

## A Digital Dictionary of Catalan Derivational Affixes\*

Elisenda Bernal  
Janet DeCesaris  
Universitat Pompeu Fabra

*This paper presents the digital dictionary of Catalan derivational affixes, awarded with the Laurence Urdang Award 2005. The aim of this dictionary is to provide a tool for lexicographers that will help them systematize the representation of the language's morphology in dictionaries, and provide an in-depth description of Catalan derivational affixes in the form of a dictionary that should be of interest both to linguists and language professionals and those seeking a model for similar projects dealing with other languages.*

### 1. Background

In 2006 the authors received the Laurence Urdang Award for their proposal entitled "A digital dictionary of Catalan derivational affixes". This dictionary has two main objectives: on the one hand, the dictionary is designed as an aid to lexicographers, to help them systematically represent the morphology of the language in dictionaries; on the other, the description of Catalan affixes necessary to prepare the dictionary should provide a better understanding of affixation and the nature of affixes in the language. We hope that the dictionary will be of interest to linguists and language professionals in general.

Derivational morphology in Catalan, like in the other Romance languages, is a productive process and involves many affixes (both prefixes and suffixes). In order to obtain a finished product within a reasonable time span, we have decided to concentrate on four groups of affixes that are representative of the productive derivational morphology in the language:

- suffixes that form nouns from verbs and adjectives: *-ada* (*fiblada* "insect bite"), *-alla* (*rialla* "laughter"), *-ança* (*enyorança* "homesickness"), *-ància* (*tolerància* "tolerance"), *-atge* (*aterratge* "landing"), *-ció* (*equivocació* "mistake"), *-dera* (*cruixidera* "crunch"), *-dissa* (*trencadissa* "scattering"), *-dura* (*soldadura* "soldering"), *-ença* (*prometença* "promise"), *-ment* (*aplaudiment* "applause");
- suffixes forming agentive nouns: *-ant estudiant* "student", *-dor caçador* "hunter"; *-er fuster* "carpenter"; *-ista pianista* "pianist";
- negative prefixes: *des-* *desfer* "undo", *in-* *inanimat* "inanimate", *a-* *atípic* "atypical";
- verb-forming native prefixes: *a-* *abaratir* "to lower the price of", *en-* *ennegrir* "to blacken".

This is not a complete list of all affixes in Catalan, but if our project is to improve current lexicographic practice and provide a more accurate description of the language it must be thorough.

The dictionary we are working aims at providing a better description of affixes. It also will list all the words that constitute a derivational series. In a language like Catalan, in which

---

\* Work on this paper was supported by the 2005 Laurence Urdang Award, which the authors gratefully acknowledge.

derivational morphology is very productive and, to a large extent, semantically transparent to speakers, it is not odd to find series of derived words that dictionaries list as synonyms:<sup>1</sup>

<b>cessació</b> f	Acció de cessar; l'efecte.	“cease to verb”
<b>cessament</b> m	Cessació.	“firing”
<b>penetració</b> f	Acció de penetrar; l'efecte.	
<b>penetrament</b> m	Penetració.	“penetration”
<b>guia</b> f	Acció de guiar.	
<b>guiament</b> m	Acció de guiar, guia.	
<b>guiatge</b> m	Acció de guiar; l'efecte.	“guiding”
<b>promesa</b> f	Acció de prometre, prometiment, prometença.	
<b>prometença</b> f	Acció de prometre, promesa.	
<b>prometiment</b> m	Acció de prometre, promesa, prometença.	“promise”

It is clear that the dictionary representation of these derivational series is problematic, for two main reasons:

- a) The definitions are circular and leave many questions unanswered. Are all of these words true synonyms, or are they partial synonyms? Are they really interchangeable in all contexts? Corpus data is able to shed some light on this question. The data from the Computerized Text Corpus of the Catalan Language (“*Corpus textual informatitzat de la llengua catalana*”), a corpus with over 50 million words compiled by the Institut d’Estudis Catalans (Catalan Academy of Sciences), show us that there is, for example, a significant difference in use of the nouns derived from the verb *prometre* “promise”:

<i>promesa</i>	1991
<i>prometença</i>	327
<i>prometiment</i>	1

Although *prometiment* (“promise”<sub>N</sub>) may have the same meaning as *promesa*, use of these two words is clearly not comparable. In that sense, the dictionary representation is rather misleading, because as written it might lead you to believe that these three words are used indistinctly by speakers, although that is clearly not the case.

- b) The coexistence of related forms in general in the Romance languages and specifically in Catalan is quite frequent, as we showed in an earlier paper on abstract nouns derived from verb stems by means of suffixation (Bernal & DeCesaris 2006). The dictionary shows that, of a total of 211 derivational series (i.e., a single verb stem having more than one nominalization meaning “action of VERB”), many verb stems have two, three and even four possible nominalizations:

	combinations	series	%
2 elements	45	173	64.29
3 elements	20	33	28.57
4 elements	5	5	7.14
<b>Total</b>	70	211	100

It is not the case that two specific suffixes are responsible for the synonymy: the column entitled “combinations” indicates that, for example, for those verb stems for which there are two derived nouns meaning “action of VERB”, there are 45 possible combinations of

<sup>1</sup> All dictionary examples are taken from the *Gran diccionari de la llengua catalana* published by Enciclopedia Catalana (hereinafter, *GDLC*) and are representative of lexicographic practice in Catalan.

suffixes. These data point to a derivational system in which many suffixes are partially synonymous, although it would seem unlikely for all of the derived forms to be fully synonymous in terms of meaning and even less likely for them to be equivalent in usage.

If the dictionary representation of derivational series is to be improved, the relevant data must be available in a format that can be consulted quickly and easily. In this sense, we hope that the dictionary that we are working on will be an aid to lexicographers and will help reduce the amount of circularity that currently characterizes dictionaries of Catalan.<sup>2</sup>

## 2. Our starting point: the DSVC

Bernal (2000) develops a model for a digital dictionary of verbalizing suffixes, and is at the core of our project.<sup>3</sup> The *Diccionari de sufixos verbalitzadors del català* (“Dictionary of Catalan verbalizing suffixes”, hereinafter DSVC) is a prototype in that it contains a partial set of suffixes, but complete in the information provided for each of the suffixes analyzed. The dictionary is structured in terms of interrelated modules: there is a module for the suffixes, a module for the derived verbs and a module for the meaning of the derived verbs. This structure allows users to search for all verbs with a semantic feature that has been identified, for example, all causative verbs, or to search for all the affixes that form causative verbs.

Each module in the dictionary contains sufficient and necessary information, which can be summarized as follows:

- a) The module for meaning has the following fields:
  - morphological operators
  - argument structure
  - generic meaning
  - definition [= paraphrase of meaning]
  - list of verbs that result from application of the word-formation rule
- b) The module with morphological operators includes all the information pertinent to the lexical entries for affixes:
  - phonological representation
  - morphological information: grammatical category, type of stem to which the affix can attach, selectional restrictions of the verb stem
  - syntactic information, including the argument structure of derived verbs
  - semantic information, including the typical meaning of the suffix
  - etymology
  - productivity
  - list of derived verbs with the affix that are contained in the DSVC

---

<sup>2</sup> We note that the same sort of circularity is commonplace in the dictionaries of Spanish we have consulted.

<sup>3</sup> Bernal’s dissertation adopts Danielle Corbin’s (1987) theory of derivational morphology, as this model takes into account the meaning of word-formation processes and is designed to be applied in the context of a dictionary. The fact that Corbin’s model relates morphological analysis and possible dictionary representation explains the use of terminology such as morphological operators (“*operadors morfològics*”) and word-formation rules (“*regles de construcció de paraules*”) in the dissertation and dictionary.

- c) The module for derived verbs has a more complex structure, because it includes all the syntactic and semantic information required for each verb. It also contains examples and semantic relations with other verbs in the same derivational series, if they exist. Below we reproduce in Catalan the entry for verbs:<sup>4</sup>

<b>lema</b> •	
Estructura morfològica:	[[base] <sub>X</sub> sufix ] <sub>V,conjugació</sub>
Estructura argumental:	argument extern: agent / causa / instrument argument intern: experimentador / meta / origen / lloc / tema
Incoatiu:	[± se]
RCM:	1. meteorològic / 2. iteratiu / 3. freqüentatiu / 4. causatiu / 5. despectiu / 6. canvi d'estat
Verbs coradicals:	RCM <sub>n</sub> verb_1    RCM <sub>n</sub> verb_2
definició parcialment formalitzada	
relacions semàntiques amb altres verbs	
idcontext 1	▶
idcontext 2	▶
...	
idcontext n	▶

Figure 1. Entry for verbs

The three modules appear in three different parts of the computer screen, so that all the information is always available, independent of the type of search that the user begins with. In this way, users can easily move from suffix to verb or to group of verbs, as the following screen shows:

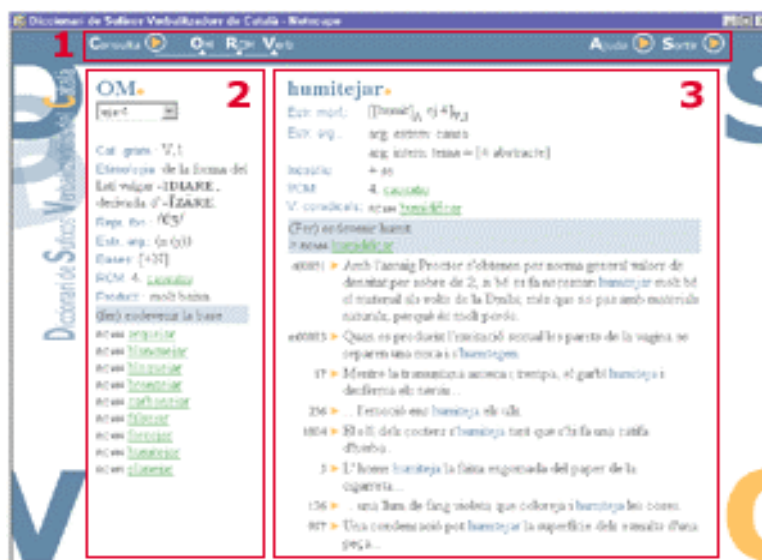


Figure 2. Modules

<sup>4</sup> English equivalents for the terminology used in the entry: Morphological structure: [[stem] X suffix]<sub>V</sub>, conjugation; Argument structure: external argument: agent/cause/instrument; internal argument: beneficiary, goal, source, locative, theme; Inchoative: [± reflexive pronoun se]; Word-formation rule: 1. weather /2. iterative / 3. repetitive /4. causative / 5. disrespectful /6. change-of-state; Verbs with the same stem; partially formalized definition; semantic relations with other verbs.

Zone 1 corresponds to the Tool Bar, which contains the three types of consultation possible, the help key, and the exit key that closes the DSVC window. Zone 2 is the consultation frame, where you see the various screens to look up the verbs listed in the dictionary. Verbs can be accessed by looking up the suffix, by looking up the word-formation rule, or by looking up the meaning, and the list of all verbs can be ordered alphabetically, by suffix or by meaning. Zone 3 contains the entry for the verb.

Navigation in the DSVC is conceived so that any subsequent consultations made by the user do not adversely affect the presentation of the data. This means that the verb shown in the verb entry zone will always appear in the verb list in Zone 2. If the user looks up a different verb, the consultation frame with the information about the suffix or the corresponding word-formation rule will also change. If the consultation frame is changed, the entry for the first verb on the list will change, as long as the verb displayed is not a part of the consultation frame.

### 3. Current status of the project and future work

At the time of this writing (March 2008), we have almost completed the description of the list of affixes that we decided to include in the project. We therefore expect to be able to make this dictionary available on the web free-of-charge next year (target completion date: June, 2009).

We are attempting to incorporate some changes from the original digital format proposed in Bernal (2000), which may be summarized as follows:

- the terminology used is not specifically tied to D. Corbin's theory; labels are more general (e.g. affixes, meaning, lemmata);
- the system should allow searches for derived stems;
- the system should represent compatibility and incompatibility of derivational processes (e.g. derived verbs with the prefix *a-* allow nominalizations with *-ment* and sometimes with *-ada*, but not with *-ció*, as in *abaratir-abaratiment-\*abaratició* "cheapen" or *afluixar-afluixament-afluixada-\*afluixació* "weaken").

We are convinced that internal coherence and systematicity are basic to dictionary-making. Our project is conceived for a digital format, which facilitates control of the data and ensures that the right information is represented in all the pertinent entries. This format allows users to access information by means of several types of queries. The fact that this dictionary has started out as a digital dictionary, as opposed to being a digitalized dictionary (in other words, it is not a dictionary designed for a paper format that has been made available in a digital format), is a conditioning factor that we believe is positive.<sup>5</sup> For example, the digital format creates its own resources by means of hypertext, so that information is not only organized into lists but rather into networks. In this sense, our project has demonstrated to us the main advantages to this type of dictionary revolve around use, data presentation, search capabilities and technical aspects. Since our project is addressed mainly to language professionals, the lack of space limitations, thus enabling us to include in the dictionary all the information we want to include, is important. A feature that should be of interest to other types of dictionary projects is that people who use our dictionary constantly create their own paths while looking up information, so that they become active (as opposed to passive) users.

As interesting point that we would like to develop in future work relates to influence on morphology in language contact. Given that speakers of Catalan are also speakers of Spanish, the identification and precise description of "legitimately" Catalan words is still an open question. Catalan and Spanish share many morphological properties, but obviously not all of them. For example, the verb *sugerir* "suggest" in Spanish yields the noun *sugerencia* "suggestion" and the adjective *sugerent* "suggestive", which have given rise to the forms *suggerència* and *suggerent* in Catalan. These new forms in Catalan are growing in usage,

---

<sup>5</sup> See Gelpí (2003) for more discussion on digital vs. digitalized dictionaries.

even though the standard grammar only accepts *suggeriment* and *suggeridor*, respectively (note the different suffixes used).

An important goal for those of us engaged in linguistic research is to facilitate access to complex linguistic analyses for the people who can most benefit from them. Given the modern workplace, it is unrealistic to expect professional lexicographers and other language professionals to have the time to research every linguistic problem that arises in their work. Our project attempts to integrate a sophisticated analysis of Catalan morphology into a format that is easy for others to use, so that they can produce better dictionaries.

## References

### *Works cited*

- Bernal, E. (2000). *Els sufixos verbalitzadors del català. Relacions semàntiques i diccionari*. Barcelona: Institut Universitari de Lingüística Aplicada – Universitat Pompeu Fabra. Also available at <http://www.tesisenxarxa.net/TDX-0404102-134006/index.html>. The DSVC is available at [www.elisendabernal.com](http://www.elisendabernal.com) (> *tesi doctoral* > *Accés al DSVC*).
- Bernal, E.; DeCesaris, J. (2006). “Nominalitzacions deverbals: distribució formal i semàntica” [CD-ROM]. In *Actes del VII Congrés de Lingüística General*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona.
- [CTILC]. Rafel, J. (dir.) (1996). *Diccionari de freqüències. I. Llengua no literària*. Barcelona: Institut d’Estudis Catalans.
- Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Max Niemeyer Verlag.
- Gelpí, C. (2003). “El estado actual de la lexicografía: los nuevos diccionarios”. In Medina Guerra, A. (coord.). *Lexicografía española*. Barcelona: Ariel. 307-327.

### *Works related to the Digital Dictionary of Catalan Derivational Affixes*

- Bernal, E. (2006). “Absències en els diccionaris: llacunes lèxiques i llacunes lexicogràfiques en els substantius deverbals del sufix *-itzar*”. *Catalan Review* 20. 51-67.
- Bernal, E. (2007). “Formació de verbs, variants morfològiques i neologismes: entre la genuïnitat i el calc”. In Lorente, M.; Estopà, R.; Freixa, J.; Martí, J.; Tebé, C. (eds.). *Estudis de lingüística i de lingüística aplicada en honor de M. Teresa Cabré Castellví. Volum II: Deixebles*. Barcelona: Institut Universitari de Lingüística Aplicada – Universitat Pompeu Fabra. 197-208.
- Bernal, E. (forthcoming). “Els prefixos verbalitzadors *a-* i *en-*: elements per a un diccionari digital”. II Congrés Internacional de Lexicografia Hispànica.
- Bernal, E. (forthcoming). “Elementos para un diccionario digital de afijos del catalán: los prefijos negativos”. XXV Congrès International de Linguistique et Philologie Romanes.
- Bernal, E.; DeCesaris, J. (forthcoming): “La lexicografía digital al servicio de la morfología: los sufijos agentivos del catalán o la rivalidad bien entendida”. III Congreso Internacional de Lexicografía Hispànica.