Cross-Disciplinary Issues in Compounding

EDITED BY

Sergio Scalise
Irene Vogel

CROSS-DISCIPLINARY ISSUES IN COMPOUNDING

Edited by

SERGIO SCALISE
University of Bologna

IRENE VOGEL
University of Delaware

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

CROSS-DISCIPLINARY ISSUES IN COMPOUNDING

Volume 311

Edited by Sergio Scalise and Irene Vogel (eds.)

Cross-Disciplinary Issues in Compounding

Acknowledgments

Cross-Disciplinary Issues in Compounding comes as a response to the strong interest in compounds and issues related to compounds that have been developing in the last few years. At the core of this interest is the question of where and how compounds fit into the grammar, and in this respect, the ways in which compounds challenge our view of the organization and architecture of the language capacity more generally.

To address such issues, a conference was organized in Bologna, Italy in June 2008, representing the conjunction of two large research projects: Componet and ENLM. The Componet project (http://componet.sslmit.unibo.it/), with its center at the University of Bologna, has offered a home to a very fruitful research effort dedicated to compounds, in particular their description across a wide range of languages. A database consisting of information on approximately 30 languages is in the process of being developed with the intention of making access available to the entire linguistics community. The European Network for Linguistic Morphology, whose center is also in Bologna, is a project that has brought together major European research centers with the goal of addressing issues of morphology in a broader sense, including psycholinguistics, language acquisition and computational approaches. It was in the context of the activities of these two projects that the idea for the 2008 Bologna conference was developed, and ultimately brought to fruition. In keeping with the broad interests of the projects, speakers were invited to participate in diverse areas with respect to languages studied as well as to areas of research- theoretical linguistics as well as issues in sign language and language acquisition, and quantitative and typological analytical approaches.

The present volume, thus, has its origins in relation to the Bologna conference. It has aimed at retaining the cross-linguistic and inter-disciplinary approach to compounds, however, it represents a significantly revised and restructured contribution to the field. A selection of the topics covered at the conference was made so as to yield a well-rounded book, not simply a series of conference papers. To this end, too, the authors of the chapters of this book substantially revised their contributions so that they were more generally authoritative with respect to their topic, providing both a strong background and interesting original research.

The Master list of references collects the references of all of the individual contributions, and thus serves as a unified, up-to-date and quite complete list of literature on compounds, from the wide variety of perspectives present in the book. This, in itself, will serve as a valuable resource for researchers interested in issues related to compounding.

Table of contents

Why compounding?

Sergio Scalise and Irene Vogel
University of Bologna and University of Delaware

1. Introduction

The study of compounds is currently at the center of attention in all areas of linguistics – both theoretical and applied. In our introduction to this book, we discuss the reasons compounds have been considered so important, and why it is interesting to advance hypotheses not only regarding the construction of compounds, but also where they fit into the model of grammar, and what aspects of compounds still present the most challenges within different areas and frameworks.

The importance of compounding to our understanding of language was very clear to Greenberg (1963: 92), who stated

(1) "There are probably no languages without either compounding, affixing, or both. In other words, there are probably no purely isolating languages. There are a considerable number of languages without inflection, perhaps none without compounding and derivation."

Although it has recently been claimed that there are languages without compounds,[1] it remains true that compounding is a fundamental process of word formation. Indeed, in some languages, it is the only one.

In what follows we will discuss the following issues with regard to compounds: (a) why they have been attracting so much attention in theoretical linguistics, (b) where they are formed in the grammar, (c) their definition, (d) their classification, (e) headedness and exocentricity and (f) the lexical categories involved.

---

1. Stekauer, Valera & Körtvélyessy (2008), for example, have recently claimed that in their corpus of 55 languages 'only' 50 languages have compounds. Languages that these authors claim lack compounds include East Dangla, Karao, West Greenlandic, Diola Fogny, and Kwak'wala.

2.   The interest of compounds

Compounds are particularly interesting linguistic constructions for a number of reasons. First, they constitute an anomaly among grammatical constructions because they are 'words', but at the same time exhibit a type of 'internal syntax'. This syntax, furthermore, is somewhat 'invisible'. This can be seen in the following three compounds:

(2)   a.   taxi driver
       b.   hard ball
       c.   poet painter

In order to interpret these compounds one must 'add' a syntactic relation between the two constituents (i.e. *driver of a taxi, a ball which is hard, poet and painter*); the 'internal syntax' is not overtly present.[2]

Compounds, furthermore, represent a contact point between several crucial linguistic and non-linguistic notions such as those in (3):

(3)   a.   syntagmatic and paradigmatic relationships
       b.   syntax and morphology
       c.   linguistic knowledge and pragmatic knowledge

As for (3a), observe that in a compound such as *taxi driver* there is not only a 'relation' between the two constituents, but a special one: *taxi* is the internal argument of the verb *drive*. The verb in effect 'selects' its own argument[3] (i.e. a form like **appledriver* would be ungrammatical), and this selection is a syntagmatic relationship.

In addition to syntagmatic relations observed in compounds, we also find the situation in which a number of compounds appear to present a type of paradigmatic class (or compound family). That is to say, the head appears to constitute a source of 'attraction' for elements leading to the construction of many other compounds. This can be seen with the Italian word *capo* 'chief', in the series of items listed in (4):

(4)   capostazione   'station master'
       capoclasse   'head of the class'
       capogruppo   'head of the group'
       caposcuola   'head of school'
       capofila   'head of the line'

With regard to (3b), it has often been observed that compounds are the morphological constructions that are closest to syntactic constructions. Consequently, there is no general agreement on which component of the grammar is responsible for their formation (see Ackema & Neeleman, this volume). In relation to (3c), a simple example

---

2.   Jackendoff (2009) observed that "Compounds can show evidence of a little internal syntactic structure," e.g. a conjunction in *[health and welfare] fund*.

3.   See Scalise, Bisetto & Guevara (2005).

will suffice. As pointed out by Jackendoff (2009), in order to understand that *bike girl* refers to 'a girl who left the bike in the vestibule', we not only need linguistic information but also contextual information so as to exclude other possible interpretations such as 'a girl who habitually goes to work by bike'.

Compounds are additionally interesting because they exhibit what Pirrelli (2002) has called "weak compositionality." That is to say, the meaning of a compound may have a range of possible meanings, as well as a range of meanings that are not possible, as observed earlier by Allen (1978).[4] Examples of such ranges of meaning with respect to *water mill* are shown in (5a) and (5b). Crucially, the range of acceptable interpretations is most often dictated by paradigmatic relations holding between members of the same compound family, rather than by combinatorial principles of syntactic composition:

(5) a. mill powered by water
   mill located by water
   mill for producing water

   b. mill which grinds water
   mill which drinks water
   mill made out of water

Moreover, the study of compounds is of interest in psycholinguistics, in particular in relation to the mental lexicon (see Gagné & Spalding, this volume). Libben (2006: IX) formulates some of the questions in this area as follows:

(6) "What are the psychological mechanisms that allow such free creation? Are the production and comprehension processes involved the same for both existing lexicalized words and novel combinations? How are these processes related to other lexical and non-lexical processes? When are they acquired? How are they compromised by damage to the brain? How might they differ across languages? What shape might compound processing take among bilinguals?"

Finally, it has recently been proposed that compounds provide insight into the early stages of language evolution, being relics of a protolanguage as Jackendoff (2009: 113) points out:

(7) "This view of modern language as 'laid over' a protolinguistic substrate leads to the intriguing possibility that the coverage is not complete: that there exist pockets of modern language that are relics of earlier stages of the language capacity. Such relics would be areas where there is only rudimentary

---

4. The range of possible meanings is much greater in some languages than in others. Delfitto & Melloni (2009: 80) for example observe that Germanic languages, as opposed to Romance languages, have a relatively large range of interpretative freedom, "whereby the semantics of compounds only depends on context-related encyclopedic information." Thus, *tree man* could refer to 'a man who is standing beside a tree'; 'a man who is sitting in a tree'; 'a man who usually sits in trees'; 'a man who defends trees or forests'; 'a man who resembles a tree', and so on.

> grammatical structure, and in which such grammatical structure as there is
> does not do much to shape semantic interpretation. Rather, we would expect
> semantic interpretation to be highly dependent on the pragmatics of the
> words being combined and on the contextual specifics of use. I suggest that
> compounding fills the bill completely."

One might also maintain, however, that compound formation fulfills a communicative
strategy that is intrinsically different from that of syntactic expressions and, therefore,
their existence could be motivated by human communication purposes.[5] That is, com-
pounding is a manifestation of the tendency towards multiword constructions such as
idioms, collocations, binomial constructions, or the so-called prefabs.

3.   Where are compounds formed?

The question of which component of grammar is responsible for the formation of
compounds is complex, and the answers that have been proposed are quite varied. This
can be seen in the following, most likely not exhaustive, list:

(8)   Compounds are formed through transformations and deletion of lexical material
                                                                            (Lees 1960)
        Compounds are formed by Morphological Rules in a specific morphological
        component   (Lexicalist Morphology, Scalise 1984).
        Compounds are formed by syntactic rules 'all the way down'
                                                                            (Harley & Noyer 1999)
        Complex words are generated in an independent morphological submodule
                                                                            (Ackema & Neeleman 2004)
        Compounding is a type of incorporation into an acategorial root, in a frame-
        work in which word-formation is treated purely syntactically
                                                                            (Distributed Morphology, Harley 2009).
        Compounds are formed by filling available slots in lexical templates
                                                                            (Construction Morphology, Booij 2009)

We will not discuss all of these positions, but we would like to point out that there is
evidence in favor of a basic framework in which morphological facts are handled by a
morphological module, or submodule, of the grammar. The relevant evidence comes
from a variety of sources: psycholinguistics (experiments show that compounds are
stored in the lexicon and storage cannot be a property of the syntactic component[6]),
neurolinguistics (aphasic studies provide evidence for the application of compound

---

5.   An attempt to answer the question 'Why are compounds part of human language?' is found
in Di Sciullo (2009).

6.   See Gagné & Spalding (2009); Plag, Kunter & Lappe (2007).

formation rules[7]), and theoretical linguistics (compounding shares many properties with derivation such as allomorphy, linking elements and furthermore stress patterns of compounds may differ systematically from those of phrases).

4.   Definition of compound

In spite of the growing interest in compounds, there is no satisfactory definition for 'compound', as in fact there is no uncontroversial definition of other basic units such as 'word' or 'sentence'. Although we cannot review all of the literature on this topic,[8] some examples of attempts to define 'compound' elucidate the difficulties we face:

(9)   a.   A word-sized unit containing two or more **roots**   (Harley 2009: 130)

     b.   A lexical unit made up of two or more **elements**, each of which can function as a **lexeme** independent of the other(s) in other contexts, and which shows some phonological and/or grammatical isolation from normal syntactic usage   (Bauer 2001: 695)

     c.   [...] a compound word contains at least two bases which are both **words**, or at any rate, **root morphemes**   (Katamba 1993: 54)

     d.   A complex lexeme that can be thought of as consisting of two or more **lexemes**   (Haspelmath 2202: 85)

     e.   Its defining property is that it consists of the combination of **lexemes** into larger words. In simple cases, compounding consists of the combination of two **words**, in which one word modifies the meaning of the other, the head   (Booij 2005: 75)

     f.   Composition [...] denotes the combining of two **free forms** or **stems** to form a new complex word referred to as compound   (Olsen 2000: 280)

     g.   [...] root compounds consist of two **stems** combined as one, with the compound as a whole bearing the category and morphosyntactic features of the right-hand stem   (Lieber 2004: 47)

     h.   When two or more **words** are combined into a morphological unit, we speak of a compound   (Marchand 1960: 11)

The following general observation can be drawn from the above definitions: in the majority of the proposals, the definition of 'compound' coincides with the definition of the units that form a compound (see Montermini, this volume). However, this fact raises new problems. First, there is no agreement about which units are the basic ones in compounding, since different authors propose different units such as stems, roots, lexemes, and words. Moreover 'stem' must be identified differently in different languages (e.g. in English stems are typically free forms; in Greek they are bound forms

---

7.   See Mondini, Jarema, Luzzatti, Burani, & Semenza (2002).

8.   See most recently a comprehensive discussion in Lieber & Stekauer (2009: 4 ff.).

(Ralli 2007)). Furthermore, words can be typically monomophemic in some languages but bi-or plurimorphemic in others, etc.

Donalies (2004: 76)[9] attempts to define compounds by combining a number of criteria. Specifically, compounds are (a) complex, (b) formed without word-formation affixes, (c) spelled together, (d) right headed, (e) inflected as a whole, (f) syntactically inseparable, (g) syntacto-semantic islands, and (h) conceptual units. In addition, they may have specific stress patterns and include linking elements.

Lieber & Stekauer (2009) show, however, that even such a long list of properties fails to define compounds unequivocally. In addition to their criticism, we might add the observation that there are counter-examples to some of the proposed defining properties of compounds. With regard to (b), we observe that compounds may contain affixes (e.g. *blue eyed*). Contrary to (d), compounds are not always right headed (e.g. in Romance languages they are left headed); and contrary to e) compounds may exhibit plural inflection on one of the constituents yet still be singular (e.g. It. *porta-lettere* 'carry letters, mailman').

From a different perspective, a recent definition of compounds has been proposed by Guevara and Scalise (2009), according to which a compound is defined in categorial terms as in (10), where X, Y and Z are lexical categories and 'r' is the (hidden) grammatical relation between the two constituents.

(10) $[\, X \, r \, Y \,]_Z$

This definition assumes that the constituents of a compound - roots, stems, lexemes or words - have a lexical category. Z may be the same as X or Y or different from both, which gives rise to the following three patterns:

(11)   a. $[\, X \, r \, Y \,]_Y$ is a compound with the head to the right
        b. $[\, X \, r \, Y \,]_X$ is a compound with the head to the left
        c. $[\, X \, r \, Y \,]_Z$ is an exocentric compound[10]

For completeness, this definition must in addition be coupled with certain prototypical features of compounding. Based on current theories, the relevant assumptions include the following: (a) that compounds observe syntactic atomicity and lexical integrity, (b) that the constituents are members of major lexical categories, and (c) that the head is lexical (while the non-head may be lexical or phrasal).

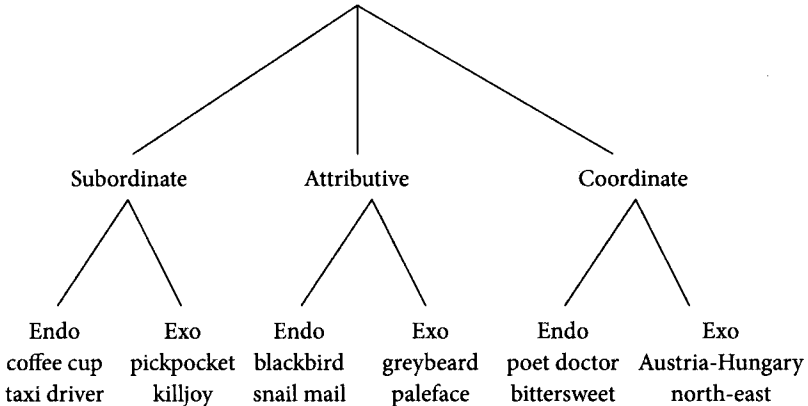**9.**   Cited in Lieber & Stekauer (2009: 6).

10.   There can be also $[\, X \, r \, Y \,]_Y$ or $[\, X \, r \, Y \,]_X$ structures that are exocentric such as $[[red]_A [skin]_N]_N$, or the Italian form $[[pelle]_N [rossa]_A]_N$ 'red skin'.

5. Classification of compounds

Classifying compounds also presents a number of challenges. In fact, every textbook of morphology seems to propose its own classification. Let us consider here a recent proposal by Scalise & Bisetto (2009) represented as follows:

(12)

	Subordinate		Attributive		Coordinate	
	Endo	Exo	Endo	Exo	Endo	Exo
	coffee cup	pickpocket	blackbird	greybeard	poet doctor	Austria-Hungary
	taxi driver	killjoy	snail mail	paleface	bittersweet	north-east

The first classificatory level in this proposal is based only on the grammatical relation between the two constituents (the 'r' seen above in (10)). The types of relations found in compounds are comparable to those in syntax:

a. **Subordinate:** the two constituents have a relation of *"complementation,"* which is particularly evident in deverbal compounds (*taxi driver*), but also in N+N compounds (*apron string*)

b. **Attributive:** the grammatical relation is of *attribution*, typically A+N or N+A structures (e.g. *high school, ice cold*). N+N structures may also be attributive. In *snail mail*, for example, the only information carried by the non head *snail* relevant for the interpretation of the compound is 'slow'. The non-head noun has thus an attributive value.

c. **Coordinate:** the relation between the two constituents is one of *coordination*, typically conjunctive coordination (*poet painter*). Within this classification, two other types of compounds can be accommodated: phrasal compounds and the so-called neoclassical compounds shown in (13a) and (13b), respectively.

(13) a. Af.   [lach of ik schiet] humor   'smile or I shoot mood'
       Eng. [floor of a birdcage] taste
     b. Eng. anthropology
       It.   odontotecnico   'lit. tooth technician'
       It.   colorificio   'lit. color factory'

The compounds in (13a) can be paraphrased as 'aggressive mood' and 'terrible taste' so their semantic interpretation is akin to what we have called 'attributive compounds'. The compounds in (13b), on the other hand, can be paraphrased as 'the study of man',

'a technician of teeth' and 'factory of colors', and can be considered subordinate compounds according to the definition given above.

6.   Headedness

There are numerous questions concerning the head of a compound, which include, but are not limited to, the following: Are morphological heads similar to syntactic heads? How can we identify the morphological head? What properties percolate from the head? We will not address these questions here since they are discussed in specific chapters of this book (e.g. Scalise & Fábregas, this volume), but instead we will comment here only on the issue of identifying the position of the head. As can be seen in (14), a number of different proposals have been advanced in this regard.

(14)   the head is on the right   (the 'Right Hand Head Rule' of Williams 1981)
        the head can be a *relativized* head   (Di Sciullo & Williams 1987)
        languages can have the head either on the right or on the left   (Scalise 1992)
        there are languages where the head can be either on the right or on the left according to the compound structure   (Packard 2000; Ceccagno & Basciano 2007)

The complexity of even this one question demonstrates that understanding the linguistic facts relating to the head of compounds will not result from a simple or sudden discovery, but instead requires continuous refinement of hypotheses, contingent on analyses of an increasing body of data.

7.   Exocentricity

At first glance it might seem that exocentricity is a marginal pattern in compounding, and may thus be relegated to the so-called 'periphery' of the language. It turns out instead, that it is not uncommon across languages, and must be considered a core component of compounding (see Bauer, this volume). The MorboComp database[11] gives the following figures regarding the position of the head, and the absence of a head - the exocentric compounds - based on a sample of 23 languages (Table 1 below).

    Although right-headed compounds are by far the most prevalent type of compound, exocentric compounds are the second most frequent type. Interestingly, different languages or linguistic groups exhibit different percentages of exocentric compounds, however, the general tendency for them to follow right-headed compounds in frequency is observed across the groups in the sample, as shown in Table 2.

---

11.   This is the morphological database developed at the University of Bologna. For details, see Footnote 1 in Scalise & Fabregas, this volume.

Table 1. Frequency of compound types with different head positions

	General %
Right	66.7
No head	16.3
Left	6.8
Both	5.9

Table 2. Frequency of compound types with different head positions in different language groups

	General %	Rom %	Germ %	Slav %	East A. %
Right	66.7	40,7	87,0	61,9	57,5
No head	16.3	31,4	8,9	12,2	17,7
Left	6.8	20,3	1,9	6,0	6,8
Both	5.9	6,8	1,3	3,1	15,0

As can be seen, the Slavic group follows the general pattern more closely than the other groups. The Romance languages exhibit a relatively high percentage of exocentric constructions, while Germanic languages are more consistently right-headed. The East Asian languages are different in allowing a relatively high percentage of compounds with two heads, most of which are coordinate compounds.

In languages such as Italian, the exocentric pattern V+N is one of the most productive processes in compound formation. In some languages we also find a pattern that has been called 'absolute exocentricity' (Scalise, Fabregas & Forza 2009), from both a categorial and semantic point of view. In such compounds, the output category is entirely different from the categories of the constituents, as illustrated in (15) for Chinese, Turkish and Italian.

(15) A+A = N   Ch.   大小 dàxiǎor   'large small = size'
      A+A = Adv Tu.   aptal aptal   'stupid stupid = in a stupid way'
      V+V = N   It.   sali scendi   'go up, go down = elevator'

Despite their relative frequency, exocentric compounds have generally been viewed as a problem for morphological theory, since it is necessary to account for information present in the whole structure that is not present in the constituents. In fact, this has led to a number of analyses in which endocentric readings have been proposed for such compounds. For example, Bisetto (1999) claims that the Romance V+N compounds such as the Italian *portalettere*, seen above, is endocentric on the assumption that there is a null nominalizing suffix after the verb expressing the meaning of *portatore di lettere* 'carrier of letters'. Booij (2005) claims that a different type of exocentric