temàtiques més freqüents, amb una freqüència igual o superior a 20 lemes, són *botànica* (20,87 %), *zoologia* (11,65 %) i *química* (9,71 %). Tan sols detectem dues marques amb una freqüència mitjana: *biologia* (8,74 %) i *medicina* (7,8 %). Les marques temàtiques que tenen una representativitat més baixa formen el grup més nombrós, i tenen una freqüència d'entre set i una ocurrència. Alguns dels exemples més representatius són *patologia* (3,4 %), *anatomia* (2,91 %), *física* (2,91 %), *lingüística* (2,43 %), *fisiologia* (2,43 %), *gramàtica* (1,94 %), etc.

Si tenim en compte les dades obtingudes en els diccionaris de llengua catalana, les marques detectades en aquest diccionari no són abundants. A més, el fenomen habitual és detectar diverses marques temàtiques amb una freqüència de repetició baixa.

#### 1.2.1.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic

A continuació consultem els lemes del DRAE21 el *Diccionario de uso del español* (DUE) de Maria Moliner.<sup>42</sup> El nombre de lemes del DRAE22 presents en el DUE és el següent:<sup>43</sup>

---

<sup>41</sup> Vegeu la taula 42 de l'Annex 4.

<sup>42</sup> Volem recordar que ni en aquest apartat sobre contrast amb corpus lexicogràfic, ni en els següents amb corpus textuals, ni en els apartats referents al mecanisme de formació, ni en els de marca temàtica, comptabilitzem per separat les marques que s'inclouen en un mateix article.

<sup>43</sup> De la mateixa manera que en l'apartat corresponent a la descripció dels diccionaris normatius catalans, el corpus lexicogràfic que seleccionem pot contenir dos entrades lexicogràfiques diferents per a un mateix lema. Per aquest motiu, en la taula es mostren les coincidències i el nombre de coincidències per entrada diferent.

<b>Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DRAE21)</b>			
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Recompte</b>	<b>Percentatge de representativitat</b>	<b>Exemples</b>
0 entrades	426	36,98 %	<i>biomedicina, limonense</i>
1 entrada	707	61,37 %	<i>balonmano, bingo</i>
2 entrades	19	1,65 %	<i>buqué, lucido</i>
<b>Total</b>	<b>1.152</b>	<b>100 %</b>	

Taula 11: Contrast entre lemes amb corpus lexicogràfic (DRAE21)

Entre els lemes nous coincidents del DRAE21, el 61,37 % té una sola entrada en el DUE. Aquesta situació és nova respecte la descripció dels diccionaris normatius catalans, que en tots els casos detectàvem com a fenomen més habitual la no-coincidència. La segona opció més habitual és la no-coincidència (36,98 %). La tercera i última situació possible és la coincidència entre un lema de nostre corpus i un lema que presenta dues entrades en el DUE. Es tracta de la situació menys habitual (1,65 %).

Mostrem, a continuació, una taula que conté la representativitat total dels lemes que presenten coincidència amb el DUE i dels que no en tenen:

<b>Representativitat dels lemes coincidents i no coincidents respecte el total de lemes (DRAE21)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Recompte</b>	<b>Percentatge de representativitat</b>
No coincidents	426	36,97 %
Coincidents	726	63,02 %
<b>Total</b>	<b>1.152</b>	<b>100 %</b>

Taula 12: Representativitat dels lemes coincidents i no coincidents respecte el total de lemes (DRAE21)

Detectem un total 22 recursos d'actualització lèxica diferents entre els 726 lemes coincidents.<sup>44</sup> Els recursos d'actualització lèxica més habituals entre els lemes coincidents són dos, la sufíxació (42,98 %) i els manlleus (13,22 %). Els recursos que tenen una representació mitjana són set. Aquests recursos representen entre un 10 % i un 2 % del total lemes coincidents i són els lemes simples (9,78 %), els lemes simples construïts (5,92 %), els compostos patrimonials (5,10 %), els compostos cultes (4,55 %) els compostos híbrids (4,55 %), la prefixació (3,72 %) i les lexicalitzacions adjectives (2,07 %). La resta

<sup>44</sup> Vegeu la taula 59 de l'Annex 4.

de recursos (16), tenen una representació inferior al 2 %. Entre aquests recursos destaquen els noms propis (1,52 %), la derivació zero (1,52 %) i les lexicalitzacions nominals (1,38 %). La resta de recursos presenten menys de 10 casos, és a dir, menys de l'1 %. Es tracta dels escurçaments, la sufixació taxonòmica, les formes sufixades, els prefixos, les formes prefixades, la derivació amb prefixoides, l'acronímia, la sintagmació amb llatinismes, els manlleus amb forma de sigla i la variació ortogràfica.

També classifiquem els lemes coincidents segons la marca temàtica<sup>45</sup>. Detectem 26 marques temàtiques i 27 seqüències de marques. Els lemes sense marca temàtica representen el 82,09 % del total analitzat. En cas que el lema coincident presenti alguna marca temàtica, les més habituals són *química* (1,65 %), *medicina* (1,38 %), *biologia* (1,24 %), *patologia* (0,96 %) i *botànica* (0,83 %). Com es pot observar, encara que aquestes marques són les més rellevants, tenen unes freqüències relatives molt baixes i unes freqüències de repetició d'entre 12 i 6 casos. La resta de marques temàtiques només es repeteixen en 3 ocasions com a màxim (la freqüència relativa més alta és del 0,41 %). Algunes d'aquestes marques són *zoologia*, *fonètica*, *cirurgia*, *fisiologia*, *marina*, *òptica*, *física*, *anatomia* o *bioquímica*. Hi ha 13 marques temàtiques del DRAE21 que no trobem representades en el DUE, com per exemple, les marques *tecnologia*, *dret* o *música*.

Quant a les seqüències de marques, entre les 27 detectades en destaquen dues, *botànica/botànica* (1,89 %) i *zoologia/zoologia* (0,83 %). La resta de seqüències, com *biologia/zoologia*, *biologia/química* o *teologia/teologia*, tenen una freqüència que no supera el 0,28 %.

#### **1.2.1.6. Contrast dels lemes dels diccionaris amb els corpus textuais**

En aquest apartat contrastem els lemes del DRAE21 en els tres corpus textuais de treball habituals.

#### ***Corpus de Referencia del Español Actual de la Real Academia Española***

El nombre de lemes coincidents entre el DRAE21 i el CREA és de 696 (6.157.185 ocurrencies) (60,41 %). Tant els diccionaris com el corpus textual pertanyen a la Real Academia Española, la mateixa institució, de manera que cal tenir en compte aquest factor per a interpretar els

---

<sup>45</sup> Vegeu la taula 66 de l'Annex 4.

resultats. Observem a continuació la representativitat dels lemes coincidents i no coincidents:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus de Referencia del Español Actual de la Real Academia Española (DRAE21)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	456	39,58 %
Coincidents	696	60,41 %
<b>Total</b>	<b>1.152</b>	<b>100 %</b>

Taula 13: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus de Referencia del Español Actual de la Real Academia Española (DRAE21)

Entre els 696 lemes coincidents trobem 18 recursos d'actualització lèxica diferents.<sup>46</sup> El recurs de formació més habitual és la sufixació (45,26 %). A continuació trobem, amb una representació mitjana, els manlleus (12,36 %) i el lemes simples (9,20 %). La resta de recursos tenen una representativitat més baixa (inferior al 6 %). Entre aquests recursos poc representatius destaquen els lemes simples construïts (5,89 %), la composició híbrida (5,46 %), la composició patrimonial (4,89 %), la composició culta (4,02 %), la prefixació (3,74 %), les lexicalitzacions adjectives (2,30 %) i les lexicalitzacions nominals (2,01 %). La resta de casos, més minoritaris (menys del 2 % de representativitat) són els noms propis (1,29 %), la derivació zero (1,15 %), la sufixació taxonòmica (1,01 %), els escurçaments (0,72 %), la derivació per prefixoide (0,29 %), els llatinismes sintagmàtics (0,14 %), les sigles manllevades (0,14 %) i les variants ortogràfiques (0,14 %). En aquest contrast detectem, un cop més, un conjunt de recursos d'actualització lèxica que no queden representats entre els lemes coincidents com, per exemple, les formes sufixades i prefixades, sufixos i prefixos, la siglació o l'acronímia.

Detectem 28 marques temàtiques diferents i 29 seqüències de marques.<sup>47</sup> El fenomen més habitual en els lemes és, un cop més, la no-inclusió de marca temàtica en el seu article lexicogràfic (85,63 %). Aquesta és una dada molt elevada si tenim en compte els resultats obtinguts amb els diccionaris de llengua catalana. Les marques temàtiques que tenen unes freqüències més altes són *química* (1,58 %), *biologia* (1,44 %) i *medicina* (1,29 %). La resta de marques representen, cadascuna d'elles, menys d'1 %. Algunes d'aquestes marques són *fisiologia*, *lingüística*, *patologia*, *òptica*, *zoologia* (0,43 %); i *bioquímica*, *esports*, *anatomia*, *impremta* o *física*, (0,29 %). Algunes

<sup>46</sup> Vegeu la taula 81 de l'Annex 4.

<sup>47</sup> Vegeu la taula 84 de l'Annex 4.



de les marques que no han quedat representades en els lemes coincidents són *cirurgia* o *bioquímica*.

Quant a les seqüències de marques, totes excepte *biologia/biologia*, presenten unes freqüències d'un sol lema, amb diverses ocurrencies en el CREA. La seqüència *biologia/biologia* la trobem en dos lemes diferents (*biosfera*, *biòtico*). Algunes de les seqüències que corresponen només a un lema (0,14 %) són *gramàtica/gramàtica/gramàtica/música*, *medicina/psicologia* o *impresma/sense marca*.

### **Corpus Tècnic de l'Institut Universitari de Lingüística Aplicada**

En aquest apartat descrivim els resultats obtinguts en la consulta dels lemes del DRAE21 en el Corpus Tècnic de l'IULA. En concret, trobem 316 lemes coincidents (15.822 ocurrencies) (27,43 %). Així doncs, el fenomen més habitual és la no-coincidència. Mostrem a continuació la representativitat dels lemes coincidents i no coincidents:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DRAE21)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	836	72,57 %
Coincidents	316	27,43 %
<b>Total</b>	<b>1.152</b>	<b>100 %</b>

Taula 14: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DRAE21)

Classifiquem els 316 lemes coincidents en 16 recursos d'actualització lèxica diferents.<sup>48</sup> Es manté constant el predomini de la sufixació (39,87 %) com el recurs més habitual, seguit dels manlleus (17,72 %). Aquests dos recursos representen més de la meitat dels lemes coincidents (57,59 %). Trobem un segon grup amb una representativitat mitjana, amb una freqüència d'entre 23 i 13 lemes. Aquests recursos són la composició híbrida (7,28 %), els lemes simples (6,96 %), els lemes simples construïts (6,01 %), la composició patrimonial (5,06 %) i la prefixació (4,11 %). Per últim, els recursos d'actualització lèxica amb una representativitat més baixa tenen una freqüència inferior als deu lemes, i són la composició culta (2,85 %), la derivació zero (2,22 %), els noms propis (2,22 %), els escurçaments (1,90 %), les lexicalitzacions nominals (1,58 %), les lexicalitzacions adjectives (1,27 %), la derivació amb prefixoide (0,32 %), les sigles manllevades (0,32 %) i els sufixos taxonòmics (0,32 %). En cap cas detectem recursos com els de l'acronímia, la siglació, els prefixos, els

<sup>48</sup> Vegeu la taula 96 de l'Annex 4.

sufixos, les formes prefixades, les formes sufixades o els lemes simples construïts en origen per sufixació.

Detectem 316 lemes coincidents que tenen marques temàtiques i que les classifiquem en 17 marques diferents i 18 seqüències de marques.<sup>49</sup> El 85,76 % dels lemes no estan marcats, de manera que és molt clara la tendència del DRAE21 d'incloure el mínim de marques temàtiques en els articles lexicogràfics. Quant als lemes que sí que estan marcats, una mateixa marca es repeteix com a màxim en quatre lemes. Les marques amb una representació més alta són *química* (1,27 %), *biologia* (0,95 %), *medicina* (0,63 %), *bioquímica* (0,63 %), *òptica* (0,63 %), *fisiologia* (0,63 %) i *lingüística* (0,63 %). Les marques menys representatives són aquelles que tenen freqüència d'1 lema (0,32 %). Algunes d'aquestes marques són *psicologia*, *patologia*, *mitologia*, *esports*, *marina* o *cirurgia*. Un cop més, hi ha una conjunt de marques que no tenen cap tipus de representació, com *geometria*, *mineria* o *música*.

Les seqüències de marques temàtiques, totes elles també presenten totes una freqüència d'1 lema. Algunes d'aquestes seqüències són *botànica/botànica/botànica*, *biologia/química*, *sense marca/anatomia/anatomia* o *medicina/psicologia*.

### **Banc BOBNEO de l'Observatori de Neologia de l'Institut Universitari de Lingüística Aplicada**

El darrer corpus que consultem és el Banc BOBNEO de l'Observatori de Neologia. En la consulta dels lemes tan sols detectem 92 lemes coincidents (401 ocurrències) (7,98 %). En la taula següent mostrem els percentatges de coincidència i no-coincidència dels lemes consultats:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DRAE21)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	1.060	92,02 %
Coincidents	92	7,98 %
<b>Total</b>	<b>1.152</b>	<b>100 %</b>

Taula 15: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DRAE21)

Classifiquem els 92 lemes coincidents segons 12 recursos d'actualització lèxica diferents. Entre aquests 12 recursos n'hi ha tres

<sup>49</sup> Vegeu la taula 102 de l'Annex 4.

que destaquen per sobre la resta: la sufixació (31,52 %), els manlleus (23,91 %) i la composició híbrida (15,22 %).<sup>50</sup> Detectem també dos recursos que tenen una representativitat mitjana i que són la prefixació (8,70 %) i la composició patrimonial (7,61 %). La resta de recursos tenen una representativitat molt escassa, amb un màxim de 3 lemes per recurs. Aquests són els lemes simples construïts (3,26 %), els lemes simples (2,17 %), els escurçaments (2,17 %), la composició culta (2,17 %), les sigles manllevades (1,09 %), les lexicalitzacions nominals (1,09 %) i les lexicalitzacions adjectives (1,09 %).

Alguns dels recursos d'actualització lèxica que no han quedat representats entre els lemes coincidents en el Banc BOBNEO són la derivació zero, els noms propis, la derivació per sufix taxonòmic o les formes prefixades i sufixades.

Descrivim a continuació els 92 lemes coincidents segons les seves marques temàtiques.<sup>51</sup> Entre aquests 92 lemes només detectem 6 articles lexicogràfics que presentin una sola marca. A més, el fenomen més habitual és que els lemes no continguin cap tipus de marca temàtica (83,70 %). Quant als articles que sí que inclouen marques temàtiques, les marques més habituals són *biologia*, *química* i *medicina*. En concret, registrem la marca *biologia* només en tres ocasions (3,26 %) i *química* i *medicina* dues vegades (2,17 %). La resta de marques i de seqüències són anecdòtiques, ja que només tenen una freqüència d'un lema (1,09 %). Les marques són *òptica*, *mitologia* i *bioquímica*. Algunes de les seqüències de marques són *informàtica/informàtica*, *impresma/sense marca* o *antropologia/biologia*.

Un cop més hi ha un conjunt de marques temàtiques que no han quedat representades entre els lemes obtinguts en la consulta del Banc BOBNEO. Algunes d'aquestes marques són *mineria*, *oceanografia*, *gramàtica*, *fonètica* o *dret*.

### 1.2.2. DRAE22

En aquest creuament obtenim 777 lemes que es recullen en la nomenclatura del DRAE22 i que no apareixen en la del DRAE21. En concret, detectem 462 lemes procedents de la lletra *b* i 315 de la lletra *l*.

---

<sup>50</sup> Vegeu la taula 114 de l'Annex 4.

<sup>51</sup> Vegeu la taula 120 de l'Annex 4.

### 1.2.2.1. Recursos d'actualització lèxica

Els lemes nous del DRAE22 estan formats a partir de 18 recursos d'actualització lèxica diferents. Els recursos que presenten una freqüència de repetició més elevada són la sufixació (51,87 %), els lemes simples (14,16 %) i els manlleus (10,94 %). Aquestes tres categories representen el 76,97 % del total de lemes.<sup>52</sup> Els recursos d'actualització lèxica que tenen una freqüència mitjana són la composició patrimonial (4,25 %), la composició culta (3,09 %), la lexicalització nominal (2,83 %), la composició híbrida (2,45 %), els lemes simples construïts (1,93 %), els noms propis (1,8 %) i la prefixació (1,42 %). Quant als recursos que tenen una freqüència baixa, aquests són els escurçaments (1,16 %), les acronímies (0,9 %), la lexicalització adjectiva (0,9 %), les formes prefixades (0,77 %), els derivats per sufixació taxonòmica (0,64 %), els compostos a la manera culta (0,39 %), la derivació zero (0,26 %) i els llatínismes sintagmàtics (0,26 %).

### 1.2.2.2. Manlleus

En el DRAE22 detectem 85 manlleus procedents de 20 llengües diferents, entre les quals n'incloem una d'origen incert. En concret, classifiquem els manlleus en 46 d'adaptats i 39 de no adaptats. Així doncs, hi ha una prevalença dels manlleus adaptats per sobre dels no adaptats.<sup>53</sup> En segon lloc, entre les 20 llengües en destaquen dues: l'anglès i el francès. L'anglès presenta 30 manlleus (35,29 %), dels quals n'hi ha 11 d'adaptats i 19 de no adaptats. Aquesta superioritat dels manlleus no adaptats no és la tendència general, ja que el més habitual és detectar-ne d'adaptats.

La resta de llengües presenten un nombre molt inferior de manlleus per a cada llengua. El nombre màxim de manlleus que detectem és quatre per al quítxua (4,71 %), tots ells adaptats. També detectem tres manlleus amb l'etiqueta d'origen incert (3,53 %), un d'ells adaptat i dos que no han patit cap modificació respecte la llengua d'origen. L'italià també presenta tres manlleus (dos adaptats i un de no adaptat). A continuació detectem un grup de llengües amb una freqüència de dos manlleus (2,35 %): l'alemany, l'aimara, el basc, el català, el filipí i el llenca. La majoria d'aquests manlleus són adaptats. Per últim, detectem un conjunt de llengües amb un sol manlleu, en la majoria de casos, també adaptats. Aquestes llengües representen només

---

<sup>52</sup> Vegeu la taula 13 de l'Annex 4.

<sup>53</sup> Vegeu la taula 19 de l'Annex 4.

l'1,18 % dels casos i són el taino, el maputxe, el japonès, el javanès, el búlgar, l'àrab, una llengua ameríndia, el maia i el cumantogo.

En la classificació dels manlleus segons si són adaptats o no adaptats en combinació amb la seva àrea temàtica, els resultats indiquen que són molt pocs els manlleus que inclouen una marca temàtica en el seu article lexicogràfic (8). I aquests vuit manlleus procedeixen només de l'anglès (50 %) i del francès (50 %).<sup>54</sup> En concret, detectem tres àrees temàtiques per al francès i quatre per a l'anglès. Les àrees temàtiques del francès són *Dep* (esports), *Fís* (física) i *Bioquím* (bioquímica). Les freqüències de repetició d'aquestes marques són de 2 manlleus adaptats en el cas dels esports (25 %), i d'un manlleu no adaptat en el cas de física (12,5 %) i de bioquímica (12,5 %). Quant a les àrees temàtiques de l'anglès, aquestes són quatre: *Inform* (informàtica, 12,5 %), *Fís* (física, 12,5 %), *Econ* (economia, 12,5 %) i *Tecnol* (tecnologia, 12,5 %). En tots els casos, excepte en *tecnologia*, els manlleus no estan adaptats.

### 1.2.2.3. Marca d'ús

En el DRAE22 detectem 12 marques d'ús diferents en 127 articles lexicogràfics. Entre aquestes 12 marques *coloq.* (col·loquial) representa gairebé el 70 % dels casos.<sup>55</sup> De la resta de marques, les més representatives són *despect.* (despectiu, 7,09 %), *vulg.* (vulgar, 6,3 %) i *p. us.* (poc utilitzat, 5,51 %). Les marques que tenen una representativitat anecdòtica són *cult.* (culte, 2,36 %), *fest.* (festiu, 2,36 %), *rur.* (rural, 1,57 %), *malson.* (malsonant, 1,57 %), *jerg.* (argot, 1,57 %), *eufem.* (eufemisme, 1,57 %), *desus.* (en desús, 0,79 %) i *irón.* (irònic, 0,79 %).

Un cop més, aquest tipus de marques són molt escasses en els articles lexicogràfics del diccionari. En la taula següent observem aquesta baixa representativitat de les marques d'ús sobre el total de lemes analitzats.

Representativitat de les marques d'ús (DRAE22)		
	Recompte	Percentatge de representativitat
Articles lexicogràfics amb marques d'ús	127	16,34 %
Articles lexicogràfics sense marques d'ús	650	83,65 %
<b>Total</b>	<b>777</b>	<b>100 %</b>

Taula 16: Representativitat de les marques d'ús (DRAE22)

<sup>54</sup> Vegeu la taula 25 de l'Annex 4.

<sup>55</sup> Vegeu la taula 37 de l'Annex 4.

#### 1.2.2.4. Marca temàtica

Les marques temàtiques que detectem en el DRAE22 són 57. En concret, detectem només 18 marques diferents.<sup>56</sup> Les marques més freqüents són *Fís* (física, 22,81 %) i *Biol* (biologia, 14,04 %). En segon lloc trobem un conjunt de marques que tenen una representativitat mitjana (entre un 11 % i un 3 %). Aquestes marques són *Quím* (química, 10,53 %), *Ling* (lingüística, 10,53 %), *Bioquím* (bioquímica, 7,02 %), *Dep* (esports, 7,02 %), *Econ* (economia, 5,26 %), *Mat* (matemàtiques, 3,51 %) i *Inform* (informàtica, 3,51 %). La resta de marques tan sols les trobem en un sol lema (1,75 %). Aquestes marques són *Arq* (arquitectura), *Arqueol* (arqueologia), *Geol* (geologia), *Zool* (zoologia), *Geom* (geometria), *Tecnol* (tecnologia), *Transp* (transports) i *Fil* (filologia).

D'acord amb el nombre i la varietat de marques temàtiques que detectem, afirmem que el DRAE22 no inclou un nombre gaire elevat d'aquest tipus de marques. A més, cap de les marques presenta una freqüència de repetició destacada, sinó que més aviat cadascuna d'elles es troba present en pocs articles lexicogràfics.

#### 1.2.2.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic

En aquest apartat consultem els lemes nous del DRAE22 en el DUE i obtenim 170 lemes coincidents. Així doncs, un cop més, la tendència habitual és la no-coincidència. (78,12 %). Observem les dades referents a la coincidència en la taula següent:

Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DRAE22)			
Coincidència i nombre d'entrades per lema coincident	Recompte	Percentatge de representativitat	Exemples
0	607	78,12 %	<i>baipás, lavarropa</i>
1	168	21,62 %	<i>bacón, limado</i>
2	2	0,26 %	<i>bue, lacería</i>
<b>Total</b>	<b>777</b>	<b>100,00 %</b>	

Taula 17: Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DRAE22)

De la mateixa manera que en el DRAE21, en cas que hi hagi coincidència només observem dues opcions: una entrada per lema

<sup>56</sup> Vegeu la taula 43 de l'Annex 4.

coincident o dues. I, entre aquestes dues opcions, la més freqüent és detectar una entrada per lema coincident.

A continuació presentem una taula que agrupa la representativitat total dels lemes coincidents amb el diccionari de contrast:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb corpus lexicogràfic (DRAE22)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Recompte</b>	<b>Percentatge de representativitat</b>
No coincidents	607	78,12 %
Coincidents	170	21,87 %
<b>Total</b>	<b>777</b>	<b>100 %</b>

Taula 18: Representativitat dels lemes coincidents i no coincidents en el contrast amb corpus lexicogràfic (DRAE22)

Registrem també les freqüències dels lemes coincidents segons la variable de recurs d'actualització lèxica.<sup>57</sup> Així, classifiquem els 170 lemes coincidents en 16 recursos d'actualització lèxica diferents. De nou, la sufixació (38,82 %) i els manlleus (24,71 %) són els recursos més habituals. Agrupem els lemes simples (7,65 %), la composició culta (7,65 %) i la composició híbrida (4,71 %) sota un mateix grup de representativitat mitjana. El darrer grup està format per els recursos d'actualització lèxica que tenen una representativitat més escassa. Entre aquests, destaquen la composició híbrida (4,65 %), els lemes simples construïts (2,35 %), la lexicalització nominal (2,35 %) i els noms propis (2,35 %). Amb una representativitat inferior al 2 % trobem els escurçaments, les formes prefixades, la composició patrimonial, l'acronímia, la sufixació taxonòmica, la lexicalització adjectiva, la prefixació i la derivació zero. Un cop més, hi ha un grup de recursos d'actualització lèxica que no queden representats entre els lemes coincidents, com la siglació, les formes sufixades, els prefixos, els sufixos o les variants ortogràfiques.

Des del punt de vista de la variable de la marca temàtica, trobem 11 marques diferents i només una seqüència de marques. Els lemes que no contenen cap tipus de marca són el 87,06 %.<sup>58</sup> A més, cap marca temàtica presenta una freqüència gaire destacada. Tot i això identifiquem cinc marques amb una freqüència superior a 1: *Biol* (biologia, 2,35 %), *Quím* (química, 1,76 %), *Fís* (física, 1,76 %), *Bioquím* (bioquímica, 1,76 %) i *Dep* (esports, 1,18 %). La resta de marques (5) tenen una freqüència de només un lema (0,59 %): *Zool* (zoologia), *Inform* (informàtica), *Geol* (geologia), *Fil* (filologia), *Econ* (economia) i *Arq* (arquitectura). Quant a les seqüències de marques,

<sup>57</sup> Vegeu la taula 61 de l'Annex 4.

<sup>58</sup> Vegeu la taula 67 de l'Annex 4.

només detectem *Fís/Geom*, amb una freqüència també d'un sol lema (0,59 %).

### 1.2.2.6. Contrast dels lemes dels diccionaris amb els corpus textuais

En aquest apartat contrastem els lemes del DRAE22 en els tres corpus textuais habituals.

#### ***Corpus de Referencia del Español Actual de la Real Academia Española***

En la consulta dels lemes del DRAE22 en el CREA, obtenim 468 lemes (31.284 ocurrències) (60,23 %). Així doncs, en aquest contrast són més habituals els lemes coincidents que els lemes no coincidents. Hem de tenir en compte que tant els diccionaris com el CREA pertanyen a la Real Academia Española, és a dir, a la mateixa institució. Aquest fet podria influir a l'hora d'actualitzar el diccionari amb els materials procedents del corpus. La taula següent mostra la representativitat dels lemes coincidents i dels no coincidents:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus de Referencia del Español Actual de la Real Academia Española (DRAE22)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	309	39,76 %
Coincidents	468	60,23 %
<b>Total</b>	<b>777</b>	<b>100 %</b>

Taula 19: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus de Referencia del Español Actual de la Real Academia Española (DRAE22)

Classifiquem els 468 lemes en 17 recursos d'actualització lèxica diferents.<sup>59</sup> Entre aquests 17 recursos en destaca especialment la sufixació (51,07 %). També són rellevants els manlleus (14,96 %) i els lemes simples (11,32 %). Quant a la resta de recursos, cap d'ells supera el 4 % de representativitat, com la composició patrimonial (3,63 %), la composició culta (3,42 %), la composició híbrida (3,21 %), la lexicalització nominal (2,35 %) i els noms propis (2,14 %). Els recursos que tenen una representativitat més escassa (inferior al 2 %), són la prefixació (1,71 %), els lemes simples construïts (1,5 %), els escurçaments (1,28 %), la lexicalització adjectiva (1,28 %), la sufixació taxonòmica (0,85 %), l'acronímia (0,43 %), la composició a la manera culta (0,43 %), els llatínismes sintagmàtics (0,21 %) i la derivació zero

<sup>59</sup> Vegeu la taula 82 de l'Annex 4.



(0,21 %). De nou, hi ha un conjunt de recursos d'actualització lèxica que no han quedat representats entre els lemes coincidents com, per exemple, la siglació, les formes sufixades i les prefixades, els sufixos i els prefixos.

Des del punt de vista de la marca temàtica detectem 16 marques diferents i una seqüència de marques. Altre cop, però, el fenomen més habitual és que l'article lexicogràfic del lema no contingui cap tipus de marca temàtica. Aquest fet es produeix en un 90,81 % dels casos. És, per tant, la situació més habitual.<sup>60</sup> En el 9,19 % dels lemes restants, la marca temàtica més representativa és *Biol* (biologia, 1,71 %) seguida de prop per *Quím* (química, 1,07 %). La resta de marques també tenen una representativitat molt escassa. Les marques que tenen més d'una repetició són *Bioquím* (bioquímica, 0,85 %), *Dep* (esports, 0,85 %), *Econ* (economia, 0,64 %) i *Inform* (informàtica, 0,43 %). La resta de marques tan sols es poden trobar en un lema (0,21 %) i són *Transp* (transportos), *Geol* (geologia), *Ling* (lingüística), *Mat* (matemàtiques), *Arq* (arquitectura), *Agr* (agricultura) i *Fil* (filologia). L'única seqüència de marques que detectem és *Fís/Geom*, en un sol lema (*baricentro*), i tan sols presenta vuit ocurrències en tot el corpus. Seguint la tendència de les consultes anteriors amb diferents corpus, no han quedat representades totes les marques possibles en la consulta que ens ocupa. En aquest cas, però, només n'ha faltat una, *Zool* (zoologia).

### **Corpus Tècnic de l'Institut Universitari de Lingüística Aplicada**

En la consulta dels lemes nous del DRAE22 en el Corpus Tècnic, detectem 306 lemes coincidents (3.296 ocurrències) (19,94 %). Aquesta dada indica que la tendència habitual és la no-coincidència (80,05 %). La taula següent inclou la representativitat dels lemes coincidents respecte el total de lemes analitzats:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DRAE22)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	622	80,05 %
Coincidents	155	19,94 %
<b>Total</b>	<b>777</b>	<b>100 %</b>

Taula 20: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DRAE22)

<sup>60</sup> Vegeu la taula 85 de l'Annex 4.

Obtenim 14 recursos diferents d'actualització lèxica.<sup>61</sup> La sufixació és, un cop més, el recurs d'actualització lèxica predominant (49,03 %). El segon recurs més representatiu és el manlleu (23,23 %). Detectem un grup de representativitat mitjana format pel recursos d'actualització següents: la composició híbrida (6,45 %), la composició culta (4,52 %), la prefixació (2,58 %) i la composició patrimonial (2,58 %). Els recursos que tenen una representativitat més escassa (inferior al 2 %) són vuit: els lemes simples construïts (1,94 %), els lemes simples (1,94 %), la lexicalització adjectiva (1,94 %), els escurçaments (1,94 %), els noms propis (1,29 %), la lexicalització nominal (0,65 %) i la derivació zero (0,65 %). No detectem recursos com l'acronímia, la siglació, els prefixos, els sufixos, les formes prefixades, les formes sufixades o els lemes simples construïts en origen per sufixació o prefixació.

D'acord amb l'anàlisi de la marca temàtica, només detectem nou marques diferents entre els 155 lemes marcats.<sup>62</sup> El fenomen més habitual, però, és que els articles no incloguin informació d'aquest tipus (86,45 %). Les marques que presenten una representativitat més alta són cinc (1,94 %), i es troben en tres lemes cadascuna d'elles. Aquestes marques són *Quím* (química), *Fís* (física), *Dep* (esports), *Bioquím* (bioquímica) i *Biol* (biologia). Les marques que presenten una representativitat més baixa són *Inform* (informàtica, 1,29 %), *Econ* (economia, 1,29 %), *Geol* (geologia, 0,65 %) i *Arq* (arqueologia, 0,65 %). No detectem en cap cas una seqüència de marques com sí ho havíem fet en les consultes amb corpus anteriors. Algunes de les marques temàtiques que tampoc han quedat representades en el articles lexicogràfics dels lemes coincidents són *Agr* (agricultura), *Fil* (filologia), *Transp* (transports), *Geom* (geometria) o *Zool* (zoologia).

### ***Banc BOBNEO de l'Observatori de Neologia de l'Institut Universitari de Lingüística Aplicada***

Per últim, consultem els lemes del DRAE22 en el Banc BOBNEO de l'Observatori de Neologia. El resultat de la consulta és 148 lemes coincidents (922 ocurrències) (19,04 %). En les consultes anteriors en aquest banc els resultats encara eren més escassos, de manera que els resultats d'aquest apartat són bastant satisfactoris. En la taula següent indiquem quina és la coincidència i la freqüència relativa dels lemes:

---

<sup>61</sup> Vegeu la taula 97 de l'Annex 4.

<sup>62</sup> Vegeu la taula 103 de l'Annex 4.

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DRAE22)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	629	80,95 %
Coincidents	148	19,04 %
<b>Total</b>	<b>777</b>	<b>100 %</b>

Taula 21: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DRAE22)

Des del punt de vista dels recursos d'actualització lèxica, classifiquem els lemes en 11 recursos diferents.<sup>63</sup> Entre aquests, destaquen dos recursos, que habitualment també destaquen en la resta de consultes: la sufixació (74 lemes, 50 %) i els manlleus (39 lemes, 26,35 %). Els recursos que considerem que tenen una representativitat mitjana són els que representen més del 2 % dels casos, com la composició híbrida (6,76 %), els lemes simples (4,05 %), la composició culta (4,05 %), la composició patrimonial (3,38 %) i la prefixació (2,03 %). Els recursos de formació que presenten uns valors més baixos són els escurçaments (1,35 %), els noms propis (0,68 %), la lexicalització adjectival (0,68 %) i la sufixació taxonòmica (0,68 %). Hi ha diversos recursos d'actualització lèxica que no han quedat representats, com la derivació zero, els lemes simples construïts, la siglació, les acronímies, els prefixos i sufixos, les formes prefixades i sufixades, etc.

Només trobem nou marques temàtiques diferents entre els 148 articles lexicogràfics dels lemes coincidents. I, un cop més, el més habitual és que els articles lexicogràfics no continguin cap marca temàtica (89,19 %).<sup>64</sup> Les marques detectades tenen una freqüència màxima de tres lemes (2,03 %), com *Econ* (economia), *Dep* (esports) i *Biol* (biologia). La resta de marques tenen una freqüència màxima de 2 lemes, i representen menys del 2 % dels casos: *Quím* (química), *Transp* (transport), *Inform* (informàtica), *Fís* (física), *Bioquím* (bioquímica) i *Agr* (agricultura). Algunes de les marques que no figuren en els articles dels lemes resultats d'aquest contrast són *Arq* (arquitectura), *Ling* (lingüística), *Geol* (geologia), *Zool* (zoologia), *Fil* (filologia) o *Tecnol* (tecnologia).

<sup>63</sup> Vegeu la taula 115 de l'Annex 4.

<sup>64</sup> Vegeu la taula 121 de l'Annex 4.

## 2. Descripció dels diccionaris d'ús

Descrivim, tot seguit, els resultats obtinguts en els creuaments en què participen diccionaris d'ús. En concret, en llengua catalana creuem els lemaris del diccionari DIEC1 i GDLC i, en llengua castellana, el DRAE22 i el DEA.

### 2.1. GDLC

En aquest apartat comprovem quines diferències presenta la nomenclatura del GDLC respecte la del DIEC1. Així, obtenim els lemes que s'inclouen en el diccionari d'ús i no en el diccionari normatiu. En concret, en el creuament DIEC1-GDLC obtenim 1.710 lemes que no figuren en el DIEC1 i sí que formen part de la nomenclatura del GDLC, dels quals 1.138 corresponen a la lletra *b* i 572 a la lletra *l*.

#### 2.1.1. Recursos d'actualització lèxica

Classifiquem els 1.710 lemes del GDLC en 27 recursos de formació diferents.<sup>65</sup> Els recursos més freqüents són la sufixació (38,60 %), la composició híbrida (12,92 %), la composició culta (11,17 %) i els manlleus (8,6 %). Aquestes categories representen, conjuntament, el 71,29 % dels lemes. Volem destacar la presència de la composició híbrida i la composició culta entre els recursos més productius. Aquests dos recursos suposen una novetat respecte els creuaments amb diccionaris normatius. En segon lloc, els recursos d'actualització lèxica que tenen una freqüència mitjana (entre un 1 % i un 5 %) són els lemes simples (4,56 %), els lemes simples construïts (4,04 %), les formes prefixades (3,33 %), la prefixació (3,16 %), la derivació per sufix taxonòmic (2,75 %), la composició patrimonial (2,51 %) i l'escurçament (1,4 %).

Per últim, el grup que té una representativitat més baixa (inferior a l'1 %), està format per les formes sufixades (0,94 %), les variants ortogràfiques (0,88 %), la derivació zero (0,82 %), les lexicalitzacions nominals (0,7 %), els símbols (0,70 %), els noms propis (0,64 %), la composició a la manera culta (0,53 %), la siglació (0,41 %), els prefixos (0,29 %), la lexicalització adjectiva (0,29 %) els lemes simples construïts en origen per sufixació (0,23 %), la sintagmació (0,18 %), l'acronímia (0,12 %), la derivació amb sufixoide (0,12 %), els sufixos (0,06 %) i els llatínismes sintagmàtics (0,06 %).

---

<sup>65</sup> Vegeu la taula 11 de l'Annex 4.

Destaquem la detecció de recursos com els símbols, la composició a la manera culta, els lemes simples construïts en origen per sufixació o la derivació amb sufixoide. Aquestes categories són poc habituals en la resta de creuaments.

### 2.1.2. Manlleus

En el GDLC detectem 147 manlleus (70 adaptats i 77 de no adaptats), que classifiquem en 40 llengües diferents.<sup>66</sup> La llengua predominant és l'anglès, amb 55 manlleus (37,41 % dels casos), seguida molt de lluny pel francès (14 manlleus, 9,52 % del total) i el castellà (10 manlleus, 6,8 %). Amb una representativitat mitjana (entre 8 i 2 casos per a cada llengua) trobem l'italià (8 manlleus, 5,44 %), l'àrab (8 manlleus, 5,44 %), una categoria d'origen incert (6 manlleus, 4,08 %), l'alemany (6 manlleus, 4,08 %), el japonès (4 manlleus, 2,72 %), el turc (3 manlleus, 2,04 %), el llatí (2 manlleus, 1,36 %) i el portuguès (2 manlleus, 1,36 %). La resta de llengües tan sols les detectem en un manlleu (0,68 %). Aquestes llengües són el basc, el batak, el beja, el bhil, el baixkir, el bembá, el bantu, el bamileke, el bhili, el balitxi, el bugui, el bambara, el buriat, el semibantu, el rus, el malai, l'italià, l'indú, l'hindi, el bohairic, el grec, el bhoties, el búlgar, l'ucraïnès, el bubí, l'afrikaans, el bobó, el bihari i l'hebreu.

També recollim les marques temàtiques que s'inclouen en els articles lexicogràfics de cada manlleu. La combinació de les variables de llengua, d'adaptació i de marca temàtica en l'estudi dels manlleus proporciona una diversitat de casos molt abundants, ja que la tendència habitual és la diversitat en la combinació de les variables però no detecció abundant de manlleus amb una combinació determinada. Així, detectem 121 combinacions diferents, resultants de combinar 35 llengües i 54 marques temàtiques diferents. La combinació més productiva és la de manlleus no adaptats de l'anglès de l'àmbit dels esports, amb nou manlleus i un 5,88 % de representativitat. Altres combinacions rellevants són els manlleus no adaptats de l'anglès de l'àmbit de música (sis manlleus, 3,92 %); els manlleus no adaptats d'origen incert, de l'àmbit de l'etnologia (4 manlleus i un 2,61 %). Detectem també un bon nombre de combinacions possibles amb una freqüència de dos casos. En la majoria d'aquestes combinacions la llengua d'origen del manlleu és l'anglès, en alguns casos adaptats i en d'altres no, i algunes de les marques temàtiques que s'inclouen en els articles d'aquests manlleus són la *dansa*, la *geomorfologia*, l'*economia*, l'*aeronàutica* o la

---

<sup>66</sup> Vegeu les taules 17 i 23 de l'Annex 4.

*publicitat*. Altres llengües que també presenten dos casos són l'àrab, el francès, l'alemany, el portuguès o el turc.

Les combinacions que tan sols presenten un cas són les més nombroses, amb 103 combinacions diferents. La llengua predominant, un cop més, és l'anglès (25 combinacions diferents, 14 de no adaptades i 11 d'adaptades). Algunes de les marques temàtiques detectades entre aquests manlleus de l'anglès són *jocs*, *meteorologia*, *transports*, *astrofísica*, *música*, *història*, *cirurgia*, *dansa* o *alimentació*. El francès, amb deu manlleus, l'italià amb set i l'alemany amb sis també tenen una representació elevada, superior a la resta de llengües. En la majoria dels casos els manlleus no han estat adaptats. Les marques temàtiques més freqüents entre els manlleus procedents del francès són *història*, *música*, *tèxtil*, *zoologia* o *alimentació*. En el cas de l'italià, algunes de les marques són *botànica*, *esports*, *història* o *música*. I, en els manlleus de l'alemany, trobem marques com *tèxtil*, *política*, *música* o *literatura*.

### 2.1.3. Marca d'ús

Només detectem dues marques d'ús en el GDLC: *dial* (dialectal) i *col·loq* (col·loquial).<sup>67</sup> La marca *dial* és la més recurrent (37 ocurrències) en front a *col·loq* (7 ocurrències). La detecció de 44 marques d'ús en els 1.710 articles lexicogràfics analitzats indica que les marques d'ús són anecdòtiques en aquest diccionari, ja que només es troben presents en un 2,57 % dels articles lexicogràfics dels lemes. Aquesta tendència es manté constant en la resta de diccionaris descrits.

### 2.1.4. Marca temàtica

Les marques temàtiques dels lemes del GDLC són 170.<sup>68</sup> Aquesta dada és molt elevada i mostra una classificació molt exhaustiva. Entre totes les marques en destaquen dues, *BOT* (botànica, 8,65 %) i *PAT* (patologia, 8,28 %). Les marques que tenen una representativitat mitjana (entre un 2 % i un 5 %) són *ZOOL* (zoologia, 4,69 %), *BIOL* (biologia, 3,23 %), *HIST* (història, 2,86 %), *QUIM ORG* (química orgànica, 2,71 %), *BIOQ* (bioquímica, 2,71 %), *ENTOM* (entomologia, 2,64 %), *MINERAL* (mineralogia, 2,57 %), *LING* (lingüística, 2,42 %). Totes juntes agrupen el 23,83 % del total de marques.

---

<sup>67</sup> Vegeu la taula 29 i 35 de l'Annex 4.

<sup>68</sup> Vegeu la taula 41 de l'Annex 4.

El 40,76 % de les marques restants tenen, cadascuna, una representativitat molt escassa. Aquest grup, però, és el més nombrós (160 categories diferents). Entre aquestes, les més representatives són *FARM* (farmàcia, indústria farmacèutica, 1,91 %), *MUS* (música, 1,76 %), *ESTRATIG* (estratigrafia, 1,76 %), *ESPORT* (esports, 1,61 %), *ALIM* (alimentació, indústries alimentàries, 1,61 %), *ETNOL* (etnologia, 1,54 %), *MICROB* (microbiologia, 1,39 %), *FISIOL* (fisiologia), *BIOTEC* (biotecnologia, 1,32 %), *ICT* (ictiologia, 1,32 %), *HERALD* (heràldica, 1,25 %), *MAR* (marina, marítim, 1,25 %), *QUÍM INORG* (química inorgànica, 1,17 %), *METROL* (metrologia, 1,17 %), *CIR* (cirurgia, 1,17 %) i *FISIOL ANIM* (fisiologia animal, 1,1 %).

### 2.1.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic

En aquest apartat comprovem quins lemes procedents del GDLC figuren en el leuari del DCVB. En la taula següent mostrem els resultats obtinguts en la comparació. Mantenim desglossats els lemes que contenen més d'una entrada en el DCVB:

<b>Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (GDLC)</b>			
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Recompte</b>	<b>Percentatge de representativitat</b>	<b>Exemples</b>
0 entrades	1.292	75,56 %	<i>benchmarking, limfografia</i>
1 entrada	363	21,23 %	<i>baronessa, limitat</i>
2 entrades	46	2,69 %	<i>batí, lleonet</i>
3 entrades	6	0,35 %	<i>benet, llorença</i>
4 entrades	2	0,12 %	<i>breç, brinca</i>
5 entrades	1	0,06 %	<i>llaça</i>
<b>Total</b>	<b>1.710</b>	<b>100 %</b>	

Taula 22: Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (GDLC)

La no-coincidència és la tendència habitual entre els lemes del GDLC i els del leuari del DCVB (75,56 %). El segon fenomen més habitual és que coincideixin lemes que només tenen una entrada en el DCVB (21,23 %). A continuació mostrem els percentatges de coincidència amb el diccionari de contrast:

<b>Representativitat dels lemes coincidents i no coincidents respecte el total de lemes (GDLC)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Recompte</b>	<b>Percentatge de representativitat</b>
No coincidents	1.292	75,56 %
Coincidents	418	24,44 %
<b>Total</b>	<b>1.710</b>	<b>100 %</b>

Taula 23: Representativitat dels lemes coincidents i no coincidents respecte el total de lemes (GDLC)

Com es pot observar, el nombre de lemes no coincidents és major que el nombre de lemes coincidents ja que, com apuntem anteriorment, hi ha una diferència temporal en la publicació de les dues obres i el GDLC té una voluntat estandarditzadora.

Classifiquem els lemes en 20 recursos d'actualització lèxica diferents.<sup>69</sup> Entre aquests 20 recursos, en destaca un de particularment productiu, la sufixació (56,94 %). La resta de recursos són menys representatius i cap d'ells mostra un percentatge gaire elevat. Per exemple, els manlleus que tenen uns percentatges de representativitat mitjana (entre un 7,89 % i un 2,39 %) són els lemes simples (7,89 %) seguits de la composició híbrida (5,5 %). Detectem també la composició culta (4,78 %), els lemes simples construïts (3,83 %), la prefixació (3,35 %), les variants ortogràfiques (3,11 %) i la composició patrimonial (2,39 %). Finalment també obtenim un conjunt de recursos d'actualització lèxica menys representatius, amb uns percentatges inferiors al 2 %. Aquests recursos són la derivació zero (1,44 %), els escurçaments (1,2 %), la lexicalització nominal (1,2 %), els lemes simples procedents de nom propi (0,72 %), els lemes simples construïts en origen per sufixació (0,72 %), la sufixació taxonòmica (0,48 %), els símbols (0,48 %), la lexicalització adjectiva (0,48 %), l'acronímia (0,24 %), la derivació per sufixoide (0,24 %) i la siglació (0,24 %). Hi ha alguns recursos de formació que no detectem entre els lemes coincidents. Alguns d'aquests recursos són les formes prefixades i sufixades, els prefixos i els sufixos, composició a la manera culta, parasíntesi, etc.

Quant a les marques temàtiques, classifiquem els lemes del GDLC en 62 marques soles en l'article lexicogràfic i 27 seqüències de marques temàtiques.<sup>70</sup> La tendència habitual de no marcar els lemes amb aquesta informació es manté també en aquest contrast com el fenomen més habitual (54,07 %).

<sup>69</sup> Vegeu la taula 59 de l'Annex 4.

<sup>70</sup> Vegeu la taula 65 de l'Annex 4.



Quant als lemes que sí que incorporen aquest tipus d'informacions, les marques més representatives, tot i tenir un percentatge relatiu baix (entre un 6,46 % i un 1,2 %), són *BOT* (botànica 6,46 %), *PAT* (patologia, 2,39 %), *MAR* (marina, marítim (2,39 %), *ZOOL* (zoologia, 1,91 %), *ENTOM* (entomologia, 1,67 %), *QUÍM INORG* (química inorgànica, 1,2 %), *BIOL* (biologia, 1,2 %) i *HIST* (història, 1,2 %). La resta de marques tenen una representativitat molt baixa, inferior als quatre lemes, és a dir, a l'1 %. Algunes d'aquestes marques són *METEOR* (meteorologia, 0,96 %), *JOCS* (jocs d'entreteniment, 0,72 %), *AGR* (agricultura, 0,24 %), *TRANSP* (transports, 0,24 %), *GASTR* (gastronomia, 0,24 %) o *psicologia* (0,24 %). A banda de les marques que detectem en aquest contrast amb el DCVB, n'hi ha que figuren en el GDLC i que no trobem entre els lemes coincidents. Algunes d'aquestes marques són *ANAT* (anatomia), *GRAF* (arts gràfiques), *JARD* (jardineria) o *ÓPT* (òptica). De les 27 seqüències de marques que detectem no n'hi ha cap que tingui una freqüència superior a un lema (0,24 %). Algunes d'aquestes seqüències són */BOT/FARM/ALIM*, */JOCS/ECON* o */FARM/QUÍM ORG*.

#### **2.1.6. Contrast dels lemes dels diccionaris amb els corpus textuais**

Presentem en aquest apartat els resultats relatius a la consulta dels lemes nous del GDLC en els tres corpus textuais de contrast habituals.

##### ***Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans***

La coincidència entre els lemes del GDLC i el CTILC és de 372 lemes (12.553 ocurrències). Així doncs, el nombre de lemes no coincidents (78,24 %) és superior al de lemes coincidents (21,75 %). S'ha de tenir en compte que el GDLC i el CTILC no pertanyen a la mateixa institució. El GDLC és una obra de l'editorial Enciclopèdia Catalana i el CTILC pertany a l'Institut d'Estudis Catalans. Presentem en la taula següent els resultats numèrics i els percentatges referents a la consulta:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans (GDLC)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	1.338	78,24 %
Coincidents	372	21,75 %
<b>Total</b>	<b>1.710</b>	<b>100 %</b>

Taula 24: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Textual Informatitzat de la Llengua Catalana de l'Institut d'Estudis Catalans (GDLC)

En la classificació dels 372 lemes coincidents segons el recurs d'actualització lèxica obtenim 22 recursos diferents.<sup>71</sup> Entre aquests, destaca, un cop més, la sufixació (36,29 %). Els recursos que també destaquen per la seva productivitat són els manlleus (12,9 %), la composició híbrida (11,29 %) i la composició culta (10,48 %). Cadascun d'aquests recursos representa més del 10 % dels casos. Els recursos que tenen una representativitat mitjana són aquells que presenten entre 24 i 4 lemes: els lemes simples (6,45 %), els lemes simples construïts (6,18 %), la prefixació (3,76 %), la derivació per sufix taxonòmic (2,69 %), la composició patrimonial (2,42 %), les variants ortogràfiques (1,61 %) i la lexicalització adjectiva (1,08 %).

Considerem que la resta de recursos tenen una representativitat anecdòtica. Aquests recursos són l'escurçament (0,81 %), la derivació zero (0,81 %), la lexicalització nominal (0,54 %), els símbols (0,54 %), els lemes simples que en origen eren noms propis (0,54 %), l'acronímia (0,27 %), la sintagmació (0,27 %), la sufixació amb sufixoide (0,27 %), la siglació (0,27 %), la composició a la manera culta (0,27 %) i els lemes simples construïts en origen per sufixació (0,27 %). Un cop més hi ha uns quants recursos d'actualització lèxica que no han quedat recollits entre els lemes coincidents en el contrast amb el CTILC. Aquests recursos són les formes prefixades i sufixades, els prefixos i sufixos o la parasíntesi.

Quant a les marques temàtiques, en detectem 83 de soles en l'article lexicogràfic i 40 seqüències de marques. En aquest cas, el 25,27 % dels lemes que no presenten cap marca temàtica en el seu article lexicogràfic.<sup>72</sup> Les marques més habituals són *PAT* (patologia, 8,6 %), *BOT* (botànica, 4,57 %), *ESTRATIG* (estratigrafia, 2,96 %), *ZOOL* (zoologia, 1,88 %), *HIST* (història, 1,88 %), *FISIOLOG ANIM* (fisiologia animal, 1,61 %), *BIOQ* (bioquímica, 1,61 %), *FARM* (farmàcia,

<sup>71</sup> Vegeu la taula 74 de l'Annex 4.

<sup>72</sup> Vegeu la taula 77 de l'Annex 4.

1,34 %), *DIAG* (diagnosi, 1,34 %) i *LING* (lingüística, 1,34 %). Algunes de les marques de representació mitjana són *JOCS* (jocs d'entreteniment, 1,08 %), *DRET* (dret, 1,08 %), *ENTOM* (entomologia, 1,08 %), *MAT* (matemàtiques, 0,54 %), *GEOMORF* (geomorfologia, 0,54 %), *PEDOL* (pedologia, 0,54 %) i *PESC* (pesca, 0,54 %). Les marques de representació baixa són aquelles que tan sols es troben presents en un sol lema (0,27 %), com *ECON* (economia), *EMBAL* (embalatge), *ESPORT* (esports), *AGR* (agricultura), *PSIC* (psiquiatria), *MED* (medicina), *SOCIOL* (sociologia) o *REL* (religió). Les seqüències de marques es troben, normalment, en un sol lema, excepte en tres ocasions: */ETNOL/LING* (5 lemes, 1,34 %), */QUÍM ORG/FARM* (3 lemes, 0,81 %) i */-ETNOL/LING* (3 lemes, 0,81 %).

### **Corpus Tècnic de l'Institut Universitari de Lingüística Aplicada**

Tot seguit presentem els resultats de la consulta dels lemes nous del GDLC en el Corpus Tècnic de l'IULA. En aquesta consulta detectem 119 lemes coincidents (2.174 ocurrències). Així doncs, els lemes coincidents representen tan sols el 6,95 % dels lemes consultats. La taula següent mostra la representativitat dels lemes coincidents i no coincidents:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (GDLC)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	1.591	93,04 %
Coincidents	119	6,95 %
<b>Total</b>	<b>1.710</b>	<b>100 %</b>

Taula 25: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (GDLC)

El recurs d'actualització lèxica més habitual segueix sent la sufixació, (31,09 %), seguida dels manlleus (28,57 %) i de la composició híbrida (11,76 %). Aquests tres recursos de formació agrupen el 71,42 % dels lemes.<sup>73</sup> Els recursos que tenen una representativitat mitjana són els lemes simples construïts (5,04 %), lemes simples (5,04 %), prefixació (4,2 %) i composició patrimonial (4,2 %). Amb una freqüència baixa, inferior a cinc lemes, detectem l'escurçament (2,52 %), la composició culta (2,52 %), els lemes simples que en origen eren noms propis (1,68 %), les lexicalitzacions adjectives (1,68 %), les lexicalitzacions nominals (0,84 %) i les acronímies (0,84 %). En cap cas trobem recursos com la siglació, els símbols, les variants ortogràfiques, la parasíntesis o les formes prefixades i sufixades.

<sup>73</sup> Vegeu la taula 95 de l'Annex 4.

Entre els 119 lemes analitzats detectem 39 marques temàtiques soles en l'article lexicogràfic i 16 seqüències de marques. Això significa que hi ha un 30,25 % dels lemes que no inclouen aquest tipus d'informació.<sup>74</sup> Les marques més freqüents dels lemes que sí que inclouen aquesta informació són *PAT* (patologia, 5,04 %), *ESPORT* (esports, 3,36 %) i *POLÍT* (política, 3,36 %).<sup>75</sup> Algunes de les marques que tenen una freqüència mitjana (apareixen a entre dos i tres lemes) són, per exemple, *ALIM* (alimentació, 2,52 %), *BIOL* (biologia, 2,52 %), *BIOTEC* (biotecnologia, 2,52 %), *JOCS* (jocs d'entreteniment, 1,68 %), *BOT* (botànica, 1,68 %) o *ZOOL* (zoologia, 1,68 %). Les marques temàtiques que presenten una freqüència més baixa són aquelles que només apareixen en un sol lema (i tenen una freqüència relativa del 0,84 %). Aquestes marques són les més nombroses. Alguns exemples són *PSIC* (psicologia), *COSM* (cosmètica), *TECNOL* (tecnologia), *TELECOM* (telecomunicacions), *MOBL* (mobiliari), *MAR* (marina, marítim), *LING* (lingüística), *INFORM* (informàtica), *INDUM* (indumentària) o *MÚS* (música). Algunes de les seqüències de marques que detectem són */-ETNOL/LING* (1,68 %), */ETNOL/LING* (1,68 %), */CIR/TECNOL* (0,84 %), */BOT/JARD* (0,84 %) o */BIOTEC/MED* (0,84 %).

### ***Banc BOBNEO de l'Observatori de Neologia de l'Institut Universitari de Lingüística Aplicada***

El darrer corpus on consultem els lemes del GDLC és el Banc BOBNEO. En concret, obtenim 117 lemes coincidents (1.534 ocurrències). Els lemes coincidents són, doncs, escassos, tan sols el 6,84 % respecte del total contrastat. En la taula següent mostrem els percentatges relatius a la coincidència:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (GDLC)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	1.593	93,15 %
Coincidents	117	6,84 %
<b>Total</b>	<b>1.710</b>	<b>100 %</b>

Taula 26: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (GDLC)

<sup>74</sup> Vegeu la taula 101 de l'Annex 4.

<sup>75</sup> Observem que l'abundant nombre de marques i de seqüències comporta que els percentatges de representativitat de cada marca o seqüència siguin baixos.

Classifiquem els 117 lemes coincidents en 17 recursos de formació diferents.<sup>76</sup> Entre aquests 17 recursos en destaquen dos: els manlleus (48,72 %), i la sufixació (24,79 %). Aquests dos recursos junts representen el 73,51 % del total de lemes coincidents. Volem destacar el fet que siguin els manlleus i no la sufixació el recurs més present entre els lemes coincidents.

La resta de recursos presenten unes freqüències molt inferiors. Els recursos que tenen una freqüència mitjana (d'entre cinc i dos lemes) són els escurçaments (4,27 %), la composició híbrida (3,42 %), els lemes simples (2,56 %), la derivació zero (2,56 %), la composició patrimonial (2,56 %), la composició culta (1,71 %), la lexicalització nominal (1,71 %) i la siglació (1,71 %). Els recursos que tenen una representativitat més baixa són els que només trobem en un lema (0,85 %): els llatínismes sintagmàtics, les variants ortogràfiques, la sufixació taxonòmica, la derivació, els noms propis que passen a noms comuns, la sintagmació i la lexicalització adjectiva. No detectem recursos com els compostos formats a la manera culta, la prefixació, la sufixació, les formes prefixades, les formes sufixades, la parasítesi, les lexicalitzacions adjectives o l'acronímia.

Segons la marca temàtica, detectem 36 marques soles en l'article lexicogràfic i 18 seqüències de marques.<sup>77</sup> El percentatge de lemes sense marques és el 27,35 %. Entre els lemes que sí que presenten alguna marca temàtica, en destaquen els dels àmbits *ESPORT* (esports, 9,40 %), *MÚS* (música, 5,13 %), *ALIM* (alimentació, 3,42 %) i *POLÍT* (política, 3,42 %). Aquelles marques que tenen una representació mitjana es troben en dos lemes (1,71 %) i són *LING* (lingüística), *PUBL* (publicitat), *ETNOL* (etnologia), *ZOOL* (zoologia), *ECON* (economia), *TAUROM* (tauromàquia) i *JOCS* (jocs d'entreteniment). Les resta de marques tan sols les trobem en un lema (0,85 %): *ESPECT* (espectacles), *EMBAL* (embalatge), *COSM* (cosmètica), *MAT* (matemàtiques), *TRANSP* (transports), *TELECOM* (telecomunicacions), *SOCIOL* (sociologia), *ISLAM* (islam), *INFORM* (informàtica) o *INDUM* (indumentària). Com ja apuntem anteriorment, també detectem seqüències de marques. Algunes d'aquestes seqüències són */MÚS/DANSA* (2,56 %), */ELECTROAC/MÚS* (1,71 %), */ECOL/GEN* (0,85 %), */ASSEG/ECON* (0,85 %) o */QUÍM ORG/FARM* (0,85 %).

---

<sup>76</sup> Vegeu la taula 113 de l'Annex 4.

<sup>77</sup> Vegeu la taula 119 de l'Annex 4.

## 2.2. DEA

En el DEA detectem 1.544 lemes que no figuren en el leuari del DRAE22. En concret, 963 lemes procedeixen de la lletra *b* i 581 de la lletra *l*. Descriuim a continuació aquests lemes segons les variables d'anàlisi habituals.

### 2.2.1. Recursos d'actualització lèxica

Deteclem 22 recursos d'actualització lèxica entre els 1.544 lemes del DEA.<sup>78</sup> Els recursos més freqüents són la sufixació (46,57 %), els manlleus (15,03 %) i els lemes simples (11,59 %). Els recursos d'actualització lèxica que presenten una freqüència de repetició mitjana són la composició híbrida (6,35 %), la composició culta (5,18 %), la composició patrimonial (2,46 %), la prefixació (2,07 %), els lemes simples construïts (2,01 %), la sufixació taxonòmica (1,75 %), la lexicalització amb adjectiu (1,49 %), els noms propis (1,49 %) i l'escurçament (0,78 %). Per últim, els recursos que tenen una freqüència més baixa són els llatinismes sintagmàtics (0,45 %), la sintagmació (0,39 %), la composició a la manera culta (0,32 %), l'acronímia (0,19 %), les formes prefixades (0,19 %), els lemes simples construïts formats en origen per sufixació (0,19 %), la siglació (0,13 %), la derivació zero (0,06 %) i la prefixació (0,06 %).

### 2.2.2. Manlleus

Classifiquem els 232 manlleus del DEA en 27 manlleus adaptats i 205 manlleus no adaptats. Així doncs, els manlleus no adaptats són més habituals que els manlleus adaptats.<sup>79</sup> Entre aquests 232 manlleus detectem 20 llengües diferents i una categoria d'origen incert. Entre aquestes 21 llengües, les més habituals són l'anglès (53,88 %) i el francès (21,98 %). La resta de llengües tenen uns percentatges de representativitat inferiors al 5 %. Entre aquestes, destaquen l'alemany (4,74 %), la categoria d'origen incert (4,74 %) i l'italià (3,45 %). És destacable el fet que totes aquestes llengües presenten més manlleus adaptats que no adaptats excepte en el cas de la categoria d'origen incert.

Les llengües que habitualment inclouen alguna marca temàtica en els seus articles són l'anglès, el francès, l'italià, l'alemany i el català. En concret, cap marca es repeteix més de quatre vegades en diferents

---

<sup>78</sup> Vegeu la taula 14 de l'Annex 4.

<sup>79</sup> Vegeu les taules 20 i 26 de l'Annex 4.

articles, i el més habitual és una sola ocurrència en un sol article. Les marques més habituals són *Informát* (informàtica, de l'anglès, 7,27 %), *Mús* (música, de l'anglès, 7,27 %), *Golf* (de l'anglès, 7,27 %) i *Mús* (música, de l'italià, 5,45 %). La marca de *música* té, doncs, un pes rellevant entre els manlleus analitzats d'aquest diccionari. Per últim, volem destacar que en un manlleu de català (*blaugrana*) hi apareix la marca *Fút* (futbol).

### 2.2.3. Marca d'ús

Detectem 470 marques d'ús en els articles lexicogràfics del DEA i les classifiquem en 13 marques diferents.<sup>80</sup> Tot i això, predominen els articles lexicogràfics sense marca d'ús (69,55 %). Les marques d'ús més freqüents són *reg* (regional, 33,19 %) i *raro* (estrany, 20,43 %). Aquestes dues marques representen el 53,62 % del total. Les marques que tenen una freqüència mitjana són *lit* (literal, 12,55 %), *col* (col·loquial, 10,64 %), *hist* (històric, 8,09 %), *jerg* (argot, 5,53 %), *hoy raro* (avui estrany, 3,40 %) i *despe* (despectiu, 3,19 %). Les marques que presenten una freqüència més baixa són *juv* (juvenil, 1,28 %), *humoríst* (humorístic, 0,85 %), *pop* (popular, 0,43 %), *vulg* (vulgar, 0,21 %), *jerg Econ* (argot d'economia, 0,21 %).

### 2.2.4. Marca temàtica

Detectem 495 marques temàtiques en els articles lexicogràfics dels lemes DEA, que classifiquem en 70 marques diferents.<sup>81</sup> En concret, aquestes 495 marques es troben en 465 articles lexicogràfics diferents (30,11 %). Les marques temàtiques més habituals en els lemes són *Med* (medicina, 17,98 %), *Biol* (biologia, 10,91 %) i *Quím* (química, 10,91 %). Aquestes tres marques juntes representen el 39,8 % del total de marques.

Les marques que tenen una freqüència mitjana (entre 38 i cinc lemes) són *Bot* (botànica, 7,68 %), *Zool* (zoologia, 4,24 %), *Anat* (anatomia, 4,04 %), *E* (especialitats varies, 3,23 %), *Geol* (geologia, 3,03 %), *Pol* (política, 2,63 %), *Mus* (música, 2,02 %), *Mar* (marina, 1,62 %), *Econ* (economia, 1,41 %), *Psicol* (psicologia, 1,21 %) o *Rel catól* (religió catòlica, 1,01 %).

La resta de marques tenen una freqüència relativa molt baixa (inferiors a l'1 %) i formen el grup més nombrós. Algunes d'aquestes marques són *Naipes* (cartes, 0,81 %), *Informat* (informàtica, 0,81 %),

<sup>80</sup> Vegeu les taules 32 i 38 de l'Annex 4.

<sup>81</sup> Vegeu la taula 44 de l'Annex 4.

*Arqueol* (arqueologia, 0,81 %), *Cerám* (ceràmica, 0,20 %), *Hist* (història, 0,20 %) o *Geometria* (0,20 %).

### 2.2.5. Contrast dels lemes dels diccionaris amb el corpus lexicogràfic

En aquest apartat consultat els lemes nous del DEA en el leuari del DUE. El resultat del contrast és el següent:

<b>Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DEA)</b>			
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Recompte</b>	<b>Percentatge de representació</b>	<b>Exemples</b>
0 entrades	1.336	86,52 %	<i>barullón,</i> <i>laissez-faire</i>
1 entrada	203	13,14 %	<i>basket,</i> <i>laserterapia</i>
2 entrades	5	0,32 %	<i>beréber,</i> <i>lepero</i>
<b>Total</b>	<b>1.544</b>	<b>100,00 %</b>	

Taula 27: Contrast dels lemes dels diccionaris amb el corpus lexicogràfic (DEA)

El fenomen més habitual és la no-coincidència entre els dos leuaris (86,52 %). La segona opció més freqüent és que el lema del DEA coincideixi amb un lema que té una sola entrada del DUE (13,14 %). En fenomen menys habitual és que un lema del DEA coincideixi amb un lema del DUE i que aquest presenti dues entrades (0,32 %).

A continuació mostrem el total de lemes que coincideixen amb el leuari del DUE i el total de lemes que no coincideixen.

<b>Representativitat dels lemes coincidents i no coincidents respecte el total de lemes (DEA)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Recompte</b>	<b>Percentatge de representativitat</b>
No coincidents	1.336	86,52 %
Coincidents	208	13,47 %
<b>Total</b>	<b>1.544</b>	<b>100,00 %</b>

Taula 28: Representativitat dels lemes coincidents i no coincidents respecte el total de lemes (DEA)

Classifiquem els lemes coincidents en 17 recursos d'actualització lèxica diferents.<sup>82</sup> Els recursos d'actualització lèxica més recurrents són els manlleus (32,21 %), la sufixació (27,40 %) i els lemes simples (13,46 %). Aquests tres recursos junts representen el 73,07 % dels casos. Volem destacar el fet que els manlleus sigui el recurs més

<sup>82</sup> Vegeu la taula 62 de l'Annex 4.



habitual per davant de la sufixació. Els recursos que tenen una representativitat mitjana són la composició culta (3,85 %), la composició híbrida (3,37 %), la composició patrimonial (3,37 %), la lexicalització adjectiva (2,88 %), els lemes simples construïts (2,40 %) i la prefixació (2,40 %). La resta de recursos tenen unes freqüències relatives inferiors al 2 % i formen el grup més nombrós: els lemes simples que eren noms propis en origen (1,92 %), la sufixació taxonòmica (1,92 %), la lexicalització nominal (1,44 %), els escurçaments (0,96 %), la sintagmació (0,96 %), els llatinismes sintagmàtics (0,48 %), la siglació (0,48 %) i les formes prefixades (0,48 %). Quant als recursos que trobem a faltar, alguns d'ells són les formes sufixades, els prefixos, els sufixos, les variants ortogràfiques, l'acronímia, la siglació, la composició a la manera culta o la parasíntesi.

Detectem 22 marques que es troben soles en l'article lexicogràfic i cinc seqüències de marques.<sup>83</sup> El percentatge de lemes sense marca és el 72,12 %. Es tracta, per tant, de la majoria dels lemes. En cas que el lema inclogui alguna marca temàtica, la més habitual és *Quím* (química, 3,85 %), seguida de *Biol* (biologia, 3,37 %), *Bot* (botànica, 2,40 %), *Med* (medicina, 1,92 %), *Mus* (música, 1,92 %), *Zool* (zoologia, 1,92 %). Aquestes marques, a més de tenir unes freqüències relatives molt baixes, es troben també en un nombre de lemes escàs (entre vuit i quatre). La resta de marques tan sols es troben en un o dos lemes. Algunes d'aquestes són *Dep* (esports, 0,96 %), *Geol* (geologia, 0,96 %), *Golf* (0,96 %), *Heràld* (heràldica, 0,96 %), *Anat* (anatomia, 0,48 %), *Ceram* (ceràmica, 0,48 %) o *Econ* (economia, 0,48 %). Tan sols trobem cinc seqüències de marques amb unes freqüències relatives del 0,48 %, com *(Golf)(Mil)(Pol)*, *(Med)(E)(E)*, *(Ferroc)(Boxeo)(Tenis)*, *(Pol)(Com)* i *(Rel catól)*.

Com és habitual, hi ha un nombre de marques que no es troben presents entre els lemes coincidents i que sí que detectem en altres lemes del DEA. Algunes d'aquestes marques són *Geom* (geometria), *Imprt* (impremta), *Mar* (marina) o *Rel* (religió).

#### **2.2.6. Contrast dels lemes dels diccionaris amb els corpus textuais**

Contrastem, en aquest apartat, els lemes nous del DEA amb els tres corpus textuais de treball habituals.

---

<sup>83</sup> Vegeu la taula 68 de l'Annex 4.

### **Corpus de Referencia del Español Actual de la Real Academia Española**

Hi ha 807 lemes coincidents entre els DEA i el CREA (38.033 ocurrences) (52,27 %). Hem de tenir en compte que el DEA i el CREA no pertanyen a la mateixa institució i, per tant, que comparteixin més de la meitat dels lemes contrastats és un fet remarcable. En la taula següent observem la representativitat dels lemes coincidents i no coincidents:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus de Referencia del Español Actual de la Real Academia Española (DEA)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	737	47,73 %
Coincidents	807	52,27 %
<b>Total</b>	<b>1.544</b>	<b>100 %</b>

Taula 29: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus de Referencia del Español Actual de la Real Academia Española (DEA)

Classifiquem els 807 lemes coincidents en 19 recursos d'actualització lèxica diferents.<sup>84</sup> El recurs d'actualització lèxica més habitual és la sufixació (40,02 %). Els manlleus també tenen un paper destacat (23,54 %). Aquests dos recursos representen el 63,56 % dels lemes coincidents. Els recursos que tenen una representativitat mitjana són els lemes simples (9,91 %), la composició híbrida (6,69 %), la composició culta (4,71 %), la composició patrimonial (2,35 %) i la derivació (2,23 %).

Els recursos restants no arribem al 2 % de representativitat: els lemes simples que eren noms propis en origen (1,98 %), la sufixació taxonòmica (1,61 %), la lexicalització adjectiva (1,36 %), els lemes simples construïts (1,36 %), la lexicalització nominal (1,36 %), els escurçaments (1,12 %), els llatinismes sintagmàtics (0,74 %), els compostos formats a la manera culta (0,25 %), la siglació (0,25 %), la sintagmació (0,25 %), l'acronímia (0,12 %) i la derivació zero (0,12 %). Els recursos que no han quedat representats entre els lemes coincidents són els habituals: les formes prefixades i sufixades, els prefixos, els sufixos i les variants ortogràfiques.

En relació a la marca temàtica, detectem 49 marques soles en els articles lexicogràfics dels lemes i 15 seqüències de marques.<sup>85</sup> El

<sup>84</sup> Vegeu la taula 83 de l'Annex 4.

<sup>85</sup> Vegeu la taula 86 de l'Annex 4.

percentatge de lemes sense marcar és el 69,39 % i és, per tant, el fenomen més habitual. Entre els lemes que sí que inclouen algun tipus de marca, les més habituals són *Med* (medicina, 6,07 %), *Quím* (química, 3,59 %), *Biol* (biologia, 3,47 %), *Mus* (música, 1,24 %), *Anat* (anatomia, 1,24 %) i *E* (especialitats varies, 1,12 %). La resta de marques són molt nombroses i presenten uns percentatges inferiors a l'1 %. Són, per tant, molt poc representatives. Algunes d'aquestes marques són *Bot* (botànica, 0,99 %), *Econ* (economia, 0,74 %), *Dep* (esports, 0,62 %), *Geol* (geologia, 0,62 %), *Pol* (política, 0,62 %), *Psicol* (psicologia, 0,62 %), *Arte* (art, 0,50 %) o *Informat* (informàtica, 0,50 %).

Quant a les seqüències de marques, totes elles només apareixen en un lema (0,12 %), tot i que aquest lema es pot repetir diverses vegades en el corpus. Algunes de les més representatives són *(Ecol)(Dep)(Anat)*, *(Coc)(Dep, Taur)(Naipes)(Naipes)(Coc)*, *(Agric)(Biol)*, *(Biol)(Biol)* o *(Mat)(Bot y Zool)*.

Tot i que detectem un nombre considerable de marques entre els lemes coincidents, n'hi ha unes quantes que no han quedat representades. Algunes d'aquestes marques també presents en altres lemes nous del DEA són *Min* (mineria), *Prehist* (prehistòria) o *Mito* (mitologia).

### **Corpus Tècnic de l'Institut Universitari de Lingüística Aplicada**

Descrivim en aquest apartat els lemes coincidents entre el DEA i el del Corpus Tècnic de l'IULA. Hi ha 215 lemes coincidents (2.348 ocurrències) (13,92 %). Per tant, el fenomen més habitual és la no-coincidència entre els lemes del DEA i del corpus (86,07 %). La taula següent mostra la representativitat dels lemes coincidents i no coincidents respecte el total de lemes consultats:

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DEA)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	1.329	86,07 %
Coincidents	215	13,92 %
<b>Total</b>	<b>1.544</b>	<b>100 %</b>

Taula 30: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Corpus Tècnic de l'IULA (DEA)

Identifiquem 15 recursos d'actualització lèxica diferents.<sup>86</sup> Els més habituals són els manlleus (36,28 %), seguits molt de prop de la sufixació (35,35 %). Aquests dos recursos representen el 71,63 % del

<sup>86</sup> Vegeu la taula 98 de l'Annex 4.

total de lemes. Els recursos que presenten una representativitat mitjana són sis: la composició híbrida (6,05 %), els lemes simples (5,12 %), la composició culta (3,72 %), la prefixació (2,79 %), els lemes simples que eren un nom propi en origen (2,33 %) i la composició patrimonial (2,33 %). Els darrers recursos són els que tenen una representativitat més baixa, inferior al 2 %. Aquests recursos són els lemes simples construïts (1,86 %), els escurçaments (1,86 %), la lexicalització nominal (0,47 %), la lexicalització adjectiva (0,47 %), la derivació zero (0,47 %), la sufixació taxonòmica (0,47 %) i l'acronímia (0,47 %). No detectem entre els lemes coincidents recursos com els símbols, la sintagmació, la siglació, les formes prefixades i sufixades, els prefixos, els sufixos, les variants ortogràfiques o els compostos formats a la manera culta.

Quant a la marca temàtica, recollim 27 marques soles en els articles dels lemes i cinc seqüències de marques.<sup>87</sup> El lemes que no inclouen cap tipus de marca són el 66,98 % del total de lemes. Les marques més freqüents són *Biol* (biologia, 5,58 %), seguida de *Med* (medicina, 2,79 %) i *Quím* (química, 2,33 %). La resta de marques presenten uns percentatges molt escassos, inferiors al 2 %. Algunes d'aquestes marques són *Dep* (esports, 1,86 %), *Mus* (música, 1,86 %), *Econ* (economia, 1,86 %), *Golf* (1,40 %), *Psicol* (psicologia, 1,40 %), *Pol* (política, 1,40 %), *E* (especialitats varies, 0,93 %), *Fon* (fonètica i fonologia, 0,93 %), *Geol* (geologia, 0,93 %) o *Taur* (tauromàquia, 0,93 %). Les seqüències de marques que detectem només les trobem en un sol lema (0,47 %), i són *(Biol)(Biol)*, *(Pol)(Com)*, *(Golf)(Mil)(Pol)*, *(Ferroc)(Boxeo)(Tenis)* i *(Med)(E)(E)*. Algunes de les marques temàtiques que no han quedat representades entre els lemes coincidents amb el Corpus Tècnic són *Zool* (zoologia), *TV* (televisió), *Enseñ* (ensenyament) o *Pesca* (pesca).

### ***Banc BOBNEO de l'Observatori de Neologia de l'Institut Universitari de Lingüística Aplicada***

El Banc BOBNEO de l'Observatori de Neologia és el darrer corpus que utilitzem per contrastar els lemes del DEA. En la taula següent mostrem la coincidència i la no-coincidència dels lemes consultats:

---

<sup>87</sup> Vegeu la taula 104 de l'Annex 4.

<b>Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DEA)</b>		
<b>Coincidència i nombre d'entrades per lema coincident</b>	<b>Lemes</b>	<b>Freqüència lemes</b>
No coincidents	1.192	77,20 %
Coincidents	352	22,79 %
<b>Total</b>	<b>1.544</b>	<b>100 %</b>

Taula 31: Representativitat dels lemes coincidents i no coincidents en el contrast amb el Banc BOBNEO (DEA)

Detectem 352 lemes coincidents (2.395 ocurrencies) entre els 1.544 lemes consultats (22,79 %). Aquest percentatge, tot i ser escàs, és més elevat que els altres contrastos d'altres diccionaris amb el Banc BOBNEO.

Des del punt de vista del recurs d'actualització lèxica classifiquem els lemes segons 17 recursos diferents.<sup>88</sup> Entre aquests, destaquen els manlleus (41,19 %) i la sufixació (30,40 %). Aquests dos recursos representen la majoria dels casos (71,59 %). Detectem també quatre recursos que presenten uns percentatges de representativitat mitjana: la composició híbrida (7,10 %), els lemes simples (5,11 %), la composició patrimonial (3,69 %) i la composició culta (3,41 %). La resta de recursos tenen uns percentatges de representativitat inferiors al 2 % i són, per tant, els més baixos: la prefixació (1,99 %), la lexicalització adjectiva (1,14 %), els lemes simples construïts (1,14 %), els llatinismes sintagmàtics (1,14 %), els escurçaments (0,85 %), els lemes simples que en origen eren noms propis (0,85 %), la sufixació taxonòmica (0,57 %), la lexicalització nominal (0,57 %), l'acronímia (0,28 %), els lemes simples construïts en origen per sufixació (0,28 %) i la sintagmació (0,28 %). Alguns dels recursos d'actualització lèxica que no han quedat representats entre els lemes coincidents són la derivació zero, la siglació, la parasíntesi, les formes prefixades i sufixades, els prefixos i els sufixos o els símbols.

En l'estudi de la marca temàtica, la situació més habitual és que els articles no incloguin marques temàtiques (69,89 %).<sup>89</sup> Tot i això, trobem 31 marques diferents que apareixen soles en l'article lexicogràfic i cinc seqüències de marques. Entre les 31 marques, les més habituals són *Med* (medicina, 5,11 %) i *Biol* (biologia, 4,83 %). Les marques que tenen una representativitat mitjana són diverses: *Mús* (música, 2,27 %), *Dep* (esports, 1,70 %), *Econ* (economia, 1,42 %), *Quím* (química, 1,14 %), *Psicol* (psicologia, 1,14 %), *Anat* (anatomia, 1,14 %) i *Informát* (informàtica, 1,14 %). Les marques

<sup>88</sup> Vegeu la taula 116 de l'Annex 4.

<sup>89</sup> Vegeu la taula 122 de l'Annex 4.

menys representatives són *Pol* (política, 0,87 %), *E* (especialitats varies, 0,85 %), *Golf* (0,57 %), *Arte* (art, 0,57 %), *Agric* (agricultura, 0,28 %) o *Der* (dret, 0,28 %). Quant a les seqüències de marques, tan sols són cinc amb unes freqüències del 0,28 %: *(Agric)(Biol)*, *(Golf)(Mil)(Pol)*, *(Med)(E)(E)*, *(Pol)(Com)* i *(Ferroc)(Boxeo)(Tenis)*. Finalment, les marques que no han quedat representades en el articles dels lemes coincidents són *Geom* (geometria), *Coc* (cuina), *Autom* (automobilisme) o *Bibl* (bibliologia, bibliografia i matèries afins).