A New Dictionary of Contemporary Catalan Pronunciation

Joan Julià-Muné and **Imma Creus**

Universitat de Lleida (Catalonia, Spain)

E-mail: jjulia@filcat.udl.es, icreus@filcat.udl.es

ABSTRACT

The authors' *Diccionari de pronúncia catalana* (*DPC*) is being compiled as a descriptive Catalan pronouncing dictionary that offers the recommended and common use pronunciation of both Central and North Western (NW) Catalan, and therefore covers the contemporary pronunciation of the majority of Catalan speakers. It provides extensive coverage of specialized vocabulary and, unlike its predecessors, includes a large number of proper names, both in Catalan and other languages, as well as acronyms and onomatopoeias, loanwords and neologisms, and NW dialectalisms. Entries also offer inflected and derived forms. IPA symbols are used to represent all pronunciations by means of relatively narrow transcriptions so that they may be useful to Catalan speakers as well as to foreign learners of Catalan. Nearly five thousand pronunciations were specially verified through usage polls.

1. INTRODUCTION

The forthcoming dictionary is the result of a research project directed by Joan Julià-Muné at the Phonetics Laboratory of the University of Lleida (Catalonia, Spain). The culmination of ten years' work, it is modelled on J.C. Wells' *Longman Pronunciation Dictionary* [17] for English as well as on the traditional research work on pronunciation undertaken at the Department of Phonetics and Linguistics, University College London.

Initially, a study was carried out of the pronunciation of the lesser-known NW variety of Catalan by conducting usage surveys to ascertain variant pronunciations according to sex, age, education and professional groups. Once a large corpus of data had been obtained, a second data-management stage was set up. It was decided that the phonetic variant which was considered standard should be presented in first position and that the other variants would follow according to usage frequency. Soon afterwards it was decided to include the Central Catalan variant for each entry in order to represent the vast majority of pronunciations.

It was also decided that the dictionary should include specialized vocabulary, proper names, loanwords and neologisms in addition to common vocabulary.

1.1. PURPOSE

The *DPC* is aimed at contributing to the knowledge of Catalan descriptive phonetics by offering contemporary pronunciation data derived from Catalan usage. The Catalan language has suffered a dearth of references on pronunciation since the multidialectal, ten-volume *Diccionari català-valencià-balear* [2], which requires updating, up to the recently compiled *Atles lingüístic del domini català* [16], whose publication is in progress. In between, few works have appeared, starting with the first Catalan pronunciation dictionary [3], which was restricted to Central Catalan and which offered just one possible pronunciation, so with no variants and few proper names, loanwords and neologisms. More recently a vocabulary [12] of about 2,000 words was published with two main pronunciations (Central and NW Catalan). It had an orthoepic purpose and used broad transcriptions aimed at Catalan speakers. Finally, a dictionary which covers the Valencia variety [8] was compiled with a large range of entries but with no proper names and just one pronunciation for each entry. The aim of the *DPC* compilers, therefore, has been to improve upon and complete these recent works and to update the earlier ones, so that Catalan users and other speakers interested in the pronunciation of Catalan, especially those involved in phonetic variation studies, might have a useful linguistic tool at their disposal.

1.2. STRUCTURE

Basically, the *DPC* includes an introduction, the corpus with variant pronunciations and a list of NW Catalan dialectalisms.

The introduction gives information on similar works on other languages such as English ([7] and [17]) and French ([9]) as well as on the antecedents of works on Catalan pronunciation. It provides extensive information on spelling-to-sound guidelines and on standard pronunciations proposed for both varieties. It also recommends further readings on the subject after giving detailed information on the pronunciation of Catalan regional varieties and the main orthoepic guidelines following the proposals of the Catalan academy of language, the *Institut d'Estudis Catalans*.

However, the bulk of the work comprises the set of lexical entries with corresponding pronunciations, whose characteristics are given below.

As a final stage, the dictionary offers an everyday NW Catalan vocabulary that is not included in general Catalan language dictionaries [5] [6]. It includes the meaning for

each entry. Several lexicographical works have been used and extensive surveys undertaken to gather data [4].

2. CONTENTS

The *DPC* contains about 70,000 entries and over 200,000 pronunciations arranged in the following manner:

2.1. CORPUS AND SOURCES
The corpus of the dictionary includes common and specialized vocabulary that accounts for approximately 85% of the entries, proper names of both Catalan and foreign origin (9%), loanwords and neologisms (3%), acronyms and onomatopoeias (2%), and dialectalisms (1%).

Common vocabulary covers most of the entries of our general dictionary [6], recent incorporations in later works [5] and the above-mentioned dialectalisms or specific NW Catalan terms.

Specialized entries includes literary, scientific and technological terms, covering computing, law, medicine, and communications.

Apart from this, the *DPC* offers good coverage of Catalan and foreign *proper names* such as:
• people (forenames, surnames, literary characters, gods, heroes, and figures of myth and legend);
• places in the so-called Catalan-speaking countries and in every continent abroad;
• commercial firms and products;
• artistic, such as classical and pop music.

At this stage, different works have been used [1] [7] [11] [17] in addition to *ad hoc* fieldwork and the authors' experience on the subject.

Fo*r loanwords* and *neologisms*, the authors have benefited from a recent work [15] and have added several entries that have not yet been incorporated into any other dictionary (e.g. *chapapote/(t)xapapote* 'oil spilled at sea which damages the coastline').

With regard to *acronyms* and *onomatopoeias*, our sources have been [10] and [13] respectively.

2.2. ENTRY CHARACTERISTICS
Most *DPC* entries appear in ordinary spelling -in alphabetical order- in lower case (i), except for proper names, which start with a capital letter (ii), and in bold.

If required, inflectional information and supplementary symbols are added, such as those following a specific entry: ® for a trademark (iii) and * for the NW dialectalisms (iv). Loanwords whose spelling has yet to be adapted to Catalan are written in italics and bold (v). Next, morphological information is given that relates either to the lexical category (verb, adjective, adverb, etc.) or to the grammatical category in noun cases (masculine

or feminine). This kind of information is abbreviated and in italics. Finally, the phonetic transcription is provided.

(i) **adjuntar, -se** *v.* [a.dʒunˈta\r.se ˈtas] ‖ [ədʒunˈta\rsə]

(ii) **Alfons, -a** *m. & f.* [aɫˈfons\ɛ] ‖ [əɫˈfons\ə]
   Barceloneta [bar.se.ɫoˈne.tɛ] ‖ [bərsəɫuˈnɛtə]

(iii) **Alfa Romeo®** [ˌaɫ.fa.roˈme.o] ‖
   [ˌaɫfaroˈmeo | ˌaɫfəruˈmɛw]
   it. [ˈalfaroˈmeo]

(iv) **empostada*** *f.* [em.posˈta.ðɛ |am.]

(v) ***ad naturam*** *loc. adj. & adv.* [ad.naˈtu.ɾam]

3. PHONETIC INFORMATION

This is the most important aspect of a pronunciation dictionary. So, users should know, to start with, the criteria that have been applied for the presentation of the transcribed forms.

3.1. TRANSCRIPTION TYPES AND PHONETIC SYMBOLS
The *DPC* includes two kinds of phonetic transcription, depending on the languages involved: (a) those in Catalan for the NW and Central varieties, and (b) occasionally those transcriptions in over thirty different languages for the original words of some foreign proper names, loanwords and neologisms. In all cases IPA symbols have been used [SIL Doulos IPA], as in other recent works in Catalan linguistics such as the *GCC* [14], depending on the phonetic characteristics of the language in question (vi). Whenever a foreign word is transcribed, the transcription available in the respective pronunciation dictionary or in further works of reference is generally used.

(vi) **Asterix** *m.* [asˈtɛ.ɾiks] ‖ [əsˈtɛɾiks]
   fr. [asteˈʁiks]

Relatively narrow transcriptions are used because of the phonetic information they carry, in order to be useful to both Catalan speakers and foreign learners. The most relevant aspects of this type of transcription are the symbols for the approximants [β ð ɣ] -note the absence of the diacritic-, that are allophonic variants of /b d g/ in complementary distribution with [b d g], and for the

velarized alveolar lateral [ɫ], which is the usual realization of /l/ in both Catalan varieties. Therefore, the sets of symbols used to transcribe Catalan entries are as follows:

vowel sounds: [i e ɛ a ə ɔ o u]
consonant sounds:

[p b t d k g f v s z ʃ ʒ ts dz tʃ dʒ m ɱ n ɲ ŋ r ɾ ɫ ʎ β ð ɣ j w]

(see the *phonemic inventory:*

/ i e ɛ a (ə) ɔ o u \ p b t d k g f s z ʃ ʒ ts dz tʃ dʒ m n ɲ (ŋ) r ɾ l ʎ j w /)

As noted in the above examples, *primary* and *secondary* stress are shown by the respective IPA symbols [ˈ] [ˌ]. *Lengthening* is marked with [ː] in specific cases, such as vowel length as a result of the contact of two identical vowel sounds ([aa] > [aː]) or the usual lengthening of bilabial and velar voiced plosives in consonant clusters -*bl-*, -*gl-* ([bːl] [gːl]) following the stressed syllable of the word. However, for lengthened sounds resulting from assimilation the geminated symbol has been preferred ([dm] > [mm]). On the other hand, few examples with raised reduced symbols will be found in Catalan. These are the case of some NW palatal approximant pronunciations before a palatal fricative: [ˈrɔʲʒɛ] (see [ˈkajʃɛ]). In other languages, such as English, this kind of symbolization is used whenever required.

It is also important to point out that the *DPC* incorporates *syllabic divisions* by using a dot, but only in the case of NW Catalan transcriptions. This is a further contribution of this dictionary, since no previous work has included this feature, except for [3] that shows syllabification on the lexical entries. Central Catalan syllabification can be easily inferred from the corresponding NW form. However, it is not shown in foreign language transcriptions because it is either difficult to get information on the subject or a controversial matter, as in the case of English ([7] and [17]).

Phonetic transcription is arranged as follows: it is written within square brackets and *segmental* variation is shown after a vertical bar (|) and *morphological* variation after a slanted bar (\). NW transcriptions appear in the first place. After a double bar (‖) Central Catalan transcriptions are shown, following the same procedure in respect of phonetic and morphological variation. Whenever a foreign language transcription is shown, the language is stated in short form and italics (e.g. *angl.*, *it.*).

3.2.   PHONETIC VARIANTS
They are arranged as follows:

a)   A single phonetic form with or without variants

This is the most unusual case. It is usually presented in this way whenever the pronunciation coincides in the two

varieties or whenever the Catalan word is of foreign origin and the pronunciation does not differ from the original (vii), or if a dialectalism is involved (viii), since NW dialectalisms are unusual words in Central Catalan.

(vii) **dimissió** *f.* [di.mi.siˈo |ˈsjo |ˈzjo (!)]

ad libitum *loc. adj. & adv.* [adˈɫi.βi.tum]

(viii) **escatxigar*** *v.* [es.ka.tʃiˈɣa |as.]

b)   One or more forms for NW Catalan followed by one or more forms in Central Catalan

The *DPC* contains a single form for each variety whenever there is no phonetic variation. This is usually the case with seldom-used, specialized words or neologisms (ix). However, the most common arrangement offers different forms for the NW variety and one or two for the Central variety (x). This last variety just incorporates the standard form, which may be twofold (see xii), and sometimes even for Central Catalan a further variant is shown if variation is extensive (xi).

(ix) **cinemòmetre** *m.* [si.neˈmɔ.me.tɾe] ‖

[sinəˈmɔmətɾə]

eurodiputat *m.* [ˌew.ɾo.ði.puˈtat] ‖

[ˌɛwɾuðipuˈtat]

(x) **alfabetitzar, -se** *v.*

[aɫ.fa.βe.tiˈdza\r.se |ˈdzas |ˈza\ˈzas] ‖
[əɫfəβəˈtidza\rsə]

(xi) **confortablement** *adv.*
[koɱ.for.ta.βɫeˈmen |.bɫe' |.pɫe'] ‖
[kuɱfurˌtabːɫəˈmen |pɫə]

It should be clear by now that whenever more than one variant is shown, the first one corresponds to the standard form at either side of the double bar. The following forms either may compare with the standard in a way, though not so recommended, or it may be considered the usual form in informal registers (xii). From time to time orthoepic observations are made and forms not (!) recommended are mentioned (xiii).

(xii) **llei** *f.* [ˈʎej] ‖ [ˈʎej | ʎɛj]

tema *m.* [ˈte.mɛ] ‖ [ˈtemə | ˈtɛmə]

ja *adv.* [dʒa | ja] ‖ [ja | ʒa]

(xiii) **atmosfera** *f.*

[ad.mosˈfe.ɾɛ |ˈmɔs. (!)]
[ədmusˈfeɾə | ədˈmɔsfəɾə (!)]

c) *One or more forms for Catalan followed by one or more forms in a foreign language*

This is usually the case for loanwords or neologisms (xiv).

(xiv) **achtung**

['ax.tuŋ | 'ah.tuŋ] *al.* ['ʔaxtʊŋ]

Folies Bergère

[fo‚ɫi.βer'ʒer] ‖ [fo‚ɫiβɛr'ʒɛr]

fr. [fɔlibɛʁ'ʒɛːʁ]

Mao Zedong / Mao Tse-Tung

[‚mao.tse'tuŋ] ‖ [‚mawtsə'tuŋ]

mand. ['maw'tsə'tuŋ]

4. CONCLUSIONS

The *DPC* aims at being essentially a pronouncing dictionary of present-day Catalan as spoken in Catalonia. It shows sufficient phonetic variation to cover current usage and a sufficient number of entries to become useful to Catalan speakers as well as to foreign learners of the language. It is expected to be compiled in machine-readable format and it will be made available in the near future as an electronic database, which may prove useful for speech technology applications such as speech recognition and speech synthesis.

ACKNOWLEDGMENTS

REFERENCES

[1] J. M. Albaigès, *Diccionari de noms de persona*, Barcelona: Edicions 62, 2000[11] [1980].

[2] A. M. Alcover and F. de B. Moll, *Diccionari català-valencià-balear* (10 vol.), Palma de Mallorca: Moll, 1979 [1926/1930-1962].

[3] J. Bruguera, *Diccionari ortogràfic i de pronúncia*, Barcelona: Enciclopèdia Catalana, 1990.

[4] I. Creus, R. González and J. Julià, "Dialectalismes i diccionaris: entre la tradició i la variació," *Sintagma*, 10, pp. 61-80, 1998.

[5] *Gran Diccionari de la llengua catalana*, Barcelona: Enciclopèdia Catalana, 1998.

[6] Institut d'Estudis Catalans, *Diccionari de la llengua catalana*, Barcelona / Palma de Mallorca / València: Edicions 3 i 4, Edicions 62, Editorial Moll, Enciclopèdia Catalana, Publicacions de l'Abadia de Montserrat, 1995.

[7] D. Jones, *English Pronouncing Dictionary*, Cambridge: Cambridge University Press, 1997[15] [1917]. Edited by P. Roach & J. Hartman.

[8] J. Lacreu (Ed.), *Diccionari valencià de pronunciació*, Alzira: Bromera, 2001.

[9] A. Martinet and H. Walter, *Dictionnaire de la prononciation française dans son usage réel*, Paris: France-Expansion, 1973.

[10] J. Martínez de Sousa, *Diccionario internacional de siglas y acrónimos*, Madrid: Pirámide, 1984[2].

[11] J. Moran, M. Batlle and J. A. Rabella, *Topònims catalans. Etimologia i pronúncia*, Barcelona: Publicacions de l'Abadia de Montserrat, 2002.

[12] D. Paloma and A. Rico, *Diccionari de pronunciació en català*, Barcelona: Edicions 62, 2000.

[13] M. Riera-Eures and M. Sanjaume, *Diccionari d'onomatopeies i mots de creació expressiva*, Barcelona: Edicions 62, 2002.

[14] J. Solà, M-R Lloret, J. Mascaró, and M. Pérez-Saldanya (Eds.), *Gramàtica del català contemporani*, Barcelona: Empúries 2002 (3 vol.). (GCC)

[15] TERMCAT, *Nou Diccionari de neologismes*, Barcelona: Edicions 62, 2001.

[16] J. Veny and L. Pons, *Atles lingüístic del domini català*, *I*, Barcelona: Institut d'Estudis Catalans, 2001.

[17] J.C. Wells, *Longman Pronunciation Dictionary*, Harlow: Longman, 2000[2] [1990].