

## Abstract

*In this paper, we will present an overview of compounding in Catalan, covering the founding theoretical-descriptive studies of the 20th century to the most recent scholarship. Specifically, it analyses the boundaries of compounding, considering the limits which make it possible to distinguish compounding from derivation and free syntagmas. We will also consider the various types of compounding documented in Catalan, as well as the syntactic and semantic relationships between the elements that make up a compound. Finally, we will analyse newly-created compounds that shed light on the productivity of this process of word formation.*

## 1. Introduction

Compounding is one of the morphological processes used by Catalan to create words. Unlike derivative processes involving one radical to which an affix (prefix or suffix) is added, compounding by definition two radicals. In Catalan, only major lexical categories (N, V, A and Adv) are likely to partake in compounding processes.<sup>1</sup>

---

1. Nevertheless, we find some compounds following the [Prep+N]<sub>N</sub> pattern, such as *sense papers* ‘illegal immigrants’ or *sense sostre* ‘homeless’, both newly introduced into the Romance languages (Roca 2004). These compound neologisms are characterised by being strongly lexicalized and by always referring to human beings.

Here, we do not take into consideration compounds that have resulted in the formation of invariable categories (adverbs, prepositions and conjunctions, basically) and which are the result of the lexicalization of a syntactic combination. Examples of this include *gairebé* [Adv+Adv]<sub>Adv</sub> ‘almost’, *enlloc* [Prep+N]<sub>Adv</sub> ‘nowhere’ and *tothora* [Adj+N]<sub>Adv</sub> ‘always’. These forms are completely unproductive on a synchronic level (see Pérez Saldanya et al. 2004).

Catalan compounds belong mostly to nominal categories, and especially to that of nouns:  $[N+A]_N$ ,  $[N+prep+N]_N$ ,  $[A+N]_N$ ,  $[V+N]_N$ ,  $[V+V]_N$ , despite the fact that some verbal and adjectival compounds do exist. Nevertheless, compounds by no means form a homogenous group: For example, from a phonologic point of view, compounds tend to maintain the respective stresses of the two radicals, as seen in the case of *portaequipatge* ‘roof rack’, where, in dialects that have vowel reduction, the secondary stress of [ɔ] is maintained in *porta* rather than being reduced to [u]. However, there are some compounds that only maintain the stress on the second radical, as seen in the case of *sanglaçar* (Pérez Saldanya et al. 2004). Morphologically, we also find variation: in some cases, both elements that form a compound are pluralized, as with *dones fatals* ‘femmes fatales’. In other cases, however, only the first element is pluralised, as seen with *homes del temps* ‘weathermen’. Finally, there are still other cases in which only the second element is pluralized, as demonstrated by *filferros* ‘wire’. From a semantic point of view, the compounding of two radicals results in a new concept, one that cannot necessarily be reduced to the compositional sum of the meanings of the two radicals that make it up. This is demonstrated by *avió de reacció* (lit. plane of jet) ‘jet plane’, which is a type of plane. *Pit-roig* (lit. chest + red) ‘robin’, on the other hand, designates a species of bird rather than a type of chest. From a syntactic point of view, note how a syntactic relation is formed between the two constituents. For example, in the compounds  $[N+A]_N$  *pèl-roig* ‘red head’ and  $[A+N]_N$  *malnom* ‘nickname’, among others, the nouns are the modifiers of the adjectives *roig* ‘red’ and *mal* ‘bad’. Meanwhile, in  $[V+N]_N$  compounds such as *penja-robes* ‘clothes hanger’, the nominal element on the right corresponds to the direct complement of the verbal radical *penjar* ‘to hang’.

We will now define the boundaries of compounding in order to delineate the way it differs from derivation and free syntagmas (Section 2). Then, we will examine the types of compounds that are documented in Catalan (Section 3), the syntactic and semantic relations that are established between the elements of the compound (Section 4), and the inflection of compounds (Section 5). Finally, we will present data regarding new words that will shed light on the productivity of this process of word formation (Section 6).

## 2. The boundaries of compounding

Compounds such as *gratacel* ‘skyscraper’ are often referred to as “lexical compounds” or “orthographic compounds” since their components are joined graphically in such a way that they share characteristics with other, non-compound words, despite the fact that they are made up of more than one lexeme. For example, they have their own unified, constant meaning with a single

reference; the elements that form these words lose their lexical independence: they must follow a fixed order (*aiguaneu* ‘sleet’ – \**neuaignua*); it is not possible to coordinate one of the constituents with another element (*gratacel* ‘skyscraper’ – \**gratacelinívol* lit. scrape sky and cloud); it is not possible to introduce other elements between the constituents of the compound (*altaveu* ‘loudspeaker’ \**altaclaraveu* lit. loud clear voice); and, finally, it is not possible to extract one of the constituents from the compound and refer to it with the syntagma (*coragre* ‘heartburn’ – \*l’agre del cor).

However, we must also be aware of the fact that we can’t solely take the graphic form into consideration because, as we will show, there are cases that are situated in a fuzzy, difficult-to-define space: in fact, one of the inherent difficulties associated with the study of compounding processes lies in establishing their boundaries regarding derivation on the one hand, and the respect of the independent clauses, on the other.

### 2.1. Compounding vs derivation

As we mentioned before, boundaries between different word formation processes are never clear. Moreover, with regard to the group of words formed with *contra-*, *sobre-*, *entre-*, etc., authors hold back in indecision when it comes to attributing them to compounding or derivation. In Catalan, this problem already exists in traditional grammars: Pompeu Fabra, systemiser of Catalan grammar and the unavoidable reference when it comes to any study related to Catalan word formation, already encountered a problem when trying to deal with prefix affixation in his (1912) grammar.<sup>2</sup> By (1875), Darmesteter had already set the concept of parasynthesis, defined as the simultaneous adjunction of a prefix and a suffix to an adjectival or noun base, therefore making it possible to explain verbs such as *aclarir* ‘to lighten’, *abraçar* ‘to hug’, *ennegrir* ‘to darken’ or *empaperar* ‘to wallpaper’ as having a ternary structure [prefix + base + suffix]. This concept was rapidly adopted by many of his contemporary scholars such as Menéndez Pidal (1904). But Fabra, regardless of the fact that he probably was aware of the existence of both works,<sup>3</sup> distanced himself from them in his (1912) grammar (§ 186), thus avoiding theoretical commitment. In section § 193, dedicated to *compound words*, we see the prefixes *ante-*, *avant-*, *ben-*, *bes-*, *contra-*, *entre-*, *in-*, *mal-*, *mig-*, *re-*, *sobre-*, *sub-* and *vice-*. In the 1956 grammar, he confirms this classification and, in a chapter dedicated to

2. In Bernal (2007) Fabra’s morphological criteria on word formation are stated in greater detail.

3. As stated in Solà (1982: 79), Fabra’s library was destroyed during the 1936–1939 civil war; consequently, no categorical assertion of this can be made.

compounding, includes compounds the first element of which is an adverb or a preposition (such as *contraprojecte* ‘counterproject’).

Apparently, there is a contradiction between what he says in § 130 and the fact that he includes words with a prefix in the chapter on compounds.<sup>4</sup> The truth is, nonetheless, that Fabra’s indecision is still valid in many modern-day theoretical-descriptive studies, as seen in Bayà (1997); likewise, Moll (1968) and Marvà (1968) deal with it within compounding, as well as Gràcia (2002) and Turon (2004) (who also highlight its bordering situation), and Padrosa (2010); on the other hand, Ruaix (1986) proposes an ambiguous title for his section dedicated to prefixes: “compounds (or derivatives) using prefixes”. Badia (1962), however, tackles prefixes in the paragraph dedicated to derivation within the chapter on word formation, thus being able to deal with prefixes and suffixes in a parallel and homogeneous manner (§ 344–349). Nevertheless, we do not find any important theoretical input until the publications of Mascará (1986) and Cabré and Rigau (1986). In all these argumentations we see a recurrent aspect of morphological base: the fact that “particles” intervening in prefixation are usually adverbs or prepositions is highlighted, yet with an added characteristic, which is that these particles modify the meaning of the constituent they are added to (Ruaix 1986).

In conclusion, when it comes to dealing with this type of words, two different trends arise. On the one hand, some identify the first element of the word with its corresponding adverbs and prepositions and see these units as compounding units, basing their argument on the fact that compounding is the combination of two or more autonomous units of the language. On the other hand, there are those who place these units under the category of derivation, claiming that these prefixes are different from the prepositions or adverbs that formally correspond to them: they are units that have oftentimes lost their phonological properties (loss of stress or vocalic reduction), as we can see in words like *entresòl* ‘mezzanine’, *entretenir* ‘to entertain’, or *contradir* ‘to contradict’, *contrarestar* ‘to counteract’, in which the pronunciation oscillates between

---

4. Concretely, in the compounding chapter, he explains prefixes as follows:

§ 153: non-stressed prefixes *a, en, in, es, des, dis, bes, re, con, per, pre, trans/tras*;

§ 154: stressed prefixes *contra, entre* and *inter, intra, sobre* and *super, super, supra, sota* and *sub, sots, infra, fora, extra, avant* and *ante, circum, ultra, semi, vice, anti, arxi, pseudo, co, ex, post, pre, no, menys, prop, quasi, ben* and *mal*.

§ 155: borrowed prefixes from Latin: *ab/abs, ad, ante, circum, con, de, dis|di, ex|e, in, inter, ob, per, pre, pro, re|red, sub, super* and *trans/tra*.

§ 156: borrowed prefixes from Greek: *anti, arxi, pseudo, a|an, neo, micro, proto, hiper, hipo, pan, deca, hecto, mono, poli, di, tri* and *tetra*; Latinate prefixes *uni, multi, bi, tri, quadri, extra, ultra, semi, infra*, etc.

The rest of the chapter’s sections (from § 157 to § 164) are dedicated to compounds that we can consider as prototypical, formed by nouns, adjectives and verbs.

maintaining the secondary stress [e] / [o] and vocalic reduction [ə] / [u])<sup>5</sup> that can also participate in the base re-categorisation phenomena, typical of derivation processes.

In this paper, we have adopted this second perspective and we consider that these cases are examples of derivation processes with a prefix. Therefore, we do not include them in the description because, alongside what has been said before and although these prepositions can exist as independent words and have their own stress introducing the secondary stress, when derivation occurs, this stressed or non-stressed character does in fact vary in some cases, depending on the dialect or idiolect, and moreover, a same prefix can appear as non-stressed in certain constructions and as stressed in others (Institut d'Estudis Catalans 2002: § 2.7), a phenomenon that never occurs in compound words.

## 2.2. *Compounding vs phrases*

The other vague boundary is the one separating syntagmatic compounding from independent clauses.<sup>6</sup> When dealing with it, we should take into account the traditional distinction between common compounds and syntagmatic or synapctic ones (Benveniste 1974):<sup>7</sup> the first ones are those that move two major lexical categories, while the second ones are the result of the lexicalisation of a syntactic sequence, such as *ull de poll* 'callus' (substantive + prepositional clause) *peix blau* 'blue fish' (substantive + adjective), *estira i arronsa* 'give and take' (phrase). All these cases are generated as independent clauses in the syntax (Cabré 1994; Lorente 2002) and some are even incorporated into the lexicon of a given language in response to denotative or connotative needs, and are given different names by different authors (syntagmatic words, lexicalized clauses, bound phrases, or synapctic or syntagmatic compounds). In this type of compound, two phenomena occur at once: syntagmation, as a formal resource in the language for the construction of clauses, and lexicalization, which formally establishes them and provides them with a specific meaning. Phraseology stands between this type of compound and the completely free combinations, adopting different names (phraseologisms, phraseological units, set or fixed expressions, idioms, set phrases) and includes elements such as *aixecar*

---

5. Due to limited space, we cannot go into analysing the stress issue in compounded words and we refer the reader to the works of Mascaró (1986, 2002) and Wheeler (1979), among others.

6. Fabra (1956: § 160) had already stated that there are certain groups of words such as *blat de moro* 'sweet corn' (lit. wheat of Moor) which, despite being written without joining their constituent elements, are nevertheless bona fide compound words.

7. Diversity in denominations and meanings when approaching compounds is obvious. See Observatori de Neologia (2004: 133–134) for a clear and short summary of these.

*la camisa* ‘to cheat’ or *fer un petó* ‘to kiss’.<sup>8</sup> Wotjak (1998) or Ruiz Gurillo (1997) argue for an autonomous space for phraseology, located at the boundary between lexicon and syntax, in which lexical and phraseological units are distributed following a gradual and theoretical continuum, ranging from the most lexicalized unit to the least lexicalized one.

When recognizing if a certain clause is independent or not, the main difficulty resides in the recognition and interpretation of its lexical or phraseological character. As Lorente (2002) states, if a speaker (or a learner) of a language fails to incorporate this type of unit into his vocabulary, he will neither be able to interpret them correctly nor to distinguish them from the independent clauses in the speech or from the so-called collocations or lexical solidarities (lexical combinations that appear recurrently in speech but which are not subject to any lexicalization process that implies a formal fixing or allocation of a specific meaning, as if to incorporate them jointly in the lexicon of the language; nor are they assigned any fixing and idiomatic properties that will convert them into phraseological units). These are set characteristics, as they cannot in fact receive any modifiers or complements independently from the whole: *carn d’olla* ‘Catalan stew made of meat, vegetables and legumes’ (lit. meat + of + pot) does not admit a modification such as \**carn d’[olla de pressió]* (lit. meat + of + pressure cooker), whereas it is possible to insert them in independent clauses, even if they have a certain degree of fixing, such as *màquina de cosir* ‘sewing machine’ vs. *màquina de [cosir i brodar]* ‘sewing and embroidery machine’ (Varela 2005).<sup>9</sup>

Compounds [N+prep+N]<sub>N</sub> or [N+A]<sub>N</sub>, among others, are originally generated in the syntax. However, for fixing purposes as well as for referentiality (besides other aspects such as the idiomatic nature, the exocentric meaning, the metaphoric motivation or the cultural elements they refer to), these are nominal syntagmatic structures that would be then situated in the furthestmost pole of lexicalization. Any Catalan speaker can undoubtedly distinguish compounds like *pi blanc* ‘Aleppo pine’, *caixa negra* ‘black box, flight recorder’, or *calaix de sastre* ‘hotchpotch’ from the independent clauses *pi vell* ‘old pine tree’, *caixa vermella* ‘red box’ or *calaix d’armari* ‘closet drawer’.<sup>10</sup>

8. See Lorente (2002) for a more detailed list.

9. Contreras and Suñer (2004) add more evidence: phonetic and phonological, morphological, syntactic, (specified in the fixing of lexical elements, the non-commutable nature of constituents, the impossibility to extract any of the constituents, and a fixed syntactic order), semantic, lexical, and pragmatic.

10. We should highlight the fact that [N+prep+N]<sub>N</sub> structure is highly present in specialised speech, due to the great need for designation (Cabrè 1992). See Lorente (2002: § 8.3.2.2) for a semantic classification of specialised syntagmatic compounds that makes a distinction

Between the lexical and the syntagmatic origin, there is a type of concrete compound: those made of two constituents belonging to the same category ([N+N]: *vagó restaurant* ‘dining car’ or *blau cel* ‘sky blue’, [A+A]: *sordmut* ‘deaf and dumb’), and showing some characteristics that set them far away from prototypical compounding. Nonetheless, Gràcia (2002) believes that whether these structures can be generated by syntax, and therefore considered as compounds of syntagmatic origin, is debatable:

- (1) *gos pigall* ‘guide dog’, *paper moneda* ‘paper money’, *cotxe bomba* ‘car bomb’, *cèl·lula mare* ‘stem cell’, *paraula clau* ‘keyword’, *rosa gerani* ‘pink geranium’, *groc llimona* ‘lemon yellow’, *verd molsa* ‘moss green’
- (2) *aiguaneu* ‘sleet’, *nord-oest* ‘northwest’

The first group (1), much more numerous and productive than the second one, belongs to the appositive type because the second constituent behaves like an apposition to the first; on the contrary, group (2) compounds are coordinated ones. This is because the semantic value of the whole is equal to the sum of constituents (although in the case of the cardinal points this is not totally accurate, because *sud-est* ‘southeast’ does not designate a point that sums up south and east, but a concrete point located between south and east).

One characteristic of appositive compounds is that they designate an entity belonging to the ones denoted in the first constituent (its head) that shows some characteristics belonging to the entities denoted by the second constituent: *gos llop* ‘German shepherd’ is a dog with physical characteristics belonging to wolves and *home orquestra* ‘one-man band’ is a man who plays simultaneously many instruments that he carries with him (like an orchestra), etc.

On the contrary, A+A compounds should be considered as specific compounds, considering that in Catalan syntax does not generate structures with two adjectives positioned at the same level:

- (3) *sordmut* ‘deaf and dumb’, *blaugrana* ‘blue and deep red’

Examples we see in (3) are clearly coordinated and are equivalent to constructions with two adjectives united by the coordinating conjunction *i* ‘and’, which establishes a mere adding relationship, although sometimes we also find an idea of direction, such as in *diccionari català-castellà* ‘Catalan-Spanish dictionary’, *frontera russotxetxena* ‘Russian-Chechen border’.

---

between exocentric meanings built on complete metaphors (*rellotge de pastor*, lit. watch of shepherd ‘mushroom’) or partial ones (*crossa de l’aorta*, lit. crutch of aorta ‘arch of aorta’), with a metaphoric head and a literal complement, or *balanci de coll de cigne* ‘swan-neck bearer’, with literal head and metaphoric complement), and endocentric literal meanings (*gat d’Angora* ‘Angora cat’).

### 3. Types of compounds

Here we will discuss compounding models in Catalan, which we will group by the type of existing relationships between their different constituents, subordinate compounds (3.1), appositive/attributive compounds (3.2), or coordinated compounds (3.3). We will then comment on the most relevant aspects for each group. Furthermore, we will add a final paragraph (3.4) with the patterns we considered clearly unproductive. In all cases, we will indicate the head,<sup>11</sup> stating if they are endocentric with a left head (L), a right head (R) or both constituents simultaneously (B), or exocentric (∅). We will then give a prototypical example for each.

#### 3.1. Subordinate compounds

Table 1

| Structure              | Head | Example                 | Origin    | Translation         |
|------------------------|------|-------------------------|-----------|---------------------|
| [N+N] <sub>N</sub>     | L    | <i>vagó restaurant</i>  |           | 'restaurant coach'  |
| [N+N] <sub>N</sub>     | ∅    | <i>mareperla</i>        |           | 'mother of pearl'   |
| [V+N] <sub>N</sub>     | ∅    | <i>pica-soques</i>      |           | 'green woodpecker'  |
| Neoclassical compounds |      |                         |           |                     |
| [sN+sN] <sub>N</sub>   | R    | <i>hemeroteca</i>       | Gr + Gr   | 'newspaper library' |
| [sN+sN] <sub>N</sub>   | R    | <i>suïcidi</i>          | Lat + Lat | 'suicide'           |
| [sN+sN] <sub>N</sub>   | R    | <i>ludoteca</i>         | Lat + Gr  | 'toy library'       |
| [sN+sN] <sub>N</sub>   | R    | <i>biocida</i>          | Gr + Lat  | 'biocide'           |
| [sN+N] <sub>N</sub>    | R    | <i>agroturisme</i>      | Gr + Cat  | 'rural tourism'     |
| [sN+N] <sub>N</sub>    | R    | <i>tricampió</i>        | Lat + Cat | 'three-time winner' |
| [N+sN] <sub>N</sub>    | R    | <i>verbívor</i>         | Cat + Lat | 'word eater'        |
| [N+sN] <sub>N</sub>    | R    | <i>devedeteca</i>       | Cat + Gr  | 'DVD library'       |
| [sN+sA] <sub>A</sub>   | R    | <i>antropomorf</i>      | Gr + Gr   | 'anthropomorphic'   |
| [sN+sA] <sub>A</sub>   | R    | <i>agricola</i>         | Lat + Lat | 'farming'           |
| [sN+sA] <sub>A</sub>   | R    | <i>hidròfug</i>         | Gr + Lat  | 'water-repellent'   |
| Phrasal compounds      |      |                         |           |                     |
| [N+[P+N] <sub>N</sub>  | L    | <i>mal de cap</i>       |           | 'headache'          |
| [N+[P+N] <sub>N</sub>  | ∅    | <i>ull de poll</i>      |           | 'corn'              |
| [N+[P+A] <sub>N</sub>  | L    | <i>diamant en brut</i>  |           | 'uncut diamond'     |
| [N+[P+V] <sub>N</sub>  | L    | <i>camisa de dormir</i> |           | 'nightdress'        |
| [A+[P+N] <sub>N</sub>  | L    | <i>curt de vista</i>    |           | 'shortsighted'      |

11. The syntactic head in Catalan tends to come first in the compound. In neoclassical compounds, the head occurs on the right hand side, both in Latinate (*fratricidi* 'fratricide') and Greek (*geologia* 'geology') bound forms. We can consider coordinate compounds to have two semantic heads, but the constituent in the canonical head location (left hand side of the



Subordinate compounds belong to the largest and, at the same time, the most productive group in Catalan, as it includes the main nominal ([N+A]<sub>N</sub>, [A+N]<sub>N</sub>, [V+N]<sub>N</sub>), as well as learned and syntagmatic compounds.

As for Greek- and Latin-derived compounds, it is worth pointing out that many entered the Romance languages already as part of a relatively formal register. Nevertheless, it is also true that they possess enormous potential. *A priori*, it is a word-forming resource we frequently see in specialised fields, yet it is by no means limited to technical vocabulary: it is also true that Romance languages (and, more generally, Indo-European ones) incorporate Latinate and Greek constituents, as well as their combination norms, in general word construction. It is basically recommended to give preference to homogeneous learned constructions over heterogeneous ones, and within learned constructions, compounds using constituents belonging to a same language of origin are recommended (Termcat 1990):

- (4) a. Greek + Greek: *hemeroteca* ‘periodicals library’, *antropomorf* ‘anthropomorphic’  
 b. Latin + Latin: *suïcidi* ‘suicide’, *agricola* ‘agricultural’  
 c. Latin + Greek: *ludoteca* ‘toy library’, *canòdrom* ‘greyhound stadium’  
 d. Greek + Latin: *biocida* ‘biocide’, *hidròfug* ‘hydrofugous’

However, not only word construction using learned constituents goes beyond the established boundaries of specialised language (to which it is allegedly limited). Rather, some compounds disregard the initial restrictions on combining Latinate and Greek bound forms separately and create hybrid Greek-Latinate compounds (sometimes even mixing them with Catalan elements), a combination that is especially productive when it comes to neologisms (Cabrè 1994; Gràcia 2002; Rull 2009):

- (5) a. Greek + Catalan: *agroturisme* ‘agritourism’  
 b. Latin + Catalan: *tricampió* ‘triple champion’  
 c. Catalan + Latin: *verbívor* ‘word eater’, *lliberticida* ‘liberty killer’  
 d. Catalan + Greek: *gandulitis* ‘colloq. lack of will to work’, *tintinòleg* ‘specialist in (comic character) Tintin’, *aromateràpia* ‘aromatherapy’

---

compound) is the most prominent from a syntactic standpoint (see 3.3). Moreover, verbal compounds with an adjective or an adverb as their first constituent also have their head in the final location: *carcomprar* ‘to pay too much for’, *malacostumar* ‘to spoil’ (see 3.4). Also to be included in this group are compounds of nominal structure [A+N]<sub>N</sub>, such as *pura sang* ‘thoroughbred’ (see 3.2).

In the cases where final learned constituents are used, it is interesting to highlight that, in their design rules, changes that bring them closer to suffixes occur. These changes are, above all, pragmatic changes, as they usually bear associated affective values that can be either positive or negative.<sup>12</sup> Along these same lines, López et al. (2002: 51) analyse *pseudo-* and *-oide* as affixes, although they define them as “constituents of learned origin used in specialised as well as in general speech”, and consider them as an example of a learned constituent’s shift to language affix.<sup>13</sup>

These changes can also be seen in initial elements in words like *aerobotiga* ‘airport shop’, *aerobús* ‘shuttle leaving from or leading to an airport’, *telenotícies* ‘television news programme’, *telemàrqueting* ‘telemarketing’, *fotoperiodisme* ‘photojournalism’, etc., in which the apparently learned constituents have lost their original features. This has brought some authors, like Vallès (2007) or García Palacios (2000–2001), to treat them as prefixoids, as they consider them undoubtedly characterised by being initial forms that are formally identical to the complex word’s fragment from which they originate and that, beyond the features they commonly share with prefixes or with learned prefixed forms (or debating whether they show more features belonging to one group or the other) speakers identify them as units. Furthermore, these units are not generally autonomous and speakers associate them with a specific meaning while also conveniently combining them with other words belonging to the language, the same way as prefixes. Now, as it was analysed in Bernal (2010), in the case of words formed with *moto* ‘motorcycle’, *tele* ‘TV’, or *foto* ‘picture’, we could also consider that the result is not learned compounds but rather non-learned compounds. This, then, is the same as a clipping: for example, note the cases

12. For a detailed analysis of transgression of the rule’s literality, (in the sense of changes in sub-categorisation) suffered by *-itis* in Catalan, see Cabré et al. (2002): the stylistic load of new non-specialised formed words, frequently recreational or ironic, is clear. A conclusion reached by Lüdeling and Evert (2005) when analysing neologisms with *-itis* in German.

13. They don’t go in further details nor do they argue more this statement, as the study focuses mainly on the analysis of the attribution of a pragmatic pejorative value to these elements. This tendency to “afixability” was observed by Bernal (1997), where the status of *-ificar*, that is treated differently in Catalan, as well as in the other languages that possess the corresponding morpheme, was discussed: likewise, for Catalan, Cabré (1994) lists *-ficar* as a Latinate compounding form and *-ificar* as a verbalising suffix by itself; Bruguera (1996) includes *-ificar* as a “Latin constituent of compounded words”; it appears in the *Taula de formants cultes* ‘Table of learned constituents’ (1997); López del Castillo (2002), on the contrary, classifies it as a learned suffix; on the other hand, Cottez (1989) includes *-fier* in his *Dictionnaire des structures du vocabulaire savant*; and in Spanish, *-ficar* appears in the DRAE (2001) as a “compounding element”. Despite this apparent unanimity when it comes to consider it as a learned constituent, *-ificar* is already a suffix, although it shows very strict combinatory restrictions: it can only combine with radicals of Latinate origin, that cannot be complex (likewise, *\*escolarificar* is not possible).

of *motocicleta* ‘motorcycle’, *televisió* ‘television’ and *fotografia* ‘photography’, which have syntactic autonomy within the language and show clear differences with forms like *dextro-* ‘right’, *histero-* ‘uterus’ or *hipso-* ‘height’.

In any case, most learned compounds are endocentric, with the head to the right, and in which the first constituent is a complement or a modifier: *nefòmetre* (cloud + measurer) ‘nephelometer’ is the instrument that measures cloudiness, *centrifug* (centre + to move away) ‘centrifugal’ means moving away from the centre, etc.

Finally, we need to highlight the fact that learned compounds need a linking vowel, or even two: *o*, if the last element is of Greek origin, or *i*, when it is Latinate. The initial forms in contact with the second bound form do not undergo too many alterations, although there may be cases of consonantal assimilations, characteristic of informal language. Some prefixes show omitted elements before a vowel, eg. *licantropia* ‘licanthropy’, *histerectomia* ‘hysterectomy’, etc.

Initial forms preceding final Latinate forms that begin with a consonant include an *i* as a linking vowel: *caducifoli* ‘deciduous’, *herbicida* ‘herbicide’. The case of final *-metria* and *-metre* is an exceptional one. They take first constituents ending with an *-i* or an *-o*: *sacarímetre* ‘sacharometer’, but *pluviòmetre* ‘pluviometer’; *planimetria* ‘planimetry’ but *craniometria* ‘craniometry’.

Regarding syntagmatic compounds, the most productive type of phrasal compounds is, without a shadow of a doubt, the one with the [N+prep+N]] structure: *cap de turc* ‘scapegoat’, *roda de premsa* ‘press conference’, *pluja d’idees* ‘brainstorming’. As we have already said, these compounds do not show differences with independent clauses and, from a semantic viewpoint, the complement is there to specify the compound’s meaning. We should always bear in mind, nonetheless, that metaphor is the origin of a great number of these compounds, and it seems that it is precisely the exocentric and metaphorical meaning that is the determining criterion making it possible to distinguish them from independent clauses (*cap de bestiar* ‘head of cattle’, *roda de canvi* ‘spare wheel’, *pluja de fang* ‘mud rain’), as well as semantically literal and compositional clauses.

### 3.2. Appositive/attributive compounds

In nominal appositive/attributive compounds, the second noun denotes some of the features embodied in the expressed entity: thus, *cotxe escombra* ‘broom wagon’ only incarnates the fact that it *picks up* runners and cycles abandoning the race, or the *peix martell*’s ‘hammerhead’ only feature is that it has a very flattened and two-lobed widened head on the sides, which reminds us of the *shape* of a hammer.

Table 2

| Structure          | Head | Example             | Translation      |
|--------------------|------|---------------------|------------------|
| [N+N] <sub>N</sub> | L    | <i>gos pigall</i>   | ‘seeing-eye dog’ |
| [N+N] <sub>N</sub> | L    | <i>verd poma</i>    | ‘apple green’    |
| [N+A] <sub>N</sub> | L    | <i>porc senglar</i> | ‘wild boar’      |
| [N+A] <sub>N</sub> | ∅    | <i>gat vell</i>     | ‘old fox’        |
| [N+A] <sub>A</sub> | R    | <i>camacurt</i>     | ‘short-legged’   |
| [A+N] <sub>N</sub> | R    | <i>malnom</i>       | ‘nickname’       |
| [A+N] <sub>N</sub> | ∅    | <i>altaveu</i>      | ‘loudspeaker’    |

There is a particular subgroup pertaining to these cases in which the apposed noun highlights or intensifies one of the properties belonging to the first noun through recurrent nouns, such as *estrella* ‘star’, *clau* ‘key’, *model* ‘model’, *membre* ‘member’, *marc* ‘framework’, etc.:

- (6) {*estat* / *país*} *membre* ‘member state’  
 {*jutge* / *client* / *pla* / *oferta* / *jugador*} *estrella* ‘star {judge / client / plan / supply / player}’  
 {*problema* / *paraula* / *dona* / *decisió*} *clau* ‘key {problem / word / woman / decision}’  
 {*visita* / *viatge*} *llampec* ‘flying {visit / trip}’

Compounds of this type, especially those belonging to case (6), make speakers feel uncertain about how to proceed when it comes to using them in plural. This issue will be analysed in Section 5.1.

In cases involving an adjective, we must underline the fact that, whether this one appears as a first or as a second element, we are mainly dealing with exocentric compounds with a very high degree of lexicalization that makes their meaning unpredictable: *baixamar* ‘low tide’ (lit. low sea), *or negre* ‘petroleum’ (lit. black gold), *càmera alta* ‘senate’ (lit. high chamber), etc.

### 3.3. Coordinate compounds

Unlike what we saw in 3.2, this group’s nominal, endocentric compounds coordinate in such a way that the juxtaposed noun denotes the entity itself: a *magistrat jutge* is a *magistrat* ‘magistrate’ with all the characteristics of a *jutge* ‘judge’, and a *sala menjador* ‘dining room’ is a *sala* ‘room’ with all the characteristics identifying it as a *menjador* ‘dining room’.

Within this group, we should highlight the case of [A+A]<sub>A</sub> compounds, like *fisicoquímic* ‘physiochemical’ or *francocanadenc* ‘French Canadian’, which

Table 3

| Structure           | Head | Example                | Translation        |
|---------------------|------|------------------------|--------------------|
| [N+N] <sub>N</sub>  | B    | <i>magistrat jutge</i> | 'magistrate'       |
| [N+N] <sub>N</sub>  | ∅    | <i>bleda-rave</i>      | 'beet'             |
| [N+N] <sub>N</sub>  | ∅    | <i>glac-desglac</i>    | 'frost-defrost'    |
| [A+A] <sub>A</sub>  | B    | <i>agredolç</i>        | 'sweet and sour'   |
| [A+A] <sub>A</sub>  | B    | <i>fisicoquímic</i>    | 'physico-chemical' |
| [sA+A] <sub>A</sub> | R    | <i>sinojaponès</i>     | 'chinese japanese' |
| [V+V] <sub>N</sub>  | ∅    | <i>alçaprem</i>        | 'turn'             |

use a linking vowel that gives them a false aspect of learned compounds (in this case, none of the two elements can be taken for a learned constituent, but we see the adjectives *fisic* 'physical' and *químic* 'chemical', or *franc* 'French' and *canadenc* 'Canadian' welded in one word). Moreover, these structures often offer a semantic double interpretation that responds to two different structures (coordinate clause (7) or subordinate clause (8)) that can only be readily disambiguated when seen in the context in which they appear:<sup>14</sup>

(7) *francocanadenc* 'between France and Canada'

[[*franco*]<sub>A</sub> [*canadenc*]<sub>A</sub>]<sub>A</sub>

(8) *francocanadenc* 'Canadian French Area'

[*franco*]<sub>A</sub> [*canadenc*]<sub>A</sub>]<sub>A</sub>

Semantically, what stands out in all coordinate compounds is the fact that both words forming the compound are coherent, which means that they belong to the same conceptual domain: mainly professions, in the case of [N+N]<sub>N</sub> compounds and adjectives related to science, political science, economy or geography in the case of [A+A]<sub>A</sub> compounds.

14. In fact, the difference in this type of compounds is contemplated in the provisional grammar published by the *Institut d'Estudis Catalans* (2002: § 4.3) in which it is stated that, although compounds in Catalan are not hyphenated (with the exception of consonant agglutinations hindering correct reading, such as *Bell-lloc*, when the first constituent bears a graphic accent such as in *pèl-roig* 'red-haired', in specialised terms, where the first constituent is a number or a symbol, like *àcid d-lisèrgic* 'd-lysergic acid', compounds with cardinal points, such as *sud-oest* 'southwest', numerals, reduplicating compounds like *bup-bup* 'bow-wow', or compounds with proper names like *comptador Geiger-Müller* 'Geiger-Müller counter'), compounds formed by two (or more) adjectives can be hyphenated, especially when there is a need to underline the independence of the expressed concepts. So, in *Els països hispanoamericans esmentats al tractat hispano-americà* 'the Hispano-American countries mentioned in the Spanish-American treaty' *hispanoamericans* is a subordinated compound meaning 'countries that speak Spanish', whereas *hispano-americà* is a coordinated compound that means 'between Spain and America'.

3.4. *Non-productive compounds*

In this last section, we will concentrate on compounds showing too few examples in the language and we will summarise them in Table 4.

Table 4

| Structure                     | Head | Example             | Translation           |
|-------------------------------|------|---------------------|-----------------------|
| <b>Subordinate</b>            |      |                     |                       |
| [A+V] <sub>V</sub>            | R    | <i>carcomprar</i>   | 'to pay too much for' |
| [N+V] <sub>V</sub>            | R    | <i>ullprendre</i>   | 'to bewitch'          |
| [N+[V+sufN]] <sub>N</sub>     | R    | <i>vianant</i>      | 'pedestrian'          |
| <b>Attributive/appositive</b> |      |                     |                       |
| [N+[V+sufA]] <sub>A</sub>     | R    | <i>palplantat</i>   | 'dead still'          |
| [Adv+V] <sub>V</sub>          | R    | <i>malacostumar</i> | 'to spoil'            |
| [Adv+[V+sufA]] <sub>A</sub>   | R    | <i>benestant</i>    | 'well-to-do'          |
| [A+[V+sufA]] <sub>A</sub>     | R    | <i>nounat</i>       | 'newborn'             |
| <b>Coordinate</b>             |      |                     |                       |
| [N+conj+N] <sub>N</sub>       | B    | <i>punt i coma</i>  | 'semicolon'           |
| [N+conj+N] <sub>N</sub>       | ∅    | <i>vetesifils</i>   | 'haberdashery'        |
| [Poss+N] <sub>N</sub>         | R    | <i>nostramo</i>     | 'master'              |
| [Num+N] <sub>N</sub>          | ∅    | <i>milhomes</i>     | 'braggart'            |
| [Adv+Adv] <sub>Adv</sub>      | ∅    | <i>també</i>        | 'also'                |

Structures in Table 4 show very little productivity nowadays. In fact, [N+V]<sub>V</sub> compounds belonging to the *camatrençar* 'to break an animal's leg' type are almost nonexistent in the rest of the Romance languages. Despite their limited use in the present day, all speakers interpret them correctly due to the syntactico-semantic relationship established between both constituents in the compound: the noun is the direct object of the verb (Mascaró 1986: 66). The strangeness lies in the fact that if it really is the internal argument, it is surprising to see it placed before the verb radical, although Gràcia and Fullana (1999) explain this claiming that the noun acts as the verb radical's modifier, which is equivalent to the so-called Greek accusative or relational accusative, and that the fact that it appears before the verb is due to a rule of movement.

#### 4. Syntactic and semantic relations between compound elements

In this section, we will analyse the syntactic or semantic traits affecting the construction of some compound types, as well as some problematic cases. We will then refer the reader to Gràcia (2002) for a more thorough description of these cases.

4.1.  $[V+N]_N$  compounds

It is probably the  $[V+N]_N$  structure (*obrellaunes* ‘can opener’, *espanta-sogres* ‘party blower’, etc.) that generated the most theoretical-descriptive works, especially regarding the grammatical category of the compound’s first constituent. In general terms, along the lines of this paper, we consider it a verbal element which, according to the different authors, is classified as stem, third-person singular, second-person singular of the imperative, etc. (Ferrater 1981; Contreras 1985; Mascaró 1986; Cabré and Rigau 1986; Gavarró 1990; Moll 1991; Padrosa 2010). However, others such as Varela (1990) and Grossmann (1986) consider them as nominal constituents with an agentive value, as there are existing forms like *busca* ‘search’ (un *buscapersones* ‘beeper’), derived from the verb without any affix adjunction and behaving like the head of a compound; based on this perspective, these then constitute endocentric compounds. On the contrary, considering that the first constituent is a verb, we must necessarily consider those as exocentric compounds, given that a compound’s grammatical and semantic properties cannot be inherited from any of the constituents that form it: *paraigua* ‘umbrella’ is a countable masculine noun related to an object and formed with the verb *parar* ‘to stop’ and *aigua* ‘water’, which is an utterly feminine noun; *somiatruites* ‘daydreamer’ is a countable masculine and feminine noun that applies to humans, formed with the verb *somiar* ‘to dream’ and *truites* ‘omelettes’, a countable feminine noun related to an object, etc.

The main characteristic of all verbs belonging to this type of compound is that they are transitive, which explains why the noun is interpreted as their internal stem argument. Verbs, however, can only be [+activity] with a subject (whether agent or instrument), and not a verb [+state], which is why it is not possible to form compounds such as *\*estimagats* (lit. to love + cats) or *\*temtempestes* (lit. to fear + storms). Along these lines, the verbal base must end in a thematic vowel and can end neither in a consonant (neither semivowel nor semiconsonant) nor a stressed vowel: *llevataps* (corkscrew), but *\*treutaps* (since the verbal base ends in a semivowel).

$[N+[V+sufN]]_N$  compounds behave in the same manner, where the noun is also the direct object of the verb (like in *terratinent* ‘landowner’) or even a prepositional complement (like in *vianant* ‘pedestrian’ (lit. walking on the road)). What is interesting here is the reverse order of appearance of the different constituents, opposite from the VO neutral order of words in Catalan, and that some authors have associated with the OV order in Latin (Benveniste 1967). This may explain its scarce productivity.

It is worth pointing out that compounds with this structure can be, in some cases, recurrently applied, such as *portaescuradents* ‘toothpick holder’, *portalligacames* ‘garter holder’ or *neteजारabrises* ‘windshield wipers’, which are structured as  $[V+[V+N]_N]_N$ .

4.2.  $[N+N]_N$  compounds

As Gràcia (2002) underlines, in subordinate as well as in appositive compounds, a relationship of dependence is established, in the sense that the second constituent modifies the first one to indicate its matter (*ferrocarril* ‘railway’), finality (*autopista* ‘motorway’) or even its status as a part of a whole (*capvespre* ‘dusk’). Furthermore, in the orthographically attached compounds, speakers of dialects with vocalic reduction pronounce these compounds with one unique stress, which indicates that speakers have lost awareness of compounding and see compounds as simple words.

Nevertheless, we do not see this happening with more recent compounds, such as *cèl·lula mare* ‘stem cell’, *peix espasa* ‘swordfish’, or *autor editor* ‘author & editor’, that behave like prototypical compounds with two stresses. It is also in these cases that metaphoric processes often intervene, a characteristic that compounds like *ferrocarril* do not have.

4.3.  $[N+V]_V$  compounds

As with  $[V+N]_N$  compounds, verbs that form this type of (at present) non-productive compounds are transitive verbs and the noun that appears in the compound is its internal argument, which is often a noun of inalienable possession: *alatrencar* ‘to break (a bird’s) wings’, *ullferir* ‘to injure the eye’, etc. Gràcia and Fullana (1999) claim that in these cases, the noun acts as a verb modifier, equivalent to Greek accusative or to relational accusative.

4.4.  $[N+A]_A$  and  $[N+[VSuf]_A]_A$  adjectival compounds

Compounds belonging to these two types are endocentric and the second constituent determines the characteristics of the resulting compound while the noun appearing in the first position acts as a modifier of the adjective or participle: *camacurt* ‘short-legged’ is a person with short legs and *palplantat* ‘dead still’ literally means ‘planted like a stick’. In the first case, the noun qualified by the compounded adjective corresponds to the argument possessor of the noun of inalienable possession that appears as the first constituent: to a *cararodona* ‘round-faced’ person, the face is the person; in the second example, the noun, qualified by the compounded adjective, is the internal argument of the verbal root: to a *palplantada* ‘dead still’ girl, the girl is the direct object of *plantar* ‘to plant’.



#### 4.5. $[A+A]_A$ compounds

We have considered, along with Mascaró (1986) and Benveniste (1967), that we are dealing with coordinated endocentric compounds such as the Indo-European dvandva compounds. However, there are authors who see them differently: Cabré and Rigau (1986) consider them exocentric, with no head, and Gavarró (1990) analyses them as compounds with a left head. This diversity responds to the fact that it is difficult to clearly establish and define the relationships established between the two constituents of the compound: it is clear that *agredolç* ‘sweet and sour’ means *agre* ‘sour’ and *dolç* ‘sweet’ at the same time, but it is no that easy to define *blaugrana* ‘blue and deep-red’ as *blau* ‘blue’ and *grana* ‘deep-red’ simultaneously.

### 5. Compounding and inflection

#### 5.1. Number

With the exception of phrasal units, Catalan compounds only pluralise the second element: *malnom* ‘nick name’ – *malnoms* / \**malsnoms*. In  $[N+N]_N$  appositive compounds and despite the existing trend to only pluralise the first constituent (which is the head (such as *escola taller* ‘vocational school’)), we find many cases where both elements are pluralized, especially in the case of coordinate compounds:

- (9) *alcalde president* ‘mayor president’ – *alcaldes presidents*  
*porc senglar* ‘wild boar’ – *porcs senglars*

Use by speakers is not that clear and we observe some indecision when it comes to pluralising this type of compounds, regardless of whether they are appositive or coordinate. A non-exhaustive search on the Internet in February 2011 showed that we are likely to find both possibilities, although the general tendency is to obtain more results with the first constituent pluralized only, as we can see in table 5.

Although results with both constituents in plural proved to be less common, this nevertheless demonstrates how these cases can generate doubts even in webpages belonging to prestigious centres and institutions. In fact, we even find both forms in documents belonging to a same webpage (both constituents in plural for *cèl·lula mare* ‘stem cell’ or *estat membre* ‘member state’ appear in the Wikipedia, page for instance). Here, we should add that, in the case of *clau* (which dictionaries insist be used as an invariable adjective), speakers are nevertheless pushed to pluralise it as they would any other adjective when they apply it to other nouns, as seen in (4).

Table 5

|                        | Total     | Pages showing the examples   |
|------------------------|-----------|--|
| <i>cèl·lules mare</i>  | 552,000   | Institut de Recerca Biomèdica de Barcelona, Observatori de Drets Sexuals i Reproductius, Hospital Clínic, etc.   |
| <i>cèl·lules mares</i> | 9,880     | <i>Diccionari enciclopèdic de medicina</i> (Enciclopèdia Catalana), Europa Press, Televisió de Catalunya, etc.   |
| <i>paraules clau</i>   | 7,590,000 | Edu365, Consorci de Biblioteques Universitàries de Catalunya, Universitat Oberta de Catalunya, etc.  |
| <i>paraules claus</i>  | 565,000   | Centre de Recursos per a l'Aprenentatge i la Investigació (Universitat de Barcelona), <i>Diccionari de paraules claus</i> (Generalitat de Catalunya), journal <i>Treballs de la Societat Catalana de Geografia</i> (Institut d'Estudis Catalans), etc. |
| <i>estats membre</i>   | 45,800    | Wikipedia, Fundació per la Pau, Europa Press, <i>Regió 7</i> , etc.  |
| <i>estats membres</i>  | 547,000   | Wikipedia, Generalitat de Catalunya, European Union, Council of Europe, etc.   |

The fact that both constituents of a compound appear together, forming orthographic compounds, does not exempt them from variation: rules state that compounds such as *capgròs* 'tadpole' (lit. big head), *figaflor* 'wimp' (lit. fig flower) or *pellroja* 'redskin' must be pluralized as *capgrossos*, *figaflors* and *pellroges*, respectively. However, all these examples often have their constituents pluralized by speakers because the speakers have lost awareness of the words' compound nature. They therefore apply plural rules to them as if they were independent clauses (*caps grossos*, *figues flors*, *pellls roges*).<sup>15</sup>

It is worth noting the case of [V+N]<sub>N</sub> compounds, where the second element tends to occur in the plural form (*trencanous* 'nutcracker', *tocadiscos* 'record player'), although we find some (albeit fewer) cases in the singular form, such as *paraigua* 'umbrella', *trencaglaç* 'icebreaker' or *tallafoc* 'firewall' which can be attributed to the grammatical nature of the noun (mass or count noun).<sup>16</sup> The truth is, however, we do find many cases of count nouns occurring in [V+N]<sub>N</sub> compounds in the singular, such as *guardabosc* 'forest warden' (cf. *guarda-costes* 'coastguard vessel', *guardaespatlles* 'bodyguard', etc.), *salta-regle* 'bevel rule' (cf. *saltataulells* 'shop assistant' (pop.), *saltaterrats* 'burglar') or *picaporta* 'doorknocker' (cf. *picamatalassos* 'carpet-beater', *picaplets* 'lawyer' (derog.)).

Finally, we should highlight the cases of coordinate compounds belonging to the *sordmut* 'deaf and dumb' type, in which, according to Moll (1968), both

15. Gràcia (2002) links the degree of lexicalisation of the compound with pluralisation: the more the compound is lexicalised, the higher the tendency to place the nominal inflection at the end.

16. The main use, though, tends to generalise the plural form of the noun: *un paraigües*, *una (porta) tallafocs*, etc.

constituents require inflexion in gender and number: *sordmut* – *sordamuda* – *sordsmuts* – *sordesmudes*, although speakers tend to apply the inflection only to the second adjective, as set in the rules.

## 5.2. Gender

In the case of endocentric compounds, no matter what type they belong to, compound gender is determined by the gender of the head element, regardless of its location. Exceptionally, though, we can find isolated cases in which this does not apply, such as *horabaixa*: [*hora*<sub>Nfem</sub>*baixa*<sub>Afem</sub>]<sub>Nmasc</sub> ‘afternoon’.<sup>17</sup>

On the contrary, in exocentric compounds, the gender tends to be masculine: *milfulls* ‘mille-feuille’, *ventafocs* ‘fire fan’, *paraigua* ‘umbrella’. Compounds applied to people are oftentimes formally invariable, even when applied to both genders: *pocatraça* ‘clumsy’.

## 6. Remarks on productivity

As a final remark, we will talk briefly about compounding in Catalan, based on documentation on neologisms retrieved from the *Observatori de Neologia*.<sup>18</sup> In this document, we will submit a productivity sample regarding compounding, with samples taken from the years 2000, 2005 and 2010<sup>19</sup>, organised in Table 6.<sup>20</sup>

As we can see, we don’t find all compounding patterns found and highlighted in Catalan. Instead, we primarily find neoclassical hybrid compounds

17. Yet, we also find the compound in feminine gender, as shown by the anaphoric pronoun: “I s’horabaixa; la, deixam passar i me mires” ‘and the afternoon we let it go and you look at me’ (*Viure sense tu* ‘Living without you’ song by Antònia Font).

18. The *Observatori de Neologia* is a research group within the *Universitat Pompeu Fabra*, headed by M. Teresa Cabré and coordinated by Judit Freixa. It analyses the phenomenon of new word occurrence in Catalan and in Spanish, based on term extractions from widespread oral and written media. The data is extracted only from written sub-corpus and the figures given do not refer directly to real productivity of the process or of the concrete structure, instead depending on the totality of the term-extracted texts in a given year and the neologisms obtained. The classification by types and sub-types is ours and we accept complete responsibility for any errors they might show.

19. 2010 figures show the state of the operational data base as of January 2011 and are, therefore, provisional.

20. See Feliu et al. (2002), Lladó and Talamino (2002), Solé and Vázquez (2002) and López et al. (2002) for a description of productivity in compounding processes in Catalan from 1989 to 2000.

Table 6

|                            |                                  | 2000       | 2005       | 2010       | total        |
|----------------------------|----------------------------------|------------|------------|------------|--------------|
| Subordinate                | [N+N] <sub>N</sub>               | 18         | 31         | 19         | 68           |
|                            | [V+N] <sub>N</sub>               | 20         | 12         | 15         | 47           |
|                            | pure [sN+sN] <sub>N</sub>        | 7          | 30         | 10         | 47           |
|                            | hybrid [(s)N+(s)N] <sub>N</sub>  | 322        | 309        | 131        | 762          |
|                            | [N+prep+N] <sub>N</sub>          | 10         | 27         | 4          | 41           |
|                            | [N+[V+sufN]] <sub>N</sub>        | –          | 1          | –          | 1            |
| Appositive/<br>attributive | [N+N] <sub>N</sub>               | 15         | 30         | 20         | 65           |
|                            | [N+A] <sub>N</sub>               | 29         | 64         | 12         | 105          |
|                            | [A+N] <sub>N</sub>               | 9          | 12         | 6          | 27           |
|                            | [N+[V+sufA]] <sub>A</sub>        | 1          | –          | –          | 1            |
| Coordinate                 | [N+N] <sub>N</sub>               | 13         | 35         | 13         | 61           |
|                            | [A+A] <sub>A</sub>               | 7          | 6          | 12         | 25           |
|                            | [A+linking vowel+A] <sub>A</sub> | 81         | 88         | 31         | 200          |
|                            | [N+conj+N] <sub>N</sub>          | –          | 2          | 2          | 4            |
| Other <sup>21</sup>        | [prep+N] <sub>N</sub>            | 1          | 3          | 1          | 5            |
|                            |                                  | <b>533</b> | <b>650</b> | <b>276</b> | <b>1,459</b> |

which use a Catalan word combined with a semi-word. We see an overwhelming production of nouns and, at a lesser level, of adjectives (most of which also imitate neoclassical construction by forming coordinate compounds with a linking vowel); there is no record of verb neology patterns resulting from compounding.

Based on this data, compounding in Catalan is a word forming productive process.<sup>22</sup> There is of course a certain bias regarding the documented types, as the corpus is fed by texts proceeding from communication media. Consequently, these texts could show variations, especially regarding the [V+N]<sub>N</sub> type with instrumental and agentive meaning and with remarkable significance in colloquial usage, as shown in spontaneous compounds like *pela-Xites* ‘colloquial name for Epilady’ (lit. to peel + Cheeta<sub>+pl.</sub> *tocacompacts* ‘compact disc player’ (lit. to play + compact discs), formed by analogy with *tocadiscos* ‘record player’, (*gel*) *matabitxos* (lit. to kill + bugs) ‘sanitizing gel’, etc., all of which do not appear in the *Observatori de Neologia* corpus because they lay far from the standard tone used by the media.<sup>23</sup>

21. These are the cases of *sense sostre* ‘homeless’ (lit. without ceiling), *sense papers* ‘illegal’ (lit. without documents), etc., than can be interpreted as syntagmatic compounds belonging to the [N+prep+N]<sub>N</sub> type, in which the first noun is elided. See Roca (2004).
22. Last years’ data show that compounding is the second most usual resource for word formation, behind derivation only. See *Observatori de Neologia* (2004), among others.
23. In this sense then, we disagree with Estopà’s statement (2010) that the [V+N]<sub>N</sub> pattern is not productive in Catalan.

## 7. Conclusions

Compounding in Catalan is, as in the rest of the Romance languages, a process of a complex and diverse word formation. The multiplicity of structures studied is an irrefutable proof of the importance it already had as a resource in Latin, as many structures are mere remnants and not presently productive at all, although speakers do not have any problems in interpreting their meaning anyway. At present, though, compounding types are minor, although they undoubtedly demonstrate their vitality.

The general scene we have presented here concerning compound words in Catalan must now be supplemented with data on uses by speakers in various communicative situations. We propose some: on the one hand, it seems adequate to go more in depth in the study of the overwhelming presence of neo-classic hybrids, at least in the communication media, linked, most probably, to the semantic transparency and to the level of formality granted to them because of their origin, a fact that could seem especially suitable depending on which communication contexts we are talking about. On the other hand, we need to undergo further research in the existing relationship between more or less conscious transgressions of certain rules which are made by speakers (use of proper nouns as a base for compounding, allegedly impossible combinations, etc.) and the recreational and expressive load they contain. Therefore, it seems appropriate to carry out a detailed analysis of the use of different patterns in syntagmatic compounds, as well as of the figurative sense contained in many examples we can find, as well as their use in specialised speech.

*Universitat Pompeu Fabra*  
elisenda.bernal@upf.edu

## References

- Badia i Margarit, Antoni M. 1962. *Gramàtica catalana*. Madrid: Gredos.
- Bayà, M. Rosa. 1997. *La prefixació en català: aproximació teoricodescriptiva*. Barcelona: Institut Universitari de Lingüística Aplicada – Universitat Pompeu Fabra.
- Benveniste, Émile. 1967. Fondements syntaxiques de la composition nominale. *Bulletin de la Société de Linguistique de Paris* 62(1). 15–31.
- Benveniste, Émile. 1974. Formes nouvelles de la composition nominale. In *Problèmes de linguistique générale* 2. Paris: Gallimard. 163–176.
- Bernal, Elisenda. 1997. *Els sufixos verbalitzadors del català. Descripció sintacticosemàntica*. Barcelona: Institut Universitari de Lingüística Aplicada – Universitat Pompeu Fabra.
- Bernal, Elisenda. 2007. Fabra i la formació de paraules: els criteris morfològics. In *La figura i l'obra de Pompeu Fabra*, 115–126. Barcelona: Universitat Pompeu Fabra.
- Bernal, Elisenda. 2010. Nuevos prefijos: implicaciones para la morfología y la lexicografía. In Maria Iliescu, Heidi Siller-Runggaldier & Paul Danler (eds.), *Actes du XXV Congrès International de Linguistique et Philologie Romanes*, VII, 361–373. Berlin: Walter de Gruyter.

- Bruguera, Jordi. 1996. *Diccionari etimològic*. Barcelona: Enciclopèdia Catalana.
- Cabré, M. Teresa. 1992. *La terminologia. La teoria, el mètode, les aplicacions*. Barcelona: Empúries.
- Cabré, M. Teresa. 1994. *A l'entorn de la paraula*. València: Universitat de València. 2 vols.
- Cabré, M. Teresa & Gemma Rigau. 1986. *Lexicologia i semàntica*. Barcelona: Enciclopèdia Catalana.
- Cabré, M. Teresa, Judit Freixa & Elisabet Solé. 2002. À la limite des mots construits possibles. In M. Teresa Cabré, Judit Freixa & Elisabet Solé (eds.), *Lèxic i neologia*, 121–138. Barcelona: Institut Universitari de Lingüística Aplicada.
- Contreras, Heles. 1985. Spanish exocentric compounds. In Frank H. Nuessel (ed.), *Current issues in hispanic phonology and morphology*, 14–26. Bloomington: IULC.
- Contreras, Joan Miquel & Avel·lina Suñer. 2004. Cuestiones generales. In Elixabete Pérez Gaztelu, Igone Zabala & Lluïsa Gràcia (eds.), *Las fronteras de la composición en lenguas románicas y en vasco*, 45–160. Donostia: Universidad de Deusto.
- Cottez, Henri. 1989. *Dictionnaire des structures du vocabulaire savant*. Paris: Le Robert.
- Darmesteter, Arsène. 1875. *Traité de la formation des mots composés dans la langue française comparée aux autres langues romanes et au latin*. Paris: Librairie A. Franck.
- [DIEC2]. Institut d'Estudis Catalans. 2007. *Diccionari de la llengua catalana*. 2<sup>nd</sup> edn. Barcelona: Edicions 62, Enciclopèdia Catalana.
- [DRAE]. Real Academia Española. 2001. *Diccionario de la lengua española*. Madrid: Espasa.
- Estopà, Rosa. 2010. La composició patrimonial en català perd representativitat. Estudi d'un corpus de neologismes de premsa i de ràdio. *Estudis Romànics* 32. 125–147.
- Fabra, Pompeu. 1912. *Gramàtica de la llengua catalana*. In Jordi Mir & Joan Solà (eds.) (2005), Pompeu Fabra. *Obres completes* /1. Barcelona: Enciclopèdia Catalana, Edicions 62, Edicions 3i4, Editorial Moll.
- Fabra, Pompeu. 1956. *Gramàtica catalana*. Barcelona: Teide.
- Feliu, Judit, Yannick Garcia & Isabel Obradós. 2002. Aspectes de composició: neologismes nom-nom. In M. Teresa Cabré, Judit Freixa, Judit & Elisabet Solé (eds.), *Lèxic i neologia*, 217–224. Barcelona: Institut Universitari de Lingüística Aplicada.
- Ferrater, Gabriel. 1981. *Sobre el llenguatge*. Barcelona: Quaderns Crema.
- Gavarró, Anna. 1990. *Syntactic theory and the grammar of catalan compounds*. Edinburgh: University of Edinburgh.
- García Palacios, Joaquín. 2000–2001. La quimera del euro. *Revista de lexicografía* VII. 33–62.
- Gràcia, Lluïsa. 2002. La composició. In Joan Solà, M. Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya (coords.), *Gramàtica del català contemporani*, I, 777–829. Barcelona: Empúries.
- Gràcia, Lluïsa & Olga Fullana. 1999. On catalan verbal compounds. *Probus* 11. 239–261.
- Grossmann, Maria. 1986. Anàlisi dels compostos catalans tipus *somiatruïtes*. *Estudis de llengua i literatura catalanes* XII. 155–169.
- Institut d'Estudis Catalans. 2002. *Gramàtica de la llengua catalana*. Barcelona: Institut d'Estudis Catalans. <http://www.iecat.net/institucio/seccions/Filologica/gramatica/default.asp> (accessed 30 November 2010).
- Lladó, Mireia & Óscar Talamino. 2002. Aspectes de composició: neologismes verb-nom. In M. Teresa Cabré, Judit Freixa & Elisabet Solé (eds.), *Lèxic i neologia*, 225–234. Barcelona: Institut Universitari de Lingüística Aplicada.
- López, Silvia, Cristina Mayoral, Montserrat Pibernat & Elisabet Solé. 2002. Dos formants pejoratiu: el prefix *pseudo-* i el sufix *-oide*. In M. Teresa Cabré, Judit Freixa & Elisabet Solé (eds.), *Lèxic i neologia*, 251–264. Barcelona: Institut Universitari de Lingüística Aplicada.
- López del Castillo, Lluís. 2002. *Diccionari de formació de paraules*. Barcelona: Edicions 62.

- Lorente, Mercè. 2002. Altres elements lèxics. In Joan Solà, M. Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya (coords.), *Gramàtica del català contemporani*, I, 831–888. Barcelona: Empúries.
- Lüdeling, Anke & Stefan Evert. 2005. The emergence of productive non-medical *-itis*: corpus evidence and qualitative analysis. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence. empirical, theoretical, and computational perspectives*, 351–370. Berlin: Mouton de Gruyter.
- Marvà, Jeroni. 1968. *Curs superior de gramàtica catalana*. Barcelona: Barcino.
- Mascaró, Joan. 1986. *Morfologia*. Barcelona: Enciclopèdia Catalana.
- Mascaró, Joan. 2002. El sistema vocàlic. Reducció vocàlica. In Joan Solà, M. Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya (coords.), *Gramàtica del català contemporani*, I, 89–123. Barcelona: Empúries.
- Menéndez Pidal, Ramón. 1904. *Manual elemental de gramàtica històrica espanyola*. Madrid: Librería general de Victoriano Suárez.
- Moll, Francesc de Borja. 1968. *Gramàtica catalana referida especialment a les Illes Balears*. Palma de Mallorca: Moll.
- Moll, Francesc de Borja. 1991. *Gramàtica històrica catalana*. Palma de Mallorca: Moll.
- Observatori de Neologia. 2004. *Llengua catalana i neologia*. Barcelona: Meteora.
- Padrosa, Susanna. 2010. *Complex word-formation and the morphology-syntax interface*. Bellaterra: Universitat Autònoma de Barcelona.
- Pérez Saldanya, Manuel, Manuel Sifre & Júlia Todolí. 2004. *Morfologia catalana*. Barcelona: Editorial UOC.
- Roca, Francesc. 2004. Neologismos nominales del tipo ‘preposición + nombre’. In Elixabete Pérez Gaztelu, Igone Zabala & Lluïsa Gràcia (eds.), *Las fronteras de la composición en lenguas románicas y en vasco*, 315–353. Donostia: Universidad de Deusto.
- Ruaix, Josep. 1986. *El català* [3. Mojà: J. Ruaix ed.
- Ruiz Gurillo, Leonor. 1997. *Aspectos de fraseología teórica espanyola*. València: Universitat de València.
- Rull, Xavier. 2009. *La composició culta en català*. Palma: Moll.
- Solà, Joan. 1982. Pompeu Fabra i la gramàtica catalana. In *Homenatge a Pompeu Fabra. Acte inaugural del curs 1982–1983*, 53–84. Barcelona: Universitat de Barcelona.
- Solé, Xavier & Mercè Vázquez. 2002. Aspectes de composició: neologismes amb estructures diferents de nom-nom i verb-nom. In M. Teresa Cabré, Judit Freixa & Elisabet Solé (eds.), *Lèxic i neologia*, 235–250. Barcelona: Institut Universitari de Lingüística Aplicada.
- Termcat. 1990. *Formació de termes amb elements cultes*. Barcelona: Termcat.
- Turon, Lúcia. 2004. Las formas prefijadas tónicas en catalán y español (*sobre-*, *sota-*, *contra-*, *entre-*). In Elixabete Pérez Gaztelu, Igone Zabala & Lluïsa Gràcia (eds.), *Las fronteras de la composición en lenguas románicas y en vasco*, 239–260. Donostia: Universidad de Deusto.
- Vallès, Teresa. 2007. Els prefixoides del català. In Mercè Lorente, Rosa Estopà, Judit Freixa, Jaume Martí & Carles Tebé (eds.), *Estudis de lingüística aplicada en honor de M. Teresa Cabré Castellví. Volum II: Deixebles*, 459–471. Barcelona: Institut Universitari de Lingüística Aplicada.
- Varela, Soledad. 1990. *Fundamentos de morfología*. Madrid: Síntesis.
- Varela, Soledad. 2005. *Morfología léxica: la formación de palabras*. Madrid: Gredos.
- Wheeler, Max. 1979. *Phonology of catalan*. Oxford: Blackwell.
- Wotjak, Gerd (ed.). 1998. *Estudios de fraseología y fraseografía del español actual*. Frankfurt am Main: Vervuert.

