

The representation and storage of lexical units. An analysis of Catalan nouns and adjectives

Aur lie Guerrero

Fabio Montermini

(CLLE-ERSS, CNRS & Universit  de Toulouse)

The goal of this paper is to discuss the format under which lexical units are stored in memory, and the way in which inflectional data are organized within morphological competence. Its main focus is Catalan adjectival and nominal inflection. Catalan presents some peculiar characteristics. In particular, some word forms are apparently overtly marked for gender and / or number inflection, while others are unmarked (cf. *prim*_{MASC SG} vs. *prime*_{FEMPL} ‘thin’). Catalan data crucially raise the question of the status to be attributed to final segments: as we will see, the same segment may sometimes explicitly mark a feminine gender inflection (as in *prima* / *prime*, where *e* and *a* both correspond to [ ]), or do not mark a gender distinction (as in *c mode*_{MASC SG} / *c moda*_{FEM SG} ‘comfortable’ or *belga*_{MASC SG} / *belga*_{FEM SG} ‘Belgian’). Whether these segments should be considered as inflectional morphemes (in some cases or in all cases), theme vowels, or whether they should be considered as part of the stem, will also be discussed.

The analysis proposed is realized in a Word-and-Paradigm realizational model of morphology (Stump 2001). We will claim that a realizational view is particularly compatible with an emergent view of morphology (cf. Blevins 2006), in which the relations between the forms are not expressed as traditional rules, but rather as generalizations on the existing lexicon. In this respect, frequency plays a major role. In the final part of this paper a quantitative analysis of some subclasses of nouns and adjectives will be presented. Another issue that will be discussed in the paper is the division of labour between phonology and morphology. We consider that the effect of phonology is limited to phonotactic rules operating automatically, while other relations between inflected forms (whether they are phonologically motivated or not) are purely morphological or lexical. An advantage of this view is that it allows to distinguish between phonological rules which are part of the speakers’ linguistic competence, and relations between forms which have a clear historical explanation, but for which it is not clear whether they still play a role in synchrony.

The paper will be organized into three parts: the first part will be devoted to a general discussion of models of lexical storage and of the kinds of units that are included in a speaker’s lexicon; in the second part we will present a global analysis of Catalan nominal and adjectival inflection within a Word-and-Paradigm model; finally, in the third part we will present a case study, namely the case of nouns and adjectives presenting a masculine singular form ending in a stressed [a].

Present-day theories of the lexicon may be divided into two main groups: those considering that the lexicon only contains idiosyncratic information (so-called impoverished theories, mainly defended within formal models of morphology, such as generative ones), and those considering that the lexicon does not necessarily store information non redundantly, and that regularly formed forms may be stored along with irregular ones (cf. e.g. Bybee 1985; Blevins 2006). Clearly, a realizational and emergent model of morphology, as the one we defend, is more compatible with the second than with the first view. In a full-entry model units does not need to be decomposed into smaller parts, and the identification of such sublexical units as morphemes, roots, stems, etc. may be viewed more as an empirical problem than as a really theoretical issue. Note that some recent developments of such symbolic models as OT recog-

nize the fact that lexemes should be memorized as a whole in the lexicon (cf. Bermúdez Otero 2009).

Catalan nouns and adjectives apparently display overt affixes for gender and number inflection. The singular number is unmarked, while the plural number is systematically marked by [s]. The masculine gender may be unmarked, but there are at least some masculine forms ending in [u] (graphically *-o*); the feminine is generally marked by [ə] (graphically *-a*). According to traditional morphemic analyses (cf. Clua, 2002 for an overview), [s] is the plural suffix, [u] is the masculine suffix for some nouns and adjectives, and [ə] is the feminine suffix. Nevertheless, there are also some masculine nouns ending in [ə] (*cotxe* ‘car’) and some feminine nouns that are unmarked (*sal* ‘salt’); the same is true for adjectives (cf. *pobre* / *pobra* ‘poor’ vs. *gran* / *gran* ‘big’). For adjectives like POBRE a phonological explanation is often invoked: in these cases, the [ə] is an epenthetic vowel that resolves a sequence otherwise impossible word-finally. However, while it is the case for POBRE, other adjectives displaying a final sequence, which would be acceptable in final position also end in [ə]: *còmode* (‘comfortable’), *belga* (‘Belgian’). There are four types of adjectives in Catalan:

- (1)
- | | | | |
|----|--------|--------|--------|
| a. | prim | prima | ‘thin’ |
| b. | flonjo | flonja | ‘soft’ |
| c. | gran | gran | ‘big’ |
| d. | pobre | pobre | ‘poor’ |

We consider that the paradigm of Catalan adjectives includes (at least) two stems, one for the masculine forms and one for the feminine forms. For some adjectives, in fact, these forms display an alternation that cannot be accounted for simply by phonological rules (cf. the data in (2)):

- (2)
- | | | | | |
|-------------|---------------|----------------|------------------|---------|
| Masc Sg | Masc Pl | Fem Sg | Fem Pl | |
| clar ['kla] | clars ['klas] | clara ['klarə] | clares ['klarəs] | ‘clear’ |
| car ['kar] | cars ['kars] | cara ['karə] | cares ['karəs] | ‘dear’ |

Under our analysis, the phoneme [ə] is part of the feminine stem and, in default cases, at least for adjectives like PRIM and CAR, the masculine (A) and the feminine (B) stems are linked by a function of the type $X - X\grave{a}$. For adjectives like CLAR, on the other hand, both stems have to be stored in the lexical representation of the lexeme. Of course, this function only holds for one class of adjectives. For the other types (represented in (1) by FLONJO and by GRAN and POBRE, respectively), other default functions are active, namely $Xu - X\grave{a}$ for the first type, and $X - X$ (identity) for the second. Thus, the final vowels are disconnected from the expression of gender, a desirable effect, since, as we have seen, they may or may not appear in forms realizing a specific gender.

To sum up, what we have in Catalan is an inflectional scheme common to all adjectives (3a), three inflectional classes, each specifying a function linking the two stems (3b), and a series of rules for the construction of actual inflected forms (3c):

- (3)
- a.
- | | | |
|----|------|-----|
| | Masc | Fem |
| Sg | A | B |
| Pl | A | B |

b.

Class I: X – Xə (Stem B = Stem A+ə)

Class II: Xu – Xə (Stem B = Stem A–u+ə)

Class III: X – X (Stem B = Stem A)

c.

Masc Sg = Stem A

Masc Pl = Stem A+s

Fem Sg = Stem B

Fem Pl = Stem B+s

What is stored in the lexical representation of Catalan adjectives and nouns is then a fully specified word form and a pattern of relations between the stems. In the default case, one form is sufficient, and the relations in (3) allow to reconstruct the whole paradigm. The final vowels have no inflectional status (they do not directly mark gender), and they are not theme vowels, since they do not indicate inflectional class membership, which is ambiguous (a masculine in consonant may belong either to Class I or to Class III, and a feminine in [ə] to any of the three classes). Class membership is in fact indicated by the relation pattern attached to the lexical representation.

In the final part of this paper we present a case study, namely the inflection of adjectives (and nouns variable in gender) whose masculine ends in a stressed [a]. The data in (4) illustrate the different types of alternation that can be found in Catalan:

| | | | | | | |
|-----|--------------|----------------|---------|------------|--------------|---------|
| (4) | clar ['kla] | clara ['klarə] | ‘clear’ | car ['kar] | cara ['karə] | ‘dear’ |
| | pla ['pla] | plana ['planə] | ‘plane’ | nan ['nan] | nana ['nanə] | ‘dwarf’ |
| | lilà [li'la] | lilà [li'la] | ‘lilac’ | | | |

Contrary to what is sometimes claimed (Mascaró 1986; Bibiloni 2002), as illustrated in (4), the relation is not phonological, at least in synchrony. We have realized an extensive research of adjectives whose masculine form ends in a stressed [a] in the *Diccionari de l'Institut d'Estudis Catalans 2* (henceforth *DIEC 2*, <http://dlc.iec.cat/>). [a]/ [anə] is indeed the most frequent pattern: in our corpus, 424 lexemes display this alternation. 381 of them (about 90%) are constructed words including a relational suffix -à:

| | | |
|-----|---|-----------------|
| (5) | from places names: Catalunya < català / catalana | ‘Catalan’ |
| | proper nouns: Shakespeare < shakespeareià / shakespeareiana | ‘Shakespearian’ |
| | common nouns: crani < cranià / craniana | ‘cranial’ |

However, the other patterns illustrated in (4), although they are less frequent, are far from marginal. Note that the sequence [an] in word final position is not phonologically excluded in Catalan, since several other noun and verb forms display it:

| | | |
|-----|-------------------------------|------------------|
| (6) | blan ['blan] / blana ['blanə] | ‘soft’ |
| | nan ['nan] / nana ['nanə] | ‘dwarf’ |
| | comprèn [kum'pren] | ‘he understands’ |

Thus, when a speaker of Catalan encounters a novel and / or unknown word ending in one of the sequences above, he / she must rely on the existing lexicon, rather on an abstract rule, to inflect it. The statistic data we present suggest that frequency plays a role in the strategies chosen by speakers to reconstruct the paradigm of unknown lexemes, a conclusion that mili-

tates in favour of a probabilistic approach to morphological organisation (cf. Baayen 2003; Albright 2009).

References

- Albright A. (2009), "Modelling analogy as probabilistic grammar" in Blevins J.P., Blevins J. (eds) *Analogy in Grammar*: p. 185-213. Oxford: Oxford University Press.
- Baayen R.H. (2003), "Probabilistic approaches to morphology". In Bod R., Hay J., Jannedy S. (eds), *Probabilistic Linguistics*: p. 229-87. Cambridge, Ma: MIT Press.
- Bermúdez Otero R. (2009), "The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features". Manuscript. University of Manchester.
- Bibiloni G., (2002), "Elisió de -N i -R, distribució de les ròtiques i altres fenòmens consonàntics en el mot". In Solà J., Lloret M.R., Mascaró J., Pérez Saldanya M. (eds.), *Gramàtica del català contemporani*. Barcelona, Editorial Empúries: 273-285.
- Blevins J.P. (2006), "Word-based morphology", *Journal of Linguistics* 42: p. 531-573.
- Bybee J.L. (1985), *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.
- Clua E. (2002), "Gènere i nombre en els noms i en els adjectius". In Solà J., Lloret M.R., Mascaró J., Pérez Saldanya M. (eds.), *Gramàtica del català contemporani*. Barcelona, Editorial Empúries: 485-534.
- Mascaró J., (1986), *Morfologia*. Barcelona, Enciclopèdia catalana.
- Stump G.T. (2001), *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.