The Catalan definite article as lexical sharing

Alex Alsina
Pompeu Fabra University

The goal of this paper is to explain a set of puzzles posed by the Catalan definite article (d-article). The main claim of the analysis is (a) that the Catalan d-article is not itself a word, but a word part, a bound morpheme that forms a compound word with a following X^0 and (b) that the derived word has the grammatical category of both component elements: D (determiner), for the d-article, and the category (N, A, C, or P) of the second element. It is thus an instance of lexical sharing (Wescoat 2002, 2005, 2007), whereby a single word instantiates two (or more) X^0 nodes in the c-structure.

The d-article (*el, la, l'*, etc) has the same broad distribution as Ds such as *aquest* 'this': they are both initial in the NP, (1a), and may immediately precede a variety of categories in the NP, including Ns, postnominal As, PPs, and relative clauses, (1b):

(1) a. *el/aquest noi* 'the/this boy', **noi el/aquest* 'boy the/this'

   b. *el/aquest (noi) dolent* 'the/this (boy) bad', *el/aquest que veus* 'the/this that you-see'

However, there are several distributional facts that distinguish the d-article from other Ds such as *aquest*. The d-article, unlike *aquest*, (i) requires a host (i.e., cannot be the sole element of its NP), (2a); (ii) if immediately followed by a PP, this PP must be headed by *de* 'of', (2b); (iii) if immediately followed by a relative clause, this clause must be introduced by *que* 'that', (2c) (cf. (1b)):

(2) a. **Veus el.* 'you-see the' vs. *Veus aquest.* 'you-see this'

   b. *el *(de pell) sense caputxa* 'the (leather) one without a hood'   (Martí 2002: 1286)
       aquest (de pell) sense caputxa 'this (leather) one without a hood'

   c. *{aquests/*els} dels quals t'he parlat* '{those/*the} of which I have spoken to you'
       la *(primera) on vaig viure 'the *(first one) where I lived'

It has other features not found in other Ds: A) it is stressless, as evidenced by the schwa in these forms (əl, lə), and forms a phonological word with the following word (*el nus* [əlnús] 'the knot', *la cua* [ləkúə] 'the tail'); B) the masculine sg. form exhibits the following alternation: *el* before stems beginning by a consonant and *l'* before stems beginning by a vowel, (*el noi/*l'noi* 'the boy' vs. *l'amic/*el amic* 'the friend'). This alternation cannot be handled by a general rule, but is restricted to a set of elements consisting of the d-article and verbal clitics, such as *em/m'* (1st p.sg.), *et/t'* (2nd p.sg.), *es/s'* (3rd refl.). In addition, the d-article is homophonous with the 3rd person verbal clitic, giving rise to NP/VP ambiguities such as *la veu* 'the voice' and '(s)he sees her'.

These phonological and morphological facts strongly suggest that the d-article should be given an analysis analogous to that of verbal clitics, which have been argued to be bound morphemes in most of the Romance languages (Miller 1992 and Miller and Sag 1997 for French, and others for Spanish, Italian, and Catalan). By extending this analysis to the d-article (also defended for French by Miller 1992:274–279), we can provide a natural explanation for the *el/l'* alternation, the fact that it is phonologically integrated in the following word, and the fact in (2a) that it needs a host.

Before adopting the lexical sharing approach to the d-article-host compound, we explore the more standard position that it has only one category: either (A) that of the host or (B) that of the d-article. However, both alternatives either fail to explain some facts or do so by complicating the description considerably. Option A fails to explain, without additional machinery, why the article-host compound, whose category is that of the host—spec-A (specificational A or postdeterminer), preN-A (prenominal A), postN-A (postnominal A), N, P, etc.—cannot occur with a D in the same NP and must be initial in the NP. Alternative B, on the other hand, does not face these problems, but it does not explain (i) why, if the host in the article-host compound is an N, the compound D cannot be followed by a spec-A (**el gat altre* 'the cat other'); (ii) why, if the host is a preN-A, the compound D needs to be followed by a noun (**el bon* 'the good', cf. *el bon amic* 'the good friend'); (iii) why, if the host is a preposition, an NP must follow (**el de* 'the of' cf. *el de la pau* 'that of the peace'), or (iv) why, if the host is a C, a clause must follow (**el que* 'the that', cf. *el que vols* 'the one you want). (Miller 1992 and Tseng 2003, among others, adopt alternative A, but, in order to make it work, introduce a set of features (EDGE features) and principles (specific LP constraints and others) that are otherwise not needed and

view it as accidental that the article-host compound appears in the position that corresponds to a D).

The problems just noted do not arise in the approach taken here, where the compound containing the d-article instantiates both a D and the X^0 category corresponding to the host (N, A, P, or C). And the facts are explained by appealing to independently needed principles. On the assumption that there is no more than one D per noun phrase, which follows from the DP hypothesis, and that the head precedes its complements, we explain that, since the article-host compound instantiates the D, there cannot be another D in the same DP and it precedes its NP complement and all words that correspond to it. Problems (i)–(iv) are explained because the article-host compound has the c-structure distribution of both the D and the category of the host: if the host is an N, it cannot be followed by a spec-A, because spec-A occupy the Spec of N and, therefore, precede the N, etc.

The restrictions illustrated in (2b,c) can be explained by imposing a constraint on the grammatical category of the host, which is to be expected in morphological operations. We assume, as a general constraint, that the host that the d-article combines with is of category N or A (or A-spec—specifiers of adjectives) mapping onto the same f-structure; this gives combinations like *el-noi, l'altre, el-bon*, etc. In addition, there are two lexically listed compounds whose first member is the d-article: d-article+*de* (such as *el de, la de*, etc.) and d-article+*que* (such as *el que, la que*, etc.); the second member of these compounds maps onto an adjunct of the f-structure of the d-article. Combinations involving the d-article that are not licensed by the general constraint or lexically listed are ill-formed, such as those shown in (2b,c). (See Brucart and Gràcia 1984 for a different explanation for these facts.)

Lexically listing compounds whose first member is the d-article allows us to explain certain combinatorial restrictions. The relative pronoun *qual* can only occur immediately preceded by the d-article, as in *el/la qual, els/les quals*. Assuming that *qual* is a spec-A and the second member of a compound with the d-article ensures that these two elements occur together and are initial in their noun phrase. (Note the spelling of the cognate forms in French: *lequel, laquelle, lesquels, duquel, desquels*, etc.). The so-called strong possessives *seu, meu, nostre*, etc. occur either immediately following the d-article or postnominally, in the standard variety: *el seu amic* 'his friend' vs. *cap amic seu* 'no friend his', but **cap seu amic* 'no his friend'. These words are alternatively postN-A or spec-A and, if the latter, part of a compound with the d-article.

An interesting consequence of the proposed treatment of the d-article is that multiple lexical sharing arises when weak prepositions such as *per, de* or *a* immediately precede a d-article. Contracted forms such as *pel, del* or *al* are argued by Wescoat 2007 to involve lexical sharing (instantiating P and D), but there are reasons to believe that these Ps are always bound morphemes, even when not contracted. Thus, a sequence such as *al peu* 'at the foot' (or *pel camí* or *dels quals*) would be a single word with three categories: P D N (or P D A and others). More complex cases of multiple lexical sharing include examples like *la dels nens* (D P D N) 'the children's' and *pels del teu* (P D P D A) 'by those of yours'. These complex cases, involving the sequence d-article–*de* (which is not possible in French or Italian), constitute a strong argument for a lexical sharing analysis of the d-article: they cannot be accounted for by an extension of the Miller/Tseng approach, where d-article and weak Ps correspond to features of the host's NP. A single word, such as *pel de la noia* 'by that of the girl' (P D P D N), may correspond to a c-structure in which a DP contains another DP. A formal proposal is presented to deal with such complex cases and ensure a correct mapping between word string, c-structure, and f-structure. The arguments for adopting a lexical sharing analysis of the d-article in Catalan indicate that, if we aim for an explanatory theory, this is the correct analysis for the d-article in French and Italian as well.

References

Brucart, José M. and Lluïsa Gràcia. 1984. I sintagmi nominali sense testa: Uno studio comparato. *Rivista di Grammatica Generativa* 11:3–32. **Martí i Girbau, Núria**. 2002. El SN: els noms, ed. Joan Solà et al., *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. 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 University. **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.