

# English compounds and the theory of abstract case

José Luis González Escribano  
Catedrático de Filología Inglesa  
Universidad de Oviedo  
Campus de Humanidades  
33011 Oviedo, España  
escr@uniovi.es

## Abstract

*The differences between English compounds and phrases arise from interaction between semantic structure, Case, and Merge. Case is the structural correlate of the relations between predicates and arguments and requires the f-head K to mediate between the two terms that contract them. At initial structure, K is a prefix of the selected term, projects a K-shell with two positions corresponding to the two K-relata, and, when overt, surfaces as prepositions, linking infixes, or inflections. In phrases, the selector surfaces as the specifier of K while the selected remains in its complement position, with K visible as the preposition of selected PPs or as a case inflection, when these exist, and invisible otherwise. Typical right-headed compounds result from phrases when K and its complement raise into a higher K-shell leaving the selector and unpronounced copies of themselves behind, with K showing or not depending on whether an adequate suffix exists. Since the K-selector is the head of the resulting construct, compounds seem head-final when raising occurs, but it need not occur, and when it does not, e.g., in Romance, compounds and phrases differ only in the covert/overt nature of K and their degree of lexicalization. The present approach requires nothing new in the grammar, but accounts for structural aspects of compounding without invoking Williams' RHR, Roeper & Siegel's FSP, or Selkirk's FOPC, and to that extent merits consideration within any 'minimalist' program.*

## 1. Introduction: The problem

Compounding has long been in a no-man's land with respect to the well defined 'modules' of morphology and syntax, partly because the criteria distinguishing compounds from phrases all leak (cf. Adams 1973: 57-60; Levi 1977; Lieber 1992: 12-13; Sadock 1998: 161-166) and the distinction is still under debate. The place of compounding in the architecture of grammar is also controversial. Traditionally, it is part of morphology, and compounding is obviously fed by derivational and inflectional processes (e.g., derived words like *immigration* enter compounds like *immigration office*, and the first term of the compound *teachers association* is clearly inflected for number), which supports the traditional view that it is 'halfway' between Morphology and Syntax (cf. Spencer 1991: 309), but any simple sequencing of compounding after derivation and inflection within morphology but before syntax, as in Allen (1978), will not do, since compounds arguably are, in their turn, subject to further inflectional processes (e.g. *teachers association+s*), their terms are free forms, and thus properly the province of syntax, rather than morphology, under Bloomfield's (1933: 207) influential distinction, and, perhaps more conclusively, because free forms may even combine into phrases before entering compounds (cf. *literature specialist* vs. *mediaeval literature specialist*), which definitely points to syntactic status (cf. Lieber 1992: 12-14, 19ff.; Carstairs-McCarthy 1993: 91, 99-100; Sadock 1998: 164-165; etc.).

Within TGG, the prevailing view is that compounds belong to the Lexicon and have special properties on that account, but there has never been agreement on how the Lexicon and Syntax interact or what role Morphology plays in compounding. In early TGG, the Lexicon was sandwiched between PS and T-rules; in late Aspects-style grammars with a 'second lexical pass', and in Generative Semantics, even T-rules fed the Lexicon, and viceversa; in EST- and P&P-inspired 'lexicalist' theories, the Lexicon fed the Syntax, with lexical Morphology in it and inflectional Morphology as a late addition in the PF branch of Syntax (cf. Anderson 1982), although possibly parallelistic models are not dismissed in Selkirk (1982) and in work by Lapointe, Borer, and others; finally, in Minimalist Theory, Morphology is in the Lexicon, which feeds the Syntax, but not viceversa (cf. Chomsky 1995).

As regards the specific issue of compounding, there has always been tension between syntacticist and lexicalist approaches. The general drift was from early syntacticist accounts like Lees (1960), to lexicalist approaches in the 70's and early 80's (e.g., Allen 1978; Roeper & Siegel 1978; Williams 1981a,b; Selkirk 1982; Lieber 1983; Di Sciullo & Williams 1987), and back to syntacticist ones since the mid 80's (e.g., Fabb 1984; Sproat 1985; Baker 1985; Roeper 1988; Lieber 1988, 1992; Liberman & Sproat 1992). Nevertheless, in practice, even within the lexicalist trend, in generative linguistics compounding was 'syntax', or at worst 'word-syntax' in opposition to 'sentence-syntax' (cf. Williams 1981a,b; Selkirk 1982), syntactic rules (e.g., PS rules in Selkirk 1982, lexical transformations in Roeper & Siegel 1978 and Roeper 1988, and syntactic transformations in Lieber

1992) played the leading part, and the principles of the gradually emerging autonomous lexicalist Morphology kept a rather low profile. Whether programmatically lexicalist or syntactacist, in such accounts it was syntactic principles like Theta Theory (Roeper & Siegel 1978; Williams 1981a,b; Lieber 1983), X-bar (Selkirk 1982), Predication (Williams 1980, 1981a,b; Selkirk 1982), Percolation (Lieber 1983), and Case (Fabb 1984; Sproat 1985) that played the leading part in constraining compounds, whereas would-be morphological principles like the ‘Right Hand Head Rule’ of Williams (1981a: 248), revised in Selkirk (1982) and relativised in Di Sciullo & Williams (1987: 26) (RHR, hereafter), the ‘No Phrase Constraint’ (NPC) of Selkirk (1982: 8), the ‘No Subject Constraint’ (NSC) of Selkirk (1982: 37), the ‘First Sister Principle’ (FSP) of Roeper & Siegel (1978: 208), the ‘First Order Projection Condition’ (FOPC) of Selkirk (1982: 37), the ‘Argument Linking Principle’ (ALP) of Lieber (1983: 262) and the whole ‘Lexical Integrity’ (LI) idea of e.g. Selkirk (1982: 70) or Di Sciullo & Williams (1987) played a secondary role, in some cases (e.g., the RHR) because they were not meant to be universal, but above all because they were either too stipulative (e.g., NSC, FSP, FOPC) or faced too many exceptions (e.g., RHR, NPC, LI). See Carstairs-McCarthy (1993: 99-120) for an excellent, but sceptical, review of syntactacist work, Sadock (1998: 161-162) *et passim* for a recent attempt to defend autonomous compounding, and Baker (1998) for a syntactacist reply.

The standard doctrine of Chomsky (1995, 1998) is the radically lexicalist one that lexical items are ‘monads’ (i.e., Lexical Integrity holds strictly) and **all** lexical processes, including **all** Morphology, precede all syntactic computation.<sup>1</sup> The lexicalist case for compounding ultimately rests on Bloomfield’s (1933) traditional assumption that only words, not phrases, occur inside compounds, (the NPC) and on the claim that typical syntactic processes like Wh-Movement, Binding, or Control do not affect elements inside compounds (cf. Williams 1981a; Di Sciullo & Williams 1987). However, as the syntactacist trend has emphasized, **other** syntactic principles (e.g., X-bar, Theta Theory, Move, Case, Predication), and core syntactic processes like movement/incorporation (cf. Baker 1985; Roeper 1988; Lieber 1992), coordination (cf. Lieber’s example *pipe and slipper husband*), and modification (cf. Fabb 1998: 75-76ff.; Sadock 1998) do apply even recursively inside compounds. In fact, there is overwhelming phonological and semantic evidence that compounds contain internal phrase structure (cf. Chomsky & Halle 1968: 22; Fabb 1984: 136, 145, 190; Sproat 1985: 199ff.; Lieber 1992: 11-14; Cinque 1993: 274ff.; Fabb 1998: 79; etc.). Furthermore, although lexicalists (e.g., Selkirk 1982: 8; Lieber 1992: 59-61) try to dismiss phrasal compounds as unrepresentative, their productiveness is undeniable (cf. Carstairs 1993: 99-100). Carstairs (1993: 91) notes that there is no upper bound on the length of a compound, and rightly concludes that, against the lexicalist claim, they cannot possibly be listed in the Lexicon (see also Sadock 1998: 164-165).

Thus, even among broadly minimalist scholars, the temptation to reduce compounding to ‘syntax’ (without adjectives) is very strong (cf. Baker 1998), and this is understandable, for that view fits Minimalism much better at bottom. Radical lexicalism and Lexical Integrity, on the contrary, are untenable, because, aside from the obvious impossibility of listing compounds, Merge and Move **need** access to word-internal features, synthetic compounding must ‘see’ the thematic properties of lexical heads, and all compounding depends on the visibility of various other phonological (cf. Katamba 1993: 292-293) and semantic features of lexemes which have largely been ignored in generative work (cf. Carstairs 1993: 94-97). In short, if compounds may contain internal phrases, syntax must ‘feed’ compounding, i.e., it must precede or be parallel to (e.g., Jackendoff 1997; Epstein et al. 1998), but cannot be later than, the Lexicon. This is inconsistent with the standard view that it starts from lexical ‘arrays’ (cf. Chomsky 1995, 1998) and follows the Lexicon. Hence, the sequential disposition Lexicon + Computational Component must be revised to allow for ‘loops’ or even unconstrained parallelistic access, probably in both directions, as Lieber (1992: 19) observes.

What has always supported the lexicalist view is that compounding is **not as productive** as *bona fide* syntax, so a certain specificity is undeniable; the question is whether it amounts to the need for an autonomous component of Morphology and/or the Lexicon. Assuming it did, the obvious way to patch up the lexicalist theory is to assume that the phrasal properties of compounds result from the pre-syntactic (i.e., lexical) equivalents of syntactic operations which e.g. Roeper & Siegel (1978) or Selkirk (1982) already count on to generate non-phrasal compounds. Generative Semantics and recent theories of lexical decomposition like Hale & Keyser (1993, 2002) do also imply a presyntactic ‘lexical syntax’ of sorts, so that machinery could easily be made to account for phrasal composition within the Lexicon.

However, that duplication, like the whole idea of ‘word-syntax’ vs. ‘sentence-syntax’ in Williams (1981a) and Selkirk (1982: 2, *passim*), as Selkirk’s own fluctuations (‘two discrete sets of principles’ but ‘the same sort of rule system’) indicate, is hardly reconcilable with the Occamian spirit of TGG, in general, and Minimalist Theory, in particular. The fact is that a constituent like *mediaeval literature* in the compound *mediaeval literature specialist* and the corresponding *bona-fide* NP in *Trevor specialises in mediaeval literature* are constructed in the same way, i.e., under the version of Bare Phrase Structure in Escribano (2004a), *literature* is merged to *mediaeval* because it satisfies its only argument, which fully saturates the adjective, leaves it syntactically inactive, and the noun projects. Predictably, what results is a NP, denotes a thing-like entity, and, except for stress differences predicted by Chomsky & Halle’s (1968: 92) Stress Rules or equivalents thereof (cf.

Cinque 1993: 272-290), has the same properties in both cases.

The genuinely ‘minimalist’ strategy, therefore, is to reinstate a strictly syntactic approach, assume that Merge, Move and the principles that constrain them (Satisfaction, Economy, Priority, Predication, etc.) apply uniformly throughout, and reject any **computational** break between Lexicon and Syntax (cf. Lieber 1992: 21, 76; Baker 1998; Spencer 2002),<sup>2</sup> but the problem, of course, productivity issues aside, is to explain why compounds and phrases should still have different **structural** properties,<sup>3</sup> in particular, 1) English compounds are mostly head-final whereas phrases are not; 2) the prepositions mediating between adjectives or nouns and their nominal complements are suppressed inside compounds; 3) certain kinds of phrases simply tend not to occur as first terms of compounds, and others lack some of their usual expansion possibilities; 4) in argument-head compounds, only one internal argument is allowed, independently of the adicity of the head; 5) certain functions, notably Agent, or ‘subject’ in general, but perhaps also Goal (cf. Selkirk 1982: 37; Fabb 1998: 74-76) seem to be unavailable to the nominals acting as first terms of compounds; and 6) the stress patterns of compounds and phrases tend to differ.<sup>4</sup>

Such issues constitute the agenda of the present work, to be developed in sections 2 and 3. Section 2 presents my theoretical framework. Section 3 describes how explananda 1-5 follow from it. Section 4, finally, contains an informal discussion of performance aspects that account for certain residues.

## 2. *The hypothesis: Abstract case and the compound/phrase distinction*

Under structure-dependent theories of grammar, the default assumption is that semantically relevant relations between constituents (e.g., Fillmore’s 1968 ‘Cases’) must be **structurally** represented. A traditional idea in this respect is that such relations are expressed ‘analytically’, with the help of ‘adpositions’, or ‘inflectionally’, by means of bound affixes, or, in typologically mixed languages, either or even both, as in the attested phases of Latin, German, or English. Within generative linguistics, a related hypothesis which enjoyed popularity during the P&P phase is that lexical material is ‘visible’ to syntax only when it satisfies abstract Case, cf. Fabb (1984), Sproat (1985) and Chomsky (1985).

Importing that P&P intuition into a current theoretical framework, I assume that semantically relevant relations like complementation and modification are encoded by means of ‘functional’ structure and, thus, that syntactic constituents **always** contain lexical material **plus** certain functional heads, although the latter are not always visible on the surface level. In particular, the structure of a syntactically computable lexical item  $L_j$  contracting an abstract ‘case’ relation with a case-selector  $L_i$  could be something like (1a), where  $K_{ij}$  expresses the structural-semantic relation between  $L_i$  and  $L_j$  and may overtly show the case of  $L_j$  in the relation  $\langle L_i, L_j \rangle$ . Of course, the resulting structure is usually related by a further K relation to a broader context, etc., as suggested by ‘...K’ in (1).

- (1) a. ...K [ $L_i$   $K_{ij}$   $L_j$ ]  
 b. ...K [ $L_i$  [ $K_{ij}$   $L_j$ ]]  
 c. ...K [[ $L_i$   $K_{ij}$ ]  $L_j$ ]

Since K encodes the abstract case relation between  $L_i$  and  $L_j$  it must act as a syntactic head for both K-relata, and  $L_i$  and  $L_j$  must be integrated into the same structural ‘domain’, a Larsonian ‘shell’, as in (1). However, as profiled in natural languages, abstract case relations are **asymmetric**, e.g., a K-selected term is the Theme, Agent, Possessor, etc. (correspondingly, a complement or specifier) **of** another, its K-selector, but not the converse, so a semantic and structural asymmetry is predictable between the two K-relata. In particular, let us assume that the K-selected term is the **first** argument of the relation K, i.e.,  $L_j$  in (1).

Furthermore, if structure is built by the binary operation Merge, ternary-branching structures like (1a) simply cannot arise. That leaves us with (1b/c) as the only *prima facie* possible configurations of any abstract case relation.

Since our Merge further requires strict **adjacency** between dependents and heads, and natural language structures must be linearized at the P interface (at least), it follows from Adjacency that K-shells, like all other shells, can have **at most two** positions, the ‘complement’ and ‘specifier’ of X in traditional X-bar theory (C1 and C2, hereafter, respectively, cf. Escribano 2004a). C1 and C2 must occur one on each side of K, but not as in (1a), for they are derivationally **ordered** as a consequence of the binary nature of Merge. Hence, K necessarily forms a syntactic constituent with C1, not with C2. That would seem to predict two possible outcomes, i.e., (2a, b).

- (2) a. C2 [K C1]  
 b. [C1 K] C2

However, according to Kayne (1994), hierarchical structure and linear order are correlated. As a consequence

of stepwise derivation *via* Merge, C2 asymmetrically c-commands and, granted Kayne's Linear Cooccurrence Axiom (LCA, hereafter), linearly precedes [K, C1]. Hence, ultimately only (2a) faithfully represents hierarchy, constituent structure, and surface order.

Observe that, even if K and C1 were both structurally trivial and therefore minimally c-commanded each other, K can only be adjacent to C2 if it **precedes** C1 (i.e., syntactic sequences like \*C2 [C1 K] are excluded by Adjacency, and [C1 K] C2 violates LCA), which in practice forces K to be a syntactic 'prefix' of C2 at initial structure. In English and similar languages, thus, K typically surfaces as a preposition, although it may be phonologically empty (e.g., between transitive verbs and their direct objects in all but a few cases involving pronouns) and invisible at the level of phonetic form (P).

Of course, even though K is a 'prefix' in 'deep' syntax, it may be a 'suffix' (i.e., an 'inflection'), or an 'infix' (e.g., a 'linking element' like German *-s-* in *Bildungsroman*, English *-o-* in *gynaecology*, English *-s-* in *craftsman*, etc.) in morphology (cf. *infra*), and in that case it cannot surface preceding  $L_j$  as in (1). Granted Adjacency, the simplest alternative would seem to be (3), where  $L_i$  and  $L_j$  simply exchange positions, but exchanging the positions of  $L_i$  and  $L_j$  also means **inverting** the K relation between the two K-related terms, which is wrong, since K was defined above as an inherently **asymmetric** relation.

(3) [ $L_j$  [K  $L_i$ ]]

The way to reconcile surface manifestation with the semantic and structural implications of the K relation in such circumstances is for K to raise and create a new K-shell in whose C2 slot  $L_j$  can be lodged. In other words, when K is morphologically a suffix,  $L_j$  must surface in a C2 (= specifier) position with respect to K if it is to precede it, but at the same time the K-selector  $L_i$  must preserve its original role in the K relation, so **two** K-shells and Move must be involved in such derivations. The relevant structure is shown in (4), an extended K-domain where, granted Chomsky's (1995: 185) concept of 'equidistance',  $L_j$  can skip the C2 slot occupied by  $L_i$  without violating Minimality.

(4) [ $L_j$  [K [ $L_i$  [ $t_k$   $t_j$ ]]]]

That, I contend, is what yields typical English compounds, most clearly argument-head (e.g. 'synthetic')<sup>5</sup> ones. Thus, *language teacher*, *dish-washer*, *oil-producing*, *north-bound*, etc., are derived from structures like (1b), which alternatively surface as X+NP/PP sequences, through derivations like (4).<sup>6</sup> In fact, movement analyses of synthetic compounds have already been proposed in Roeper & Siegel (1978), Roeper (1988) or Lieber (1992: 59-61), but, as will be shown in detail in section 3, all have undesirable empirical consequences and contain stipulations that make them incompatible with current minimalist theory and inferior to the present account.

Of course, the first term of a compound may alternatively be a 'modifier' of the head, of virtually any category and with almost any semantic relation holding between the two terms, (e.g., *ice-cream*, *Sunday closing*, *call girl*, *think tank*, *hardworking*, *underestimate*, etc., cf. Marchand 1969; Levi 1977; Selkirk 1982), but, as shown in Escribano (2004a), a modifier-head relation is simply a head-argument one in which the head gets saturated and other principles force the argument to remain active in the syntactic computation. Essentially: the modifier is a **predicate** and the modified is its **subject** and satisfies its selection feature. Since, granted Williams' (1980) theory of Predication, subjects are the latest arguments to be satisfied, a modifier never remains unsaturated after it merges with its modified, and therefore is never syntactically active after Merge. In such cases, then, the subject projects, typical cases being A+N or N+PP and Adv+V or V+PP combinations yielding NPs and VPs, respectively, and the same occurs in modifier-modified compounds.

However, the derivation of a modifier-modified compound may in principle be either just (1b) or (4), depending on whether the modifying term may take its modified as a complement or not. Adjectives, prepositions, and verbs certainly can, and yield A+N, P+N, and V+N compounds like *greenhouse*, *red hot*, *overwrite*, *draw bridge*, *swearword*, *think tank*, etc., as assumed in Selkirk (1982), or Lieber (1992). Whether nouns can also directly function as modifiers of other categories inside compounds is controversial (cf. Bauer 1998; Spencer 2002), although Selkirk (1982), Lieber (1992), Cinque (1993), and Carstairs (1993), among others, argue that they can. In that case, derivation (1b) suffices, but, even if they cannot, derivation (4) accounts for their structural properties.

So far we have considered only cases in which the two K-relata are word-level items, but, of course, under Bare Phrase Structure, any non-projecting head is *eo ipso* 'maximal', so C1 may itself be a phrase dominating one or more structures like (1b). Thus, we expect configurations like *specialist in mediaeval literature*, a structure like (1b) where K is the governed preposition *in* (cf. *He specialises in/\*on/\*about mediaeval literature*), *specialist* is C2 of K, and *mediaeval literature* is C1, as in (5).

(5) [C2: specialist [K: in [C1: mediaeval literature]]]

Correspondingly, we can also expect as possible (although, of course, not predict as actually occurring) the phrasal compound *mediaeval literature specialist* of (6), derived according to the pattern in (4), where K and *mediaeval literature* have both raised into the head and the C2 of the higher K-shell, respectively, although, there being no suffixal realization for [K: *in*] in English, K is invisible in its landing site. In (6), EC stands for (phonologically) ‘empty category’ in K, and the low copies in parentheses are not pronounced, as usual with traces.

(6) [C2: **mediaeval literature** [K: EC [C2: specialist [(K: in) [C1: (**mediaeval literature**)]]]]]]

Observe that in case (6) a simple derivation like (1b) is impossible, since *mediaeval literature* is **not** a lexical head and cannot take *specialist* as its C1. *Mediaeval* does take *literature* as its C1 in *mediaeval literature*, but that saturates *mediaeval*, which becomes syntactically inactive. Thus, all phrasal compounds are derived through raising, whereas W+W compounds may or may not involve Move.

Of course, to continue syntactically visible, NPs like (5) and (6) have to contract further abstract case relations with other lexical heads and are thus bound to be inserted as C1s of further K heads, as suggested by the presence of ...K in (1). For example, in both (5) and (6), *specialist* is the head and projects (observe that *in mediaeval literature* is an argument of *specialist*), and both constitute suitable complements of so-called ‘determiners’ like *a*, *this*, or *the* (in fact, subtypes of modifiers, under the analysis proposed in Escribano 2004c), but, granted the Visibility constraint, *the* and its NP argument must themselves enter a K relation, which requires them to become C2 and C1, respectively, of a new K-shell, as in the simplified representation (7). In their turn, the NPs of (7) will have to become arguments of still other heads, and will accordingly be inserted as C1s of still further K-shells, etc., until no further semantic/structural case relations are established and a fully saturated ‘propositional’ expression results (cf. Escribano 2004a).

(7) a. [C2: the [K: EC [specialist .....literature]]]  
b. [C2: the [K: EC [mediaeval literature ....specialist ...]]]

As is well known, recursion within argument-head compounds is impossible (cf. \**children gift giving*, \**employee mortgage offer*), traditionally as a consequence of the FSP (cf. Roeper & Siegel 1978: 208), of Selkirk’s (1982: 37) FOPC, of Lieber’s (1983: 262) ALP, or of the Case Filter, as in Sproat (1985) (see *infra* for a principled explanation of such generalizations) but modifier-modified compounds may be recursive subject to the same constraints (roughly, ‘predicability’) that govern the stacking of phrasal modifiers (cf. *London call-girl*, *silk party dress*, *wood towel rack*, *Oxford University library facilities*, etc.).

Such compounds, however, are unproblematic: they simply involve recursion of structures like (1b), if the higher modifier can take an NP as its C1, or derivations like (4), if it inherently cannot or is already phrasal. Thus, in *silk party dress* the low compound *party dress* contracts a further case relation with *silk*. As a consequence, the higher compound might have a structure like (1b), if nouns like *silk* can directly modify nominals, as Bauer (1998) claims. Alternatively, all noun modifiers that figure as first terms in compounds start as [K, NP] structures, i.e., *London call-girls* etc. result from something like *call-girls K: from/at London* through a derivation like (4).<sup>7</sup> On the contrary, in *Oxford University library facilities*, *Oxford University*, being phrasal, clearly cannot take the compound *library facilities* as its C1 and the derivation **must** be like (4) (cf. *the library facilities of Oxford University*).

Other aspects of the internal structure of compounds are straightforward and will be discussed here in very cursory terms. For example, both L terms K-related in a compound may have resulted from previous derivation, inflection, or compounding. Since such processes reduce to feature-satisfaction through Merge, and this is available throughout, we do expect, and find, compounds like *immigration officer*, where *immigration* and *officer* are themselves derived before contracting the K relation. In cases like *arms dealer*, *systems analyst*, *Arts Council*, *parks commissioner*, etc., the first term results *via* inflection previous to merging with K. In cases like *teachers association*, both derivation and inflection apply to the first term, in that order, before it contracts the K relation with *association*, which is itself derived. In phrasal compounds like *Oxford University admissions office*, both immediate constituents result from previous compounding possibly preceded by derivation and inflection (cf. *admissions*). Finally, when terms are multiply inflected (e.g., Ns and As, inflected for gender, number, and case in Indo-European), case, if visible, must be the outmost inflection, or K will not be adjacent to the two K-relata. The only differences between inflections like gender or number and abstract case are a) that the former are monadic predicates, whereas abstract case is inherently relational, and b) that, given their adicity, they may be (and are) suffixes, instead of obligatorily (deep) prefixes.

Abstract case is visible to very different degrees across languages. In modern English, it is visible *in situ*, as in structure (1b) *supra*, through the prepositions that introduce selected complements of verbs, nouns, or adjectives, which, by assumption, simply ‘spell out’ abstract case features as required by their respective

selectors (cf. Fillmore 1968).<sup>8</sup> On the other hand, case is visible in the higher K head of structure (4), as a morphological suffix, in what remains of case in the pronominal system (*he/him/his, who/whom/whose, etc.*),<sup>9</sup> in the '-s' suffix of pronominal NPs (cf. 'descriptive genitives' like *men's room, bachelor's flat*, and ordinary thematic genitives, as in *John's house, John's death, John's decision, etc.*), in 'linking elements' like -s- (e.g., *craftsman, swordsman, clansman, cf. Allen 1978 on the non-plural nature of that infix*),<sup>10</sup> or the -o- (an earlier genitive suffix) of 'neoclassical' compounds like *geography, gynaecology, etc.* (cf. Bauer 1983: 213-214), and, outside the nominal system, in the inflections of verbal forms selected in various syntactic contexts, i.e., a) zero after *to*, modals except *ought*, and *do*, b) active past participle after *have*, c) present or passive past participle after *be* and in absolute uses, and 4) tensed forms elsewhere. In other circumstances (adjectives, etc., cf. Fabb 1984), K has no surface realization.

### 3. Discussion: Advantages of the present analysis

Needless to say, the abstract case theory sketched here does not explain **all** the differences between phrase-building and compounding, or the infinite casuistry of the latter, and it is not supposed to (cf. endnotes 3 and 4), but it already has substantial implications at the structural-computational level, and certainly does account for most of the differences between compounds and phrases listed as explananda in section 1 above.

For example, it explains why most English compounds are (apparently) head-final whereas near equivalent phrases are head initial, i.e., explanandum 1) above (= the RHR). The reason is that, in the case of argument-head compounds, C1 has raised leftwards leaving C2 and an empty K head behind,<sup>11</sup> whereas, in that of modifier-modified ones, there may or may not be raising, but, in either case, the head is the modified term, as in **all** cases of modification (cf. Escribano 2004a).

Of course, movement-based accounts of synthetic compounds have been proposed before, e.g. in Roeper & Siegel (1978), Roeper (1988) and especially in Lieber (1992), but there are important differences that make the present proposal neatly superior. Essentially, since none of those authors assume Case to be relevant **inside** compounds, they cannot depend on derivations like (4), and their simple substitution or adjunction accounts yield inaccurate predictions, require *ad hoc* stipulations, and include devices which cause conceptual problems or are unavailable in the much more restrictive theories favoured nowadays.

For example, Roeper & Siegel's (1978: 208) T-rule allows only N, V or A to move, hence wrongly predicts that phrasal compounds like *American Literature teacher* will not exist. On the other hand, the substitution they propose, into a (redundantly) unique empty slot in front of the head, is a **lowering** rule (i.e., illicit, as trace must be c-commanded). Similarly, what Lieber (1992: 59-61) proposes as a derivation for a synthetic compound like *thirst quencher* is that the complement *thirst* starts as a right sister of the N *quencher* and left-adjoins to the head N. Yet, that derivation was problematic at the time and is even more so in a current framework, on multiple accounts: a) that movement also violates the c-command condition on traces, since the N complement of N is **lowered** into N; b) *thirst* must be a non-maximal complement of *quencher* when the compound is formed, which violates both X-bar and current Bare Phrase Structure; c) on the contrary, *thirst* must be maximal for the phrase *quencher of thirst* to arise, an arbitrary stipulation which, again, would now violate BPS; d) Lieber's N-movement is adjunction to N and violates Uniformity anyway, for, if *thirst* is minimal in situ, as she assumes (otherwise it would stay and form an *of* phrase), it becomes a non-head (hence maximal) in its landing site, and if *thirst* is maximal *in situ* (as it must under X-bar or BPS), then it **cannot** move and what arises is the *of* phrase; e) Lieber's account, thus, like Roeper & Siegel's and Roeper (1988), incorrectly predicts that no phrases will ever appear inside synthetic compounds, and cases like *American literature teacher* should be ill-formed; and finally f) Lieber aims at, but does not succeed in finding a motivation for N-movement, since case can always be supplied even to a bare N complement (assuming it differs from an NP!). To use case as a trigger, Lieber would have to stipulate that an NP exclusively dominating a bare N **cannot** receive case *in situ* (whereas a bigger NP can) and so moves, but this is very implausible (multiple counterexamples immediately come to mind), and makes no sense, anyway, for the noun does **not** move in order to satisfy its case needs, since there is no case licenser in its new landing site. Apparently, it moves to *avoid* being case-licensed by *of* (and forming a phrase instead of a compound), i.e., it moves in order to form the compound, but this does not make sense either, since phrases constitute the unmarked option, whereas compounds are the exception. Lieber's (1992: 62-63) remark that the N+N pattern emerged in OE (e.g., *man-slaga, gold-gifa, etc.*) and stayed productive is correct, and that circumstance can be invoked to explain why compounds are **allowed**, but the availability of a pattern is not a trigger of movement, and Lieber's explanation is circular.

The present account of compounds, on the contrary, faces **none** of the objections a-e). It certainly does not determine what triggers raising, but there is no reason why it should: nowadays compound building is a marked option, an exceptional pattern inherited from an earlier stage of the language. It need **not** occur, most of the time it does not (i.e., phrases are generated, instead of compounds), and when it occurs it simply is **not** predictable. All that need be done, therefore, is to establish how it is **possible** when it happens and what its limitations are.

As to the occasional left-headed compounds of English, e.g., phrasal verbs like *look up*, *take over*, cases like *passer by*, *notary public*, *mother-in-law*, etc. (and the left-headed compounds of Romance, cf. Scalise 1984: 125; Lieber 1992: 64-68), the present theory says simply that they are generated exactly like phrases, i.e., according to pattern (1b) or, more generally, (2a), although their non-compositional interpretation converts them into listemes and often subtracts them from the effect of certain syntactic rules (e.g., the PP complements often cannot be modified, etc.; cf. Fabb 1998: 76ff.).

Secondly, though a deep 'prepositional' K can raise as in (4), in English there are almost no overt suffixes to manifest it in its landing site, which directly accounts for why usually nothing is visible at surface level between the two terms of N+N or N+A compounds, explanandum 2) above. The only productive K suffixes available, mainly the 'genitive' suffix 's, the earlier Greek genitive *-o-*, the *-s-* infix of *sportsman*, etc., therefore, syncretize many different semantic relations, a correct prediction (cf. Escribano 2004b on different types of genitives).

Thirdly, since the constituent [K, C1] is neither minimal nor maximal, it cannot be affected by *Move in toto*: whether it raises into a head or into a specifier it will violate the Uniformity condition on chains (which follows from the Copy Theory of Movement, cf. Chomsky 1995, chapter 4). That automatically explains why non-lexicalized PPs do not occur as first terms inside compounds and dismisses an important part of explanandum 3) above. The PPs that do occur within compounds have first become formulaic modifiers (i.e., lexicalizations) and their Ps, therefore, are no longer exponents of K (e.g., *up-to-date* or *round-the-clock* in *up-to-date book*, *round-the-clock room service*, etc.).

Furthermore, the additional K-shell hypothesized here explains why case indicators, when overt, always occur **between** the two K-related items, typically as selected Ps or linking suffixes like 's, *-o-*, etc., although this fact may be obscured at surface level by the morphological appearance of K when it is a suffix and the obligatory leftward raising of  $L_i$  in case-marked nominals like *his*, *him*, *whom*, *whose*, etc. and in all inflected verbs.

Finally, this theory offers a nice explanation for another long-standing puzzle, our explanandum 4) above, i.e., the fact that argument-head compounds allow only one internal argument even though their head is derived from a predicate of higher adicity, cf. *\*employee hotel offer*, *\*hotel employee offer*, *\*guest comfort offering*, *\*comfort guest offering* etc., all ill-formed if the terms in bold are interpreted as arguments of *offer*. This used to be stipulated by means of descriptive generalizations or pseudoprinciples like the FSP (Roeper & Siegel 1978: 208), Selkirk's (1982: 36-38) FOPC, or Lieber's (1983: 62) ALP. Alternatively, Sproat (1985) derived it from the Case Filter, under the assumption that only one NP can be case-marked by the deverbal head, and Roeper (1988) from Theta Theory, by claiming that a head can only assign a theta role to one adjacent argument.

However, all such proposals face empirical and conceptual difficulties or contravene current assumptions. Roeper & Siegel's, for example, directly stipulates that only one sister of the head will move leftward by allowing for just one empty slot and, as shown, disallows phrasal movement and violates the trace principle. Lieber (1983) base-generates such compounds, but stipulates that only one left sister is allowed. Of course, this follows from binary branching, but there is no reason why the head should not project a word-internal specifier to lodge a second internal argument and thereby satisfy Argument Linking. Finally, Lieber's predictions are inaccurate, for a) phrasal compounds are, again, wrongly forbidden, and b) strictly speaking, under Lieber's theory, nothing prohibits multiple Ns (cf. *\*employee hotel offer*, *\*children story reading*) if they are interpreted as modifiers instead of arguments. Finally, both Sproat's and Roeper's solutions are dubious, for polyadic predicates (e.g., verbs like *give*) assign inherent Case to a **second** NP, cf. Chomsky (1985: 193), and their theta-marking power is **not** limited to their sister. Under standard analyses, they raise into higher shells and theta-mark additional **specifiers** (which are **not** their sisters).

On the contrary, the relevant restriction follows naturally from the present account of K-structure. The reason is straightforward: abstract case is a **two-place** relation between a head and its dependent, **not a three-place** relation between a head and **two** dependents. If *offer* is C2 of K and *an hotel* is C1, we expect *offer of (an) hotel*, according to (1b), or, if K is an empty suffix, *hotel offer*, via a derivation like (4), but there is no way to generate a **third** K-shell in which *employee* (or *hotel*) can be C2. K cannot simply raise again out of the second K-shell of (4) into a third one to acquire a third specifier, as in (8), for although it may support the Theme case relation between *hotel* and *offer*, or, arguably, the Goal case relation between *employee* and *offer*, since one of the K-relata changes, the same K cannot simultaneously encode both.

(8) \***[K [L<sub>j</sub> [t<sub>k2</sub> [L<sub>i</sub> [t<sub>k1</sub> t<sub>j</sub>]]]]]**

Granted the Copy Theory of Movement, neither can the selector  $L_i$  sitting in the low C2, e.g., the noun *offer* in examples like *hotel offer*, *?employee offer*, *\*hotel employee offer*, etc., raise into the C2 of the third K-shell of (8) and take the whole previously built C1 (i.e., [L<sub>j</sub> [t<sub>k2</sub> [L<sub>i</sub> [t<sub>k1</sub> t<sub>j</sub>]]]]) as its new case-selected term, cf. (9), without destroying its first K relation with  $L_j$ .

- (9) \*
- $[L_i [K [L_j [t_{k2} [t_{Li} [t_{k1} t_j]]]]]]$

Thus, whereas more than two shells and specifiers can be built whenever a head takes three or more arguments (e.g., *offer*), there is no three-place K relation, multiple K-shell structures **cannot** be built, and compounds with several internal arguments cannot arise, which makes FSP and FOPC redundant.

For parallel reasons, the present theory also directly accounts for much of the import of the No Subject Constraint (Selkirk 1982: 37), our explanandum (5) in section 1. The only explanation of the NSC offered in the literature (to my knowledge) hinges on Williams' (1980, 1981a,b) Predication Theory and is roughly as follows: a) only maximal projections have subjects, b) maximal projections are impossible inside compounds, hence c) compounds contain no subjects. Unfortunately, b) is unlikely to be right, or rather, under Bare Phrase Structure, it **cannot** be right, for any non-projecting term is *eo ipso* maximal. Thus, if NSC is really relevant, which is questionable (cf. *infra*), a new explanation is needed. Anyway, under the present theory, cases like (10a) are parallel to the cases of (8-9) and their ungrammaticality is immediately accounted for.

- (10) a. \*husband wife beating
- 
- b. \*husband beating of wives

However, case (10b), where only **one** argument of *beating* is realized internally to the compound, requires more careful discussion. Under the structure that Lieber (1983) would assign to it, i.e., (11a), its ungrammatical status is immediately predictable from our theory: the derivation involved a) violates the Thematic Hierarchy, since Agent must be projected higher than Theme and it is not, and b) its linear order violates Kayne's LCA, since *of wives* asymmetrically c-commands and should precede *husband beating*. What must be excluded, therefore, is the alternative structure (11b), which satisfies the Thematic Hierarchy.

- (11) a. \*[[husband beating] of wives]
- 
- b. \*[husband [beating of wives]]

Observe, though, that if *husband* has the Agent relation to *beating*, it should appear as the C1 of a K-shell in which *beating* is C2 and *husband* is C1 of K. Yet, that is impossible, for *beating* already has a K-relation with a C1 (i.e., *wives*). Even if *beating* raised into the specifier of a new K-shell, as suggested in (12), its C1 would **not** be *husband*, but the subtree [husband [(beating) of wives], which is not a possible K-relatum at all, since it is neither an argument nor a predicate of *beating*.

- (12) \*beating [K [husband [(beating) of wives]]]

Thus, when the head is transitive and both arguments are realized, the No-Subject Condition follows automatically from the present theory. There remain to be considered cases in which only a subject-like argument is realized within the compound.

Compounds like (13), where the head is an unergative verb *sing*, *smoke* are ill-formed, but this is also expected under present assumptions if such verbs derive from dyadic structures [[small v] [K *song/smoke*]] through incorporation of the cognate object *song* or *smoke* to small v, as Hale & Keyser (1993, 2002) maintain, for, under such an analysis, the Agent *bird* or *man* is a **third** argument of K, and we are back in the type of case excluded in (11).

- (13) a. \*birdsinging (cf. the singing of birds)
- 
- b. \*mansmoking (the smoking of the men)

On the contrary, compounds like (14), based on monadic predicates (or dyadic ones with optional complements), contain Agents as their first terms, and yet they seem acceptable, contravening the NSC, as expected under the present theory if they arise *via* derivation (4) from structures like (2a) which alternatively surface as phrases like *the reaction of (the) students*, *the complaints of (the) customers*, etc.

- (14) a. student action (attendance, reaction, etc.)
- 
- b. customer report (complaint)
- 
- c. army rule (control, action)
- 
- f. government decision (opposition)

The NSC also fails to predict cases like (15a-c), where the head is monadic and unaccusative and the subject is a Theme (a deep object). Examples like (15a-c) are well-formed, as expected under present assumptions, since they too can arise *via* derivations like (4). However, problematically, parallel cases like (15d-e) are unexpectedly

bad. That is what remains in support of the NSC, if such a principle exists, but, if it does, it must also account for innumerable exceptions like (14) and (15a-c).

- (15) a. sunrise  
 b. earthquake  
 c. bus stop  
 c. \*man death  
 d. \*?bridge fall

Of course, if the NSC is taken to apply only to ‘synthetic’ compounds, the issues reduce to whether *action*, *decision*, *rise*, *stop*, etc. count as deverbal heads and such compounds are synthetic or not. In fact, scholars disagree in that respect, but, clearly, e.g. *decision* does not seem any less deverbal than *death*. Within Lieber’s (1983) approach, she could always claim that the acceptable examples of (14) and (15) are good because they are interpreted as root compounds, hence exempt from NSC. However, her Argument Linking Principle is **not** meant to exclude root compounds, and, anyway, if that view is accepted, **any** subject should occur as a modifier of any head of any adicity, which is clearly wrong.

In my view, the distinction between root and synthetic compounds is either spurious or just a low-level and fuzzy one (cf. Allen 1978; Carstairs 1993: 118-119). Even if it explains the possibly different interpretive strategy of the speakers (about which Carstairs is sceptical), it should have no effects on the structural matter at hand. If nothing blocks external arguments in root compounds, synthetic ones should probably have them too, but as a matter of fact some, but not all, root compounds allow subjects, and many, but not all synthetic ones reject them; a messy situation. As shown, when the two arguments are visibly realized, e.g., in (11), the NSC is redundant, and when only one is, many counterexamples are just apparent, e.g., (13), and the rest of the evidence is murky. Thus, the status of the No Subject Constraint is dubious and, even if it exists, is unclear whether it needs a structural explanation. On the contrary, the erratic patterns illustrated in (14-15) suggest that the phenomenon, insofar as it subsists, might well be due to non-structural reasons.<sup>12</sup>

In sum, the great majority of the explananda listed at the end of section 1 follow rather naturally, at **no** cost, from the theory of abstract case defended above. There certainly are a few residues, but, as briefly shown in section 4, they are unlikely to be due to high-level computational principles.

#### 4. Epilogue: Non-structural aspects of compounding

The present theory of abstract case, of course, does not entirely dismiss our explanandum 3) above, particularly why, whereas PPs and even full CPs occur inside compounds (e.g., cases like *over the fence gossip*, *a slept all day look*, *a pipe and slipper husband*, *a who’s the boss wink*, etc., cf. Lieber 1992: 11), NPs containing determiners do not, cf. (16a,b) vs. (16c), or why phrases inside compounds often lack some of their usual expansion possibilities, cf. (17a) vs. (17b, c). Neither does it predict any stress differences between compounds and phrases, cf. (18), i.e., our explanandum 6) above. Yet, such matters probably belong with those mentioned in endnote 3 and need not be settled by the theory of syntax in the sense in which it is understood here. On the complex issues of stress, nothing will be added to the short remarks and references of footnote 4, but some discussion of how and where the residues of explanandum 3) above are to be approached may be a proper way to close this paper.

- (16) a. \*(a) this opera lover  
 b. \*(an/the) an opera lover  
 c. opera lover
- (17) a. \*?very hot water resistant  
 b. ?hot water resistant  
 c. water resistant
- (18) a. piáno music  
 b. músic for the piáno

The high-level principle to bear in mind with respect to matters like these is Economy, in the broad, pre-minimalist sense. Even though the computational machinery (perhaps as a residue of an earlier SOV stage, as Lieber 1992 and Baker 1998 say) allows for compounds to arise through derivations like (4), in the present state of English, compounding is clearly a ‘marked’ option. The reason is simple: native case inflections have virtually disappeared, and imported ones sometimes remain but are unrecognizable to speakers (e.g., Greek –o-

in *gynaec-o-logy*), so, in practice, when building compounds, the options the speaker faces are either to overuse the only productive native affix still available, ‘genitive’ ‘-s’, at the cost of syncretising a great diversity of semantic functions, or, plainly, since English has not developed new **suffixes** to encode the meanings of K heads raised as in (4), to leave such semantic differences unexpressed, to be inferred from context. The latter is arguably an ‘economic’ strategy from the point of view of the speaker, but also one with obvious disadvantages for the hearer, since it is clearly bound to induce ambiguity (e.g., a *security zone* may be an area where there is security or one where there is no security at all and people are not allowed, an *abortion campaign* may be a campaign for or against abortion, a *mosquito-net* and a *butterfly-net* serve to keep mosquitoes and butterflies without/within the net, etc.). Even from the point of view of the speaker, compounds, by definition, imply additional effort since, if they are listemes, they have to be noticed, learned, and stored in the Lexicon, and, if they are simply built on the spur of the moment, they still imply a costlier derivation, i.e., (4) vs. (2a). Hence, speakers can reasonably be expected to exploit the compounding option only when the tradeoff is clearly positive. Although deciding when that is the case implies complex multifactorial calculation that speakers cannot (consciously) perform, in general, Economy dictates that at least formulaic compounds must earn (and keep) their place in the Lexicon. Thus, only sufficiently general-purpose, socially approved of, and ‘effort-saving’ categorizations deserve the extra cost implied in pre-computing and storing compact, ready-made designators. Unless there are good reasons (stylistic effects, special rules, expectations operative in a certain jargon, etc.), building the equivalent phrase requires less effort, minimizes ambiguity, and offers more possibilities of expression (through right-branching recursion).<sup>13</sup> Hence, although, for diachronic reasons, a considerable number of compounds already belong to the core of the language (*apple pie*, *garden party*, etc.) and speakers constantly coin new ones, for compounding remains productive and its output usually is immediately understandable in context, most of the newly occurring compounds are nonce formations, and the few that remain long enough to spread around arise mostly in special registers (e.g., journalistic, advertising or technical jargon). In short, for understandable reasons, many possible compounds never occur, and of those that do only a tiny minority survives anyway (just as nearly all sentences are ephemeral). Most are not worth their keep. With this background in place, let’s now turn to the residues of explanandum 3) in (16-18).

Granted Bloomfield’s (1933: 207) definitions of Morphology and Syntax, the contrast in (16) has always been approached in lexicalist morphology (e.g., Roeper & Siegel 1978; Selkirk 1982; Sproat 1985; Di Sciullo & Williams 1987) in terms of a presumed inability of XPs to occur inside compounds, which would consist exclusively of word-level units (the ‘No Phrase Constraint’). However, the existence of phrasal compounds is well established (cf. Bauer 1983: 163-4 206-207; Lieber 1992: 11ff.). On the other hand, under current assumptions, such an approach is unavailable, for a) within any theory subscribing to Abney’s (1987) DP hypothesis, NPs are maximal even **without** determiners, and b) under Chomsky’s (1995, 1998) dynamically defined concepts of ‘maximality’, any non-projecting term is *eo ipso* maximal by definition. Here, Abney’s hypothesis is not assumed, so NPs are ‘just NPs’ and determiners and quantifiers are subclasses of modifiers (cf. Escribano 2004b on the reasons for such an analysis), but the maximality of any non-projecting term is crucial, so the ungrammatical status of (16a) cannot be due to any inability of maximal NPs to occur inside compounds.

It is not, and neither is it due to the often associated (cf. Spencer 1991: 312) inability of the first term to have specific reference (the distinctive property of Abney’s DPs). Observe, first, that what causes the contrast in (16) is not specific reference in (16a) vs. the generic reference of *opera* in (16c), for (as noted in Sadock 1998: 164) many proper names, obviously with specific reference, are possible in compounds, cf. (19).

- (19) a. Chomsky hierarchy  
 b. Madonna fan  
 c. Oedipus complex  
 d. \*?Tom problem

However, the proper names of (19) denote very popular individuals to whom it is possible to refer out of context, and the resulting compounds denote well-established general categories which have the same property. As a consequence, the proper names in such contexts have classificatory, rather than determining force. Observe that compounds like (19) behave as **common** nouns (e.g., they are compatible with real determiners, cf. *the Chomsky hierarchy*, *a Madonna fan*, *his Oedipus complex*). As the popularity of the individuals referred to decreases, on the contrary, cf. (19d), and therefore non-contextualized reference/interpretation turn impossible, such compounds become unacceptable, or acceptable only in special pragmatic contexts (e.g., a girl could perhaps say to Tom’s girlfriend something like *Do you have Tom problem again?*). Thus, terms with specific reference may enter compounds provided they refer to a sufficiently prominent individual to warrant automatic recognition in the speech community and the meaning of the compound as a whole is a conventionally established concept that can be invoked out of context. The problem with ‘determiners’ is that they are used to refer in a specific way to entities (otherwise named by **common** nouns) whose identification is strictly **context-dependent**. Thus, the interpretation of *this* in *this opera* must be resolved in particular speech situations, and as a

consequence a deictic cannot (or is extremely unlikely to) contribute to the semantic composition of a **general** predicate, which is what lexicalized common nouns worth having usually denote. The same consideration applies to the definite article, which, except in rare cases (cf. *the lion is a feline*, etc.) also requires a strictly context-dependent interpretation.

As to the indefinite article *a*, the problem is that it has not lost its original quantificational meaning (i.e., it still has the force of numeral *one* in many contexts). Consequently, a compound like (16b), *\*an opera lover*, under the analysis [[an opera] lover], would denote a very implausible category, i.e., a class of people who love one, and just one opera. Independently of the syntactic conflicts that the existence of such a common noun would induce (cf. *\*the an opera lover*, *\*all an opera lovers*, etc.), speakers are not as wasteful as to invent/learn categories and denominations for categories of that type.

The reason why number-inflected words are rare as first terms of compounds, cf. (20), is also non-structural, and essentially of the same type. There is no categorical ban on number-inflected words in arguments or modifiers inside compounds (recall *arms dealer*, *systems analyst*, *teachers association*, mentioned above, as well as appositional compounds like *women doctors*, etc.), but the plurality of the entities involved is often unlikely to add anything relevant to the semantic composition of a general property, which is what a compound noun denotes, and adding a number inflection showing plurality is usually uneconomic, since plurality or generality can be presupposed, as in (20b) (no car manufacturer could afford to produce just **one** car!) or even inappropriate, as in (20a) (where *factories worker* would apply only to an employee of more than one factory, an unusual state of affairs, and therefore an implausible category).<sup>14</sup>

- (20) a. *\*factories worker* (cf. *factory worker*)  
 b. *\*cars manufacturer* (cf. *car manufacturer*)

On the contrary, more specific quantification than mere plurality, i.e., cardinality, may well be relevant to the definition of properties of general interest to speakers, especially in technical jargon or advertising, but in that case it is encoded by means of specific quantifiers, i.e., numerals, and these do occur inside compounds, cf. *ten-year old*, *six-cylinder engine*, etc.

For parallel reasons, the ungrammatical status of (17a) cannot be due to any **structural** constraint on the modifiability of adjectives by *very* etc. inside NPs figuring as first terms of compounds. Neither is there the least difficulty in interpreting such compounds, so the problem is not semantic, either. The problem is, again, one of economy of lexical resources, which, being ‘formulae’ (cf. Jespersen 1924: 18-24), are ‘exceptional’ and must gain their place in the Lexicon. It is clear that there is a gradient of acceptability in (17) and similar paradigms, i.e., the more restrictive the denotation of the argument of *resistant* becomes, the less likely the compound is as a **general** predicate. If *water resistant* is plausible enough as a property of interest to ordinary speakers to merit a compact ready-made denomination (cf. *water resistant paint/sunbathing lotion/wristwatch*, etc.), *?hot water resistant*, although still fairly plausible, is already less likely to be a property of general interest to speakers to merit being stored as a lexical formula, and of course *\*?very hot water resistant* has only a remote chance of becoming a property of almost any entity sufficiently recurrent in general discourse to deserve an exceptional item in the lexicon of ordinary speakers, although it could perhaps become relevant in some specialized jargon, in a peculiar discourse context, etc. Such tendencies follow from the rather prosaic fact that most established compounds, like most lexical items, have reached that status because they are general predicates, and number distinctions, intensifiers, overly constraining second-order attributes or absolutely particularizing deictics, simply are unlikely to play any role in the semantic structure of such predicates.

In general, then, it is true that inflected words and phrases containing arguments or modifiers of a highly restrictive type or, in the limit, deictics requiring context-dependent interpretation (e.g., articles, demonstratives, possessives, intensifiers), are less likely than stems, simple words, or generic NPs to become first terms of compounds, so we do not in general expect full VPs, NPs with deictics, PPs, or CPs in such positions, but there are many exceptions, either of a highly formulaic nature (cf. *a dog in the manger attitude*, *a we’ll cross that bridge when we come to it*, *a what-you-see is what you get graphics interface*), or, on the contrary, *hapax legomena* built by imaginative speakers on the spur of the moment, and thus the mechanism of syntax must **allow** them, which our proposal, contrary to Lieber’s etc. does. However, contrary to e.g., the restriction on multiple arguments, whether and when such curiosities occur or not is not a matter that a theory of the computational component of Human Language need take care of.

## NOTES

<sup>1</sup> That is not the only view under investigation within current Chomskyan linguistics, though. An influential alternative is Halle & Marantz’s (1993) ‘Distributed Morphology’ hypothesis, and there are various hybrid alternatives around, including the earlier view of all morphology in PF, as well as parallelistic theories like

Jackendoff's (1997) or Montagovian rule-by-rule architectures like Epstein et al. (1998), etc. Detailed discussion of the alternatives is unnecessary here.

<sup>2</sup> That is compatible with the view that certain outputs of Merge/Move, e.g., all idioms and **some** compounds, belong to the Lexicon, but **most** occurring compounds do **not** become lexicalized at all (cf. Carstairs 1993: 91; Sadock 1998: 164-166)

<sup>3</sup> Obviously, other aspects of compounding, such as which of the infinite possible combinations occur or fail to occur as compounds, which actually occurring compounds become 'listemes' in Di Sciullo & Williams' (1987) sense, how the non-compositional aspects of their meaning are determined, etc., depend on historical, social, psychological, or contextual factors that cannot be covered by a general theory of the computational component of the grammar. The present proposal is meant to handle the **structural** properties of productive compounding, not to predict the vagaries of lexicalization.

<sup>4</sup> Although, in general, compounds get primary stress on their first term (cf. Bloomfield 1933: 228; Chomsky & Halle 1968: 15-18, 91-92), semantic and pragmatic factors intervene to make the actual details of usage quite complex, cf. *carrot cake* vs. *apple pie*, *rice pudding*, *Óxford Street* but *London Road* or *Penny Lane*, the stress difference between synthetic and appositional readings of *apprentice instructor*, the stress of appositional compounds like *child prodigy*, *woman doctor*, etc. The casuistry is irrelevant for present purposes and will not be discussed here. See Marchand (1969: esp. chapter II), Adams (1973: 59-60), Bauer (1983: 102-112), Spencer (1991: 319-321), Liberman & Sproat (1992), Cinque (1993: 272ff.), and the useful appendix in Spencer (2002).

<sup>5</sup> The established terms 'synthetic/verbal/secondary compound' will be replaced here with the more perspicuous and general one 'argument-head compound', since not all argument-head compounds are deverbal. Correspondingly, instead of the term 'root/primary compound' I prefer the clearer label 'modifier-head compound', since many modifiers are not roots but derived stems, inflected words, compounds, full phrases, or sentences.

<sup>6</sup> Carstairs (1993: 118-119) notes that the correspondence between X+XP patterns and XP+X ones is far from perfect, and fails in both directions in cases like *?giver of care* vs. *care-giver*, *?race in motors* vs. *motor-racing*, *?keeper of the door* vs. *door keeper*, or in the opposite direction in *keeper of a mistress* vs. *?mistress-keeper*, *taking offence* vs. *offence-taking*, etc., which should be embarrassing for syntacticist accounts. However, under the present account we need not assume derivations like (4) in all cases. Once a paradigm is established, speakers may coin new members even if the correlations the system rests upon do not obtain in all cases. Back-formations are a well-known case in point.

<sup>7</sup> That kind of derivation extends not only to 'descriptive' genitives like *bachelor's flat*, but also to 'determiner' genitives in ordinary NPs like *My father's house*, *my father's death*, or *my father's decision*, related to *the house of my father*, *the death of my father*, and *the decision of my father*, respectively, as claimed in Abney (1987) and subsequent work, but such cases need not be discussed here, cf. Escribano (2004b).

<sup>8</sup> Not all prepositions are mere context-sensitive exponents of K, though. Those that introduce modifiers of nouns, verbs, etc., of course, are real predicates, with independent semantic content and their own thematic domains, cf. Escribano (2004a). In such cases, the preposition behaves as a transitive head, i.e., it selects an NP which must be in a case relation with it, as in structure (1b).

<sup>9</sup> It follows that inflected forms like *him* and *died* are derived via raising of nominal and verbal L items, respectively, into the specifiers of higher K-shells headed by K-suffixes which do exceptionally have overt realizations *m* and (*e*)*d*. In the latter case, verbal case is usually associated with Infl nodes into which verbs raise either overtly (French) or at LF (English).

<sup>10</sup> Evidence that such suffixes, even if homonymous of former case inflections, have special status is the 'Fugungselement' *-s* after feminine nouns (which lack *-s* genitives) in German compounds like *Weihnacht-s-oratorium*, *Bildung-s-roman*, etc.

<sup>11</sup> This does not preclude the possibility that the argument+head order of most Modern English compounds should ultimately have a diachronic explanation, perhaps as a remnant of the earlier SOV phase of IE, as Baker (1998: 194-195) suggests, but that knowledge is not available to the average speaker and therefore does not help explain how the system works for him.

<sup>12</sup> Other presumed 'gaps' in the thematic possibilities of the first terms of synthetic compounds also seem questionable, or at least unlikely to be due to structural factors. Fabb (1998: 74) following Levi (1977), claims that, whereas Source is attested, as in *heaven sent* (= *sent from heaven*), Goal is not, but at the cost of offering unconvincing paraphrases for cases like *church-goer*, *party-goer*, *sea-going*, *north-bound*, *London-bound*, etc., where Goal seems the most plausible interpretation for the first term to have.

<sup>13</sup> Lieber (1992: 61) cites language acquisition evidence relevant in this respect, i.e., children first produce *\*kicker ball* or *\*builder wall* (according to the current SVO pattern) and only later learn to say things like *wall builder*.

<sup>14</sup> Of course, number **may** be relevant, and in that case we expect contrasts, cf. *park fence* vs. *parks commissioner*, and other factors may intervene (e.g., analogy, once a certain pattern is established). Furthermore,

the import of number clearly varies across languages. Spanish has *lavacoches*, *sacacorchos*, etc. instead of *\*lavacoche*, *\*sacacorcho*, whereas French has *tirebouchon*, English has *corkscrew*, etc. The matter should be looked at in detail, but such ‘counterexamples’ do not really overthrow the present explanation, for whereas in English there is no semantic contrast between singular and plural common count nouns, as the former is simply ungrammatical (cf. *\*Do you have car? \*Did you bring car?*), in Spanish, singular and plural do contrast, cf. *¿Tienes coche(s)/corcho(s)*, and only the plural has the generic interpretation required in *sacacorchos*, *lavacoches*, etc.

## REFERENCES

- Abney, S. 1987. *The English Noun Phrase in its Sentential Aspects*. Ph.D. dissertation, MIT.
- Adams, V. 1973. *An Introduction to Modern English Word-formation*. London: Longman.
- Allen, M. 1978. *Morphological Investigations*. Ph.D. dissertation, University of Connecticut.
- Anderson, S. 1982. 'Where's Morphology?'. *Linguistic Inquiry* 13: 571-612.
- Baker, M. 1985. 'The mirror principle and morphosyntactic explanation.' *Linguistic Inquiry* 16: 373-416.
- Baker, M. 1998. 'Comments on the paper by Sadock.' In Lapointe et al. eds. 188-212.
- Bauer, L. 1983. *English Word-formation*. Cambridge: Cambridge University Press.
- Bauer, L. 1998. 'When is a sequence of two nouns a compound?' *Journal of English Language and Linguistics* 2: 65-86.
- Bloomfield, L. 1933. *Language*. London: Allen & Unwin.
- Carstairs McCarthy, A. 1993. *Current Morphology*. London: Routledge.
- Chomsky, N. 1985. *Knowledge of Language*. New York: Praeger.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N. 1998. Minimalist inquiries: The framework. *MIT Occasional Papers in Linguistics* 15.
- Chomsky, N. and M. Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Cinque, G. 1993. 'A null theory of phrase and compound stress.' *Linguistic Inquiry* 24: 239-297.
- Di Sciullo, A. and E. Williams. 1987. *On the Definition of Word*. Cambridge, MA: MIT Press.
- Epstein, S. D. et al. 1998. *A Derivational Approach to Syntactic Relations*. Oxford: OUP.
- Escribano, J. L. G. 2004a. 'Head-final effects and the nature of modification.' *Journal of Linguistics* 40: 1-43.
- Escribano, J. L. G. 2004b. 'NPs as just NPs.' Ms. submitted to *Lingua*.
- Fabb, N. 1984. *Syntactic affixation*. Ph.D. dissertation, MIT.
- Fabb, N. 1998. 'Compounding.' In A. Spencer and A. M. Zwicky, eds. *The Handbook of Morphology*. Oxford: Blackwell. 66-83.
- Fillmore, C. 1968. 'The case for Case.' In E. Bach and R. T. Harms (eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart & Winston. 1-88.
- Hale, K. and S. J. Keyser. 1993. 'Argument structure and the lexical expression of syntactic relations.' In K. Hale and S. J. Keyser, eds., *The View from Building 20*, Cambridge, MA: MIT Press, 53-110.
- Hale, K. and S. J. Keyser. 2002. *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: MIT Press.
- Halle, M. and A. Marantz. 1993. 'Distributed morphology and the pieces of inflection.' In K. Hale and S. J. Keyser, eds., *The View from Building 20*, Cambridge, MA: MIT Press, 111-176.
- Jackendoff, R. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- Jespersen, O. 1924. *The Philosophy of Grammar*. London: Allen & Unwin.
- Katamba, F. 1993. *Morphology*. London: Macmillan.
- Kayne, R. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
- Halle, M. and A. Marantz. 1993. 'Distributed morphology and the pieces of inflection.' In K. Hale and S. J. Keyser (eds.) *The View from Building 20*. Cambridge, MA: MIT Press. 111-176.
- Lapointe, S. et al., eds. 1998. *Morphology and its Relation to Phonology and Syntax*. Stanford: CSLI.
- Lees, R. B. 1960. *The Grammar of English Nominalizations*. The Hague: Mouton.
- Levi, J. 1977. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.
- Liberman, M. and R. Sproat. 1992. 'The stress and structure of modified noun phrases in English.' In I. Sag and A. Szabolcsi (eds.), *Lexical Matters*. Stanford: CSLI. 131-181.
- Lieber, R. 1983. 'Argument linking and compounds in English.' *Linguistic Inquiry* 14: 251-285.
- Lieber, R. 1988. 'Phrasal compounds in English and the morphology-syntax interface.' *CLS* 24: 202-222.
- Lieber, R. 1992. *Deconstructing Morphology*. Chicago & London: The University of Chicago Press.
- Marchand, H. 1969. *The Categories and Types of Present-day English Word-formation* (2<sup>nd</sup>. Ed.) München: Beck.
- Roeper, T. 1988. 'Compound syntax and head movement.' *Yearbook of Morphology* 1: 187-228.
- Roeper, T. and M. Siegel. 1978. 'A lexical transformation for verbal compounds'. *Linguistic Inquiry* 9: 197-260.
- Sadock, J. M. 1998. 'On the autonomy of compounding morphology'. In Lapointe et al. (eds.). 1998. 161-187.
- Scalise, S. 1984. *Generative Morphology*. Dordrecht: Foris.
- Selkirk, E. O. 1982. *The Syntax of Words*. Cambridge, MA: MIT Press.
- Spencer, A. 1991. *Morphological Theory*. Oxford: Blackwell.
- Spencer, A. 2002. 'Does English have productive compounding?' Paper in Spencer's web page.
- Sproat, R. 1985. *On Deriving the Lexicon*. Ph. D. Diss. MIT.
- Williams, E. 1980. 'Predication.' *Linguistic Inquiry* 11: 208-238.
- Williams, E. 1981a. 'On the notions 'lexically related' and 'head of a word'.' *Linguistic Inquiry* 12: 245-274.

**Williams, E.** 1981b. 'Argument structure and morphology'. *The Linguistic Review* 1: 81-114.