Fernández-Domínguez, Jesús

N+N compounding in English : semantic categories and the weight of modifiers

Brno studies in English. 2010, vol. 36, iss. 1, pp. [47]-76

ISSN 0524-6881 (print); ISSN 1805-0867 (online)

Stable URL (handle): https://hdl.handle.net/11222.digilib/105088
Access Date: 17. 02. 2024
Version: 20220831

Brno Studies in English
Volume 36, No. 1, 2010
ISSN 0524-6881

Jesús Fernández-Domínguez

N+N Compounding in English: Semantic Categories and the Weight of Modifiers*

Abstract

Based on a 3,093-item corpus, this paper delves into the meaning relationship between the two constituents of N+N compounds. After tackling theoretical questions such as semantic categories and prototypes of compounds, some methodological details are considered and explanations provided on the collection of the corpus. Next, the experimental section examines the semantics of N+N compounding and, by means of various computations, it describes and analyzes to what extent the presence of a given modifier influences the overall meaning of these lexemes. Finally, the aspects from the theoretical and the practical sections are combined, and future prospects on the topic assessed.

Key words

N+N compounding; interpretation of compound; modifier; semantic category

1. Introduction

Noun+Noun (hereafter N+N) compounds are frequent items of everyday language, constantly used and understood by average speakers. They are the result of one of the most profitable word-formation processes of Contemporary English and, perhaps as a consequence of their high rate of occurrence, one is often unaware of the complexities that they hide, in particular with regard to their semantics. If N+N units are closely examined, complex attributes can be discovered beneath the mask of seeming structural simplicity, and this has been the object of discussion for over thirty years now.

There is agreement, at the outset, that words[1] are coined in response to a given naming need, which implies that no ambiguity is transmitted, in principle, by the language user as regards the denotation of the new unit, because "[…] *on the coiner's side*, a new form corresponds to a single meaning" (Štekauer 2005: xv; my emphasis). However, empirical research has suggested that semantic interpretation is not always a straightforward process when it comes to N+N units and that, for various reasons, several possible readings may surface for the same lexeme, which poses a potential problem for communication.

In this article, I would like to echo the need for an account of the variability of meaning of N+N compounds and, for this, I make use of a list of semantic predicates that allows discerning trends in the reading of N+N compounding. To this end, the foremost approaches to the semantics of N+N compounds are here revisited and the influence of modifiers over compounds is then elaborated on. Subsequently, various statistical measures are used to correlate the semantic analysis of N+N units with the modifier that they carry, and so detect whether one fact can be connected to the other.

In particular, this article wishes to attain a number of specific objectives:

(i) Examining how the semantics of N+N compounds has been traditionally categorized and providing an account for the lack of consensus in this sphere.

(ii) Analyzing the relationship between groups of units with a common modifier and how they are ascribed to Levi's (1978) *Recoverably Deletable Predicates* (hereafter RDPs).

(iii) Considering whether the occurrence of a given modifier across different N+N compounds influences their overall meaning.

The structure of the article is as follows: section 2 is a literature survey which focuses on the semantics of N+N compounds (2.1 and 2.2) and on the occurrence of these units between morphology and syntax (2.3). Data preparation is explained in 3, and 4 deals with the influence of modifiers in N+N compounding in descriptive terms. Finally, a number of conclusion and assumptions are discussed in 5.

2. The nature of N+N compounds

A compound word is customarily characterized as "[…] one composed of two (or occasionally more) smaller bases" (Bauer and Huddleston 2002: 1644), a definition which may be more or less precise depending on the subtype of compound in question[2], but which is fully operative in the case of N+N compounds.

A considerable amount of research has focused on this word-formation process in the history of modern linguistics, especially since Lees' (1960) seminal work and his inspection of compounds from a generative point of view. Many have followed Lees' monograph in discussing the matter and, in view of the general tendencies, the main problems with N+N constructions seem to be related to two specific fields: their internal semantic configuration and the fact that their formal makeup is identical with that of some syntactic phrases (those where a noun is premodified by another noun).

This section debates both issues by sketching out the most representative positions on the topic and, where pertinent, offers suggestions for progress in these fields. In particular, section 2.1 is devoted to the semantics of N+N constructions, section 2.2 is about the semantic categorization of compounds, and section 2.3 concentrates on the morphology-syntax interface. The examples are taken from the study corpus unless otherwise stated.

2.1. The classification of semantic categories: an overview

As far as their composition is concerned, N+N units essentially consist in the concatenation of two nouns, with no visible element binding them together. N+N compounding, thus, is different from other types of compounding in that it allows encoding complex concepts through an extremely compressed format, making use of as little space as language enables to. The paraphrases of the following units show that what is expressed by a clause can be worded also through a compound:

(1)	a. eye-rhyme	‘thing which appears to the eye to be a rhyme’
	b. footpath	‘path designed for people who are on foot’
	c. liferaft	‘raft designed to be used for saving life’
	d. timberline	‘apparent line formed by the highest extent of timber growth’ (Bauer and Huddleston 2002: 1647)

The profits of conciseness, however, are not troublefree because compression is done at the cost of removing semantic material from the corresponding sentence. If we compare, for instance, *footpath* with its paraphrasis, it emerges that certain lexical items from the clause have been omitted in the compound, namely the ACTION (*to design*) and the BENEFACTIVE (*people*). It follows that, for a speaker unfamiliar with this word, *any* ACTION and *any* BENEFACTIVE can replace the original ones, so *footpath* can have, in principle, as many interpretations as the pertinent context and cultural education permit (Girju et al. 2005, Štekauer 2005: 224, 2009, Körtvélyessy 2008).

The default of components to bind together the grammatical relations in N+N compounding has led numerous analysts to study this morphological process for an answer to the semantic interaction between its constituents. As a representative of psycholinguistically-oriented views, Gagné proposes her model of *Competition Among Relations in Nominals* (hereafter CARIN), where the issue is approached by comparing the participant’s processing speed with thematically frequent modifiers of the head noun (Gagné 2002, Gagné and Shoben 1997, 2002, Gagné and Spalding 2006a, 2006b). By contrast to schema-based approaches (e.g. Murphy 1988, 1990, Wisniewski and Murphy 2005), founded on the meaning dimensions of the head noun, the CARIN model is a relation-based alternative that takes into account not only which elements occur more often in the modifiers vs. head position, but also variables like word-order and recent exposure to similar N+N combinations. Even if her conclusions have varied in time, Gagné's experiments place major importance on comprehension times and make evident the central weight of left-hand elements in N+N units (Gagné and Shoben 2002: 643)[3]. Research of this type has significantly increased in recent times, thus showing the relevance that N+N compounding has also outside the domains of morphological theory (see Gagné 2009 for an overview of psycholinguistic research into compounding).

As regards studies produced strictly within theoretical linguistics, we must go back to Jespersen (1942), who perceives six different types of N+N compounds: 'B modified by A' (2a), 'A modified by B' (2b), 'A plus B' (2c), 'at the same time A and B' (2d), 'Bahuvrihi-compounds' (2e), and 'Type *son-in-law*' (2f).

(2) a. gas-light
b. tiptoe
c. Austria-Hungary
d. servant-girl
e. red-coat
f. lady-in-waiting

Jespersen's (1942: 143–157) lengthy analysis is primarily a semantic one and, perhaps for this reason, it results in a fine-grained discussion about the possible senses of the left- and right-hand member of the unit, e.g. COMPOSITION, INSTRUMENT, MATERIAL, POSSESSION, SOURCE, etc. In spite of his thorough inspection, Jespersen remarks that the number of possible logical relations of this word-formation process is infinite, and that the chief aim of his classification is to show the difficulty of their arrangement rather than to provide a definitive taxonomy.

In the following years, a number of authors proposed similar classifications, an example of which is Hatcher (1960) who, after disapproving of Jespersen's (1942) scheme for its alleged limitations, proposes a fourfold typology that arranges N+N compounds into the following categories: 'α is in β' (3a), 'β is in α' (3b), 'α is the goal of β' (3c) and 'α is the source of β' (3d):

(3) a. doghouse
b. house cat
c. sugar cane
d. cane sugar

Hatcher's (1960) controversial proposal meant a landmark in compounding studies because it proposed an innovative set of relationships to answer for the semantics of compounds, but it was later criticised because of its narrow number of categories (see Soegaard's 2005 account of the *reductionist theories*). Botha

(1968), one of the most critical linguists in this sense, argues that inventories like Hatcher's are justified on the grounds of their simplicity, logic or elegance, but are actually based on arbitrary criteria and their validity cannot be confirmed empirically.

Even so, proposals like Jespersen (1942) and Hatcher (1960) foreshadow one of the most influential monographs on compounding, in this case framed within generative grammar: Lees (1960). In this work, noun compounds are analyzed as deriving from a full structure at the deep level, and their constituents are linked with the thematic roles from a previously existing full sentence[4]. Later, in Lees (1970), a number of possible configurations are listed into whose slots the elements of compounds naturally fit. Lees' aim is to provide a range of configurations that encompasses all structural variants of compounds, such that it covers all the steps from a complete sentence to a regular N+N compound. An example is *motor car*, catalogued under *V-Object-Instrument* by following the paraphrasis 'the car (INSTRUMENT) uses (V) a motor (OBJECT)', as is done also with various widespread verbs (see ten Hacken 2009: 55–63). Other well-known works of this kind are Li (1971), Kay and Zimmer (1976) and Warren (1978).

Apart from these scholars, one of the first authors to refer to the importance of modifiers is Allen (1978), who tries to answer for the multiple interpretations of compounds through her *Variable R Condition* (henceforth VRC). This rule takes into consideration the semantic features of the two constituents of a compound and aims at answering for its multiple meanings, so that those semantic characteristics that aptly complement may create a possible compound, while impossible combinations prevent coinages. The VRC provides a range of possible readings for N+N compounds and considers the hierarchy of semantic features of their constituents so, when compatible attributes of both constituents fit, an interpretation can be reached (Allen 1978: 114–115, Murphy 1988, 1990, Abdullah and Frost 2007). The VRC, valid for semantically regular compounds, justifies why only some of the readings provided below for *water-mill* are possible:

(4) a. mill powered by water
b. mill located near water
c. *mill which lives near water
d. *mill which grinds water
(Allen 1978: 92)

But perhaps the most widespread model within generative semantics is Levi (1978) with her RDPs, a set of logico-semantic relations which, similarly to Lees' generalized verbs, intend to portray the inner semantics in N+N compounds. According to Levi (1978), RDPs occur in underlying relative sentences then transformed to create the complex nominal, from which the RDP can be retrieved for proper understanding. Only the following nine predicates can be deleted and then recovered:

CAUSE	tear gas	viral infection
HAVE	picture book	government land
MAKE	honeybee	snowball
USE	voice vote	–
BE	consonantal segment	–
IN	field mouse	–
FOR	horse doctor	–
FROM	olive oil	–
ABOUT	tax law	–

Figure 1. Levi's (1978) RDPs

RDPs roughly correspond to traditional semantic categories: ABOUT can be linked with topic, BE with essive or appositional, CAUSE with causative, FOR with purposive/benefactive, FROM with source/ablative, HAVE with possessive/dative, IN with location, MAKE with productive/ compositional, and USE with instrument. In theory, they embrace all the potential semantic associations which N+N compounds can portray. Specifically, Levi's predicates are aimed at "[…] nonlexicalized, nonspecialized, nonidiomatic, and […] nonmetaphorical forms" (1978: 8), precisely the constitution of the study corpus in this paper (see section 3), and they are supposed to surface naturally if the N+N compound is reworded:

tear gas	'the gas *causes* tears'
picture book	'the book *has* pictures'
honeybee	'the bee *makes* honey'
voice vote	'the vote *uses* the voice'
consonantal segment	'the segment *is* a consonant'
field mouse	'the mouse lives *in* the field'
horse doctor	'the doctor is *for* horses'
olive oil	'the oil comes *from* olives'
tax law	'the law is *about* taxes'

Figure 2. Paraphrases of RDPs

At this point, a glance at the literature will reveal two major trends. There is, on the one hand, a significant number of authors who have provided inventories of semantic relations that supposedly capture the main semantic connections within N+N compounds. Here belong the authors discussed up to now.

A different trend, on the other hand, is embodied by those who reject finite inventories holding that it is impossible to encompass all existing compound relations given their huge variety. Zimmer, for example, notices that traditional semantic lists always imply a *positive* characterization of compounds, i.e. "[…] a pairing of surface compounds with some sort of underlying structure, with the ultimate goal that any acceptable compound must conform to one of the listed pairings" (1971: C9). This he finds problematic due to the vast number of semantic variants of N+N compounds, and proposes to approach the issue from a *negative* perspective, that is, by listing which relationships cannot underlie compounds. Zimmer offers various examples of impossible readings, like *knife box* as 'a box which typically has no knives' or *war man* as 'man who dislikes, denounces, etc. war'. According to him, if the forbidden readings are spotted, the remaining interpretations will all be acceptable and we will thus have a solution to the problem of N+N compounds semantics.

Closely related to this, similar criticism goes that lists like Levi's are too vague and general, as it is often the case that the same compound can fit into several categories simultaneously, which stands as a serious weakness (see Soegaard 2005: 320). This is what happens, for example, with RDPs BE (meaning 'essive/appositional') and MAKE (meaning 'constitutive/compositional'), which partly overlap and slip across each other in units like *blood stream*, *artisan community* or *cheese slice*. The problem here is that it is perfectly possible to reword a compound like *blood stream* both as 'the stream *is* blood' and as 'the stream is *made* of blood', a problem of analytic indeterminacy which is in fact acknowledged by Levi herself (1978: 8–10). This degree of ambiguity, however, is deemed by the author to be sufficiently restricted for a hearer to identify the relation intended by a speaker by recourse to lexical or encyclopaedic knowledge, while still allowing for the semantic flexibility that undoubtedly characterizes compound nouns.

A third downside is that finite lists like Hatcher (1960) depict semantic categories deficiently, oversimplifying the shades of meaning of compounding, like in *headache pills* vs. *fertility pills*. As Downing (1977) says, these two units can be argued to feature the meaning of purpose (in fact, Levi includes both of them under FOR), but their semantics is by no means identical: in *headache pills* the aim of the medicament is to *relieve* the headache, while in *fertility pills* the aim is to *aid* in fertility. If FOR underlies these two units, it seems that it leaves a negative hint in *headache pills*, and a positive one in *fertility pills*. What is clear, in any case, is that there is a considerable difference in the meaning of both items and, what is more relevant, that the head of the compound is the same in both units, so the modifier seems to be the cause for the change in the overall semantics.

Viewing these criticisms from perspective, the subject of compound relations can be felt to be greatly exposed to the influence of categorization, a process coming from Aristotelian thinking whereby ideas and objects can be defined by a set of univocal, unambiguous and differentiated attributes. Frege (1970[1903]: 159), as an adherent of this philosophy, makes the point clear that categories should have impenetrable boundaries:

> Thus there must not be any object as regards which the definition leaves in doubt whether it falls under the concept; though for us men, with our defective knowledge, the question may not always be decidable. We may express this metaphorically as follows: the concept must have a sharp boundary. […] Any object Δ that you choose to take either falls under the concept Φ or does not fall under it; *tertium non datur*.

As can be seen, what linguists like Downing (1977) or Soegaard (2005) actually disapprove of is the categorical indeterminacy of the semantic categories in Hatcher (1960), Lees (1970) and Levi (1978). Critics of these lists believe that an N+N compound should be allotable to one and no other category in the set so, where ambiguity occurs, they automatically presume that the set is ill-defined and a different grouping is needed. Such is also the case of Beard's (1995: 391–395) *Universal Set of Nominal Grammatical Functions*, which recognizes up to 44 categories by opting for very subtle nuances in their distinction[5]. In this case, the author provides highly detailed semantic relationships, which involves that there are more categories but also fewer members within each of them.

In view of the dilemmas about semantic categories, one safe choice is to use taxonomies where the association between the constituents of the compound is an unspecified association, a more general perception than a series of explicit categories. Zimmer, for example, argues that the classification of N+N compounds heavily depends on the distinction between naming and describing, as "[a]nything at all can be described, but only relevant categories are given names" (1971: C15). In an attempt to overcome these problems he elaborates on relevant categories, and introduces the notion of *Appropriately Classificatory* (hereafter AC) *Relationship*, which reads: "A noun A has an AC relationship to a noun B if this relationship is regarded by a speaker as significant for his classification – rather than description – of B" (1972: 4). In this view, AB is a hyponym of B, and the type of categorization C is left unspecified for a looser relationship between the constituents of N+N compounds.

Highly evocative of Zimmer's contribution is Bauer's (2006: 494–496) *Mnemonic Theory*, which accounts for compound relations through the wide-ranging paraphrasis 'A type of element-2 efficiently brought to mind by mention of element-1', where the semantic relationship must be "[…] positive, non-modal, and inherent or permanent". Similarly to Zimmer's, this theory implies that, for a unit like *picture book*, only one possible reading (5a) can occur at once and interpretations that go beyond the limits of the compound's semantics are, for this reason, improbable:

(5) a. a book which has pictures
b. *a book without pictures
c. *a book which may contain pictures
d. *a book which contains pictures just today
(Bauer 2006: 495–496)

The main advantage of this paraphrasis is that it is valid for all N+N compounds thanks to its generality, which automatically rules out forced or unnatural meanings, and ensures that the reading will be easily perceived by the hearer. Similar attempts are found in Selkirk (1982) and Lieber (2004), where user knowledge, context and pragmatics take special significance and explicit semantic categories are discarded.

As can be seen, universal conceptions like Zimmer (1971) and Bauer (2006) are an attractive option in that they can answer for the semantics of all N+N compounds, and for the fact that, once coined, multiple interpretations of a compound are rarer (Allen 1978: 88). This undoubtedly eradicates the vagueness of co-existing semantic categories and provides an accurate theoretical statement for N+N compounds.

It also true, however, that Zimmer's and Bauer's alternatives offer universal readings by basing on world knowledge and the individual's background, which is beneficial thanks to their broad range of performance, but also exposes them to vagueness and personal interpretations (the compound can mean anything as long as it stays within the limits of the paraphrasis). This automatically discards these options for the study of individual N+N compounds, since the same unit may have one reading in a given context and a different reading in a different context for the same speaker. The impression with the above schemes is that what is gained in terms of meaning interpretation is lost in terms of specificity, with the result that they do not truly give a factual account of compound meaning.

2.2. Delimiting categories of N+N units

In view of the situation in 2.1, it seems clear that, as in many other areas of language theory, the issue of compound relations requires taking a stand for or against generality while being aware that each option has weak and strong points (see Bolinger 1961 for a discussion on the gradience of categories).

Finite lists like Lees (1970) or Levi (1978), on the one hand, introduce wide categories intended to simplify the multiplicity of meanings of N+N compounds, but for that very reason categories often intersect and have areas in common. On the other hand, fine-grained directories provide a full semantic coverage, thus comprising numerous entries with very restricted semantic features, but their effectiveness may prove unsatisfactory precisely because of their size. Then, considering the amalgam of opinions as to the semantics of N+N compounds, an initial question should be, perhaps, not how many categories of compound relations can be distinguished, but which the nature itself of these categories is. Put simply, what are the features of the category *semantic relationship of N+N compounds*?

As has been shown in the previous section, recurrent complaints in this field go that traditional semantic categories are not well-delimited and that their demarcation is not accurate, but one has the impression that new proposals have often fallen into the same errors as the models which they reacted to (see 2.1).

It may be advisable, in view of these problems, to review the existing types of categories by going back, for example, to Labov (1973) who, in studying categories in cognition, perceives three types of categories depending on how *hard* (+) or *soft* (–) their limits are:

Category types	Boundary	Gradation
1	+	–
2	+	+
3	–	+

Table 1. Types of categories

The first type of category is bounded and non-graded, "[…] with distinct and invariable outer limits, characterized by a fixed set of necessary attributes and simple yes-or-no membership" (Aarts et al. 2004: 5). This definition corresponds to the concept of category in the hardest Aristotelian sense, as in Frege's (1970[1903]: 159) quotation above. The second category type is both bounded and graded, still with well-defined limits, but in this case some members are "[…] better examples of the category than others". In this case there is a certain margin for variability within a single class, but class membership is still exclusive. In the third place, other categories are unbounded and graded because it is not easy to pin down clear-cut edges in each class but, rather, "[…] they fade off into each other, with the more peripheral members having more in common with members of other (contrasting categories)".

The question, then, is: which type of categorization does the semantics of compounds require? Does it need a type of category one, with solid inflexible edges, or rather a type of category three, with fuzzy intermixing partitions?

We may now remember that one criticism against Lees (1970) and Levi (1978) is that they propose overlapping categories that make their classifications subjective and ambiguous at certain points. Yet, Labov's typology (Table 1) suggests that the problem lies, not in the nature itself of these categories, but rather in how they are perceived, as some authors prefer to have clear-cut edges, while others argue for fuzzy limits. It is, so to say, a matter of preference in the understanding of concepts.

Given the nature of the issue, it seems difficult to grasp all nuances of semantic connections by means of brief catalogues, but this may be a sensible starting point for the study of the categorization of compound relations, because it provides an idealized picture of the phenomenon. It is no wonder that the characterization of semantic categories can be performed in an extremely detailed manner, recognizing a large number of variants, but it seems to me that this would bring about problems concerning their nature and span, even more when research in this field seems to lack a minimum degree of consensus yet, as reported in Soegaard (2005). Despite the fact that detailed categories are beneficial because more precise categories are covered, they can bring about the unexpected effect of overlapping, caused by their high degree of specificity. This is a complex topic which can be perhaps also approached through *ad hoc* categories, so that they resemble category type one in some areas and category type two in others.

In view of the alternatives and implications of this subject area, a matter to be resolved is the proper definition of the semantic categories of compounds, as was stated at the beginning of this section. My opinion is that only once the category of compound relations is identified, can we properly define their number, nature and scope. As has been shown, scholars often suggest that the semantics of N+N compounding is better approached by admitting within-category gradience, which would explain the limited number of semantic predicates in traditional lists (often below ten) as well as why some N+N compounds can match into several categories simultaneously (see MAKE vs. BE above). This makes it possible to catalogue problematical and tricky cases while having a high ratio of members per category.

This paper, thus, uses Levi's (1978) RDPs because, despite the above-mentioned criticisms, its semantic categories are reasonably detailed and it stands as one of the most common and reliable sources for the study of complex nominals. Not in vain it has been continually quoted as a reference in first-line contributions since its publication thirty years ago (e.g. Gagné 2002, 2009, Gagné and Shoben 1997, 2002, Štekauer 2005, 2009, Wisniewski and Murphy 2005, Plag et al. 2008). We are of course aware that further semantic relations can be distinguished than this nine-item set, but it is also true that the present article requires wide-ranging categories for its computations, and Levi (1978) fits these needs suitably. Note that this does not entail an unconditional defence of RDPs, but an acknowledgement of their suitability for this experiment.

2.3. N+N units on the cusp between morphology and syntax

One of the most complex areas of study in contemporary morphology is where N+N constructions belong, given that there is no agreement on whether they fall under the domains of morphology or of syntax. As Libben says, "[…] compound words are structures at the crossroads between words and sentences reflecting both the properties of linguistic representation in the mind and grammatical processing" (2006: 3). Part of the difficulties of this subject probably lies in the fact that the opposition compound noun vs. nominal phrase is founded "[…] more on an ideological basis than on an empirical one" (Bisetto and Scalise 1999: 47). In this section we concentrate on the implications which the morphology-syntax interface has for the modifiers in N+N constructions, by looking particularly at the levels of semantics and morphology.

First, as concerns the meaning of N+N compounds, their bewildering array of relations has been claimed to bring about meaning unpredictability, regarded an inherent property of compounds for two main reasons: on the one hand, for the semantic shift they often undergo and, on the other, for their lack of structural elements (e.g. prepositions) to explain the relationships between the members of the compound. Traditionally, lack of meaning predictability has been attributed to the absence of a verbal element, although some have also pointed to lexicalization, a phenomenon that obliges the hearer to interpret compounds based on the context, as opposed to syntactic phrases, where meaning is considered to be compositional (see Lipka 2002: 97). As a token, Downing's (1977) classical example *apple-juice seat* shows that, by putting two nouns in succession, compounds may be created whose referent depends almost entirely on the context:

(6) a. 'a seat for drinking apple juice'
b. 'a seat in the colour of apple juice'
c. 'a seat with apple-juice spilled on it'
(Downing 1977: 818)

One weakness of this view lies in that lexicalization affects compounding, but also non-compositional syntax, as in idioms, a process labelled *desyntactization* (Corbin 1997). A second counterargument is that specialization of meaning seems not to be caused by frequent usage solely, but also by first use. For example, in *apple-juice seat* a number of different interpretations are possible but, if one of them is picked up by the coiner, the rest of them do not have equal opportunities of being selected in the future.

Another alleged factor to distinguish morphology from syntax is the full productivity of phrase-structure rules (due in part to their unlimited recursivity), which is not always the case with compounds (more limited productivity). If so, listedness can be a straightforward method to separate morphology from syntax, since syntactic constructions are in principle created endlessly while morphological constructions are lexically listed. Once more, there are drawbacks to this proposal, e.g. that listedness does not always correspond with occurrence in dictionaries, and also that compounds with various constituents are *statistically unlikely* to be listed, which may mislead research (see Bauer 1998: 70).

However, the most serious disadvantage of ascribing listedness as a property exclusive of compounds is that this would deny their productivity, thus contradicting their unquestionable profitability of this process. As is well-known (Di Sciullo and Williams 1987: 14, Plag 1999: 6–8), the lexicon is where irregular/unproductive items and processes are stored, features which do not quite characterize N+N compounding, and this is why an absolute relationship between listedness and word status is often rejected as a valid criterion for the differentiation of morphology vs. syntax (see Carstairs-McCarthy 1992: 30–31).

Second, with regard to the morphology of N+N compounds, it has been put forth that they are single lexemes, thus standing out against phrases, which can include various lexemes themselves. If compounds are single lexemes, then inflectional marking should not be allowed for their internal elements, but it should be placed only at the end, as in indecomposable units. Adams (1988), for example, argues that first constituents of compounds are grammatically neuter because plural marking (7a), genitive case (7b) and verbal inflection (7c) are always absent from them:

(7)	a. tear gas	‘gas which causes tears’
	b. pigtail	‘pig’s tail’
	c. watchdog	‘dog that watches’

(Adams 1988)

According to Adams, the most usual is for the element at the end to bear the plural mark and, even where the first constituent has a final -*s* as an independent lexeme, it is dropped in the compound, as in (8):

(8) a. trousers
b. trouser[s]-press
(Bauer 2006: 720)

Against these statements is the case of *genitive compounds*, sometimes perceived as lexicalized syntactic phrases (e.g. Shimamura 1998). Genitive compounds have been often reported to encompass two distinct subtypes: on the one hand, there are units where the first noun is inflected for genitive case and it is the determiner of the head, as in (9). On the other hand, in other units the left-hand noun bears a genitive mark, but it is understood as a member of a compound due to its contribution to the meaning of the unit, as in (10), in which case it can be a common or a proper noun (see Rosenbach 2006, 2007):

(9) a. the mayor’s house
b. Fred’s car
c. your brother’s picture
(Shimamura 1998: 1)

(10) a. chef’s salad
children’s rights
director’s chair
legionnaire’s disease
(Shimamura 1998: 1)

b. Achille’s heel
Down’s syndrome
Murphy’s law
Rubik’s cube

As with other instances of N+N units, the problem with genitive constructions is that their structure mirrors that of syntactic phrases, whereas their meaning seems to go beyond mere modification (which is typical of syntax). From the above paradigms, it becomes clear how difficult it is to present an accurate definition of N+N compounding while sidestepping the morphology-syntax interface. The fundamental dilemma with this demarcation lies not in the complexity of distinguishing isolated units, but in establishing a systematic procedure for telling compounds from phrases. A unique type of structure, a halfway point between lexeme and phrase, may also be defended but, if so, would it be of a morphological or of a syntactic nature? A possible solution is to call these constructions *N-bars* (see Selkirk 1982, Di Sciullo and Williams 1987), a term with a strong generative influence, and so reflect their parallel structures. The main drawback to this possibility is that it does not really address the differences in use and meaning between compounds and phrases, as it provides a term without further specifications of usage, meaning or productivity.

Yet another way out could be to define, first, the concept of *phrase*, and then to exclude the remaining elements assuming that they must be compounds but, to the best of my knowledge, no specific proposals have been made in relation to this beyond Payne and Huddleston (2002). Whatever the choice, a fundamental idea is that, to attain a real division between these two grammar components, the distinction between morphology and syntax should be based on a combination of tests, but not on only one of them: "any distinction drawn on the basis of just one of these criteria is simply a random division of noun+noun constructions, not a strongly motivated borderline between syntax and the lexicon" (Bauer 1998: 78).

2.4. Summary

This section has reviewed the main references to the modification of N+N compounds and, from their number and quality, it seems safe to assert that the issue holds a central position in morphological studies at present. The specialized literature in this field has shown that left-hand constituents of compounds play an essential role for the overall sense of the unit, such that they are often decisive as to how the compound is understood. Experiments like those conducted by Gagné (2002, Gagné and Shoben 1997, 2002) provide empirical proof that the semantic reading of N+N units is decidedly influenced by modifying elements, thus corroborating the premises sketched in section 1.

Once the theoretical assumptions on the topic have been explained, the remainder of the article turns to the experimental side of the subject, namely the assessment of the weight of modifiers for N+N root compounds. There, our aim is to contrast the aforementioned premises with the corpus entries and so check to which extent their meaning is justifiable through the theory of compound modifiers.

3. Data preparation

The experiments in this paper are carried out on a 3,093-unit corpus compiled from the BNC Sampler by using *Oxford WordSmith Tools version 4.0.0.376* (Scott 2004; hereafter *WordSmith Tools*). Structurally, N+N compounds are known to take on three forms: open (*alarm clock*), hyphenated (*bus-bench*) and solid (*creaseline*), depending on whether their two elements appear separated, linked by a hyphen, or written together. *WordSmith Tools* was used for the retrieval of these three variants, for which several steps are required.

3.1. Retrieval of compounds

Because the aim of this experiment is to assess the influence of modifiers within N+N compounds, an option is to apply an exhaustive coverage to a subpart of the BNC Sampler rather than to carry out a randomized sampling. In this case, it was decided to focus on all units starting with letters *a*, *b* and *c*, which were analyzed and, when irrelevant, discarded from the study. In this manner, it is possible to retrieve N+N constructions with identical first constituents, something essential for the study of how modifiers interact with heads in compound semantics.

As was explained in section 2, the ultimate meaning of a compound is not a fixed schema, but varies depending on the context and, most relevant, it is affected by both its left- and right-hand constituents. This can be checked in (11), which features compounds with different meanings, regardless of the fact that their left-hand constituents are the same (see 3.2):

(11) a. city centre
b. city policy
c. city romantics
d. city style shorts
e. city vote

Different stages were needed for the recovery of the various types of compounds. For the generation of the wordlist of solid and hyphenated compounds, first, the parts-of-speech (hereafter POS) tagging of the BNC Sampler is used to identify all nouns of the corpus, e.g. <w NN1>, <w NNL 1> and <w NN2>. At this stage, non-compound entries and irrelevant items are removed from the study, e.g. units which may look like compounds but are not (*championship*, *chin chimney*), neoclassical compounds (*anthropology*, *biodegradable*) or simple lexemes in general (*journey*, *cranberry*). After this inspection, a preliminary list remains of 4,653 items.
For the retrieval of open N+N compounds, second, a choice has to be made because the format of these units places them very close to syntactic constructions (see 2.3), which makes their retrieval a key methodological decision. Depending on one's theoretical stance, a unit like *stone bridge* can be seen as a syntactic object or as a morphological one, and this obviously affects the results of any experiment. In this case, it was decided to initially retrieve all open constructions, and then apply filters to preserve only those with a sufficient load of meaning to be regarded lexemes.

The problem is that, unlike solid and hyphenated units, the open N+N constructions in the BNC Sampler do not carry a POS tag, which implies that the function *WordList* cannot be used for their retrieval. Instead, it was decided to employ the function *Concord*, for which POS tags have to be combined so as to guarantee that no N+N compound is overlooked. For the creation of the list of open N+N strings, hence, the *Concord* is operated by using all possible tag sequences until all entries are exhausted and the list completed.

After the preliminary files have been obtained, their results are merged into a list that includes all open N+N constructions and consists of 33,454 entries. The next stage goes manually through all 33,454 items for retrieval of the ones to be kept for the experiment. After the inspection of the units starting with *a*, *b* and *c*, 4,504 open constructions are retrieved from the BNC Sampler.

Once all open, hyphenated and solid constructions were retrieved, it was necessary to adapt their format accordingly. This required, for example, unifying the orthography and spelling of all units because the same lexeme can occur in singular or in plural, with or without hyphens, in which cases the meaning of the unit is the same, only it takes different word-forms. When this happens, the spelling of the word-form with the highest frequency is retained and the frequencies of the rest of units added to it. If all units have the same frequency, the spelling used here is the singular or, in its default, the most widespread form. As a token, (12) exemplifies a compound for which two variants were found: open (a) and hyphenated (b). Here, the open alternative has the highest frequency, so that spelling is kept and the two separate frequencies (16 and 2) added up[6].

(12) a. bog garden 16
b. bog-garden 2
c. bog garden 18

The final step involves the screening of entries that fall out of the scope of RDPs (see Levi 1978: 8) and units other than N+N compounds: synthetic compounds (13a), lexicalized formations (13b), exocentric compounds (13c) or N+N constructions that include reference to a proper name (13d).

(13) a. air freshener
b. chestnut
c. couch potato
d. Coronation opera

During this stage, 1,411 entries are removed from the experiment, after which 3,093 entries conform the definitive study corpus. The next phase inspects these units from a semantic perspective.

3.2. Semantic analysis

The steps in 3.1 provide all compounds for this study, but it is impossible to conduct the experiment on the entries as such given the formal makeup of N+N compounding. The reason is that this word-formation process operates by concatenating two nouns, and no other formal mark is left in the construction to indicate that derivation has taken place. As a token, the meaning of *acid attack* is ‘an attack that is made using acid’, but there is no signal in the N+N unit for the language user to notice that a new word has been created from *acid* plus *attack*.

The opposite happens in processes like affixation because the affix occurs only in the derivative, so it shows that a new lexeme has been created (e.g. the suffix -*able* in *renewable* indicates derivation from *renew*).

This means an added difficulty for the analysis of N+N compounds, since one further step has to be taken to properly perceive their semantics. A solution is to use a list of predicates like Levi's (1978), to build homogeneous categories of compounds, so that all units with analogue meanings are allotted to the same RDP. This step proves essential because, without it, the corpus is nothing more than a flood of N+N compounds that cannot be separated according to any semantic criterion.

Hence, with the study corpus at hand, all 3,093 entries are examined and each of them assigned to one RDP. This is done after observing the occurence of each item in the BNC Sampler, which is essential in that lexemes are examined in their natural context, not as isolated units. Such operation is carried out for each corpus entry and proves essential because, due to the lack of a verb, N+N compound may have various possible readings.

This semantic analysis naturally implies considering certain methodological issues. First, given the partial similarity of BE and MAKE for some units, it was decided to assign a unit under BE when its meaning is closer to 'composition', and to MAKE when it is closer to 'production'. This may be better illustrated by means of some examples: units like *air space*, *barber-surgeon* and *creator god* fall under BE because they do not imply a production process ('the air is space', 'the barber is a surgeon', 'the creator is a god'). By contrast, compounds like *alabaster tomb*, *brass pole* or *cotton shirt* carry an implicit sense of manufacturing ('the tomb is made of alabaster', 'the pole is made of brass', 'the shirt is made of cotton'), so they are included under MAKE.

Such a semantic analysis is performed on all corpus entries, after which it is possible to refer not just to N+N compounds in general but also to N+N CAUSE compounds or to N+N FROM compounds, a fundamental condition for the present study. The following are the occurrences of the N+N compounds across RDPs:

Predicate	Occurrences
ABOUT	602
BE	310
CAUSE	31
FOR	895
FROM	116
HAVE	188
IN	700
MAKE	161
USE	90
Total	3,093

Table 2. Distribution of entries across RDPs

Once these methodological steps have been followed, it is possible to tackle the experimental body of this paper.

4. The role of modifiers in N+N compounding

Linguistic research, it has been explained, has shown that the meaning of modifiers plays a dominant role in determining how the relationship between the modifier and the head will be understood and that, although the head is the principal constituent in terms of inflection, agreement and feature percolation (Lieber 1981), both members of the compound have a bearing on the ultimate meaning of the lexeme. In this section, the study corpus is used to illustrate how the semantics of N+N lexemes is highly influenced by the specific meaning of its modifier.

4.1. The picture of current tendencies

One of the aims of this paper is to check if the adscription of an N+N compound to one RDP or other is affected by its left-hand member. Looking at the corpus entries after the semantic analysis in 3.2, there seems to be a strong tendency for units with a given modifier to fall under the same RDP. In fact, it often seems that the modifier is more relevant to the overall semantics than the head itself, which goes against well-grounded assumptions about feature and attribute percolation (see Lieber 1981, Selkirk 1982). Observe the following corpus entries and their corresponding RDPs:

(14)	a. accident book	ABOUT	‘the book is about accidents’
	b. accident case	BE	‘the case is an accident’
	c. accident department	ABOUT	‘the department is about accidents’
	d. accident form	ABOUT	‘the form is about accidents’
	e. accident rate	ABOUT	‘the rate is about accidents’
	f. accident record	ABOUT	‘the record is about accidents’
	g. accident report	ABOUT	‘the report is about accidents’

The above are all corpus entries starting with *accident*. Six out of seven units (85.7%) belong in ABOUT, an unpredicted outcome because the heads of these units are very different among themselves (*book*, *department*, *record*, etc.). This list exemplifies that the mere occurrence of a particular modifier is central for the overall meaning of the compound, as only in (14b) the natural rewording requires BE instead of ABOUT. From these examples, it may be maintained that *accident* fosters a semantic reading of TOPIC for the units where it occurs. This view is corroborated by the CARIN model, where it is claimed that “[…] the ease with which the appropriate relation can be found depends on both the strength of the to-be-selected relation and on the strength of the alternatives” (Gagné and Shoben 1997: 81). One would perhaps expect various RDPs to occur for the above entries, based on the classical view that several meanings underlie N+N compounds and that alternative readings exist but, quite on the contrary, strong agreement is found among the various interpretations of these units, and one wonders why this is so. We come back to this point below in this section.

Example (14) has disclosed a link between the semantics of the heads of N+N units and that of their modifiers by focusing on elements with the same left-hand member. The opposite can also be done if we compare results by picking up all corpus entries where the right-hand member is the same, for example *book*. There are altogether fifteen compounds in the corpus whose head is *book* and, of these, six are included under ABOUT (40%), three under FOR (20%), three under HAVE (20%) and three under IN (20%)[7]. As it seems evident, these results clearly differ from the case of *accident*, as they provide a more uniform distribution of RDPs, which suggests that modifiers are more influential for the meaning of the compound than has been traditionally implied. These figures, nevertheless, derive from individual examples, so we may now explore the patterns of distribution of compounds to RDPs in the whole corpus to confirm these assumptions.

This can be done by grouping all items with the same modifier and then comparing to which extent the occurrence of one modifier influences which RDP surfaces for that unit. Out of all 3,093 corpus entries, 2,511 have a modifier that appears in at least another different entry. Each of these sets of compounds with common modifiers is hereafter referred to as a *cluster*; in contrast, 582 entries do not share their modifier and thus conform 1-item clusters. This is remarkable in that 81.1% of the data displays a tendency towards grouping, which indicates that an important proportion of the corpus size is influenced by the appearance or absence of X modifier within the lexeme.

This 81.1% of clustered units can be subdivided into groups of compounds with identical modifiers. By doing so, 415 clusters are obtained, two of which are (15) and (16):

(15) a. abortion issue
b. abortion law
c. abortion right

(16) a. advice bureau
b. advice centre
c. advice chair
d. advice column

The average of entries per cluster is 6.05, although an important number of clusters (51.5% of the total) have either two or three members, as in (15), which confirms Zipf's law that the frequency of any word is inversely proportional to its rank in frequency. In this case, the clusters with these two specific variables (having two or three items) comprise more than half the contents of the data.

An initial hypothesis of this paper is that the number of members of a cluster will influence their semantic interpretation, based on the assumption that the more members a cluster has the lower the proportion of them that will fall under the same RDP. Such premise derives from examples like the above, and can be confirmed by *Pearson product-moment correlation coefficient* (*r*), whereby it is possible to measure a correlation between two variables, *x* and *y*. The following formula allows calculating the degree of covariance between *x* and *y*, i.e. the strength with which a variable can be accounted for by another one. For these purposes, let *x* be the total number of members of a cluster and *y* the number of members of that cluster with the same RDP:

$$(17)\quad r = \frac{\sum Z_x Z_y}{N}$$

In *r*, the nearer the resulting value is to 1 or to –1 the closer the correlation between the two variables in a positive or a negative way (in a positive correlation, *x* augments as *y* does; in a negative one, *x* falls as *y* augments). After applying this formula to the corpus values, the resulting figure is 0.911, which confirms a particularly remarkable correlation between the two variables under study and is interpreted by saying that almost all entries *y* vary according to changes in *x*. This is a first sign that the number of members of a cluster influences how many of them embrace certain semantic relationships.

R is perhaps better appreciated through graphical output, the other side of the same coin. Chart 1 is a scattergram which allows observing areas of greater cluster build-up and confirms that most dots occur on the area of two and three variants, thus hinting at sections where modifiers are more influential. Each dot represents a cluster, which makes results visually noticeable[8]:

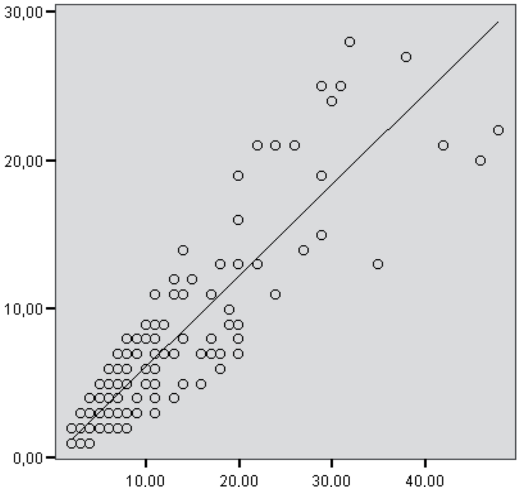

Chart 1. Cluster distribution according to their modifier

The cloud of dots stretching from the bottom left-hand corner (0 in the *x* and *y* axes) to the opposite end of the chart confirms that the positions of all values develop in an analogous manner in both axes, as dots with high *x* values have high *y* values, and vice versa. This stems from the high *r* value of the experiment (0.911) and implies that the two variables of the study robustly depend on one another. Also note that most clusters are crowded around values below 5 for both axes, which is why fewer dots occur in zones of higher membership, for example, in the upper right-hand corner. Moreover, a significant number of clusters appear next to the bottom left-hand corner and they disperse as we go up in the values, in such a way that dots are far from each other towards the end of the line.

In a similar vein, clusters with over 20 units on the *y* axis (those at the top of the chart) tend to appear more towards the right as we go up the scale, which means that the more units a cluster comprises the fewer of them will belong to the same RDP. If no correlation occurred between *x* and *y*, the individual dots would appear towards the top but not towards the right, with a scattered layout. They would, essentially, occur in separate and irreconcilable positions. Also note that, although this scattergram has a high *r* value, there are outliers too, for example when a cluster appears low and towards the right and when a cluster is high towards the left.

Once the results of *r* have been interpreted, we may square its value and obtain r^2, the *coefficient of determination*, which reveals the percentage of explained variation between *x* and *y*. This coefficient makes it possible to predict outcomes based on information which is already available, in this case the corpus entries (*x*), their adscription to clusters and RDPs (*y*), and their frequency figures.

In this experiment, r^2 is 0.83, a significant figure that implies that 83% of the variability in the number of different clusters can be explained by how many compounds in each of them share a given RDP. In other words, 83% of the changes of a shared RDP is accounted for by the total number of elements carrying a common modifier. This result implies that the corpus entries are highly influenced in their adscription to a given RDP by the modifier they carry, so that the occurrence of one modifier or another determines the semantic analysis of the entire unit[9].

This finding is especially relevant in that it reveals a *distribution effect* in compounds: clusters with not many units behave neutrally with regard to RDPs but, as more items fall into a cluster of similar modifiers, the tendency increases for a lower proportion of it to belong to the same RDP. Put differently, the more units a cluster has the lower the chances that other similar members are attracted towards them for an analogous semantic reading. Hence the fact that in groups with fewer members the readings of compounds will tend to be analogical among them, as shown also by the results in Štekauer's research on meaning predictability: "While there are many potential readings of novel, context-free naming units, it is usually only one or two that are significant in terms of meaning predictability" (2005: 257).

These results seem to go against certain aspects of the CARIN model (Gagné 2002, Gagné and Shoben 1997), but it must be stressed that this article is founded on corpus-based experiments, while Gagné's tests are carried out from a psycholinguistic perspective and incorporate human interaction with data. In her model, for example, the researcher "[…] manipulates a factor (or factors) that he/she thinks might influence processing and then examines whether there is an observable change in a dependent variable" (Gagné 2009: 260), an option which is neither viable nor relevant here; it is, therefore, reasonable that the results here may partly contradict those from Gagné, even if they have points in common.

After the observation of broad trends, the next step comes closer to specific clusters to examine their behaviour. Table 3 displays the top ten clusters by descending order of membership, as well as the total number of members of the cluster, the number of members with that modifier which share the same RDP (*Shared RDP*) and the percentage of members which fall under the same RDP:

Modifier	Total members	Shared RDP	
		Total	%
computer	48	22	45.83
air	46	20	43.47
business	42	21	50
community	38	27	71.05
car	35	13	37.14
Christmas	32	28	87.50
church	31	25	80.64
county	30	24	80
country	29	25	86.20
city	29	19	65.51
		Mean %	64.73

Table 3. Top ten clusters according to membership

The clusters with most lexemes are those where the modifiers are *computer* (48 items), *air* (46 items) and *business* (42 items). The RDPs involved in the clusters above are ABOUT (*business*, *computer*) and IN (*air*, *car*, *Christmas*, *church*, *city*, *community*, *county*, *country*), in accordance with Levi's (1978: 77) statement that they are fertile predicates both in terms of attestation and of productivity (see Fernández-Domínguez 2009: 154–168).

One relevant fact is the substantial number of units per cluster (36 on average), a high figure considering the corpus size. This ratio indicates that, once a modifier has been used for a compound, it is easy for speakers to employ it again in the future, and this promotes further uses of that unit. Table 3, thus, confirms that occurrence of a given modifier is a crucial factor for the overall meaning of a unit and, in turn, for its adscription to a RDP. A remarkable example is the case of *Christmas*, a cluster with 32 compounds, 28 of which fall under ABOUT, which indicates that 87.5% of the N+N compounds with that modifier have very close meanings, presumably due to their modifier in common.

The key piece of information, however, is provided by the last column because it reveals which percentage of members of a cluster belongs to the same RDP. As presumed in Chart 1, the tendency can be observed for clusters with many compounds to have lower percentages of shared RDP, with figures that reach a proportion of almost 90% when membership is not very high (*Christmas* and *country*). There are, nevertheless, cases where the situation is the opposite, and clusters of shared modifiers with many members show a high percentage of shared RDP, as in *community* (71.05%).

Part of the reasoning for this is that a noun like *computer* has more available slots according to Allen's (1978) VRC, so that it is easier for slots of other nouns to suitably match into this noun to create a new compound. A supporting piece of evidence for this statement is the fact that the cluster of units with the modifier *computer* features up to seven different RDPs:

(18) a. computer analysis BE
b. computer terminal ABOUT
c. computer disk FOR
d. computer aid FROM
e. computer store HAVE
g. computer company MAKE
h. computer people USE

By contrast, where the percentage of shared RDP is higher, this means that the modifier in question has fewer accessible slots and, because of this, it is more difficult for another noun to successfully fit into one of them, as in *Christmas*. When this happens, more members within the set share their RDP, with the result that there is less variation in the range of occurring predicates; in the case of *Christmas*, for instance, only three different RDPs are found (BE, FOR, IN), a sharp contrast to *computer*. From this, a preliminary conclusion emerges that availability of Allen's (1978) slots increases as a cluster comprises more compounds. This is implied in the above results, and is further considered below.

Although there are values of various types, Table 3 confirms our assumption that homogeneity within clusters of N+N compounds is lower as more members occur. Even though the mean percentage of shared RDP is 64.73%, it is apparent that clusters where the modifier is *country* or *church* contribute more positively to this value than those where the modifier is *computer* or *air* (compare the difference in the percentages of these clusters in Table 3). The same idea is portrayed in Chart 2, where the horizontal axis represents the number of members of a cluster and the vertical axis the percentage of members under the same RDP[10]:

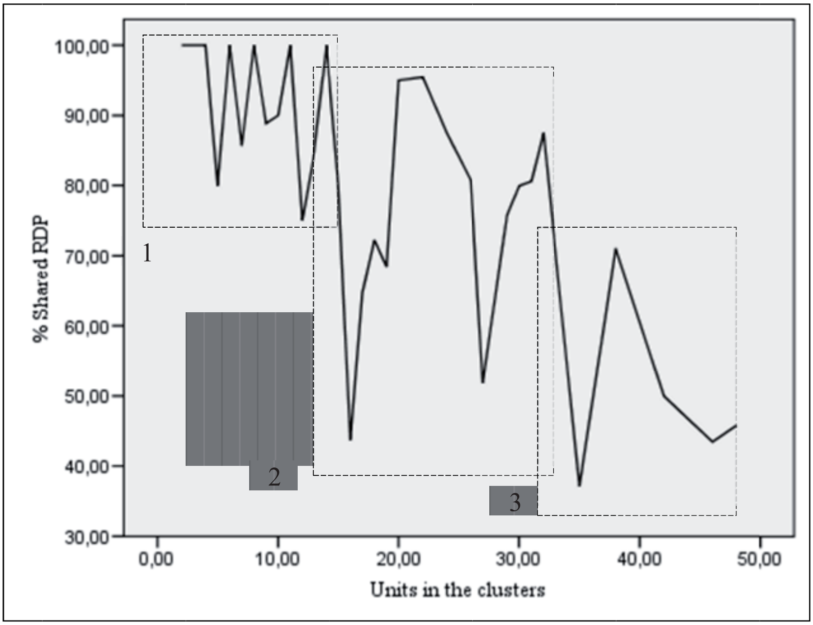

Chart 2. Evolution of shared RDP for increasing membership in the corpus

On the basis of the line on Chart 2, it can be asserted that the highest percentages occur in clusters with fewer members, and that there is a point at which the figures stabilize and remain around 50%. Three general areas can be distinguished on the chart (each marked with dashes): area 1 comprises clusters with less than 15 members and is noteworthy in that its percentages are always very high (no less than 70%), which in itself denotes the positive significance of low membership for a cluster. Next, the clusters in the area 2 have between 17 and 35 members, and are characterized by constant ups and downs in the development of the line. This implies a stark contrast between clusters where the rate of shared RDP is above 89% and others which fall to 40%, a reflection of inconsistencies as more compounds are considered for the experiment. The percentages in zone 3, finally, drop drastically from 75% to 35% and, after one rise, remain around 45%. The irregularity in zone 3 shows similarities with that in zone 2, where inconsistencies are also frequent, with the difference that the global percentages are lower in zone 3 than in 2, in line with the point under discussion here.

Chart 2, overall, depicts an accurate match between the staggered fall in the percentages of shared RDP and the number of members of a cluster, in such a way that clusters with not many members display high percentages, and precision decreases as more units are added. Furthermore, the line on the chart suggests that percentages are less stable in clusters with many members than they are in clusters with fewer members, as evidenced by the inconsistency of the line towards the right-hand side of the chart.

In sum, this section has explored the corpus in connection with groups of N+N compounds that share the same modifier vs. the percentage of them that are analyzed under the same RDP. It has been shown that there exists an obvious link between these two variables, as has been corroborated not only by observation of the corpus figures, but also by tests using *r*, r^2 and Charts 1 and 2, which are conclusive as regards the connection between *x* and *y*.

4.2. Summary

In 4.1, this study has surveyed which factors may affect the semantics of N+N compounds and, after some tests, it has confirmed that modifiers represent a powerful feature of N+N units and that they influence, at least to the same extent as heads, the overall meaning of these constructions.

In accordance with the above experiments, it can be asserted that theoretical models like Allen (1978), despite their strong generative influence, are well-grounded as far as the above empirical results are concerned. The corpus entries vary in their allocation to different RDPs in a manner that is closely related to the number of compounds that occurs in each cluster. Allen's principle of slot filling, therefore, seems to interact with the quantity of cluster membership, so that a higher figure of members leads to more empty slots, and vice versa. This explains why clusters with more members display a more varied array of RDPs, while clusters with few members tend to be analyzed under just one or two RDPs (see 4.1).

It is, then, to be expected that changes in the modification of a given head will make the compound in question susceptible to present a different semantic reading and that, for this very reason, the processing of N+N units is highly dependent on which the left-hand member is (see Gagné 2002).

5. Conclusions

The main aim of this paper has been to demonstrate that the overall meaning of N+N compounds is influenced not only by their heads, but also, and perhaps even to a greater extent, by their modifiers. Section 4.1 explained how the global meaning of an N+N compound is strongly bound up with which modifier it carries. There, a number of statistical tests confirmed that, where a given modifier appears across different N+N units, this highly determines the meaning of the entire unit, in this case by using Levi's (1978) RDPs. Both numerical (Table 3) and graphical evidence (Charts 1 and 2) support this statement.

The outcome of these tests leads us to a number of generalizations:

(i) N+N compounding is a highly speaker-oriented process, as proved by the fact that it is often possible to interpret these units in various different ways. Precisely because of this, it is usually easier to produce than to decode N+N compounds, which places the weight of disambiguation on the part of the hearer.

(ii) Modifiers are fundamental to the meaning of N+N compounds. Counter to the trend for which heads are the ultimate node in the structure of compounds (Selkirk 1982, Murphy 1988, 1990), it has been here shown that modifiers are at least as crucial as heads for the semantics of a compound, as they decidedly have a bearing on which RDP underlies the construction.

(iii) The semantics of N+N compounds is readily influenced by how many compounds share the same modifier. This experiment has found that there is an indirect relationship between the number of units with the same modifier and their adscription to RDPs so that, the more N+N compounds share a modifier, the lower the proportion of them that will be analyzed under the same predicate. This can be explained by Allen's (1978) compound slots, because the fact that a cluster has more compounds means having more different slots, and this leads to a higher variety of semantic readings and, hence, to a lower compound per cluster ratio.

An asset of these experiments is that they are technically applicable to other corpora, for the reason that statistical operations make it possible to employ variables that can be then replaced by any other figure, regardless of aspects like corpus size, hapaxes or type frequency. This article, I hope, is a contribution to the understanding of N+N compounds and the role of modifiers within them.

Notes

* I must acknowledge the help of several colleagues during the preparation of this article: Juan Santana (University of Granada) and Christina Gagné (University of Alberta) pointed to readings of special relevance for compound processing, Juan Antonio Becerra (University of Jaén) offered advice on statistical operations, and Salvador Valera (University of Jaén) and Pavol Štekauer (Pavol Jozef Šafárik University) gave me feedback on the content of the investigation. Needless to say, the remaining errors are only mine.

[1] Scholars have distinguished diverse types of words, such as phonological words, lexical words, grammatical words, orthographical words, word-forms, etc. Particularly, word-formation is said to be concerned with *naming units*, *lexemes* or *listemes*, depending on the author one turns to (see Di Sciullo and Williams 1987: 3, Lipka 2002: 72–73). In this paper, the term *word* is used in a simpler sense as a synonym of *lexeme*, and any use different from this one is noted.

[2] It is arguable, for instance, whether synthetic compounds (like *coffee-maker*) or neoclassical compounds (like *neurolymphomatosis*) are indeed composed of various smaller *bases*, or whether they should be analyzed as comprising a more complex inner structure (see Allen 1978: 246, Selkirk 1982: 244–252).

[3] Gagné and Shoben show, as a token, that the fact that *mountain* fosters a LOCATIVE relation highly influences that N+N compounds where *mountain* is a modifier tend to be interpreted through that reading than through any other one. This explains why participants find it easier to interpret *mountain bird* ('a bird in the mountains') than *mountain magazine* ('a magazine about mountains').

[4] Some of the tenets in Lees (1960) are already implicit in Bühler's (1934) concept of *Sachsteuerung*, i.e. how things are mentally represented for understanding. Bühler maintains that the extralinguistic context (the *Zeigfeld*) is essential for the speaker to be able manipulate an utterance. In the case of N+N compounds, this is especially significant in that the comprehension of these units requires a certain amount of encyclopaedic knowledge. That is why the lack of contextual information often leads to meaning ambiguity and, in turn, to an inability to fully understand N+N compounds.

[5] Under SPATIAL in the *Primary Declensional Categories*, for example, there are LOCATION (*workshop*), TEMPORAL (*evening edition*), GOAL (*Belgrade train*) and ORIGIN (*Belgrade train*). As for *Secondary Declensional Categories*, SPATIAL includes 21 variants, among them ANTERIORITY (*anteroom*), POSTERIORITY (*postwar*), TRANSESSION (*overseas*), OPPOSITION (*anti-aircraft*) or PERLATION (*throughway*).

[6] This paper has also contemplated the possibility of having polysemous cases among compounds. Thus, in applying the above procedure, the context of each occurrence was checked to confirm that they all have the same meaning, and it can be said that no instance of polysemy occurs in the study corpus.

[7] ABOUT: *accident book*, *anatomy book*, *area book*, *bird book*, *club book*, *cookery book*; FOR: *action book*, *children's book*, *control book*; HAVE: *address book*, *carol book*, *control log book*; IN: *bank book*, *cellar book*, *childhood story book*.

[8] Note that there are in fact more clusters than circles are displayed on this figure. The reason is that, when creating the chart in ©SPSS 15.0, clusters with identical values overlap, so what looks like one circle is often a series of them placed on top of each other. Observe, however, that where the lines of the figure are thicker, this indicates that various circles are superimposed, which happens especially in the areas closer to 0 in both axes. The total number of clusters in Chart 1 is 415.

[9] A regression line has been added in Chart 1 for easier observation of r^2. In a perfect correlation between x and y (value 1), the line would go from one corner right to the opposite one, meaning that there is 100% correlation between both variables. Here, the fact that r^2 is 0.83 implies that x corresponds to y to a high extent, hence the line is almost straight, thus hinting a close relationship between the two groups.

[10] This chart considers all clusters in the corpus, therefore the values on the horizontal axis range from 2 to 48 (the minimum and maximum number of members in the clusters, respectively), and the vertical axis displays the percentages of members with the same RDP.

References

Aarts, Bas, David Denison, Evelien Keizer and Gergana Popova (2004) 'Introduction: the nature of grammatical categories and their representation'. In: Aarts, Bas, Daniel Denison, Evelien Keizer and Gergana Popova (eds.) *Fuzzy Grammar: A Reader*. Oxford: Oxford University Press, 1–28.

Abdullah, Nabil and Richard A. Frost (2007) 'Rethinking the semantics of complex nominals'. *Proceedings of the 20th Conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence*. Berlin and Heidelberg: Springer-Verlag, 502–513.

Adams, Valerie (1988) *An Introduction to Modern English Word Formation* (2nd ed.). London: Longman.

Allen, Margaret R. (1978) *Morphological Investigations*. University of Connecticut, CT: Storrs.

Bauer, Laurie. 1998 'When is a sequence of two nouns a compound in English?' *English Language and Linguistics* 2, 65–86.

Bauer, Laurie (2006) 'Compounds and minor word-formation types'. In: Aarts, Bas and April McMahon (eds.) *Handbook of English Linguistics*. Oxford: Basil Blackwell, 483–506.

Bauer, Laurie and Rodney D. Huddleston (2002) ‘Lexical word-formation’. In: Huddleston, Rodney D. and Geoffrey K. Pullum (eds.), *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 1621–1723.

Beard, Robert (1995) *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word-Formation*. Albany, NY: State University of New York Press.

Bisetto, Antonietta and Sergio Scalise (1999) ‘Compounding: Morphology and/or syntax?’. In: Mereu, Lunella (Ed.) *Boundaries of Morphology and Syntax*. Amsterdam: John Benjamins, 31–48.

Bolinger, Dwight (1961) *Generality, Gradience and the All-or-None*. ’s-Gravenhage: Mouton de Gruyter.

Botha, Rudolf P. (1968) *The Function of the Lexicon in Transformational Generative Grammar*. The Hague: Mouton

Bühler, Karl (1934) *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Jena: Gustav Fischer.

Carstairs-McCarthy, Andrew (1992) *Current Morphology*. London: Routledge.

Corbin, Danielle (1997) *Le Lexique Construit. Méthodologie d’Analyse*. Paris: Armand Colin.

Di Sciullo, Anna Maria and Edwin Williams (1987) *On the Definition of Word*. Cambridge, MA and London: MIT Press.

Downing, Pamela (1977) ‘On the creation and use of English compound nouns’. *Language* 53, 810–842.

Fernández-Domínguez, Jesús (2009) *Productivity in Word-Formation. An Approach to N+N Compounding*. Bern: Peter Lang.

Frege, Gottlob (1970[1903]) *Grundgesetze der Arithmetik*, vol. ii. In: Geach, Peter and Max Black (eds.) *Translations from the Philosophical Writings of Gottlob Frege*, section 56. Oxford: Basil Blackwell.

Gagné, Christina L. (2002) ‘Lexical and relational influences on the processing of novel compounds’. *Brain and Language* 81, 723–735.

Gagné, Christina L. (2009) ‘Psycholinguistic perspectives’. In: Lieber, Rochelle and Pavol Štekauer (eds.) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 255–271.

Gagné, Christina L. and Edward J. Shoben (1997) ‘Influence of thematic relations on the comprehension of modifier-noun combinations’. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23(1), 71–87.

Gagné, Christina L. and Edward J. Shoben (2002) ‘Priming relations in ambiguous noun-noun compounds’. *Memory and Cognition* 30(4), 637–646.

Gagné, Christina L. and Thomas L. Spalding (2006a) ‘Using conceptual combination research to better understand novel compound words’. *SKASE Journal of Theoretical Linguistics* 3(2), 9–16. http://www.pulib.sk/skase/Volumes/JTL06/2.pdf. Accessed 30 December 2009.

Gagné, C. L. and Thomas L. Spalding (2006b) ‘Relation availability was not confounded with familiarity or plausibility in Gagné and Shoben (1997): Comment on Wisniewski and Murphy (2005)’. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32, 1431–1437; discussion 1438–1442.

Girju, Roxana, Dan Moldovan, Marta Tatu and Daniel Antohe (2005) ‘On the semantics of noun compounds’. *Computer Speech and Language* 19(4), 479–496.

Hatcher, Anna G. (1960) ‘An introduction to the analysis of English noun compounds’. *Word* 16, 356–373.

Jespersen, Otto (1942) *A Modern English Grammar on Historical Principles*. Part VI, *Morphology*. London and Copenhagen: Munksgaard.

Kay, Paul and Karl E. Zimmer (1976) ‘On the semantics of compounds and genitives in English’. Paper presented at the Sixth California Linguistics Association, San Diego State University.

Körtvélyessy, L. (2008) *Vplyv sociolingvistických faktorov na produktivitu v slovotvorbe*. Ph.D dissertation.

Labov, William (1973) ‘The boundaries of words and their meanings’. In: Bailey, Charles-James and Roger W. Shuy (eds.) *New Ways of Analyzing Variation in English*. Washington, DC: Georgetown University Press, 340–373.

Lees, Robert B. (1960) *The Grammar of English Nominalizations*. Bloomington, IN: Indiana University Press.

Lees, Robert B. (1970) 'Problems in the grammatical analysis of English nominal compounds'. In: Bierwisch, Manfred and Karl E. Heidolph (eds.) *Progress in Linguistics*. The Hague: Mouton de Gruyter, 174–186.

Levi, Judith N. (1978) *The Syntax and Semantics of Complex Nominals*. New York, NY: Academic Press.

Li, Charles (1971) *Semantics and the Structure of Compounds in Chinese*. Doctoral dissertation, University of California.

Libben, Gary (2006) 'Why study compound processing? An overview of the issues'. In: Libben, Gary and Gonia Jarema (eds.) *The Representation and Processing of Compound Words*. Cambridge: Cambridge University Press, 3–22.

Lieber, Rochelle (1981) *On the Organisation of the Lexicon*. Doctoral dissertation, Indiana University Linguistics Club.

Lieber, Rochelle (2004) *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.

Lipka, Leonhard (2002) *English Lexicology: Lexical Structure, Word Semantics and Word-Formation*. Tübingen: Gunter Narr.

Murphy, Gregory L. (1988) 'Comprehending complex concepts'. *Cognitive Science* 12, 529–562.

Murphy, Gregory L. (1990) 'Noun phrase interpretation and conceptual combination'. *Journal of Memory and Language* 29(3), 259–288.

Payne, John and Rodney D. Huddleston (2002) 'Nouns and noun phrases'. In: Huddleston, Rodney D. and Geoffrey K. Pullum (eds.) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 323–523.

Plag, Ingo (1999) *Morphological Productivity: Structural Constraints in English Derivation*. Berlin and New York, NY: Mouton de Gruyter.

Plag, Ingo, Gero Kunter, Sabine Lappe and Maria Braun (2008) 'The role of semantics, argument structure and lexicalization in compound stress assignment in English'. *Language* 84, 760–794.

Rosenbach, Anette (2006) 'Descriptive genitives in English: a case study on constructional gradience'. *English Language and Linguistics* 10, 77–118.

Rosenbach, Anette (2007) 'Emerging variation: determiner genitives and noun modifiers in English'. *English Language and Linguistics* 11, 143–189.

Scott, Mike (2004) *Oxford Wordsmith Tools, version 4.0.0.376*. Oxford: Oxford University Press. http://www.lexically.net/wordsmith/version4/index.htm. Accessed 25 April 2009.

Selkirk, Elisabeth O. (1982) *The Syntax of Words*. Cambridge, MA and London: MIT Press.

Shimamura, Reiko (1998) 'Lexicalization of syntactic phrases: the case of genitive compounds like *woman's magazine*'. http://coe-sun.kuis.ac.jp/coe/public/paper/outside/ shimamura2.pdf. Accessed 24 February 2010.

Soegaard, Anders (2005) 'Compounding theories and linguistic diversity'. In: Frajzyngier, Zygmunt, Adam Hodges and David S. Rood (eds.) *Linguistic Diversity and Language Theories*. Amsterdam: John Benjamins, 319–337.

Štekauer, Pavol (2000) *English Word-Formation. A History of Research (1960–1995)*. Tübingen: Gunter Narr.

Štekauer, Pavol (2005) *Meaning Predictability in Word-Formation: Novel, Context-Free Naming Units*. Amsterdam: John Benjamins.

Štekauer, Pavol (2009) 'Meaning predictability of novel context-free compounds'. In: Lieber, Rochelle and Pavol Štekauer (eds.) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 272–297.

ten Hacken, Pius (2009) 'Early generative approaches'. In: Lieber, Rochelle and Pavol Štekauer (eds.) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 54–77.

Warren, Beatrice (1978) *Semantic Patterns of Noun-Noun Compounds*. Acta Universitatis Gothoburgensis: Göteborg.

Wisniewski, Edward J. and Gregory L. Murphy (2005) 'Frequency of relation type as a determinant of conceptual combination: A reanalysis'. *Journal of Experimental Psychology: Learning, Memory and Cognition* 31,169–174.

Zimmer, Karl E. (1971) 'Some general observations about nominal compounds'. *Working Papers on Language Universals* 5, 1–21.

Zimmer, Karl E. (1972) 'Appropriateness conditions for nominal compounds'. *Working Papers on Language Universals* 8, 3–20.

Jesús Fernández-Domínguez Jesús Fernández-Domínguez received an MA and PhD in English linguistics from the University of Jaén (Spain), where he teaches subjects related to English linguistics. His work has been published in various peer-reviewed journals as well as in the monograph *Productivity in word-formation. An approach to N+N compounding* (Peter Lang, 2009). Dr. Fernández-Domínguez has delivered talks at a number of international meetings and has been the co-convenor of workshops at major conferences. He currently specializes in English word-formation and terminology.

Address: Jesús Fernández-Domínguez, Universidad de Jaén, Campus Las Lagunillas, s.n., 23071 Jaén, Spain. [email: jesusferdom@gmail.com]