# ANÀLISI DE RESULTATS EXTRETS DEL DICCIONARI DE FREQÜÈNCIES DE L'INSTITUT D'ESTUDIS CATALANS

## 1. Introducció

El corpus de dades emprat per a la realització de l'estudi del *Corpus textual informatitzat de la llengua catalana* (CTILC) i del *Diccionari de freqüències* (Rafel 1998) té una extensió considerable, i està format per 14.613 textos de tipus literari i 11.508 de no literaris (en total 26.121). Els textos literaris estan formats per obres de teatre, narrativa, poesia i assaig. Els textos no literaris estan formats per: articles científics, textos legals, premsa, tractats, manuals, cartes personals, etc.

Tots els textos s'han entrat lematitzats a través d'un procediment manual. Recordem que la lematització consisteix en la categorització morfosintàctica, rarament semàntica, de les diferents ocurrències de cada forma gràfica que apareix en el text, i en la seva associació a una referència lèxica convencional anomenada lema. Per exemple: la forma d'infinitiu d'un verb o la forma d'un masculí singular d'un adjectiu variable.

En el corpus s'han entrat 109.159 lemes no literaris i 64.449 lemes literaris: el total de lemes és 173.608. Aquests lemes se'ls ha associat el nombre de vegades que han aparegut en el text i, per tant, tindrem les ocurrències parcials per al cas literari (23.108.691) i pel cas no literari (29.266.353), en total seran 52.375.044 ocurrències.

Aquestes dades entrades al corpus lematitzat s'han classificat pel tipus morfològic i s'han ordenat per freqüències. Cada lema té, doncs, associat el nombre de cops que ha aparegut en tot el corpus i així podem fer ordenacions descendents o ascendents. Anomenem freqüència absoluta d'un lema al nombre d'aparicions, a través de qualsevol de les formes que té associades, o bé de les formes dels lemes secundaris que hi ha estat relacionats. Per exemple: Un verb i les seves variacions de temps i persona.

En les dades també apareix el concepte de freqüència relativa, és a dir la relació d'ocurrències d'un lema respecte als altres lemes. F(relativa)= F(absoluta)/total ocurrències). Cal notar que a causa del nombre elevat de xifres, la freqüència relativa

s'ha multiplicat per 100, per així donar millor interpretació als lemes que apareixen poques vegades. A continuació veurem una mostra dels diferents grups analitzats (mostrem els 10 primers de cada llista, *cf.* Puig 1994).

## 2. Adverbis

RELACIÓ D'ADVERBIS MÉS FREQÜENTS ORDENATS DE MÉS GRAN A MÉS PETIT (DESCENDENT)

| LEMA-ADVERBI | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA |
|---|---|---|
| no | 646.184 | 1,2607 |
| com | 320.320 | 0,6249 |
| més | 265.522 | 0,5180 |
| ja | 117.832 | 0,2299 |
| quan | 100.576 | 0,1962 |
| molt | 89.305 | 0,1742 |
| també | 75.733 | 0,1477 |
| encara | 72.973 | 0,1423 |
| tan | 72.706 | 0,1418 |
| on | 65.376 | 0,1275 |

2.1 ADVERBIS MÉS FREQÜENTS ACABATS EN *MENT*

| ADVERBI ACABAT EN -MENT | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA |
|---|---|---|
| solament | 11.777 | 0,0229 |
| precisament | 6.539 | 0,0127 |
| especialment | 6.080 | 0,0118 |
| finalment | 5.673 | 0,0110 |
| generalment | 4.829 | 0,0094 |
| naturalment | 4.822 | 0,0094 |
| igualment | 4.382 | 0,0085 |
| realment | 4.291 | 0,0083 |
| completament | 3.862 | 0,0075 |
| perfectament | 3.751 | 0,0073 |

## 3. Lemes

Exposem la relació de lemes segons el número de lletres que el componen. La columna de l'esquerra són les lletres que formen el lema, i a la dreta hi ha la quantitat de lemes que hem trobat amb el número de lletres. La mitjana de les lletres és 13, i la

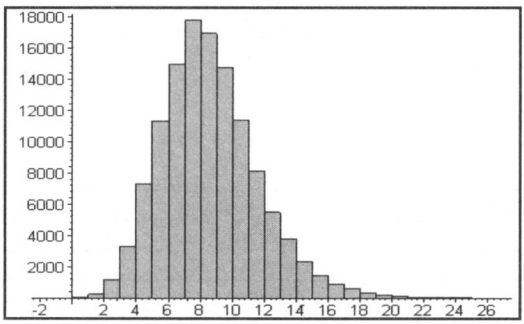
mitjana dels lemes és 4905,88 Podem observar en la taula, que el número de lemes 17.780 li correspon als formats per 8 lletres, Això ens diu que són els lemes més habituals en el corpus analitzat (Vallverdú 2001).

RELACIÓ DE LEMES SEGONS EL NOMBRE DE LLETRES

| LLETRES | LEMES | | LLETRES | LEMES |
|---|---|---|---|---|
| 1 | 19 | | 14 | 3.787 |
| 2 | 239 | | 15 | 2.337 |
| 3 | 1.160 | | 16 | 1.461 |
| 4 | 3.305 | | 17 | 915 |
| 5 | 7.313 | | 18 | 577 |
| 6 | 11.330 | | 19 | 334 |
| 7 | 14.958 | | 20 | 195 |
| 8 | 17.780 | | 21 | 123 |
| 9 | 16.934 | | 22 | 70 |
| 10 | 14.750 | | 23 | 54 |
| 11 | 11.351 | | 24 | 26 |
| 12 | 8.118 | | 25 i+ | 37 |
| 13 | 5.475 | | | |

La gràfica que correspon a les últimes dades serà:

Lemes



## 4. Adjectius

| LEMA | TIPUS D'ADJECTIU | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA |
|---|---|---|---|
| aquest | Adjectiu | 306.318 | 0,597 |
| seu | Adjectiu | 277.213 | 0,540 |
| tot | Adjectiu | 239.907 | 0,468 |
| altre | Adjectiu | 169.096 | 0,329 |
| aquell | Adjectiu | 117.932 | 0,230 |
| mateix | Adjectiu | 92.764 | 0,180 |
| nostre | Adjectiu | 92.006 | 0,179 |
| gran | Adjectiu invariant | 84.111 | 0,164 |
| son | Adjectiu | 82.456 | 0,160 |
| qual | Adjectiu invariant | 79.580 | 0,155 |

## 5. Conjuncions

| LEMA-CONJUNCIÓ | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA |
|---|---|---|
| i | 1.703.937 | 3,324 |
| que | 744.061 | 1,451 |
| o | 208.429 | 0,406 |
| si | 180.750 | 0,352 |
| però | 162.520 | 0,317 |
| perquè | 87.192 | 0,170 |
| ni | 81.288 | 0,158 |
| sinó | 41.831 | 0,081 |
| doncs | 34.936 | 0,068 |
| bé | 26.608 | 0,051 |

## 6. Preposicions

| LEMA-PREPOSICIÓ | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA |
|---|---|---|
| De | 3.276.521 | 6,392 |
| A | 1.237.435 | 2,414 |
| En | 854.864 | 1,667 |
| per | 671.944 | 1,311 |
| amb | 453.297 | 0,884 |
| sense | 84.703 | 0,165 |

| | | |
|---|---|---|
| entre | 72.118 | 0,140 |
| sobre | 58.008 | 0,113 |
| fins | 55.470 | 0,108 |
| des | 37.548 | 0,073 |

## 7. Interjeccions

| INTERJECCIONS | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA |
|---|---|---|
| Oh | 12.886 | 0,0251 |
| Ai | 8.233 | 0,0160 |
| Ah | 7.585 | 0,0147 |
| vaja | 2.699 | 0,0052 |
| oi | 2.531 | 0,0049 |
| eh | 2.042 | 0,0039 |
| o | 1.534 | 0,0029 |
| apa | 1.272 | 0,0024 |
| ha | 998 | 0,0019 |
| ca | 968 | 0,0018 |

## 8. Adjectius numerals

| LEMA-ADJECTIU NUMERAL | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA |
|---|---|---|
| Dos | 80.418 | 0,156 |
| Tres | 30.724 | 0,059 |
| Quatre | 17.761 | 0,034 |
| Mil | 9.075 | 0,017 |
| Cinc | 8.520 | 0,016 |
| Cent | 7.615 | 0,014 |
| Deu | 6.108 | 0,012 |
| Sis | 5.848 | 0,011 |
| Set | 4.715 | 0,009 |
| Vuit | 4.003 | 0,007 |

## 9. Pronoms

| LEMA-PRONOM | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA |
|---|---|---|
| ell | 1.485.513 | 2,898 |
| que | 850.047 | 1,658 |
| jo | 474.544 | 0,925 |
| hi | 223.968 | 0,436 |
| tu | 178.938 | 0,349 |
| en | 150.706 | 0,294 |
| ho | 118.297 | 0,230 |
| què | 99.632 | 0,194 |
| això | 87.712 | 0,171 |
| qui | 70.920 | 0,138 |

## 10. Relació de les freqüències de les vocals

| VOCALS | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA |
|---|---|---|
| e | 30.232.283 | 0,3210 |
| a | 27.268.594 | 0,2895 |
| i | 15.243.681 | 0,1618 |
| o | 12.451.700 | 0,1322 |
| u | 9.095.143 | 0,0965 |

## 11. Relació de freqüències de les consonants

| CONSONANTS | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA |
|---|---|---|
| s | 18382379 | 0.14936 |
| r | 14721771 | 0.11961 |
| l | 14677661 | 0.11925 |
| n | 14351861 | 0.11661 |
| t | 13370389 | 0.10863 |
| d | 8584238 | 0.06974 |
| c | 7889502 | 0.06410 |
| m | 6883929 | 0.05593 |
| p | 5944888 | 0.04830 |
| v | 3037242 | 0.02467 |
| q | 2953244 | 0.02399 |

| | | |
|---|---|---|
| b | 2878137 | 0.02338 |
| g | 2798913 | 0.02274 |
| f | 2197853 | 0.01785 |
| h | 1566665 | 0.01272 |
| x | 1150436 | 0.00934 |
| y | 870628 | 0.00707 |
| j | 651241 | 0.00529 |
| ç | 238431 | 0.00193 |
| z | 145577 | 0.00118 |
| k | 9409 | 0.00007 |
| ñ | 5267 | 0.00004 |
| w | 2594 | 0.00002 |

## 12. Lemes acabats amb -ll

| LEMA | FREQÜÈNCIA ABSOLUTA | FREQÜÈNCIA RELATIVA | CATEGORIA GRAMATICAL |
|---|---|---|---|
| ell | 1.485.513 | 2,898 | Pronom |
| aquell | 117.932 | 0,230 | Adjectiu |
| ull | 32.928 | 0,064 | Masculí |
| fill | 29.500 | 0,057 | Masculí |
| treball | 22.906 | 0,044 | Masculí |
| vell | 15.039 | 0,029 | Adjectiu |
| bell | 12.567 | 0,024 | Adjectiu |
| consell | 8.378 | 0,016 | Masculí |
| nivell | 7.786 | 0,015 | Masculí |
| perill | 7.371 | 0,014 | Masculí |

## 13. Lemes de més de 24 lletres

| LEMA DE 24 O MÉS LLETRES | LLETRES | CATEG. GRAMATICAL | FREQÜÈNCIA ABSOLUTA |
|---|---|---|---|
| Contenciosoadministratiu | 24 | Adjectiu | 51 |
| Diclorodifeniltricloroetà | 25 | Masculí | 39 |
| Desmettilclortetraciclina | 24 | Femení | 6 |
| Cientificotecnicoprofessional | 29 | Adjectiu invariant | 2 |
| Degenerativoconstitucional | 26 | Adjectiu invariant | 2 |
| Generativotransformacional | 26 | Adjectiu invariant | 2 |
| Cirrocúmulus lenticularis | 24 | Masculí | 2 |

| | | | |
|---|---|---|---|
| Monarquicoconstitucional | 24 | Adjectiu invariant | 2 |
| Academicorealistanaturalista | 28 | Adjectiu invariant | 1 |
| Aristotelicohegelianomarxista | 29 | Adjectiu invariant | 1 |

## 14. Lemes més freqüents (per categories gramaticals)

| LEMA I FORMES | CAT. GRAMATICAL | FREQ. ABSOLUTA | FREQ. RELATIVA |
|---|---|---|---|
| el / la / les/ els/ als / 'l/ | Article | 5.140.416 | 10,029 |
| de/ d' /d'a/ da | Preposició | 3.276.521 | 6,392 |
| i | Conjunció | 1.703.937 | 3,324 |
| ell / el/ 'l/ els / ella/ ells/ | Pronom | 1.485.513 | 2,898 |
| a | Preposició | 1.237.435 | 2,414 |
| un /una /unes /uns/ | Article | 1.184.909 | 2,311 |
| ésser / ser / (temps verbals) | Verb intransitiu | 1.000.352 | 1,951 |
| en /ne / ' n / n' | Preposició | 854 .864 | 1,667 |
| que / què / qué/..... | Pronom | 850.047 | 1,658 |
| del / dels / d'els / de'l | Contracció | 779.944 | 1,521 |

## 15. Verbs més freqüents acabats amb -ar

| LEMA-VERB | TIPUS DE VERB | FREQ. ABSOLUTA | FREQ. RELATIVA |
|---|---|---|---|
| anar | Verb auxiliar | 194.259 | 0,379 |
| anar | Verb intrans. | 114.340 | 0,223 |
| estar | Trans., intrans. pronom. | 111.112 | 0,216 |
| donar | Trans. pronom. | 91.254 | 0,178 |
| trobar | Trans. intrans. pronom. | 74.743 | 0,145 |
| passar | Trans., intrans. pronom. | 64.499 | 0,125 |
| deixar | Trans. pronom. | 59.859 | 0,116 |
| posar | Trans., intrans. pronom. | 52.973 | 0,103 |
| arribar | Intrans. pronom. | 51.420 | 0,100 |
| semblar | Intrans. | 47.037 | 0,091 |

## 16. Verbs més freqüents acabats en -re

| LEMA–VERB | TIPUS DE VERB | FREQ. ABSOLUTA | FREQ. RELATIVA |
|---|---|---|---|
| veure | Trans. pronom. | 128.496 | 0,250 |
| caldre | Intrans. | 48.152 | 0,093 |
| creure | Trans., intrans. pronom. | 38.851 | 0,075 |
| deure | Trans. | 34.596 | 0,067 |
| prendre | Trans., intrans. pronom. | 32.348 | 0,061 |
| viure | Trans., intrans. | 29.157 | 0,056 |
| treure | Trans. | 22.286 | 0,043 |
| perdre | Trans., intrans. pronom. | 20.108 | 0,039 |
| caure | Intrans. | 19.363 | 0,037 |
| escriure | Trans. pronom. | 17.915 | 0,034 |

## 17. Verbs més freqüents acabats amb -er

| LEMA/VERB | TIPUS DE VERB | FREQ. ABSOLUTA | FREQ. RELATIVA |
|---|---|---|---|
| ésser | Intrans. | 1.000.352 | 1,951 |
| haver | Intrans. auxiliar | 655.393 | 1,278 |
| fer | Trans., intrans. pronom. | 338.070 | 0,659 |
| poder | Trans. | 211.688 | 0,413 |
| voler | Trans. | 98.675 | 0,192 |
| saber | Trans. intrans. | 90.019 | 0,175 |
| conèixer | Trans., intrans. pronom. | 30.248 | 0,059 |
| valer | Trans., intrans. pronom. | 13.195 | 0,026 |
| córrer | Trans. intrans. | 12.908 | 0,025 |
| aparèixer | Intrans. pronom. | 12.731 | 0,024 |

## 18. Verbs més freqüents acabats amb -ir

| LEMA-VERB | TIPUS DE VERB | FREQ. ABSOLUTA | FREQ. RELATIVA |
|---|---|---|---|
| tenir | Trans. pronom. | 236.795 | 0,462 |
| dir | Trans., intrans. pronom. | 222.380 | 0,433 |
| venir | Intrans. | 56.518 | 0,110 |
| sentir | Trans. pronom. | 50.096 | 0,097 |
| sortir | Intrans. | 33.074 | 0,064 |
| seguir | Trans. pronom. | 30.014 | 0,058 |

| produir | Trans. pronom. | 20.691 | 0,040 |
|---|---|---|---|
| servir | Trans. intrans. pronom. | 19.763 | 0,038 |
| morir | Intrans. pronom. | 18.027 | 0,035 |
| obrir | Trans., intrans. pronom. | 16.488 | 0,032 |

## 19. Conclusions

De tots aquests resultats observem, en primer lloc, que en el cas dels verbs no apareix el verb *estar*, a causa de la procedència dels textos emprats. En segon lloc, s'observa que els lemes més freqüents tenen 8 lletres. De les preposicions podem veure que la més emprada és *de*. De les conjuncions cal remarcar que la més freqüent és *i*, marcant diferència amb les que li van al darrera. Observem que *doncs* està més enrere i en canvi en el llenguatge parlat s'empra molt més. En el cas de les vocals s'ha comprovat que la *e* és la primera de la llista, en les consonants es veu que la més freqüent és la *s*. Hem analitzat els lemes acabats en *-ll*. S'han enumerat uns quants lemes amb més de 24 lletres. Podem veure que són poc freqüents. Com a categories gramaticals es veu que la més freqüent és l'article. Una proposta de futur seria poder fer una comparació amb el llenguatge oral, on es veurien les diferències més importants.

Anna Puig Montada
Escola Tècnica Superior d'Enginyeria Indústrial de Terrassa,
Universitat Politècnica de Catalunya

## REFERÈNCIES BIBLIOGRÀFIQUES

Rafel 1998: Joaquim Rafel Fontanals, *Diccionari de freqüències,* Barcelona, IEC, 1998 (CD-ROM).

Puig 1994: Anna Puig Montada, *Tractament de corpus textuals lematitzats i estudi comparatiu del llenguatge científic amb la prosa estàndard*, tesi doctoral, Barcelona, Universitat Politècnica de Catalunya.

Vallverdú 2001: *Enciclopèdia de la llengua catalana*, ed. Francesc Vallverdú, Barcelona, Edicions 62.