

**THE CATALAN DEFINITE ARTICLE AS LEXICAL  
SHARING**

Alex Alsina

Pompeu Fabra University

Proceedings of the LFG10 Conference

Miriam Butt and Tracy Holloway King (Editors)

2010

CSLI Publications

<http://csli-publications.stanford.edu/>

## Abstract

This paper argues for the claim that the definite article (d-article) in Catalan is a bound morpheme that adds its information about grammatical category to that of the host it forms a compound with. In this way, the derived compound satisfies two  $X^0$  categories in the c-structure, in an instance of lexical sharing (Wescoat 2002). The present proposal is shown to be superior to alternatives that, while assuming bound morpheme status for the d-article, assume a single category for the word that contains the d-article (as in Miller 1992 for French). The crucial evidence is provided by potentially boundless sequences in which the d-article and weak prepositions are adjacent to each other. Only a theory that recognizes the status of these elements as bound morphemes that add a grammatical category to the word they are part of correctly accounts for the relevant evidence.

## 1 Introduction

The goal of this paper is to argue for the claim that the definite article (d-article) in Catalan is a bound morpheme and that the word that the d-article is part of corresponds to two  $X^0$  categories in the c-structure. This is an instance of the theoretical construct of *lexical sharing*, proposed by Wescoat (2002), whereby a single word (or vocabulary item) instantiates, or corresponds to, more than one  $X^0$  category. The paper likewise argues for the claim that weak prepositions in Catalan are also lexical sharing morphemes. A consequence of these claims is that, since the d-article and the weak preposition *de* can be adjacent to each other in potentially boundless sequences, they give rise to instances of multiple lexical sharing (words corresponding to many  $X^0$  categories). The evidence provided suggests that lexical sharing is the right way to handle “phrasal affixes” in general, over alternative analyses.<sup>†</sup>

The paper starts out by outlining the structure of the DP in Catalan and arguing for the existence of two classes of determiners. In the context of the Catalan determiner system, in section 3, the d-article is shown to have some properties common to all determiners, but also some special properties that set it aside from both classes of determiners. In section 4, it is proposed that the explanation for these properties involves assuming (a) that it is a bound morpheme (following similar proposals by Miller 1992 for French) and (b) that the word containing this morpheme has the grammatical categories of both its component elements—an instance of lexical sharing. Next, in section 5, it is shown that so-called weak prepositions in Catalan are also bound morphemes that attach to a following word yielding a word with the grammatical categories of its several component elements. Since the d-article can attach to the weak preposition *de* to its right and this item can attach to

---

<sup>†</sup> This paper has been supported in part by research grant HUM2007–61916FILO. I thank Joan Bresnan and participants in LFG10 for observations, Martin Forst for providing many examples from the web that have made me change and refine my analysis, and Michael Wescoat for detailed comments on earlier versions of the paper. All errors are due to me.

the d-article to its right, complex words consisting of sequences of d-articles and weak prepositions are formed: these words have a list of the categories contributed by their morphemes. It will be argued that this approach is superior to alternative approaches without lexical sharing. The present approach will be shown to provide a simple way of accounting for co-occurrence restrictions involving particular lexical items, in section 6. Finally, conclusions will be presented.

## 2 The structure of the Catalan DP

This section proposes an analysis of the Catalan DP: it includes the assumption of two classes of determiners (subsection 2.1), a proposal about its c-structure (2.2), and a proposal about its f-structure and the mapping between the two structures (2.3).

### 2.1 Two classes of determiners

If we use the working definition of determiner as a word belonging to a closed class that can only appear in the initial position of a noun phrase, we find that, in Catalan, the class of determiners splits into two classes on the basis of the syntactic behavior of these words. We shall refer to the two classes of determiners as *semipronominal* and *nonpronominal* determiners; some examples of words belonging to these two classes are given in (1):

- (1) a. Semipronominal Ds: *aquest* ‘this’, *aquell* ‘that’, *quin* ‘which’, *quant* ‘how much/many’, *cap* ‘no/none/any’, *un* ‘one/a/an’, *algun* ‘some/any’, *tant* ‘so much/many’, *molt* ‘many/much’, etc.
- b. Nonpronominal Ds: *cada* ‘each’, *sengles* ‘each his own’, *qualque* ‘some’, *cert* ‘certain’, *mon* ‘my’, *ton* ‘your’, *son* ‘his/her’, *llur* ‘their’, *tot* ‘every’ (as in *tot estudiant* ‘every student’, not as in *tots els estudiants* ‘all the students’), etc.

The difference between the two classes is that, whereas nonpronominal Ds must head a DP with a noun heading its sister NP, semipronominal Ds may occur without this noun. In other words, semipronominal Ds may introduce headless noun phrases, whereas nonpronominal Ds may not. Taking *aquella* ‘that(F,S)’ and *cada* ‘each’ as examples of semipronominal and nonpronominal Ds respectively, we see that both can introduce noun phrases with overt nouns, but only *aquella* can introduce a noun phrase lacking an overt noun, that is, consisting only of the determiner, or of the D and a PP, or of the D and an AP, or of the D and a relative clause:

- (2) a. Hem insistit en aquella situació (compromesa)  
we-have insisted on that(F,S) situation(F,S) delicate(F,S)  
(que es va presentar)  
that arose  
‘We insisted on that (delicate) situation (that arose).’

- b. Hem insistit en aquella.  
we-have insisted on that(F,S) ('We insisted on that one.')
  - c. ... en aquella compromesa.  
... on that(F,S) delicate(F,S) ('... on that delicate one.')
  - d. ... en aquella que es va presentar.  
... on that(F,S) that arose ('... on that one that arose.')
- (3) a. Hem insistit en cada situació (compromesa) (que  
we-have insisted on each situation(F,S) delicate(F,S) that  
es va presentar)  
arose  
'We insisted on each (delicate) situation (that arose).'
- b. \* Hem insistit en cada.  
we-have insisted on each ('We insisted on each one.')
  - c. \* ... en cada compromesa.  
... on each delicate(F,S) ('... on each delicate one.')
  - d. \* ... en cada que es va presentar.  
... on each that arose ('... on each one that arose.')

## 2.2 The c-structure of the DP

We will assume that both classes of determiners belong to category D and that the category D projects a DP, adopting the DP hypothesis of Abney (1987) and others, which has been incorporated into the LFG framework by Bresnan (2001), among others. The structure of phrases, and specifically the structure of DP, conforms to X-bar Theory, as expressed in (4):

- (4) X-bar Theory:
- a.  $X'' \rightarrow YP X^{(i)}$
  - b.  $X' \rightarrow X^{(i)} ZP$

We take XP to be equivalent to  $X''$ . The sister of D, if there is one, is an NP. Spec of DP is reserved for definite *tot/s* 'all' and adverbs or adverbial phrases such as *fins i tot* 'even', *exactament* 'exactly', *només* 'only', *almenys* 'at least', etc. Examples containing some of these specifiers of DP would be: *tots els llibres* 'all the books', *només aquelles notes* 'only those notes', etc.

Spec of NP is reserved for a special class of adjectives that we can call specificational adjectives, or SpecA, some examples of which would be: *altre* 'other', *dos* 'two', *tres* 'three', *primer* 'first', *segon* 'second', *últim* 'last', *seu* 'his', *nostre* 'our', etc.

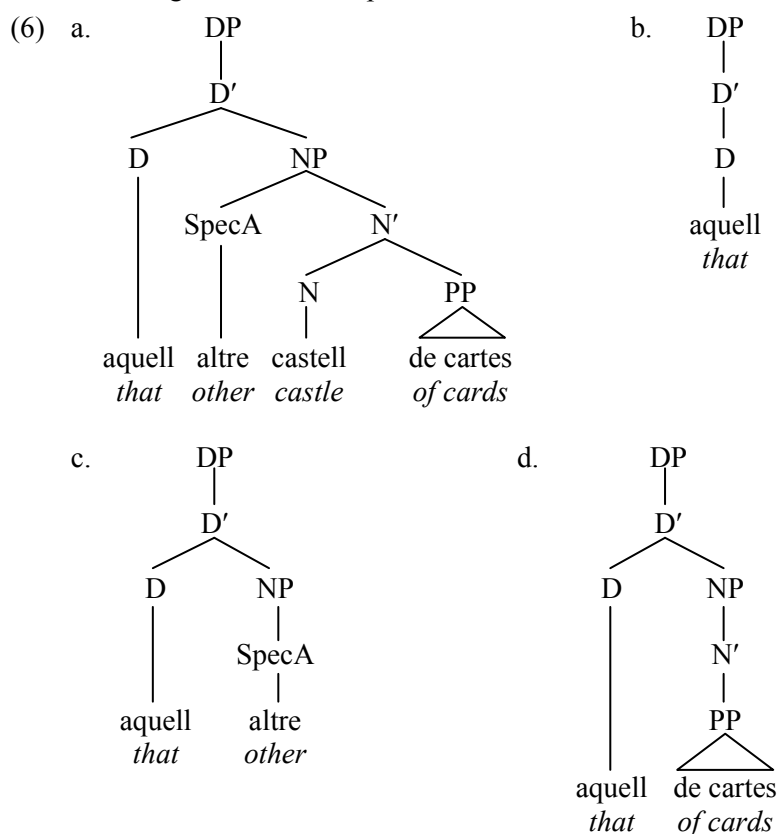
In addition, phrase structures are subject to Bresnan's (2001) Economy of Expression, which allows any category licensed by (4) to be missing:

- (5) Economy of Expression:

All c-structure nodes are optional and are only used if needed for semantic reasons or to satisfy well-formedness conditions.

According to this principle, there can be DPs that do not include an NP and NPs that do not include an N' or an N. Nonhead categories can likewise be missing. There are noun phrases without a noun, but there are no null nouns. Economy of Expression rules out the null noun: if there were a null noun, there would have to be an N node in the c-structure that would not be needed for semantic reasons or to satisfy a well-formedness condition.

Following are some examples of c-structures of DPs:



### 2.3 The f-structure of the DP

The combination of (4) and (5) correctly allows headless noun phrases, that is, DPs that either lack an NP altogether, as in (6b), or include an NP that lacks the head N, as in (6c,d). Headless noun phrases are well-formed provided the determiner belongs to the class of semipronominal Ds. But the c-structure principles given so far do not distinguish between the two classes of Ds. So, by these principles we expect to find headless noun phrases with nonpronominal Ds. Headless noun phrases such as those given in (3b–d) are ungrammatical and yet are assigned a well-formed c-structure.

The reason why the noun phrases in (3b–d) are ungrammatical is because they violate a well-formedness condition on the f-structure,

specifically, Completeness. Completeness, as stated in Bresnan (2001), has two parts: the requirement that an f-structure have all the grammatical functions specified in its local PRED feature, with which we will not be concerned here, and the requirement in (7):

(7) Completeness:

Every f-structure with a thematic role must have a PRED feature.

Words of categories N, V, and A generally have a PRED feature. Therefore, a headed NP will normally satisfy condition (7) because the head noun provides the necessary PRED feature. Ds split into two groups: some do not have this feature—the nonpronominal Ds—and some have an optional [PRED ‘pro’] feature—the semipronominal Ds.<sup>1</sup> See vocabulary entries of a D of each type:

- (8) a. *aquell*: D  $\left[ \begin{array}{l} \text{SPEC} \text{ [DEM +]} \\ \text{NUM} \text{ sg} \\ \text{GEN} \text{ masc} \\ \text{(PRED 'pro')} \end{array} \right]$
- b. *cada* D  $\left[ \begin{array}{l} \text{SPEC} \text{ [DISTR +]} \\ \text{NUM} \text{ sg} \end{array} \right]$

In the mapping between f-structure and c-structure, the following conditions apply: (a) The f-structure of a phrasal category is that of its head, and (b) the f-structure of a functional category is that of its complement phrase. Thus, the phrasal categories DP and NP map onto the same f-structure as their head D and N respectively. Given that D is a functional category (like C and I), its f-structure must be that of its NP complement. Consequently, the f-structure of a DP is the unification of the f-structure of the head D and of the N head of the complement NP.

A semipronominal D can head a well-formed DP without an NP sister or with a headless NP, as it has the PRED feature needed to satisfy Completeness. A nonpronominal D cannot do so, as the DP would lack a PRED feature and so violate Completeness. The optionality of the feature [PRED ‘pro’] in (8a) allows a semipronominal D such as *aquell* to co-occur with a headed NP complement, in which case that feature is not used, or with a headless NP complement, in which case that feature is used. The absence of that feature in (8b) forces a nonpronominal D such as *cada* to co-occur with a headed NP, as that is the only way for the DP to satisfy Completeness.

This provides the explanation for the contrast between a grammatical DP consisting only of the word *aquell* and an ungrammatical DP consisting only of the word *cada*. [<sub>DP</sub> *aquell*] maps onto a feature structure containing all the features in (8a), including the PRED feature, whereas [<sub>DP</sub> *cada*] maps onto the feature structure in (8b), which lacks this feature.

<sup>1</sup> There is a third type of D that we are not concerned with here: pronominal Ds, which always have the [PRED ‘pro’] feature, such as *algú* ‘someone’, *qui* ‘who’, *res* ‘nothing’, *ell* ‘he’, etc.

### 3 Properties of the d-article

Having outlined the structure of the DP and having established the existence of two classes of determiners in Catalan, we will see why it is reasonable to consider d-articles determiners, in 3.1, and how they differ from other determiners syntactically, in 3.2, and morphophonologically, in 3.3.

#### 3.1 The d-article as a determiner

The d-article is like a determiner, in that it is initial in the noun phrase:

- (9) a. el pas ràpid ('the fast pace')  
the(M,S) pace(M,S) fast(M,S)  
b. \*pas el ràpid / \*pas ràpid el
- (10) a. la primera veu ('the first voice')  
the(F,S) first(F,S) voice(F,S)  
b. \*primera la veu / \*primera veu la
- (11) a. les peces de ferro ('the iron parts')  
the(F,P) parts(F,P) of iron  
b. \*peces les de ferro / \*peces de ferro les

Consequently, it cannot co-occur with another determiner in the same noun phrase, as the two would be competing for initial position: *\*cada la situació* 'each the situation', *\*la cada situació* 'the each situation', *\*l'aquella situació* 'the that situation', *\*aquella la situació* 'that the situation', etc.

The d-article is like a semipronominal determiner, in that it may introduce not only a noun phrase with a head noun, as in (9)–(11), but also a noun phrase without a head noun, as in (12) (the relevant phrase in boldface):

- (12) a. D'aquestes peces, només vull **les de ferro**.  
of these(F,P) parts(F,P), only I-want the(F,P) of iron  
'Of those parts, I only want the iron ones.'
- b. La segona pàgina és més interessant  
the(F,S) second(F,S) page(F,S) is more interesting  
que **la primera**.  
than the(F,S) first(F,S)  
'The second page is more interesting than the first one.'
- c. Aquest cavall és **el que corre més**.  
this(M,S) horse(M,S) is the(M,S) that runs the most  
'This horse is the one that runs the most.'

In these examples, the d-article is followed by an NP consisting only of a PP, as in (12a), of a SpecA, as in (12b), or of a relative clause, as in (12c).

#### 3.2 Special syntactic properties of the d-article

The d-article differs from ordinary semipronominal determiners in that it must be followed by some lexical element in its NP sister:

- (13) a. Mentre esperava els amics, pensava en  
 while he-waited the(M,P) friends(M,P) he-thought in  
 els \*(de la infantesa).  
 the(M,P) of the childhood  
 ‘While he waited for his friends, he thought about {\*them/his  
 childhood friends}.’
- b. Parlant de llibres, ja he llegit el  
 talking of books(M,P) already I-have read the(M,S)  
 \*(que recomanes).  
 that you recommend  
 ‘Talking about books, I have already read the one \*(that you  
 recommend).’

This indicates that the d-article needs a host: there must be some element in the NP it introduces that it can attach to. At this point, we could say: the d-article is a semipronominal determiner and what is special about it is that it cliticizes onto its NP sister (see Brucart and Gràcia 1984 for a similar claim). In other words, there has to be some lexical item in the NP sister of the d-article that the d-article can depend on phonologically.

However, not just anything in the NP serves this purpose. It is useful to distinguish between the *prenominal sector* of the NP—everything in the NP that precedes the head noun (or would precede it if it is absent)—and the *postnominal sector*—everything in the NP that follows the head noun (or would follow it if it is absent). The d-article can attach to the head noun, as in (9a) and (11a), and to anything in the prenominal sector: to a SpecA, whether the head noun is present, as in (10a), or not, as in (12b); to a prenominal adjective, or PreN-A, as in (14); to a modifier of a following adjective or noun, which can be an intensifier or an adverbial phrase, as in (15a–c), a PP headed by any preposition,<sup>2</sup> as in (15d–e), and even a clause, as in (15f).

- (14) a. l’ antiga fàbrica (‘the old factory’)  
 the old(F,S) factory(F,S)
- b. el pobre gos (‘the poor dog’)  
 the poor(M,S) dog(M,S)
- (15) a. la molt celebrada resposta del ministre  
 the very celebrated reply of-the minister
- b. els increïblement intensos minuts de concert  
 the incredibly intense minutes of concert
- c. l’ encara president del govern  
 the as of yet president of-the government

<sup>2</sup> The importance of this class of examples was brought to my attention by Martin Forst, who culled many examples of d-articles followed by PPs from the web. The grammatical instances of PPs other than *de*-PPs were prenominal (see below for discussion of postnominal PPs).



- d. el, des del mes passat, millor cuiner del món  
the from the month past best cook of-the world
- e. el {sens dubte / per descomptat} principal factor econòmic  
the without doubt/ of course main factor economic
- f. la, que jo sàpiga, única solució al problema<sup>3</sup>  
the that I know only solution to-the problem

There does not seem to be anything in the prenominal sector that the d-article cannot attach to. Crucially, in all of the cases illustrated in (15), the word or phrase that immediately follows the d-article is a modifier of a following word or phrase, which may be an adjective, a noun, or an N'.

We consider next what can follow the d-article in the postnominal sector. It may seem that it may be followed by any phrase that can appear postnominally: a PP, as seen in (12a) and (13a), a clause (or CP), as shown in (12c) and (13b), or an AP of the postnominal variety. PPs and CPs such as those of (16) and (17) are always postnominal:

- (16) a. la filla dels veïns  
the daughter of-the neighbors  
'the neighbors' daughter'
- b. \* la dels veïns filla
- (17) a. el llibre que et vaig mostrar  
the book that you I-PST show  
'the book that I showed you'
- b. \* el que et vaig mostrar llibre

APs, as modifiers of a noun, can appear prenominally or postnominally. Some adjectives can appear in either position without a difference in meaning. However, some can only appear in one of the two positions or have a different meaning depending on the position they appear in. For example, *pobre* in prenominal position, as in (14b), does not mean 'lacking in financial resources', as it does in postnominal position, but signals an object of pity on the part of the speaker. The adjectives *francès/francesa* 'French' or *industrial* are postnominal adjectives (PostN-A) and can only appear postnominally:

- (18) a. la muntanya francesa  
the mountain(F,S) French(F,S)  
'the French mountain'
- b. \* la francesa muntanya
- (19) a. l' activitat industrial  
the activity(F,S) industrial(S)  
'industrial activity'

---

<sup>3</sup> When the modifier to which the d-article is adjacent is of some structural complexity, as in (15d,f), it is generally written between commas.

b. \* la industrial activitat

The fact that the d-article can immediately precede these adjectives, as in (20), indicates that the d-article can be adjacent to a postnominal AP:

- (20) a. M' agrada tant la muntanya italiana com **la francesa**.  
 me likes as much the mountain Italian as the French  
 'I like the Italian mountain as much as the French one.'
- b. L' activitat comercial ha superat **la industrial**.  
 the activity comercial has exceeded the industrial  
 'Comercial activity has exceeded industrial activity.'

When we say that the d-article may be adjacent to a *postnominal* phrase (PP, CP, or AP), it does not mean that the phrase follows a noun. It only means that it is a phrase of the kind that follows a noun if a noun is present. The phrase is licensed by rule (4b): this rule licenses a maximal projection following a head X or X'. Since both nodes are optional by Economy of Expression, the rule may license only the nonhead constituent, which is the case in the boldfaced sequences in (20).

Whereas the d-article may be adjacent to any AP, there are significant restrictions on the postnominal PPs and CPs that may follow the d-article. A postnominal PP must be headed by the preposition *de*, if adjacent to the d-article (see Martí (2002: 1286); see also Lobeck (2006: 159) concerning the same restriction in Spanish).<sup>4</sup> Other semipronominal Ds can be followed by the same range of PPs as nouns, but not the d-article:

- (21) No vull aquelles sabates, sinó  
 not I-want those(F,P) shoes(F,P), but  
 aquestes / \*les amb cordons blancs.  
 these(F,P) / \*the(F,P) with laces white
- (22) a. L' abric / aquest / el de pell sense caputxa  
 the coat / this /the of leather without hood

---

<sup>4</sup> An anonymous reviewer indicated apparent counterexamples to this generalization:

- (i) a. els sense sostre (lit. *the without ceiling*) 'the homeless'  
 b. el sense nom (lit. *the without name*) 'the nameless one'  
 c. els sense feina (lit. *the without job*) 'the jobless'

The sequences like *sense sostre* in these examples have to be analyzed as lexical units of category N. Unlike true headless NPs, the NPs in (i) are not interpreted as if there were a missing N whose semantics is filled by identity with some N in the discourse. Those sequences can be preceded by PreN-As like *mer* 'mere' or *antic* 'former', which is only possible with Ns: *un mer sense sostre* 'a mere homeless person'. They cannot be altered in any way: *\*els sense aquest nom* 'those without this name', *\*els sense cap sostre* 'those without any ceiling', etc. A few other PPs not headed by *de* are marginally acceptable after the d-article (e.g., *?els en vies de desenvolupament* (lit. *the in track of development*), where *en vies de desenvolupament* = 'developing'), but they have to be fixed phrases that are plausibly reanalyzed as single words, such as an adjective.

- b. L'abric / aquest / \*el sense caputxa de pell  
 the coat / this / \*the without hood of leather  
 (Martí 2002: 1286)

The d-article can be immediately followed by a clause, only if the clause is introduced by *que* (see (12c)),<sup>5</sup> contrasting with other semipronominal Ds or with the d-article adjacent to a SpecA, which can be followed by the same range of relative clauses as nouns, as shown in (23):

- (23) a. aquell/ \*el de qui et vaig parlar  
 that(M,S)/ \*the(M,S) of whom you I-PST speak  
 'that/\*the one I spoke to you about'
- b. aquests/ \*els en els quals ens trobem  
 these(M,P)/ \*the(M,P) in the which we find ourselves  
 'those/\*the ones in which we find ourselves'
- c. la \*(primera) on vaig viure  
 the(F,S) \*(first(F,S)) where I-PST live  
 'the \*(first) one where I lived'
- d. l'altre / \*el per on passarem  
 the other / \*the through where we will pass  
 'the \*(other) one that we will pass through'

### 3.3 Special morphophonological properties of the d-article

There are some morphophonological properties that make the d-article unique among determiners:

(1) The d-article is stressless, unlike most other Ds.<sup>6</sup> Evidence for this is found in the schwa in these forms (*el* [əl], *la* [lə], *els* [əls], *les* [ləs]) and in that they form a phonological word with the following word, as shown by the homophony of single words with d-article-word combinations:

- (24) a. [láβil]: *làbil* 'labile' / *l'hàbil* 'the skilled one'  
 b. [ləkáβə]: *lacava* 'lacquer (past)' / *la cava* 'the cellar'  
 c. [ləmén]: *lament* 'lament' / *la ment* 'the mind'  
 d. [lətén]: *latent* 'latent' / *l'atent* 'the attentive one'

(2) There is a syllabic/asyllabic alternation: the masculine singular form is *el* [əl] before a consonant and *l'* [l] before a vowel: *el noi* / \**l' noi* 'the boy', *l'amic* / \**el amic* 'the friend'. There is a similar alternation in the feminine singular form (*la/l'*); however, whereas this alternation could be accounted

<sup>5</sup> Some dialects and registers also allow a relative clause immediately following a d-article to begin with *qui*, signalling a human referent (e.g., *els qui vindran* 'those who will come'). I will ignore this option here for ease of exposition.

<sup>6</sup> The only other Ds that are stressless are the so-called weak possessives—*mon/ma* 'my', *ton/ta* 'your', *son/sa* 'his/her/its'—which are nonpronominal determiners and very restricted in their use.

for by a regular phonological rule of deletion of schwa adjacent to another vowel, no general rule can account for the *el/l'* alternation.

(3) The d-article patterns with verbal clitics morphophonologically. The syllabic/asyllabic alternation just noted is only found in so-called clitics, such as *em/m'* (1st p.sg.), *et/t'* (2nd p.sg.), *es/s'* (3rd refl.), *el/l'* (3rd p.masc.sg): *em veus* 'you see me', *m'ajudes* 'you help me', *el sentim* 'we hear him', *l'escoltem* 'we listen to him', etc. The d-article is homophonous with the 3rd person clitic, giving rise to VP/NP ambiguities such as: *la veu* 'the voice' / '(s)he sees her', *les cordes* 'the ropes' / 'you tie them', *el viu* 'the live one' / '(s)he lives it', etc.

### 3.4 Summary of properties of the d-article

The d-article:

- is initial in the noun phrase, like a D;
- may occur without the head noun, like a semipronominal D;
- must precede some element in its noun phrase, unlike a semipronominal D (i.e., it needs a host);
- may be adjacent to any element in the NP; however,
  - if adjacent to a postnominal PP, this PP must be headed by *de*;
  - if adjacent to a postnominal clause, it must be a *que*-clause;
- is morphophonologically unlike other Ds:
  - it is stressless: its vowel is schwa, if any, and it forms a phonological word with the following word;
  - it shows a syllabic/asyllabic alternation, only found in verbal clitics.

## 4 Explaining these properties

Simply assuming the d-article is a D, of the semipronominal kind, would explain some facts, but would leave the special properties of the d-article unexplained: the need for a host, the syntactic restrictions on the host, and the morphophonological properties. We will adopt the idea defended by Miller 1992 for French that the d-article is a bound morpheme, forming a word with the following word.<sup>7</sup> This is the treatment also given to verbal clitics in French by Miller 1992 and Miller and Sag 1997 and widely adopted for the other Romance languages. This analysis provides a way to explain those special properties of the d-article. If we assume that the d-article is the same type of morphological element as verbal clitics, it follows that they should all exhibit the same syllabic/asyllabic alternation, that they are all stressless and form a phonological word with the following word.

---

<sup>7</sup> I prefer to think of the d-article as a member of a compound, rather than an affix. The fact that it can take inflectional affixes itself (feminine and plural morphs: e.g., *les*: [[l][a][z]]) is inconsistent with its being an affix.

However, what is the syntactic category of the compound containing the d-article (the d-compound)? Is it a D, ignoring the category of the host, or is it the category of the host, obliterating the D? Both alternatives face serious problems or complications.

The analysis of the d-compound as a D explains the fact that it is initial in the DP, but does not explain: (a) why, if the host in the d-compound is a noun, it cannot be followed by a SpecA like *altre* ‘other’: \**el-gat altre* ‘the cat other’; (b) why, if the host is a PreN-A like *bon* ‘good’, it needs to be followed by a noun: \**el-bon* ‘the good’, cf. *el bon amic* ‘the good friend’, \**l’amic bon*; (c) why, if the host is a P, an NP must follow: *el de \*(ferro)* ‘the iron one’; (d) why, if the host is a C, a clause must follow: *el que \*(surt)* ‘the one that is coming out’.

On the other hand, the analysis that assigns the d-compound the category of the host does not have these problems, but cannot explain (without additional machinery) why the d-compound cannot occur with a D in the same DP and must be initial in the noun phrase. If the d-compound had the category of the host (whether N, PreN-A, PostN-A, etc.), it should be possible for it to occur with a D (\**aquell l’amic* ‘that the friend’, \**cada l’hora* ‘each the hour’, etc.) and for it to be preceded by a word that can precede the host (\**bon l’amic* ‘good the friend’, \**altre el gat* ‘other the cat’, etc.). Miller 1992 adopts a version of this analysis and enriches the theory with an EDGE feature and with principles, including Linear Precedence constraints that linearize constituents with the appropriate EDGE feature such as the d-compound to the left of other constituents (see also Tseng 2003).

Both analyses fail to recognize that the categories of both components of the d-compound are needed to account for its distribution. If the d-compound has the category of the d-article (D), it follows that it cannot co-occur with another D (only one D per DP) and cannot be preceded by a SpecA or any other category in the NP (the D precedes its sister NP). If it has the category of the second member of the compound, it has to appear in the position reserved for this category: If this category is N, it cannot be followed by a SpecA like *altre* (SpecA precedes N); if it is a PreN-A like *bon*, it must be followed by a noun; if it is the preposition *de*, it must be followed by a noun phrase; if it is the complementizer *que*, it must be followed by a tensed clause. The conclusion is that the d-compound has the categories of both components of the compound.

This is possible only if we allow two different  $X^0$  categories in the c-structure to be instantiated by the same word. The theory of Lexical Sharing (Wescoat 2002, 2005, 2007) allows this possibility. All we need is to allow a single word to fill (or map onto) two or more  $X^0$  nodes in the c-structure. If we assume that words are not part of the c-structure, but are part of a separate level—the word string (or lexical-structure, in Wescoat’s terms)—where words are linearly ordered, there is a mapping, or correspondence, between terminal nodes ( $X^0$  nodes) in the c-structure and words in the word string.

This mapping, by which all terminal nodes are instantiated by words and all words instantiate terminal nodes, is constrained by Wescoat's Order Preservation Axiom (Wescoat 2007: 444–445):

(25) Order Preservation Axiom:

For all c-structure nodes X and Y, if the word that instantiates X precedes the word that instantiates Y, then X precedes Y.

By this constraint, words with multiple categories can only appear in structures where the various linked categories are adjacent.

The lexical entry of all words includes a list of categories, which, in the simplest case, consists of only one member. The order of the categories in this list matches the order of the  $X^0$  nodes in the c-structure that the word maps onto. The d-article, as a lexical item below the word level, carries information about morphology, c-structure, and f-structure.

(26) Lexical information of the d-article:

**Morphology:** the d-article is a bound stem and attaches to a word to its right to form a word (the d-compound).

**C-structure:** the list of categories of the d-compound is the concatenation of the category D (for the d-article) and the list of categories of the 2<sup>nd</sup> member of the compound.

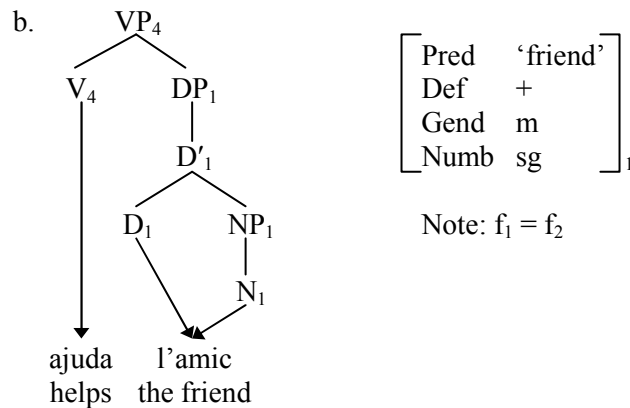
**F-structure:** the f-structure of the 2<sup>nd</sup> member is included in the f-structure of the first member (where f-structure A includes f-structure B iff  $B = A$  or is contained in A).

**Restriction:** if the category of the 2<sup>nd</sup> member is dominated by a PP or CP adjunct of the d-article, the 2<sup>nd</sup> member is *de* or *que* respectively.

The morphological status of the d-article allows us to explain the article's properties in 3.3. The c-structure information in (26) provides the explanation for the distribution of the d-compound: it depends on the category of both members of the compound. The f-structure information captures the observation that the d-article needs a host that belongs in the same DP. The restriction in (26) accounts for the observation that, if the d-article is immediately followed by a PP adjunct of the noun phrase, the head of this PP is the preposition *de*, and if the d-article is immediately followed by a CP adjunct of the noun phrase, this CP begins with the complementizer *que*.

As an illustration of this theory, consider the d-compound *l'amic*. (27a) shows the vocabulary entry of this compound, with a list of categories consisting of a D and an N and the f-structures corresponding to each category, with indices signaling the correspondence. (27b) shows the c-structure and f-structure of the sentence *ajuda l'amic*, where arrows, rather than lines, represent the correspondence between c-structure and the word string.

(27) a. l'amic:  $\langle D_1, N_2 \rangle$   $\left[ \begin{array}{l} \text{Def} \quad + \\ \text{Gend} \quad m \\ \text{Numb} \quad \text{sg} \end{array} \right]_1$   $\left[ \begin{array}{l} \text{Pred} \quad \text{'friend'} \\ \text{Gend} \quad m \\ \text{Numb} \quad \text{sg} \end{array} \right]_2$



Notice that these structures satisfy all the constraints proposed, specifically: the d-article is the first part of a word that has the two categories of its component elements in a list, and the f-structure of the second member is included in (here, is equal to) the f-structure of the first member. The restriction in (26) applies vacuously as there is no PP or CP adjunct involved.

The morphological and f-structure requirements in (26) ensure that the d-article is followed by some element, explaining contrasts such as those in (13), and that this following element is part of the noun phrase introduced by the d-article, ruling out examples such as the following:

- (28) a. \* Pensava en  $[\text{DP} \text{ els } ]$  sovint.  
           he-thought in    the(M,P)    often
- b. \* Consideren  $[\text{DP} \text{ la } ]$  una bona amiga.  
           they-consider    the(F,S) a(F,S) good(F,S) friend(F,S)

The restriction in (26) explains the contrasts illustrated in (21) and (22): if the second member of the d-compound heads a PP adjunct of the d-article, it must be the preposition *de*, accounting for the contrast between *les de ferro* (*the of iron*) and *\*les amb cordons* (*the with laces*). This condition does not prevent the d-article from attaching to a preposition other than *de*, provided this P heads a PP that is not an adjunct of the d-article. This is what we see in (15d,e), where the P adjacent to the d-article heads an adjunct of the following adjective or noun. That restriction also explains the contrasts in (23): in these examples, the second member of the compound is part of a CP adjunct of the d-article, but it is not *que*, and therefore they are ruled out.<sup>8</sup>

<sup>8</sup> The fact that the restriction in (26) is a separate statement from the other statements about the d-article is not a negative feature of the analysis, but a positive one, as there are languages that have a d-article as a lexical sharing bound morpheme, but lack this restriction or have a different one. Spanish, from the 15th to the 17th centuries, allowed constructions very much

Whereas a d-article can be followed by a restrictive relative clause beginning with *que*, it cannot be followed by a nonrestrictive relative clause, even if it begins with *que*:

- (29) M'agrada aquell./ \*el, que em vas ensenyar.  
 I-like that(M,S)/ \*the(M,S) that me you-PST show  
 'I like that/\*the, which you showed me.'

A nonrestrictive relative clause is distinguished from the word that immediately precedes it by an intonational break. The demonstrative *aquest* in (29) can be followed by a nonrestrictive relative clause and an intonational break separates *aquest* from *que*, which begins the relative clause. The d-article has to form a compound with the following word, but if this is the first element of a nonrestrictive relative clause, there would have to be an intonational break inside a word. An intonational break can separate a word from another word, but not a word part from another word part.<sup>9</sup> This is an additional argument for the claim that the d-article forms a word with a following word.

## 5 Weak prepositions and multiple lexical sharing

This section argues that weak prepositions—*per*, *de*, *en*, *a*, *amb*—are also lexical sharing bound morphemes. After presenting evidence for this claim, we will see that the preposition *de* and the d-article can alternate in indefinitely long sequences. The implications of these structures will be considered for choosing among alternative theories of the d-article.

Evidence that weak prepositions in Catalan are bound morphemes includes the following observations:

(1) Weak prepositions need a host. Other prepositions can function without a following NP, i.e. they can be the sole element in the PP. This is the case of *sense* 'without', *sota* 'under', *contra* 'against', among others: *Vols el te amb sucre o sense?* 'Do you want your tea with sugar or without?'; cf. *\*Vols el te sense sucre o amb?* 'Do you want your tea without sugar or with?'

(2) The choice between the prepositions *a* and *en* is partly conditioned by phonology: e.g., *he caigut {en un forat / al forat / en aquell forat / a dins del forat}* 'I have fallen {in a hole / in the hole / in that hole / inside the hole}'. The rule governing this alternation can be stated as: use *en* if followed by a

---

like those with the d-article in (23). A search in CORDE—the historical corpus of Spanish texts—reveals many examples from that period such as *la por quien muero* (lit. the for whom I die) 'the one I die for'. (The modern Spanish corpus CRAE, on the other hand, has no example like these.) To deal with different grammars, the restriction in (26) can be redefined or eliminated altogether, without affecting the remaining statements about the d-article.

<sup>9</sup> The intonational break that occurs at the boundary of an appositional or restrictive clause is completely different from the intonation we find in expletive insertion cases (e.g., *fan-fucking-tastic*) or in coordination of prefixes in English and German (e.g., *pro- and anti-abortion activists*; *ein- und aussteigen* 'to get on and off' from Forst, King, and Laczko 2010).



vowel; otherwise, use *a*. By the Principle of Phonology-Free Syntax (Zwicky 1969; Zwicky and Pullum 1986; Miller, Pullum, and Zwicky 1997), this cannot be an instance of a phonologically conditioned choice of lexical items. But, according to this principle, it can be an instance of phonologically conditioned allomorphy or prefix suppletion.

(3) The prepositions *de*, *per*, and *a* alternate with contracted forms involving the d-article: *del*, *dels*; *pel*, *pels*; *al*, *als*. The contracted forms are used when the d-article morph /l/ is not in a syllable onset; otherwise, the contracted forms are used: *per la noia* ‘by the girl’, *per l’amic* ‘by the friend’, *pel noi* ‘by the boy’, *pels amics* ‘by the friends’.<sup>10</sup>

(4) Certain phonological properties of these prepositions also argue for their status as bound morphemes:

- a) They are stressless: their vowel is schwa: [pər], [də], [ən], [ə], [əm].
- b) *Amb* ‘with’ is not pronounced [əm] in all contexts (which would be expected if it were a word, given the rule of word-final stop deletion after a homorganic consonant), but is pronounced [əmb] when followed by a vowel: [əm lə má] ‘with the hand’, [əmb əl dí] ‘with the finger’.
- c) *Per* ‘by, for’ triggers the deletion of the following vowel if it is part of a demonstrative: *per (a)quest camí* ‘by this path’, *per (a)quella raó* ‘for that reason’, *per (ai)xò* ‘because of this’, *per (a)llà* ‘that way’ (lit. ‘by there’), etc. This vowel does not drop in other contexts. And *per* does not cause the dropping of a following schwa in other cases. Again, it is a case of allomorphy or a morphologically conditioned phonological rule.

These facts support the claim that weak prepositions are bound morphemes.<sup>11</sup> The reasons for assuming that there is a P in the c-structure corresponding to the word they are part of include the following: (a) a phrase beginning with a weak preposition occupies the same positions as a phrase beginning with an independent word of category P; (b) the c-structure rule stating that a P is initial in the PP applies both to independent words of category P and to words containing a weak preposition, requiring them to be initial in their PP; (c) a weak preposition can scope over a coordinate NP, just as an independent P

<sup>10</sup> Lexical sharing arises not only in the contracted forms (as assumed in Wescoat (2007)), but in all occurrences of weak Ps. The choice between the contracted and uncontracted forms does not depend on syntactic conditions, but on phonological conditions. Assuming phonology-free syntax, the alternation cannot be a choice between different words. If weak Ps are always lexical sharing prefixes, the alternation is phonologically conditioned allomorphy. (Similar facts occur in French: *à l’homme* ‘to the man’ / *au bon homme* ‘to the good man’.)

<sup>11</sup> Wescoat (personal communication) notes that the bound morpheme status of weak prepositions can be traced back to Latin, at least. The Latin so-called clitic *-que* ‘and’ attaches to the end of the first word of a conjunct, but monosyllabic prepositions frequently do not count for the purpose of *-que* placement (see Gildersleeve and Lodge 1895: 300), as the following examples (from Carlson 1983: 81) illustrate: *in forōque* ‘and in the forum’, *ob eāsque rēs* ‘and because of these achievements’. The reason prepositions such as *in* and *ob* do not count as the first word of the conjunct is that they are not separate words, but part of a word together with the following stem.



- (33) el de París va arribar ahir  
 the(M,P) of Paris 3sg-PST arrive yesterday  
 ‘the one from Paris arrived yesterday’

Whereas Catalan and Spanish allow the sequence d-article+*de*, (e.g., (30)), French and Italian do not (e.g., (32)). By assuming the structure in (32b), with a null noun, Miller predicts the illformedness of this sequence in all four languages. In the present approach, where there are no null nouns, all we need to assume to explain the contrast among these languages are different constraints on the host of the d-article:

- Spanish and Catalan allow the d-article to form a compound with a following preposition *de*.
- Italian and French do not allow the d-article to attach to any preposition.<sup>12</sup>

Such restrictions in word formation are to be expected and are expected to vary from language to language.

The presence of the null noun in headless NPs in Miller 1992 makes it hard to explain the cross-linguistic variation noted. It would be very strange to assume that some languages allow inflection on a null stem, whereas others do not. Even if we made this assumption, we would still need to explain that there are restrictions on what can follow the null stem when it is inflected with the d-article: in Spanish and Catalan, it must be followed by some element in the same NP; if followed by a P that heads an adjunct of the d-article, it must be *de*. So, we would need to make two unnatural assumptions: that the null stem can have inflections in some languages and that the inflections impose restrictions on what follows the null stem. This supports the conclusion that both the d-article and weak prepositions should be treated as lexical sharing bound morphemes, rather than as inflections.

## 6 Listed compounds

D-compounds are for the most part unlisted: they are formed on-line, satisfying the constraints in the lexical entry of the d-article in (26), and are not listed in the lexicon or in a repository of words. However, since d-compounds are words, the possibility exists that certain d-compounds are listed. Certain words can only function as the second component of a d-compound. The clearest case is *qual*, which can only be used immediately following the d-article forming the “compound relative”: *el qual*, *la qual*, *els quals*, *les quals*. Other combinations involving *qual* are impossible: *\*un qual*, *\*aquell qual*, *\*la primera qual opció* (cf. *la qual primera opció* ‘which first option’), *\*amb qual*, etc. All we need to assume is that *qual* is a SpecA (like

<sup>12</sup> This constraint can probably be subsumed under a more general statement that would also exclude other combinations of the Italian and French d-article with postnominal elements, such as with a relative clause (e.g., French *\*le qui parle* (lit. *the who speaks*) vs. *celui qui parle* ‘he who speaks’).

*altre* ‘other’, *primer* ‘first’, etc.) that is lexically listed as forming a compound with the d-article.<sup>13</sup>

Another relevant case is the group of so-called strong possessives: *meu/meva* (1st pers. sg.), *teu/teva* (2nd pers. sg.), *seu/seva* (3rd pers.), etc. They are used either as SpecA or as PostN-A. (The use as postnominal adjectives, PostN-A, is found in noun phrases such as *amic meu* ‘my friend’, *molts coneguts teus* ‘many acquaintances of yours’, and (34c).) However, for many speakers, whose judgments are reported here, they can only be used as SpecA if immediately preceded by the d-article:

- (34) a. el meu amic ‘my friend’  
           the(M,S) my(M,S) friend(M,S)
- b. \*aquell meu amic ‘that friend of mine’  
       that(M,S) my(M,S) friend(M,S)
- c. aquell amic meu ‘that friend of mine’  
       that(M,S) friend(M,S) my(M,S)
- d. els meus dos gossos ‘my two dogs’  
       the(M,P) my(M,P) two dogs(M,P)
- e. \*els dos meus gossos ‘my two dogs’  
       the(M,P) two my(M,P) dogs(M,P)

Strong possessives have a dual classification as SpecA and PostN-A. But, as SpecA, they are the second member of a d-compound (i.e., must form a compound with the d-article). The claim that the d-article forms a compound with the following word affords a simple and natural way to account for the marked distribution of certain words: relative *qual* and strong possessives when used as SpecA are listed as the second member of a d-compound.

## 7 Conclusions

In this paper we have seen compelling evidence for the claims: (a) that the d-article in Catalan (as in closely related languages) is a bound morpheme that attaches to a word to its right to form a word (the d-compound); (b) that the d-compound instantiates two or more terminal nodes in the c-structure corresponding to the categories contributed by its component elements: an instance of lexical sharing; (c) that weak prepositions in Catalan are also bound morphemes of the lexical sharing type; and (d) that the number of categories that a word can instantiate can be greater than two and has no upper bound.

Lexical sharing, i.e. the hypothesis that a word maps onto more than one terminal node in the c-structure, proves to be the correct way of accounting for what has sometimes been known as phrasal affixes (bound morphemes whose distribution depends on phrasal information). Alternative

<sup>13</sup> Notice that the cognate form in French is spelled as a single word together with the d-article: *lequel, laquelle, duquel, auxquelles, desquels*, etc.

approaches that do not allow lexical sharing account for the same (or similar) facts by complicating the theory considerably or, simply, cannot account for the relevant facts (as argued about Miller's (1992) approach). Many phenomena that seem to straddle the morphology-syntax boundary would benefit from an analysis that incorporates lexical sharing.

## References

- Abney, Steven. 1987. The English noun phrase in its sentential aspect. Ph.D. dissertation, MIT.
- Bresnan, Joan. 2001. *Lexical-functional syntax*. Oxford: Blackwell.
- Brucart, José and Lluïsa Gràcia. 1984. I sintagmi nominali senza testa. *Rivista di Grammatica Generativa* 11: 3–32.
- Carlson, Greg N. 1983. Marking constituents. Frank Heny and Barry Richards (eds.), *Linguistic categories: Auxiliaries and related puzzles*, 69–98. Dordrecht: Reidel.
- Forst, Martin, Tracy Holloway King, and Tibor Laczko. 2010. Particle verbs in computational LFGs: Issues from English, German, and Hungarian. *Proceedings of LFG'10*. CSLI Publications.
- Gildersleeve, Basil L. and Gonzalez Lodge. 1895. *Gildersleeve's Latin Grammar*. London: Macmillan.
- Lobeck, Anne. 2006. Ellipsis in DP. Martin Everaert and Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, vol. 2, 145–173. Malden, MA: Blackwell.
- Martí i Girbau, Núria. 2002. El SN: els noms. Joan Solà et al. (eds.), *Gramàtica del català contemporani*, vol. 2, 1281–1339. Barcelona: Empúries.
- Miller, Philip H. 1992. *Clitics and constituents in phrase structure grammar*. New York: Garland.
- Miller, Philip H, Geoffrey K. Pullum, and Arnold M. Zwicky. 1997. The Principle of Phonology-Free Syntax: four apparent counterexamples in French. *Journal of Linguistics* 33:67–90.
- Miller, Philip H. and Ivan A. Sag. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* 15:573–639.
- Tseng, Jesse. 2003. Phrasal affixes and French morphosyntax. *Proceedings of Formal Grammar 2003*, 177–188. Technische Universität Wien.
- Wescoat, Michael T. 2002. On Lexical Sharing. Ph.D. dissertation, Stanford Univ.
- Wescoat, Michael T. 2005. English nonsyllabic auxiliary contractions: An analysis in LFG with lexical sharing. *Proceedings of LFG'05*, 468–486. CSLI Publications.
- Wescoat, Michael T. 2007. Preposition-determiner contractions: An analysis in Optimality-Theoretic Lexical-Functional Grammar with lexical sharing. *Proceedings of LFG'07*, 439–459. CSLI Publications.
- Zwicky, Arnold M. 1969. Phonological constraints in syntactic descriptions. *Papers in Linguistics* 1:411–453.
- Zwicky, Arnold M. and Geoffrey K. Pullum. 1986. The Principle of Phonology-Free Syntax: Introductory remarks. *Working Papers in Linguistics* 32:63–91. Columbus, OH: the Ohio State University.