

# A Descriptive Dictionary of Contemporary Catalan: the DDLC project

Joaquim Rafel i Fontanals

Joan Soler i Bou

Institut d'Estudis Catalans

Carne, 47

E – 08001 Barcelona

jrafel@iec.cat

jsoler@iec.cat

## Abstract

The main goal of this paper is to present a complete overview of the lexicographical project *Diccionari Descriptiu de la Llengua Catalana* (DDLC), a corpus-based Catalan dictionary that is currently being prepared at the Institut d'Estudis Catalans (IEC). The DDLC is conceived as a *descriptive* dictionary, which means that its goal is to give a real and complete definition of each lexical item, without any restrictions based on prescriptive criteria. From this point of view, a descriptive dictionary is opposed to a standard prescriptive dictionary, in the sense that the former contains the current real uses and the latter includes only those words and meanings which are approved by the established standard. In this contribution we will present the main distinctive features of the DDLC as a lexicographical project: the data structure of the dictionary, a review of the work already done and the present stage of development.

## 1 Introduction

The main goal of this paper is to present a complete overview of the lexicographical project *Diccionari Descriptiu de la Llengua Catalana* (DDLC). The DDLC is a corpus-based Catalan dictionary that is currently being written at the Institut d'Estudis Catalans (IEC). The preparatory work of DDLC started in 1998, and by March 2006 we have already reached the figure of more than 27,000 dictionary articles. This figure represents about a third of the estimated final size of the dictionary.

Up to this moment, Catalan lexicography has produced outstanding dictionaries covering different aspects of language: historical and dialectal (DCVB), etymological (DECLC), prescriptive standard (DIEC, among many other similar works), and so on. The DDLC differs from these previous works because it is conceived as a synchronic *descriptive* dictionary, which means that its aim is to give a real and complete definition of each lexical item, without any restrictions based on prescriptive criteria. From this point of view, a descriptive dictionary is opposed to a standard prescriptive (normative) dictionary, in the sense that the former contains the real uses and the latter includes only those lexical units (single and multi-word expressions) and those values (meanings and uses) which are approved by the established standard of a given language.

- In order to present this overview of DDLC, this contribution is divided into three parts:
- a) The main distinctive features of the DDLC as a lexicographical project.
  - b) The structure of the DDLC.
  - c) The review of the work already completed and the present stage of development

## 2 Features of DDLC

Before presenting some of the most distinctive features of the dictionary, it should be said that one of the primary duties of the IEC is to act as the Academy for the Catalan Language. That means that one of its major responsibilities is to produce, maintain and update the standardisation criteria for the social use of the language. The initiative of launching a descriptive, instead of prescriptive, lexicographical work should thus be interpreted as a strategic decision in order to produce relevant data for the development of the IEC's ongoing standardisation tasks (cf. Rafel, 2000).

The DDLC is neither a conventional dictionary nor a theoretical or computational one. In terms of Geeraerts (1985), it could be defined as a *linguistic dictionary*. Thus, even though the DDLC does not pretend to be a theoretical dictionary, it shares some features with this kind of linguistic resource: it does not have a pragmatic nature and it does not pursue a pedagogical goal, it is conceived on the basis of *ex novo* research, it is based upon the highest degree of scientific precision, its formulation is aimed to ensure a maximum degree of explicitness and it uses a particular formalization in its vocabulary and presentation.

Due to the above-mentioned reasons, the ideal DDLC users are linguists and language professionals in general. Nevertheless, this work is not only addressed to language specialists, but also to people with an average cultural level and with a shared general knowledge who are interested as users in the problems concerning language. The DDLC presents its contents combining descriptive precision with expository clarity and ease of comprehension.

The dictionary as a whole is conceived as a multifunctional (Zampolli, 1991a, 1991b) database that enables a number of different uses related to many different applications such as, just to mention two quite different examples, linguistic research and the accessibility and dissemination of the dictionary's contents. As to the diffusion of its contents, the DDLC permits a wide range of possibilities: it may be (as in fact it is being) published via Internet as an on-line dictionary, and also it may produce different kinds of published products, in paper or in any of the current electronic media.

The DDLC list of entries is obtained entirely from the reference corpus (CTILC<sup>1</sup> from now on). The organisation of data is quite complex because it consists of two different kinds of elements:

- a) Proper entries, whose description is contained in the article to which they belong.
- b) Subsidiary entries, whose description is included in the main article of a proper entry. Subsidiary entries may be classified into two clear-cut categories: a) derived words

---

<sup>1</sup> CTILC is the acronym of *Corpus Textual Informatitzat de la Llengua Catalana*; cf. Rafel (1994), and Rafel (1996-1998).

formed by one of the seven systematic derivative processes established in the general criteria of the dictionary, and *b*) formal variants documented within the CTILC that may be associated to another form which is considered as a primary entry.

Complex lexical units do not constitute entries by themselves and they are not linked to any particular sense included in the article. They are grouped together in a specific section of the article they belong to in accordance with explicit criteria.

The splitting of entries is entirely based upon strictly formal criteria, which are applied throughout the dictionary with no exception. These criteria are based on the combination of the three structural elements of the word-entry: canonical form, category and inflectional properties. A slight variation in any of these three attributes determines the existence of a differentiated entry, independently of any semantic or diachronic consideration.

Two kinds of constituents are distinguished in the definitions in the DDLC: intrinsic and extrinsic constituents (Rafel, 1989; Soler, 2001). Intrinsic constituents contain the strictly defining elements; while extrinsic constituents contain the selectional restrictions or conditions for the use of the defined form. All extrinsic elements are codified with reference to the syntactic pattern of the definiendum, and they correspond to a certain extent with the concepts of *entourage* (Rey-Debove, 1971) and *contorno* (Seco, 1978, 1979).

For each entry, the DDLC contains quantitative information concerning the real use of the unit in the CTILC. This information is graphically simplified and it indicates one of five possible preestablished ranks.

### 3 Structure of the DDLC

The DDLC is made up of a finite number of structural elements organized in a database. Each of these elements is linked in different ways to a reference item: the entry. This link allows us to present these elements structured in a dictionary article, which is the most common way (but not the only one) to present them in an organised structure. We will follow the standard article pattern (shown in Figure 1) as a guide to present the content and the structure of the dictionary.

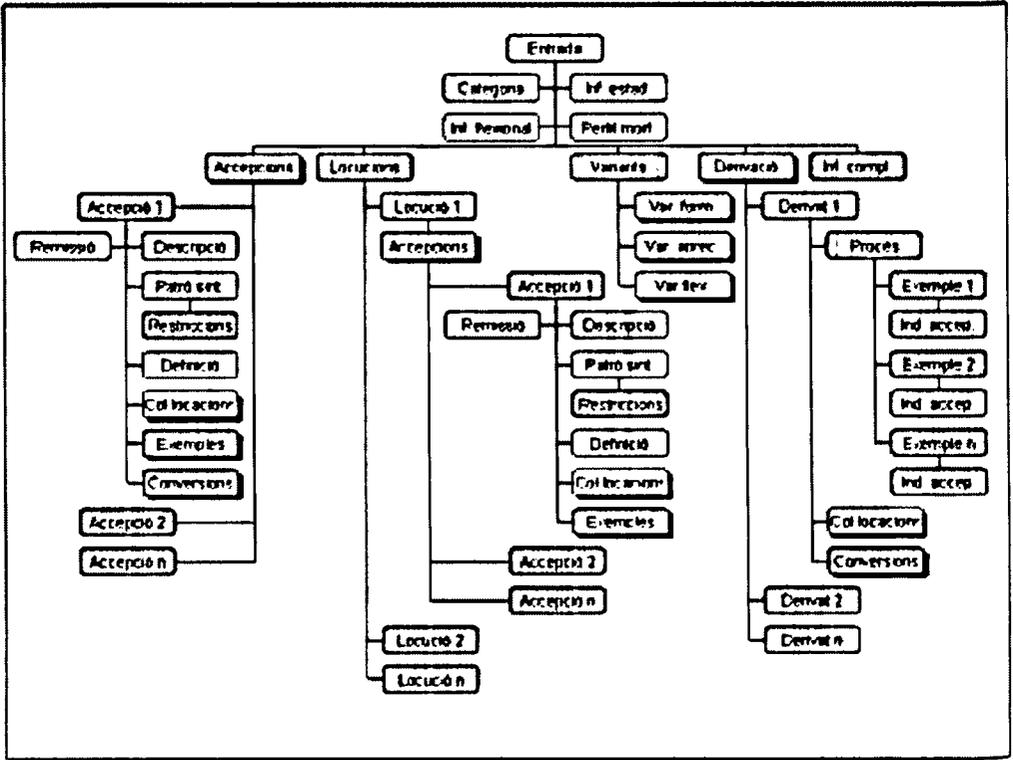


Figure 1. Standard DDLC article architecture.

The whole article consists of a number of elements that make up the heading (*entry* “entrada”, *category* “categoria”, *inflectional information* “inf. flexional”, *statistical information* “inf. estad.” and *morphological profile*<sup>2</sup> “perfil morf.”) and also of some other elements that constitute the body of the article. In the following example we can see two examples of headings of different articles for the words *adob m.* ‘fertilizer’ and *amollar v.* ‘to let go of’:

<sup>2</sup> The morphological profile is indicated for nouns and verbs, but only for those entries with ranks from 1 to 3.

adob m. (adob) ●●○○

N	51	49
---	----	----

amollar v. CANTAR ●●○○

F	63	37						
NP	61	23	16					
M	81	9	9					
T	29	23	24	4	1	6	3	9
P	5	9	63	3	2	18		

These elements are not equally mandatory or dependent on other elements within the dictionary. Consequently, the dictionary articles may have different degrees of complexity; some articles may be very complex, especially if we consider the potential recursiveness of some of their components, and others may be very simple. There are, nevertheless, a few mandatory elements which must be always present in a DDLC article. As to the heading, the word-form of the entry, the category, the inflectional information (that can be either explicit or implicit), and the statistical information are required. As to the body, every article must have at least one *sense* (“accepció”). This sense may be conveyed by means of: *a*) a lexicographical description section (“descripció”), which entails a set of associated descriptive elements (syntactic pattern, collocations, definition, etc.), or *b*) a *cross-reference* (“remissió”) referring to a relevant article or to a descriptive sense in another article of the dictionary.

It should be said that in the DDLC structure a cross-reference is a reference introduced by the indication “Vegeu” (= *Vid.*, *See*). This type of explicit reference must not be confused with the implicit references that usually appear under the form of synonymic or synthetic definitions and which are usually named in conventional dictionaries in the same way (“remissions” in Catalan). Both (cross-references and definitions by synonym) are a part of the referential structure of the dictionary, but they have a very different value: while cross-references are found instead of a description set, a synonym stands for the definition element in this set; thus, they are related to two completely different functions. The existence of cross-references in the DDLC – in this strict sense – is a consequence of the complex nomenclature that we have referred to when dealing with the general characteristics of the dictionary.

In the body of the article, the *sense* section has a certain complexity. The description set of a sense consists of the following elements: *syntactic pattern* (“patró sintàctic”), *semantic constraints* (“restriccions”), *definition* (“definició”), *collocations* (“col·locacions”), one or more *examples* (“exemples”), and *syntactic conversions* (“conversions”).

On the one hand, the syntactic patterns are expressed by a set of positions filled by categorial symbol strings (N, V, Adj, etc.), which can incorporate modifiers in the form of sub-specifications (NCOMPT, VPRON, etc.). The elements integrating the pattern can also have the form of subscripts indicating the coreference (or the lack of coreference) between two or more positions. Finally, every symbol of the pattern can be introduced by a literal word (a preposition, an article, etc.). On the other hand, a few regular syntactic alternations (as for example, a causative transitive construction and the corresponding non-causative intransitive

one) are described in a related way. Additionally, the lexical or semantic restrictions that apply to the elements of the pattern are also indicated, except for the core element of the pattern that stands for the defined word. We have already referred above to the nature of the definition; but it is worth mentioning at this point that the extrinsic elements included in the definition correspond to specific positions of the syntactic pattern and are used either to attain the semantic refinement of the defining hypernym for the defined word or to syntactically adjust the definition with regard to its related arguments.

Collocations are understood as a group of at least two lexical units frequently co-occurring in the CTILC due to a certain lexical attraction between them. Given that the meaning of a collocation is a function of the meanings of its components, collocations do not need a specific semantic description as idiomatic constructions.

The examples are always quotations extracted from the CTILC, and they are reproduced without any adaptation nor modification. The selection of the most adequate example to illustrate a specific meaning is based upon a complex series of criteria that we cannot explain in detail in this paper. However, to briefly summarise our procedure, we shall say that it rigorously follows the general principle that the example must contain the necessary and sufficient information to illustrate a particular sense, and that it must not be ambiguous from any point of view. If one sense has multiple patterns, an example for each one is required. The examples are accompanied and identified by a simplified textual reference that provides preliminary information on the origin of the quotation. In the electronic version of the DDLC, the elements coded by this simplified reference can be displayed interactively and the whole bibliographic reference of the work can be accessed on demand.

Syntactic conversions do not entail any kind of complementary description, but they have an example associated to them. The types of conversions that are described in this section are: *infinitive noun*, *participle adjective*, *interjection noun* and *noun interjection*.

As an example of this structural part of the article, we can show the first two senses of the article *encetar* v 'to begin':

1. [N1 V (N2)] ↔ [N2 VPRON] (N1[humà, esdeveniment]; N2[acció, comunicació]) Iniciar [una acció, una comunicació, que tenen una certa durada]2 ↔ [Una acció, una comunicació, que tenen una certa durada]2 iniciar-se. ⇒ fi ~ (una) conversa, ~ el diàleg. *Volem encetar aquesta pàgina de crítiques de llibres [...] amb un breu comentari de tres llibres de l'editorial LaSal [...].* [Aixa (1985): 41, niv. 4, p. 8]. *Veiam! parlem d'aquesta guerra. Qui enceta?* [Tolzà (1979): T, p. 27]. *La fina brisa nova dóna el seu perfum a la peça extraordinària amb què la plàstica gòtica s'enceta a les terres catalanes: la Mare de Déu [...] obra del Mestre Bertomeu [...].* [Cirici (1957): 73, p. 98].

⇒ adj. *La sublevació de Vicàlvaro només donà dos anys de govern progressista, encetats a Catalunya amb les pertorbacions que ressenyem més endavant [...].* [Romeva (1952): 94, p. 321].

2. [N1 VPRON] (N1[període]) [Un període de temps] començar. *Vàrem passar la nit i va encetar-se el dia.* [Escuder (1980): 79, p. 62].

⇒ adj. [...] *el matí tot just encetat es mostrava lluent de sol, rentat de fresc per la rosada de la nit [...].* [Faner (1984): N, p. 160].

Another section of the DDLC article structure (*phrases* "locucions") is constituted by phraseology and idiomatic constructions. This section contains any combination of lexical

elements the meaning of which cannot be directly inferred from that of its components. Idioms included in the dictionary display several syntactic functions; there are nominal, adjectival, verbal, adverbial, prepositional, conjunctive and interjectional idioms. These units are treated as subordinate pieces of information within a main entry framework. More precisely, they are described in the article corresponding to their entry of reference; every idiom or phrase, on the other hand, can have one or more senses and every sense consists of the same (except for the syntactic conversions) subordinate structural elements as the senses directly related to the entry. This section of the article can be seen in the example of **amollar** v. 'to let go of':

• **amollar el xot** *loc. verb.* [N1 LOC] (N1[humà]) Expressar vehementment els pensaments continguts. [...] *no s'hy podian acostar de mitg' hõra lluny, porque tot ho passave per un desaire, emb una llengo de fõc, qui teyave cantons, y en amollar en es xõt tot eran botas y carratells, y altres hẽrbas [...].* [Aguiló (1834): N, p. 10].

• **amollar la corda** *loc. verb.* [N1 LOC (a N2)] (N1[humà]; N2[humà, facultat humana]) Afluixar la corda (*loc.*). *Ab la temensa que vos duya l'ermità, si vós li haguessiu posat carota, s'hauria aguantat una passada... pot èssere [...]. Mes vós li amollareu deseguida tota la corda y...* [Albert Paradís2 (1905): N, p. 291].

• **amollar les brides** *loc. verb.* [N1 LOC (a N2)] (N1[humà]; N2[humà, facultat humana]) Afluixar la corda (*loc.*). *Portem el nostre tresor en vasos trencadissos; fora, doncs, imprudència entrar pel camí de la vida amollant les brides a les nostres passions.* [Comerma (1933): 23, p. 101]. *Saber ser jove no val gaires penes, | ni té mèrit que deixi satisfet; | sols amollant la brida de les venes | l'empenta dels pocs anys t'ho dõna fet.* [Sagarra (1928): P, p. 186].

• **no amollar prenda** *loc. verb.* [N1 LOC] (N1[humà]) No dir res comprometedor. *Lo que ell no amolla prenda. Bé n'hi he tirades, d'indirectes, però com si no-res: o no ho entén, o no ho vol entendre.* [Tous1 (1947): T, p. 90].

The section of *variants* ("variants") encompasses three types of subelements: the *formal variants* of the entry ("var. formals", i. e. regional, social, etc. variants of the entry attested in the CTILC); the *appreciative* or *intensive derivatives* ("var. apreciatives", i. e. diminutive, augmentative, etc. forms of the entry), and the inflectional variants not foreseen in the inflectional pattern associated with the entry. Examples of the different kinds of variants can be seen in the articles of **albercoc** *m.* 'apricot' (formal variants), **agradar** *v.* 'to like' (formal and inflectional variants) and **abric** *m.* 'coat' (formal and appreciative/intensive variants):

VAR.: • **abercoc**, • **abrecoc**, • **abricoc**, • **albaricoc**, • **alberoc**, • **arbreccoc**, • **bercoc**, • **obrecoc**.

VAR.: • **gradar**; *agraes* 2pi, *agrà* 3pi, *agraen* 6pi, *agradavi* 3ii, *agradavin* 6ii, *agràdia* 3ps, *agrae* 3ps, *agrat* 3ps, *agradassen* 6is.

VAR.: • **abrigo**; *abrigàs*, *abrigó*, *abrigot*, *abriguet*.

The section on *derivation* "derivació" includes all the highly regular morphological derivatives sharing the same stem (or the classical form of the same stem, in the case of neo-

classical formations) as the word entry. Every derivative unit included has a maximally predictable meaning, because it is the result of an entirely regular derivative process (in terms of semantics, morphosyntax and phonology). Seven derivative processes with the specification of the affixes that can perform them have been established in order to avoid subjective decisions on the part of the lexicographers. For this reason, each derivative process is clearly defined on the basis of the type of transformation that it performs, the category and the meaning of the derivative it produces by means of a formulaic definition. Each unit is tagged with the number corresponding to its derivative process (“procés”), and as many examples as different senses conveyed by the derivative are taken as illustrating input coming from the CTILC. This section (if necessary) also contains collocations and conversions. An example of this section can be seen in the article of *amollar* v. ‘to let go of’, with two nominalizations:

- **amollada** f. (procés 6) *Concretament s’amollaren 10 perdius de les 20 aconseguides entre ICONA i la societat – la diferència es deu a les baixes produïdes quan s’estaven preparant per l’amollada* –. [Llum (1981): 41, niv. 1, p. 8] (acc. 2c). *Contemplant el moviment de corrioles que pujaven i baixaven, la isjada i l’amollada, l’entrada i la sortida dels carros, passava hores entretingudes*. [Pla (1934): N, p. 76] (acc. 4a).
- **amollament** m. (procés 6) *La humilitat de Verdaguer, en qui vosaltres i jo pensem arribats aquí, la seva terrible i documentada confessió [...], passa llampejant, llesta i maligna com d’un gran espatxí, de l’amollament a l’atac*. [Triadú (1961): 86, p. 23] (acc. 5c).

Finally, the article includes in a specific section of *complementary information* (“Inf. compl.”) all the lexical information related to the meanings and phraseology of an entry that has not been documented in the CTILC, but appear in one or more of the 13 dictionaries taken into account as subsidiary sources of the DDLIC. These dictionaries, which were selected from the most significant Catalan dictionaries from XIXth and XXth centuries, constitute the “lexicographical corpus” of the DDLIC and have been integrated in the BDLex database (Sanromà, 2004). The relevant list of dictionaries, with their abbreviations, is the following:

- DMFC** Febrer i Cardona, A. “Diccionári menorquí espanyol francès y llatí”. [manuscript, beginning of XIXth century].
- DEBJ** Esteve, J.; Bellvitges, J.; Juglà, A. *Diccionario catalan-castellano-latino*. 1803-1805.
- DLCL** Labèrnia, P. *Diccionari de la llengua catalana: ab la correspondencia castellana y llatina*. 1839-1840.
- DMCF** Figuera, P. A. *Diccionari mallorqui-castella*. 1840.
- DVCE** Escrig, J. *Diccionario valenciano-castellano*. 1851.
- NDMA** Amengual, J. J. *Nuevo diccionario mallorquin-castellano-latin*. 1858.
- DMCT** [Tarongí i Cortès, J.]. *Diccionari mallorquí-castellà*. 1878.
- DGMG** Martí i Gadea, J. *Novísimo diccionario general valenciano-castellano*. 1891.
- DCVB** Alcover, A. i Moll, F. de B. *Diccionari Català-Valencià-Balear*, 1926-1962.
- DPCV** Vallès, E. *Pal·las: diccionari català-castellà-francès: amb vocabularis castellà-català francès-català*. [1927].
- DGLC** Fabra, P. *Diccionari general de la llengua catalana*. Barcelona, 1932.

DGFP Ferrer Pastor, F. *Diccionari general*, 1985.  
 DIEC Institut d'Estudis Catalans. *Diccionari de la llengua catalana*, 1995.

Each item of complementary information is accompanied by one of the above abbreviations indicating the dictionary from where it was taken. In the cases in which the information included does not match with the entry of DDLC, the entry of the dictionary where the information can be looked it up is also indicated. We can see an example of this section in the article *amollar* v. 'to let go of':

COMPL.: [El bestiar] pasturar furtivament en [un camp on encara s'ha de collir la collita] [DGMG]; [El vi] esbravar-se [DCVB]; [El temps] abonçar-se [NDMA]; [Un arbre] abaixar-se de les branques per l'excés de fruita o per altra causa [DCVB]; [Un cos, una superfície tous] cedir a la força o a la pressió [NDMA]; Adquirir desimboltura o soltesa [a fer una cosa, en l'expressió, en una activitat] [DCVB, NDMA]; Arrencar a córrer [DCVB, DGMG, NDMA]; Deixar anar la vela fins que tingui el moviment necessari per a prendre el vent com cal [DCVB]; Deixar l'atenció al govern d'una nau quan va vent en popa [DGFP]; Deixar sense subjectar [un estrai, una burda] per donar-li pintura o quitrà [DCVB]; Descomptar [a algú] [una part del que havia de pagar] [DCVB]; Dir mentides [DCVB]; Disparar [una arma] [NDMA]; Jugar [una carta inferior a la que va jugada] quan se'n tenen de superiors, en el joc del revessí [DGMG, NDMA, DVCE, DLCL]; Perdre la vergonya o un altre motiu d'aturament [DCVB, NDMA]; Plegar de la feina [DCVB]; Relaxar [una imposició, una norma] [NDMA]; **amollar el xot** *loc. verb.* Adquirir desimboltura o soltesa [NDMA]; **amollar el xot** *loc. verb.* Lliurar-se a un vici, a una passió, a un sentiment [NDMA]; **amollar el xot** *loc. verb.* Plorar abundantment per un patiment contingut [NDMA]; **amollar la cadelleta** *loc. verb.* Xerrar molt [DMCT, NDMA, DMCF]; **amollar les rodetes** *loc. verb.* Parlar massa [DGMG]; **no amollar** *loc. verb.* Resistir, no servir respecte [a algú] [DMCT, NDMA, DMCF]

In order to complete this general review of the DDLC structure, we should mention two tags: the *non-standard* and the *not-documented* labels present in the CTILC. The non-standard label (which has the form of "•") indicates that the descriptive information included in and below the marked element is not accepted by (or simply not included in) the current standard norm (as established in the DIEC). This label may affect several elements of an article, and, moreover, it is always inherited by all the elements in lower levels of any of the informational structures already described. Thus, if the non-standard label is placed in front of the entry, it affects the entire article, if placed in front of a sense, it affects all the elements of the sense; however, when placed in front of a syntactic pattern, it affects only this pattern, and, if it is indicated just before a variant, it affects only this variant.

The purpose of the not-documented information mark is to include information that, even though it does not appear in the CTILC, is necessary for the descriptive coherence of the dictionary. Generally, the use of this label is limited to the function of completing syntactic regular alternations with a syntactic pattern (for example: [N1 V (N2)] ↔ °[N2 VPRON]), or of including the feminine/masculine form of a noun entry denoting people (for example: **creditor** *m.* °**creditora** *f.* ). The use of this label is infrequent, because the cases in which this specific circumstance occurs are few.

## 4 Work already done on the DDLC and present stage of development

### 4.1 Preparatory phase

The preliminary work on the DDLC started in 1998, once the compilation of linguistic resources (CTILC and BDLeX) was finished. The following was done:

a) Establishment of the list of CTILC lemmas to be included in the DDLC nomenclature. This list contains about 95,000 lemmas retrieved from the CTILC which were selected according to frequency and distribution criteria. The final number of entries may be slightly different because of the necessary adjustments between both entities (lemmas and dictionary entries) which imply some incorporations or suppressions during the writing process.

b) Systematization of the lexicographical criteria to be applied in the compilation of the dictionary.

c) Drafting of sample articles. Parallel to the systematization of lexicographical criteria, we wrote 600 samples that constituted a workbench for the testing of the initial criteria and the computer tools that were being implemented as a part of the lexicographical workstation at that time. This interaction allowed us to adjust these criteria to make them more suitable to the production phase of the writing task. Obviously, we tried to include a varied and exhaustive series of prototype articles in order to constitute a valid representation of all the issues involved in the lexicographic description.

d) The constitution of the lexicographical workstation. Together with the two previous activities, we developed the lexicographical workstation. This platform is the result of the modular integration of different components, such as databases, user-friendly tools for writing the dictionary, resources allowing the control and the validation of the whole process, computer programs used for assistance, modules of internal data transfer, operating systems, and so on. Thus, the lexicographers had, and still have, immediate access from their workstation to all the necessary elements to perform their individual work: the tools for writing as well as the linguistic resources.

### 4.2 Dictionary-making process

Having finished the preparatory phase, in 1999 we started the systematic writing of the dictionary. One of the most important features that distinguishes the DDLC from the majority of conventional dictionaries is that the dictionary-making process does not follow a pure alphabetic progression. Starting from alphabetic criteria as a primary selection filter for the entries to be described, in the same work package we also deal with the entries that are morphologically and/or semantically related to the selected entry. Thus, each lexicographer simultaneously deals with all the entries which are related according to these criteria; for example, when, by following alphabetical order, the entry *adulació* 'adulation' was reached, a list of 45 entries were added to this one in order to constitute the entire work package; these series of entries, assigned to a single lexicographer, contained the morphologically related entries of *adulació* (i. e. *adulador*, *adulaire*, *adular*, *adulatori*), and other entries that share at least one sense with *adulació* (i. e. *afalagar*, *alabar*, *elogiar*, *lloar*, and their morphological derivatives). From a methodological point of view, this is a successful mechanism in order to avoid or, at least, to dramatically reduce, the lack of coherence often characterising dictionaries.

One of the major concerns for the project managers is also to monitor the production results during the writing procedure, both from the quantitative and the qualitative point of view. This control is carried out by two means: *a)* through weekly meetings of every lexicographer with the coordinator and the editor of the dictionary, in order to discuss the issues raised during the preparation and the writing process of the articles; and *b)* through the statistical control of the rate of production of articles; which allows the project managers to follow up (with daily updated data) on the quantitative results of the work both for the team as a whole and for each individual lexicographer.

Finally, another important task for the project follow-up is the structural validation of the written entries. This is an operation carried out on the articles already written and consists of the systematic checking of the application of the criteria and also of the validation of the structural coherence of the whole dictionary.

#### ***4.3 The electronic version of DDLC***

From the beginning of the project, we have made the part of the dictionary already written available to the public. This strategy of disseminating the work fits in perfectly with the general philosophy of the project. Electronic and dynamic publication of results that increase daily and free on-line access to the DDLC data are fully consonant with our methodology, the agenda and the above-mentioned innovative approach. After having obtained quite a large amount of data, in February 2005 we put the dictionary contents on-line, at the address <http://dcc.iec.cat/ddlc>.

The electronic edition contains a core module – with the dictionary articles – and some complementary modules such as a brief presentation and a detailed user's guide. All queries directly apply to the dictionary database and, for this reason, the electronic version is constantly updated: the number of articles increases as the writing of a work package is finished. Moreover, the homepage also includes a contact link waiting for the users' comments about the articles published in the DDLC, which will be used in order to answer all questions with regard to the structure and criteria used in the dictionary.

One of the major achievements of the electronic version is to give a public and dynamic access to the DDLC reflecting not only the relationship among elements belonging to a single article or to different articles, but also the references to complementary external items (such as descriptive criteria, corpus quotations references, etc.) by means of active links. In this sense, the electronic version of the DDLC is the most suitable format to dynamically publish the results during the writing process of the dictionary. It is also the most suitable format for browsing through articles according to their the data structure (in that it allows some of the internal links in the dictionary to be seen externally) and also because it enables final users to interact with the data and gives them access to a real-time updated online version of the work.

#### **5 Conclusions**

The general review of the main characteristics of this project allows us to conclude that the DDLC has important differences with regard to dictionaries produced by conventional

lexicography. Since the DDLC was conceived from the very beginning as a multifunctional lexical database, the dictionary that we have presented in this paper must be considered as just one of the several possible outputs of the project. As a lexical database, the DDLC goes beyond a conventional dictionary, and it satisfies all the requirements of linguistic resources (accessibility, reusability, documentation, formalization of data, maximum explicitness, etc.). Furthermore, this strategy incorporates some innovations in dictionary making such as: a permanently updated version of all the already written dictionary articles accessible via Internet, structural validation of the data in the dictionary-making process, and quantitative control of the work done, among others. Thus, the DDLC can be considered an advanced lexicographical project that benefits from descriptive linguistics and computational lexicography.

## References

### A. Dictionaries

- DCVB = ALCOVER, A. M. & MOLL, F. de B., *Diccionari català-valencià-balear* (10 vol.), Palma de Mallorca: Editorial Moll, 1926-1962.
- DGLC = FABRA, P., *Diccionari general de la llengua catalana*, Barcelona, Llibreria Catalònia, 1932.
- DIEC = INSTITUT D'ESTUDIS CATALANS, *Diccionari de la llengua catalana*, Barcelona [etc.]: Edicions 62 [etc.], 1995.
- DECLC = COROMINES, J., *Diccionari etimològic i complementari de la llengua catalana* (9 vol.), Barcelona: Curial / Caixa de Pensions "La Caixa", 1980-1981.

### B. Other Literature

- Alsina, V. & DeCesaris, J. (eds.) (2001), *Estudis de lexicografia*, Barcelona, Universitat Pompeu Fabra.
- Battaner, P. & DeCesaris, J. (eds.) (2004), *De lexicografia. Actes del I Symposium Internacional de Lexicografia* (Barcelona, 16-18 de maig de 2002), Barcelona, Universitat Pompeu Fabra.
- Geeraerts, D. (1985), 'Les données stéréotypiques, prototypiques et encyclopédiques dans le dictionnaire', *Cahiers de Lexicologie*, 46, 1, pp. 27-43.
- Rafel, J. (1989), 'Sobre la normalització dels diccionaris catalans: els elements intrínsecs i extrínsecs en les definicions lexicogràfiques', in *La Corona d'Aragó i les llengües romàniques. Miscel·lània d'homenatge per a Germà Colon*, Tübingen: Gunter Narr.
- Rafel, J. (1994), 'Un corpus general de referència de la llengua catalana', *Caplletra*, 17, pp. 219-250.
- Rafel J. (2000), 'El Corpus Textual Informatitzat de la Llengua Catalana i l'activitat lexicogràfica de l'Institut. Aspectes descriptius i aspectes normatius', *Jornades per a la Cooperació en l'Estandardització Lingüística*, Barcelona, Institut d'Estudis Catalans, pp. 197-205.
- Rafel, J. (2005), *Lexicografia*, Barcelona, Universitat Oberta de Catalunya.
- Rafel, J. (dir.) (1996-1998), *Diccionari de freqüències*. 3 vols + 2 CD-ROM, Barcelona: Institut d'Estudis Catalans.
- Rey-Debove, J. (1971), *Étude linguistique et sémiotique des dictionnaires français contemporains*, The Hague, Mouton.
- Sanromà, R. (2004), 'Constitució i explotació d'una base de dades lexicogràfica (BDLex)', in Battaner, P. & DeCesaris, J. (eds.) (2004), pp. 741-754.
- Seco, M. (1978), 'Problemas formales de la definición lexicográfica', in *Estudios ofrecidos a Emilio Alarcos Llorach, II*, Oviedo, 217-39 [reprinted in Seco, M. (1987), 15-34].
- Seco, M. (1979), 'El 'contorno' en la definición lexicográfica', in *Homenaje a Samuel Gili Gaya (in memoriam)*, Barcelona, pp. 183-191 [reprinted in Seco, M. (1987), 35-45].
- Seco, M. (1987) *Estudios de lexicografía española*, Madrid: Paraninfo.
- Soler, J. (2000), 'Producing LR's in Parallel with Lexicographic Description: the DCC Project', in

- LREC 2000. Second International Conference on Language Resources and Evaluation. Proceedings*, Athens, European Language Resources Association, pp. 341-346.
- Soler, J. (2001), 'Elements extrínsecs i definicions adjectivals', in Alsina, V. & DeCesaris, J. (eds.) (2001), pp. 49-67.
- Soler, J. (2006), *Definició lexicogràfica i estructura del diccionari*, Barcelona: Institut d'Estudis Catalans (in press).
- Zampolli, A. (1991a), 'Los bancos de datos léxicos: Bases multifuncionales de datos léxicos', in Vidal Beneyto, J. (ed.) (1991), pp. 127-146.
- Zampolli, A. (1991b), 'Hacia bases multifuncionales de datos léxicos', in Vidal Beneyto, J. (ed.) (1991), pp. 185-202.
- Vidal Beneyto, J. (ed.) (1991), *Las industrias de la lengua*, Madrid: Fundación Germán Sánchez Rupérez.