Empirical and theoretical aspects of phrasal compounds:
against the "syntax explains it all" attitude

Carola Trips
University of Mannheim
ctrips@rumms.uni-mannheim.de

1. Introduction

In the literature, some attention has been paid to phrasal compounds (PC) especially since they seem to pose problems for a theory of morphology that is based on a classical generative, syntactocentric framework. If instances of Noun+Noun compounding (NNC) are compared with instances of PCs (examples in (1) vs. (2)) structural differences become evident immediately:

(1)   a.   peanut butter
        b.   love letter

(2)   a.   She also knows that the media tendency to lump together women singer-songwriters in a *"gee whiz, gosh, women are now making it'' syndrome* is patronising, if not pernicious. (BNC, A7S190)
        b.   Bombay-based Anil put India's failure to exploit its manpower and mind power and its lack of excellence in sport, economics and the arts down to a *"Learn what is there and don't question it'' attitude* (BNC, HAE4088)

What makes these compounds so special is that the left-hand member is a complex, maximal phrase: as in the examples given above, it can be a whole sentence like an IP (or CP depending on the analysis), which clearly sets them apart from NNCs, the left-hand member of which is non-phrasal and thus an entity on the word level.

Concerning the theoretical analysis of PCs, it has been shown that they are problematic in a generative framework, which assumes that on the one hand language can be clearly divided into a lexicon and a grammar and that on the other hand syntax determines the combinatorial properties of phonology and semantics (Chomsky:1965, Chomsky:1981). *The No Phrase* constraint (Botha:1981) and the *Lexical Integrity Hypothesis* (Lapointe:1980) are a result of these assumptions on the level of morphology which means that PCs will always violate these constraints as they show that syntactic phrases do have access to word structure. I will show that an analysis based on the semantics of PCs and allowing for a parallel production and processing of semantic, phonological and semantic structure like Jackendoff's model of Parallel Architecture (PA, Jackendoff 1997, 2002, Culicover & Jackendoff 2005) fares much better in this respect than analyses hitherto proposed (e.g. Botha 1981, Gallmann 1990, Lieber 1992, Wiese 1996, Ackema & Neleman 2004, Štekauer & Lieber 2009).

The aim of the paper is twofold: on the one hand I am going to fill a gap in the literature and present a qualitative and quantitative analysis of PCs found in the *British National Corpus* (BNC) which shows which types of PCs actually occur in English. This also includes an analysis of the distribution patterns of PCs across demographic and textual features such as mode, register (e.g. academic prose, spoken conversation) and text domain. This analysis will allow us to gain further insights into the question of why PCs are built at all by speakers/writers and why they are sometimes preferred over other options. Since in my opinion this relates to their morphopragmatic character, or if you will, their expressive flavour, I will discuss Meibauer's (2007) empirical study in this context and compare his results with mine. I will further discuss his approach based on his study of German PCs

and show that it should be integrated into a new approach, which I will introduce in the last part of the paper. As mentioned above, my analysis will be based on generative principles but part company with standard generative approaches to PCs since in my opinion they can be better explained if a conceptual-semantic structure as basis is assumed.

The outline of the paper is as follows: Section two describes the PCs found in the BNC and classifies the different types. In section three the distribution of PCs is investigated in terms of the demographic and textual features mentioned above. Meibauer's (2003, 2007) interesting observations concerning the morphopragmatic character of PCs will be taken into account here. Section four introduces a new analysis in the framework of Jackendoff's model of Parallel Architecture (PA). Section five concludes.

2. A qualitative analysis of phrasal compounds in English

The collection of PCs was gathered by exploiting the BNC. I chose to use the *BNCweb* via the Lancaster interface since it allows the user to work with the *cqp* (corpus query processor) language in a convenient way. Further, it provides statistical information like the distribution of the phenomenon across categories, e.g. mode, text type, age of author etc. Before defining the corpus query, it was necessary to check how PCs are actually spelt, i.e., how the phrasal part of these compounds is represented in written speech (see section 3). I found that in English, or at least in the BNC, the phrasal part of PCs predominantly occurs in the form of a quotation. Although I also found PCs where the phrasal parts are separated by dashes, they were not as frequent as those presented in quotations. Since at present I could not find a technical solution to the problem of covering all hits for PCs with dashes, I confined the search to quotations. I defined the query as follows: "search for any string that ends in an N and is preceded by 3 to 10 words which are embraced by quotation marks". The results gained by this query search was then manually checked, non-PCs were weeded out, and true PCs were categorised according to the word categories of their non-head and head using *grep* (global regular expression print) and other unix shell tools.

In the following, I will classify the total amount of instances that can be called (true) PCs into two broad classes, non-verbal and verbal PCs. Non-verbal PCs are those that do not contain verbal material in their phrasal non-head, whereas verbal PCs are those that do. I decided to opt for these two classes since, as I will show below, there is a relevant semantic difference between PCs containing a verb and PCs that lack verbs. In section four I will provide theoretical support for this assumption.

Within these two broad classes a number of different patterns occur, I will describe these patterns in the tables below and give some examples[1]. What will become evident is that in English there is a plethora of possibilities to modify nominal heads in PCs. This does not come as a surprise and has been found in other Germanic languages as well as e.g. in German (Meibauer 2003, 2007).

In Table 1 and 2 the patterns for the non-verbal type are presented. In the left-hand column the patterns, presented by the part of speech tags of the BNC, are given, in the right-hand column relevant examples of that pattern are given. Table 1 and 2 present the patterns with the highest frequencies in the corpus, Table 3 further below presents marginal, less frequent, patterns.

The first pattern is the *Nominal-Preposition-Nominal* (Nom-Prep-Nom) one, also including other types of nominals: The first and/or the second noun can also be modified by determiners and adjectives, and the noun can also be a proper noun. Preposition stands for any preposition except of the preposition *of*, which is part of a pattern that builds its

---

1   All examples appear in the original graphic representation of the BNC.

own subclass (see below). As can be seen from the second column, the examples found could again be classified into several types, e.g., lexicalised (in the sense of having been stored in the mental lexicon as a unit), and non-lexicalised ones, and the former type again into titles, idioms, cliches, etc. For the moment however, I will refrain from doing so since it presupposes an interpretation which I would like to provide later on. Apart from the nominals occurring in first position in these phrasal non-heads, adverbs can occur as for example *the "away from oil" policy.* Prepositions can also be found in the rather marginal pattern *Adj-Prep-Adj* as in *a "medium to high" probability*, and the pattern *Adj-Prep-Nom* as in *these "kind to hair" curlers* (so in short in AdjPs as non-heads of PCs).

As mentioned above, another highly frequent pattern is the *Nominal-of-Nominal* one containing the preposition *of.* Here again, subtypes can be identified: the first and/or the second noun can be modified by a determiner or adjective, and either noun can occur as proper noun. Other phrases can contain the preposition of too (see the example *the "out of touch" policy* where the adverb is the head of the phrase).

Table 1: Patterns of non-verbal PCs in the BNC, part I (most frequent patterns)

Nom-Prep-Nom	a "cost per case" basis, the "Brothers in Arms World", "Highway to Hell" album, the "toe in the water" stage, "milk for Spain" appeals, the "sex in shiny packets" literature, this "gentle rain from heaven" process, the "famous for fifteen minutes" type, a "medium to high" probability
Num-Prep-Num	the "first in last out" policy, a "ten to two" position, the "two for the price of one" sales
Adv-Prep-Nom	the "away from oil" policy, the "straight for English" policy, a "once upon a time" approach, the "always on the top" option
Adj-Prep-Nom	these "kind to hair" curlers, the "tieless in Soho" fashion, the "Famous for fifteen minutes" type, a "free for life" card, the "ready for action" look
Prep-NomP	a "with a run" proviso, this "at a glance" guide, the "In such a night" speech, a "by appointment only" notice, the "around the next corner" syndrome, the "of no consequence" category
Nom-*of*-Nom	the "language of thought" thesis, the "Prince of Thieves" film, a "state of the nation" novel, a "sword of Damocles" hanging, the "representational theory of mind" sense, the "Dream of Blue Turtles" album, a "top end of the market" service, a "bunch of sixes" opportunity
Adv-*of*-Nom	the "out of touch" policy

Table 2: Patterns of non-verbal PCs in the BNC, part II (most frequent patterns)

Nom-*and/or*-Nom	a "chicken and egg" situation, the "year and a day" rule, the "Jekyll and Hyde" behaviour, the "Beowulf and brooches" approach, the "Copyright and performing Right" article, the "Race and Poll Tax" workshop, the "warts and all" school, the "peace and goodwill to all men" attitude, a "sun, sea, sex and sangria" story
Pron-*and/or*-Pron	the "them and us" syndrome, the "I and Thou" relationship
Adv-*and/or*-Adv	the "little and often" principle, the "then and now" variety
Prep-*and/or*-Prep	a "before and after" basis

The next pattern showing a high frequency of occurrence is the *Nominal-and/or-Nominal* pattern where two nominals are conjoined by *and/or*. These nominals either have the shape of nouns or proper nouns, the former of which can again be modified by determiners and/or adjectives, or occur as personal pronouns as in *the "them and us" syndrome*. Examples like *the "peace and goodwill to all men" attitude* and *a "sun, sea, sex and sangria" story* further show that postnominal modification by a PP and the listing of nouns is also found (and hence possible). Other subtypes are phrases where two adverbs (*the "then and now" variety*) or prepositions (*a "before and after" basis*) are conjoined.

In Table 3, marginal patterns are presented:

Table 3: Patterns of non-verbal PCs in the BNC (marginal patterns)

List of NPs (with and without punctuation)	the "no alcohol, no limits" campaign, a "first author, last author" citation, a "one member, one vote" system, a "no pain no gain" position
NP-*only*	a "one attempt only" scenario, the "Black women only" space, a "personal use only" agreement
Phrases showing ellipses of words (indicated by ...)	the "one--pause--two" sequences, the "if ... only" form, the "if ... then" technique
Det-Adj-Nom	the "not enough memory" line, those "Three Blind Mice" exercises, a "no first use" pledge

Here we find the pattern *Nominal-as-Nominal* as for example *the "little and often" principle*, phrases of PCs which contain a list either indicated by commas or not as in *the "no alcohol, no limits" campaign* or in *a "no pain no gain" position*, the *NP-only* pattern where a full NP is modified by a postnominal only denoting 'no one except one particular thing or person' as in *a "personal use only" agreement*, and the pattern where the phrasal part shows ellipsis indicated by "..." as in *the "if ... then" technique*. Interestingly, the pattern where a full NP consisting of a determiner, adjective and noun instantiates the phrasal part of the PC is rarely found, contrary to intuition as for example in *the "not enough memory" line*. That is why it is also found among the marginal patterns here.

Next, I am going to discuss the verbal type of PCs in the BNC. I will start out with the VPs as phrasal heads, the relevant patterns of which are given in Table 4:

Table 4: Patterns of verbal PCs in the BNC: VPs

Verb(+Obj)-*and/or*-Verb(+Obj)	a "wait and see" mentality, a "make or break" time, a "Bring and Buy" coffee, a "take it or leave it" attitude, a "Come and Try It" day, the "see and be seen" generation, the "don't drink and drive" message
Verb-Object (including phrasal verbs)	a "Guess the Weight" cake, the "Cut the Crap" LP, a "wait a minute" gesture, a "think of a number" game, a "search personal history" category, the "show the shirt" routine, the "Try your strength" machine, a "follow my leader" type
Verb-prep-NP	the "Reach for the Sky" Appeal, a "smack in the face" offer, a "wait for mail" situation
Verb-Object-Modifier	a "slow it down" mode, the "get your hands dirty" philosophy, the "recycle at all costs" policy, "could try harder" comment
V-ing-Object	that "powdering my nose" act, the "Retaining the angle" exercises, the "Rocking The Forest" mini-LP
Verb-*to*-Inf	a "return to learn" entitlement, a "need to know" basis
List of VPs (with and without commas)	the "hear no evil, see no evil" brigade, the "won't pay, can't pay" campaign, a "first come, first served" basis

The first and most frequent pattern is the coordination of two (transitive) verbs with or without object, as for example in *a "take it or leave it" attitude* or *a "wait and see" mentality*.

The pattern that is about as frequent as this one is just one VP, which is a transitive verb and its object. Examples found are *a "wait a minute" gesture* or *the "Cut the Crap" LP*, where the VP refers to the title of an LP (by the Clash). It also includes phrasal verbs, and full NP objects that are modified by determiners and/or adjectives, or prenominal possessives.

We have seen above that the most frequent pattern for the non-verbal type is the one where a nominal is followed by a preposition and a nominal. Concerning the verbal pattern, we find verbs that are modified by prepositional phrases (PP) with the function of modifier adjunct or prepositional object[2]. Some examples are *a "smack in the face" offer* and *a "wait for mail" situation*. A similar pattern are VPs with implicit or explicit object that show adverbial modification as in *a "could try harder" comment* or verbs in resultative constructions where the adjective denotes a final state resulting from the action of the verb as *the "get your hands dirty" philosophy*.

Marginal patterns are first of all VPs expressed by a verb in the progressive form and its object (e.g. *that "powdering my nose" act*), a verb followed by a *to*-infinitive (*a "need to know" basis*), a list of VPs that appear with and without commas as in *the "hear no evil, see no evil" brigade*, and participles modified by PPs as in *the "made in Japan" tag*. In the latter case, most of the time the head noun is an object like a tag or sticker which is the bearer of the utterance "made in X" and hence seems to be a quotation.

Apart from VPs and their different patterns, the majority of the verbal PCs are those that contain a full sentence, including the subject. Again, the different patterns of this type are given in a Table:

---

2 As concerns the type *Verb-prep-NP* type where the forms *reach*, *smack*, etc. could potentially also have the status of nouns, I decided to apply the semantic criterion of conversion and assigned verbal status to all occurrences.

Table 5: Patterns of verbal PCs in the BNC: sentences (part I)

Full subject NP + predicate	the "dog eat dog" view, the "expected price equals marginal cost" principle, , a "gee whiz, gosh, women are now making it" syndrome, the "little old lady who's lost her ticket" routine, the "straw that broke the camel's back" syndrome, the "polluter must pay" argument
Subject pronoun+predicate	the "I knew as much" smirk, the "I cannot read that" instruction, the "I want to lick you all over" variety, "I hate Kylie Minogue" T-shirts, the "we know best" philosophy, the "I'll go away and think about it" response, the "you had to be there to totally appreciate it" vein, the "I've been attacked" routine, the "Oh yeah man, we was stoned" book, the "I am not going to miss out on the fun" brigade, a "we are marching and that's that" position, the "Whoops, sorry, we forgot you" Oscars, a "once we've sold it, we forget it" attitude
Copula construction	the "small is beautiful" brigade, a "Meat Is Murder" placard, the "blood is thicker than water" explanation, this "Steffi is Great" attitude, the "less is more" effect, a "Weather hot, cricket wonderful" postcard, the "slippery, when wet" floors
Directives	the "Look At Me" category, the "Go for Gold" scheme, a "kick me please" type, a "Learn what is there and don't question it" attitude, a "NOT TO BE TAKEN AWAY" sticker, that "see me privately, or not at all" ultimatum, the "Say No To Strangers" campaign, the "do and be rewarded" training, a "don't touch me" statement, a "catch me if you can" game, the "Now Print!" mechanism
Directives with *let*	a "let it happen" attitude, the "let's get away from it all" kind, a "let's get a sunlounger and lie on the sand" sort of resort

The first two patterns of this type are declarative sentences with either a full NP subject or a subject pronoun. These in turn can either have the shape of simple sentences (subject-verb-object) where none of the units are modified as in *the "dog eat dog" view*, or they can show modification of its elements, as for example in *the "little old lady who's lost her ticket" routine*, where the "little old lady" is modified by a relative clause. We even find interjections like "gee whiz", "gosh" or "whoops" as part of the PCs as in *a "gee whiz, gosh, women are now making it" syndrome* or *"Whoops, sorry, we forgot you" Oscars*. These interjections clearly give PCs a flavour of informal spoken speech. We will come back to this point in the discussion of the sociolinguistic factors that determine the distribution of PCs in section 3.

Complex sentences showing coordination or subordination are also found, as for example in *the "I'll go away and think about it" response* or *a "once we've sold it, we forget it" attitude*. Most of the time, the form of the subject pronoun is first person singular or plural, there are some sporadic examples with second person singular (*the "you had to be there to totally appreciate it" vein*). In all of these cases, full verbs, modals and auxiliaries can occur specified for all tenses (present, past, perfect, future), predominantly in indicative mood and active voice.

Another quite prominent pattern are copula constructions where the copula is *be* and the subject complement can either be an adjective (*the "small is beautiful" brigade, this "Steffi is Great" attitude*) or a noun (*a "Meat is Murder" placard*). There are examples as *a "Weather hot, cricket wonderful" postcard* where the copula has been left out ("weather is hot, cricket is wonderful"). This also applies to the PC *the "slippery, when wet" floors* ("floors are slippery when wet).

The next type I would like to discuss are directives occurring in the form of an imperative sentence, i.e., it lacks a subject and generally has a verb in the base form[3]. Again, all kinds of sentences are possible, simple ones as in *the "Look At Me" category*, complex ones as in *a "Learn what is there and don't question it" attitude*, negative imperatives as in *a "don't touch me" statement* and directives with *let* as in *"let's get a sunlounger and lie on the sand" sort of resort*. Questions are also quite frequent as phrasal parts of PCs as Table 6 shows:

Table 6: Patterns of verbal PCs in the BNC: sentences (part II)

Yes/No-questions with *be* or *have*	the "Is West Belfast Working?" Conference, the "isn't it a nice day?" stage, the "Are you sure" field, the "Is this a dagger?" speech, the "are ya doin' awright?" approach, the "has he or hasn't he?" riddle, the "Have you heard the Good News" routine, the "but my, aren't they dangerous?" tag
Yes/No-questions with *modals* or *do*	that "Would you like to sit on my knee?" nonsense, the "Shall we go to the pub?" variety, the "does he take sugar?" approach, a "Did you know" section, the "Do you like housework?" question, a "gee-whiz, would-you-believe-it?" fashion
Wh-question	a "What's on" column, the "what is whisky?" debate, the "What should we do now?" variety, the "Where are you now?" Directory, the "who duz wot" bit, the "who's that dying on the runway" stuff, the "Who The Hell Did This One, Then?" party, the "how did I ever live without it" variety
Echo question, *what* in *so* expression	those "he did what?" examples, the "So what?" syndrome

Here we find on the one hand Yes/No questions with *be* (*the "isn't is a nice day" stage*), *have* (*the "Have you heard the Good News" routine*) as well as with modals (*that "Would you like to sit on my knee?" nonsense*) and *do* (*the "Do you like housework?" question*). On the other hand *wh*-questions occur where arguments and adjuncts are questioned as for example in *the "What should we do now?" variety, the "who duz wot" bit, the "Where are you now?" Directory* and *the "how did I ever live without it" variety*. Note that the question is not always graphically marked as a question (the question mark is sometimes lacking). Echo questions are a marginal type (*those "he did what?" examples*) as well as *so* expressions introduced by *what*. Although being marginal, i.e., occurring only sporadically, they nevertheless are indicative of the informal spoken flavour that has been attributed to PCs. Table 7 summarises the types of PCs found in the BNC (numbers given are absolute frequencies):

Table 7: Types of PCs found in the BNC

NP+N	694
VP+N	650
AdvP+N	32
PP+N	15
AdjP+N	6

---

3 Verbal phrases are categorised as directives if it was clear that the omitted subject is the 2nd person (this can be demonstrated by adding a tag question, e.g., "look at me, will you?").

Out of a total of 1397 PCs in the *BNCweb* (where the phrasal part is marked by quotation marks), the most frequent type is the nominal type with a total of 694 occurrences (i.e. tokens). The next highest frequency is found for the verbal type, including sentences, questions, etc. with a total of 650 occurrences. Furthermore, PCs with AdvPs (32 occurrences), PPs (15 occurrences) and AdjPs (6 occurrences) were found, but their frequency is much smaller than the nominal and verbal type. In the following section, it will be investigated if the frequencies of the NP+N and VP+N type correlates with features like mode, register, text type, age and gender.

3. Demographic and textual features determining the distribution of PCs and the expressive flavour of PCs

Since the BNC is a mixed general-purpose corpus of (British) English it should represent written and spoken language in a balanced way. What we find if we survey the composition of the corpus is however, that the spoken component constitutes approximately 10 per cent (10 million words) whereas the written component constitutes 90 per cent (90 million words), which means that we do not have equal proportions here. Although users can yield valuable empirical statistical data for the spoken and the written part, numbers showing overall results and their interpretation have to be treated with caution.

In the following, results concerning the distribution of PCs across demographic and textual features will be presented in a number of tables. The number of words, the number of hits and the frequency per million words (pmw) will be given and serve as a basis for the interpretation of the data. The definitions of the features to be discussed can be found in the *Reference Guide for the British National Corpus* (http://www.natcorp.ox.ac.uk/docs/URG/index.html).

329

Let us start with the distribution of PCs with respect to mode. Table 8 provides the results gained:

Table 8: Distribution of PCs across mode of speech

Spoken or Written:			
Category	**No, of words**	**No, of hits**	**Frequency per million words**
Written	87,903,571	1394	15.86
Spoken	10,409,858	3	0.29
total	98,313,429	1397	14.21

Out of a total of 1397 instances of PCs, 1394 were found in written speech, only 3 in spoken speech. At first sight, this seems to be a clear result, but a number of methodological problems that might have led to it, must be addressed here. First of all, as noted above, the written component constitutes 90 per cent of the whole of the BNC, thus it is not surprising that many more occurrences of PCs are found there as opposed to the few instances found in the spoken component. However, the frequency count (per millions words, pmw) confirms this result: 15.86 of PCs occur in written speech, only 0.29 occur in spoken speech. Second, at the beginning of section two the process of finding the "right" corpus query was described, and it was said that the most frequent pattern of PCs in the corpus was the one where the phrasal part is marked by quotations. Nevertheless, I also found instances where the phrasal parts were separated by dashes, e.g. "... the tired, *out-of-touch* grandfather in the Elysée" (BNC, HXU 1689). So although a large number of PCs are indeed marked by quotation marks and not by hyphens, it is methodologically extremely hard to exclude that there are other possibilities, for example no marking at all. To account

for the latter case, I cross-checked highly frequent patterns like e.g. a determiner followed by the personal pronoun referring to first person singular as in

(3) We are left with the fun loving (overgrown kids) and the *"I am not going to miss out on the fun'' brigade.* (HP61079)

but without quotation marks. The result was that I did find sporadic examples as

(4) In fact, the day before *the I Love Lucy St Patrick's Day*, she'd been told of a good job with the British Council in the Gilbert Islands. (A0L 1497)
(5) Van den Berghs' marketing director Bill Young took the UK marketer of the year trophy in recognition of his deft handling of *the I Can't Believe It's Not Butter controversy*, which showed that even the monolithic Unilever can be quick on its feet when its brands are under threat. (BNH 916)

but they are quite rare. However, to fully account for all the patterns in a corpus like the BNC one would have to find a way to retrieve all of the occurrences which definitely is a task that should be taken on in future research. What can be said at the moment is that the majority of PCs are indeed marked by quotation marks so that we get quite a good picture about the patterns that occur, albeit not a comprehensive, conclusive one.

Third, if the three occurrences from the spoken part of the BNC are investigated, we see that they are represented just like the ones in the written part:

(6) The other kind of camp on is erm *the "ring no reply'' camp on*.
(KS6 508 PS6KK)
(7) Which is why it's a very good idea when you next go into your offices this afternoon, to rock the receiver, like this, because any camp ons put on the wrong extension on *the "ring no reply '' camp   on* will be matured, on a first come first served basis, so that's how people get to talk to you. (KS6 513 PS6KK)

(8) It's full of *the "So what ?'' syndrome*. (KRP 986 PS62R)

The first two examples which illustrate the same type of PC - *the "ring no reply" camp on* - are from a dialogue during a telephone system training, the example of *the "So what?" syndrome* was uttered during a dialogue at the Environmental Health Officers' conference. Again, I cross-checked the most frequent patterns, but this time I retained no hits, which is probably a matter of coincidence. For spoken speech, what one would intuitively expect to occur is actually a pause after the determiner and probably also before the head noun to indicate a complex morphological unit in the flow of speech, for example:

(9) The other kind of camp is <pause> *the "ring no reply" camp on* <pause> which can be selected by the costumer.

Although pauses are encoded in the corpus, they cannot be queried since they are not tagged. What would be interesting to see is if the assumption that speakers indeed embrace PCs by pauses in the flow of speech is borne out, again I have to leave this aspect for further research.

Having discussed these problems, let us return to the frequencies found for PCs in spoken and written speech. If we apply a statistical hypothesis test to establish the significance of a comparison of the frequencies given in Table 8, we gain the result of $x^2 = 151.70368$ which means that the difference between the two different types of mode is

significant at p < .001[4]. To pursue this result in more detail, next, I would like to discuss the results gained concerning the distribution of PCs across derived text type[5]. Since only three occurrences of PCs were found in spoken speech, the following discussion will predominantly refer to the 1.394 hits in written speech.

The results presented in Table 9 show that the highest number of hits (427) are among the text type labelled "Other published written material" which is an exclusion of all the other categories relating to written material given in the table. It also represents the highest frequency (23,82) of PCs within these text types. The next highest number of hits (398) is found in the text type labelled "Non-academic prose and biography" with a frequency of 16.46, followed by 255 hits (16.16) in the text type "Academic prose". The next highest frequency occurs in the text type "Unpublished written material" with 20.82 pmw. As concerns the three examples in spoken speech, all three occurred in the derived text type labelled "Other spoken material".

Table 9: Distribution of PCs across derived text type

Derived text type:			
Category	No, of words	No, of hits	Frequency per million words
Other published written material	17,924,109	427	23.82
Unpublished written material	4,466,673	93	20.82
Academic prose	15,778,028	255	16.16
Newspapers	9,412,174	163	17.32
Non-academic prose and biography	24,178,674	398	16.46
Fiction and verse	16,143,913	58	3.59
Other spoken material	6,175,896	3	0.49
Spoken conversation	4,233,962	0	0.00
total	98,313,429	1397	14.21

331

A look at the distribution of PCs across text type given in Table 10 reveals that the highest number of hits (1232) occurs in "Written books and periodicals" followed by 159 hits in the category of "Written miscellaneous".

Table 10: Distribution of PCs across text type

Text type:			
Category	No, of words	No, of hits	Frequency per million words
Written miscellaneous	7,437,161	159	21.38
Written books and periodicals	79,187,792	1232	15.56
Written-to-be-spoken	1,278,618	3	2.35
Context-governed	6,175,896	3	0.49
Demographically sampled	4,233,962	0	0.00
total	98,313,429	1397	14.21

---

[4] I have used the *Corpus Frequency Test Wizard* online available at sigil.collocations.de/wizard.html.

[5] "Derived text type" refers to larger units of genre which have been defined by David Lee, one of the designers of the corpus.

Since the designers of the corpus aimed at proportions of 60 per cent from books and 30 per cent of periodicals (the latter of which include about 250 issues of newspapers), and only 10 per cent from miscellaneous sources (published, unpublished, and written-to-be spoken), the text type "Written books and periodicals" with 79,187,792 words shows a lower frequency of PCs (15.56) than the type "Written miscellaneous" (21.38) with a total of 7,437,161 words. The three examples in spoken speech occurred in the context-governed part, which is the part that consists of more formal encounters as for example meetings, lectures and the like, complementing the demographic component of the corpus which represents more informal encounters defined by age, sex, social class and geographic region. By looking at the examples, we have seen that all of them were produced in a rather formal context, which might be a surprising finding.

As an interim conclusion, we can say that PCs are a written phenomenon, they predominantly occur in books and periodicals, but also in publicity leaflets, brochures, fact sheets, school and university essays or letters. For the latter case (i.e. written miscellaneous) I provide some examples below:

(10)   By day it's very much *a "let's get a sunlounger and lie on the sand" sort of resort* and the sandy beach is well equipped with everything you'll need in the way of bars, restaurants and shade, and there are several places to waterski, windsurf and parascend or try your hand on the wet bikes (AM0 1208, Club 18-30 summer holiday brochure 1990).

(11)   It is a real link, not just *a "we'll be thinking of you from time to time" relationship* (CC1 205, Queen's Park Baptist Church Magazines)

(12)   We are left with the fun loving (overgrown kids) and *the "I am not going to miss out on the fun" brigade* (HP6 1079, Scottish Amicable Newsletter)

To get further insights into the motivation to express something by means of a PC, next I would like to turn to the distribution of PCs across text domain. In the BNC, a general distinction has been made between imaginative and informative. Since again representativeness of the corpus was one of the main goals of the designers so that the corpus could be regarded as "a microcosm of current British English in its entirety"[6], with respect to the difference between imaginative and informative they chose to draw 25 per cent from imaginative texts and 75 per cent from informative texts.

332

Table 11: Distribution of PCs across text domain

Text Domain:			
Category	**No, of words**	**No, of hits**	**Frequency per million words**
Informative: Arts	6,574,853	208	31.64
Informative: Commerce and finance	7,341,009	180	24.52
Informative: Social science	14,025,538	267	19.04
Informative: Applied science	7,173,003	141	19.66
Informative: Belief and thought	3,037,532	57	18.77
Informative: Leisure	12,191,902	229	18.78
Informative: World affairs	17,244,523	209	12.12
Informative: Natural and pure sciences	3,818,803	41	10.74
Imaginative prose	16,496,408	62	3.76
total	87,903,571	1394	15.86

---

[6] Source of quotation: http://www.natcorp.ox.ac.uk/docs/URG/BNCdes.html.

The former category refers to fictional texts and texts perceived to be literary or creative. They are not classified according to field of subject. All other texts are labelled informative and are classified according to the eight domains listed in Table 11. The highest number of hits occurs in the domain of "Social science" (267), followed by "Leisure" (229) and "World Affairs" (209). A look at the frequency per million reveals that they actually occur most frequently in the domain of "Arts" (31.64), hence it is not the number of hits but the frequency pmw that we should take into account here. Apart from this result, the table also shows that PCs only amount to 3.76 of the total in imaginative prose. Again, I give some examples to provide the contexts in which these PCs occur.

Examples from imaginative prose:

(13)   They can't fool me with *that "powdering my nose" act* (A0D 1728, from the book "A classic English crime").

(14)   I decided to try *the "little old lady who's lost her ticket" routine*, but I was rumbled immediately and directed politely but firmly to the station manager's office (A0F 1448, from the book "Part of the furniture).

Examples from informative prose:

(15)   Relatively weak description of *the "he was very glad of my arrival" sort* is cut, and we are left with the histrionic handclasp of Stepan Verkhovensky the actor manqué whom no reader of The Possessed will ever forget (A18 1440, from the book "Dostojewski")

(16)   She also knows that the media tendency to lump together women singer-songwriters in *a "gee whiz, gosh, women are now making it" syndrome* is patronising, if not pernicious (A7S 190, from "The Guardian", electronic edition of 1989-11-08: Arts section)

(17)   Most eventually got honorary Lifetime Achievement Awards — alias *the "Whoops, sorry, we forgot you" Oscars*, or even "Whoops, sorry, we didn't know you were still around", as happened to Sophia Loren in January, thirty years after she won Best Actress for Two Women (ABS 2601, from the "Esquire")

By looking at the results we have gained so far, it seems that PCs predominantly occur in periodicals and magazines and in informative text domains like "Arts" and "Commerce and finance". The question is if there is a correlation between the medium, i.e. periodicals, in which PCs are most frequently found and the text domains mentioned. Although a number of papers have dealt with PCs from a qualitative point of view (cf. Botha 1980, Lieber 1988, Lawrenz 1996), only one study of German PCs has particularly dealt with the question of why this type of compound is produced in the first place. Meibauer (2007) sees PCs as a marked phenomenon that can be explained if their expressive character is taken into account. Let us consider example (16) from above: the sentence shows a high degree of lexical density, expressed by compounds like *media tendency* and *women singer-songwriters*, and finally the PC *"gee whiz, gosh, women are now making it" syndrome*. If we pondered over an alternative for the PC, we would probably come up with an NNC like *superwoman syndrome* which would come quite close to the semantics of the PC but would nevertheless be less expressive or witty. The properties of expressivity or wittiness have been attributed to (marginal) morphology, e.g., by Zwicky & Pullum (1987:335) who have stated that a derivation like *laserteria,* meaning "a specialist retail outlet for laser equipment", could not be uttered in a business meeting without raising chuckles, and that

these words are "... whimsical coinages, carefully contrived for dubbing commercial enterprises, and carry an effect lacking in plain derivational morphology", which could also be defined as pragmatic effect (see also Bauer 1997, 2002). According to Meibauer the expressivity of PCs is caused by a conflict between two pragmatic principles belonging to the theory of generalised conversational implicatures, the I(nformativeness)-Principle and the Q(uantity)-Principle (cf. Levinson 2000). As long as a speaker produces an NNC like *superwoman syndrome* he or she adheres to the I-principle since minimal linguistic information is produced with the result that the recipient has to infer the underspecified information from the context of the utterance. As soon as a speaker however produces a PC like *"gee whiz, gosh, women are now making it" syndrome*, for him or her adhering to the Q-principle is more important since he or she has provided the strongest statement possible in that context, i.e., a statement that is more informative than one expressed by a NNC. Meibauer assumes that this is the case because PCs include sentences which contain propositions, have a set of entailments, and are bearers of illocutions. The conflict that arises between the two principles is stated by Meibauer as follows:

(18) *Expressivity in CP phrasal compounds*
Expressivity of phrasal compounds stems from a conflict between a principle that requires enrichment of a minimal and underdetermined structure in normal compounds (e.g. the I principle) and a principle that requires maximal informativity (e.g. the Q principle) and leads to the integration of a phrase into word structure.

(Meibauer, 2007, 248)

To find proof for his assumptions, Meibauer conducted a number of experiments with students who had to evaluate a PC as well as a number of alternatives concerning their understandability and wittiness. Meibauer defined understandability as the case when the effort of enrichment is too big, and wittiness as the case when incongruity on the word level occurs, which means that the integration of a phrasal meaning into a word meaning is surprising for recipients. The latter aspect implies that a PC is wittier than an NNC. For his study he used the following material:
    A non-lexicalised PC in context:

(19) Während diese Zeilen entstehen, werden mehrere hundert laminierte **"Kaufe-Ihr-Auto-Kärtchen"** hinter die Hubscheibenwischer alter Mittelklasse-Mercedes geklemmt. Dabei würden deren Besitzer viel lieber an den freundlichen jungen Mann verkaufen, der sich so rührend um seine anderen alten Autos kümmert.
[Youngtimer 2/06,S.55] 'While these lines are written, several hundreds of laminated buy-your-car cards are stuck behind the lift windscreen wipers of old middle class Mercedes. Yet their owners would prefer to buy their cars to the friendly young man who is so very solicitous towards his other old cars.'

(Meibauer, 2007, 250)

Alternatives to the PC:

(20) a. Autokärtchen
          car cardDIM
     b. Kaufkärtchen
          buyV/N cardDIM
     c. Kaufe-Ihr-Auto Kärtchen
          buy1.PS.SG-your-car cardDIM
     d. Kärtchen "Kaufe Ihr Auto"

cardDIM "buy1.PS.SG your car"
e.   Kärtchen mit der Aufschrift "Kaufe Ihr Auto"
cardDIM with the writing "buy1.PS.SG your car"
f.   Kärtchen, auf denen "Kaufe Ihr Auto"
cardDIM on which "buy1.PS.SG your car" is written
(Meibauer, 2007, 250)

The task of the informants was to rate the PC and its rivals in context and in isolation in terms on understandability and wittiness on a 5 point scale (high vs. low degree of understandability/wittiness). As predicted, the PC reached the highest value for understandability and wittiness in context. Since it was not clear if it was the whole context that was perceived as being witty or the PC itself, the same experiment was conducted with the PC and its rivals in isolation. The results were similar but the values for both understandability and wittiness are lower than in task one. This means that the context indeed contributes to the overall interpretation in terms of these two features.

If we adopt Meibauer's assumptions about the pragmatic effect PCs have, and especially the role understandability and wittiness play, we might explain why they are predominantly found in text types like periodicals and magazines but also in leaflets and brochures. On the one hand, writers of these media do not have as much space as writers of books, so they have to write in a compact, yet entertaining and appealing fashion. If they produce PCs (instead of NNCs) they are likely to attain an effect of wittiness and at the same time a high degree of understandability, which is much more important than in other text types like e.g. fiction and verse. As concerns the distinction between imaginative and informative, it seems to be plausible that in the latter domain at least a high degree of understandability is more important than in the former domain (compare the examples given above). The high(er) number of hits in the subdomains "Commerce and finance" can then be explained along the same lines: since in selling something it is extremely important to attract the potential buyer, to achieve a high degree of understandability by using a PC is a good strategy. And since wittiness is also an attractive feature, PCs reach both effects automatically. Thus, Meibauer's explanation of the occurrence of PCs based on pragmatic grounds serves quite well to (partly) explain the quantitative results of the study presented. Nevertheless, this effect could also be due to the fact that different types of language situations lead to differences in the distribution of word classes: in their reference grammar, Biber et al (1999) point out that nouns are much more common in newspapers and academic texts than they are in fiction and conversation. Since PCs are nouns, such a distribution would be expected. As concerns the reason of why there is a considerable difference between written and spoken speech, we could assume that it is probably a matter of processing but this is only an ad hoc guess and of course would have to be investigated in depth.

As mentioned above, the aim of this study was to provide a qualitative, and to a limited degree, quantitative basis to explore the occurrence of PCs and their distribution across demographic and textual features. We have seen that the latter type of features do indeed, at least to some degree, determine their occurrence. Now, to complete the picture, I will present the results gained by looking at the distribution of PCs across age of author:

Table 12: Distribution of PCs across age of author

Age of Author:			
Category	No, of words	No, of hits	Frequency per million words
0-14	59,559	3	50.37
15-24	542,578	5	9.22
25-34	2,267,024	44	19.41
35-44	6,726,929	71	10.55
45-59	7,230,584	57	7.88
60+	5,126,298	31	6.05
total	21,952,972	211	9.61

As concerns the highest number of hits, authors between 35 and 44 produced most of the PCs in the corpus (71), followed by the age group 45-59, which is again followed by the age group 25-34 years (44). In relation to the total number of words, the age group between 0 and 14 years of age with a frequency of only three hits has the highest frequency pmw (50.37) but it is obvious that a result based on three cases only is not very reliable. If we wanted to find out whether the difference between the age categories 1 (0-14) and 3 (25-34) is really a significant difference, and not just an accident, again we would have to apply a statistical hypothesis test. The result is that the difference is not significant (x^2 = 1.43453). However, if the frequencies of age category 3 and 4 (71 hits) are compared, we get the result that in this case the difference found is significant at p < .01 (G^2 = 9.48800). The same applies to the comparison of frequency of age category 3 and 5 (57 hits), it is significant at p < .001 (G^2 = 18.818885).

If we finally take into account the distribution of PCs across gender of author presented in Table 13, we see that with respect to the number of hits male authors used 424 PCs whereas female authors only 111 (I will leave out the interpretation of mixed authors here with mixed being defined as more than one author of different gender), so there is a difference in frequency between 13.83 and 7.61 pmw. The statistical hypothesis test is applied again to establish the significance of a comparison of the frequencies given in the table with the result of x^2 = 31.81616 which means that the difference between male and female producers of PCs is significant at p < .001. Both results, concerning age and gender, could maybe explained with the same factor, namely that more men between the age 25 to 34 write newspaper articles because they are more often hired than women.

Table 13: Distribution of PCs across gender of author

Gender of Author:			
Category	No, of words	No, of hits	Frequency per million words
Male	30,662,031	424	13.83
Mixed	6,538,929	97	14.83
Female	14,588,254	111	7.61
total	51,789,214	632	12.20

Before we turn to the theoretical part of the paper in section 4, I would like to discuss the frequency breakdown of PCs in the BNC, which actually brings me back to what I have said at the beginning of section 2 as concerns the difference between lexicalised and non-lexicalised forms. It has been assumed by a number of authors (cf. e.g. Baayen 1993, Plag 2003) that words with a high frequency can be correlated with their being stored (as whole words) in the mental lexicon, i.e. with their status of being lexicalised. Words with a low frequency, on the other hand, are not likely to be stored, i.e., they do not have an entry

in the mental lexicon. According to this line of reasoning, *Hapax legomena* (items occurring only once in a corpus) are a good indicator to define which words are stored (being based on non-productive rules) and which words are not stored and hence being built productively. What we would expect to find for PCs is that the preponderance of them shows a low frequency since most of them are of the verbal type which is non-lexicalised and built on the fly. We further expect to find a number of PCs with higher frequency numbers because they include established concepts of world knowledge like titles, clichés, etc. The results presented in Table 14 confirm our expectations: first of all, none of the PCs in the BNC occurs with a high frequency, the highest frequency found is 11 for the phrasal non-head *"harm to interest"*, followed by *"Is West Belfast Working?"* with 8 occurrences, followed by the phrasal non-heads *"Reach for the Sky"* (7), *"first come, first served"* (6), *"What's On"* (5), *"What's Happening"* (5), *"response to injury"* (4) and *"law and order"*, which is a total of 50 cases and thus 3.6% of all PCs. Six different types of phrasal non-heads occur three times (total of 18 cases), two of which are given in the table. 59 different types of phrasal non-heads occur two times in the corpus (total of 118 cases), five of these are given in the table. The rest occurs only once, which means that 86.7% of all PCs in the corpus are *Hapax legomena* (1211 cases), six of which are given in Table 14. Of course, it could be due to mere chance that these PCs occur only once in the corpus, but two points speak against this conclusion: first, the BNC is a huge corpus and the likelihood that the result is due to chance is very small. Second, we could use the observation that those PCs with higher frequencies show more than one different type of nominal head as an indicator of lexicalisation. This assumption seems to be borne out, most of the PCs with the highest frequencies in the corpus (between 11 and 2 tokens) do indeed occur with different nominal heads: for example *"harm to interest"*, the phrasal non-head with the highest frequency, occurs with *theory*, *theorist*, and *principle*. There are even cases like the phrase *"law and order"* that shows five different nominal heads: *cases*, *campaign*, *language*, *attorney*, and *rhetoric*. The only two cases that speak against this assumption are the PCs with the second and third highest frequencies. If these PCs are searched for in the corpus, it becomes obvious why they show deviations from the "rule": both have been mentioned in one text several times and therefore should probably only be counted once (*the "Reach for the Sky" Appeal* in the RAFA journal (A67), and *the "Is West Belfast Working?" conference* in the EFD periodical (EFD)). This small investigation then shows that it is not the assumption that is flawed but the quantitative statistics used (which is a general problem, of course, and not particular to this study).

337

Table 14: Frequency breakdown of PCs in the BNC

	Frequency breakdown of PCs in the BNC	
	phrasal non-heads	**N-heads**
11	"harm to interests"	theory, theorist, principle
8	"Is West Belfast Working?"	Conference
7	"Reach for the Sky"	Appeal
6	"first come, first served"	basis, principle, stands
5	"What's On"	leaflet, section
5	"What's Happening''	listings, section, pages
4	"response to injury"	hypothesis
4	"law and order"	cases, campaign, language, attorney, rhetoric
3	"wait and see"	mentality, group, attitude
3	"small is beautiful "	brigade, centre, rule
2	"workshop of the world "	tag, type
2	"warts and all"	school, closeness
2	"them and us"	syndrome, attitude
2	"take it or leave it"	attitude, basis
2	"sword of Damocles"	hanging
1	"women suffer a great deal through their husbands"	sensuality
1	"women speak more standard"	rule
1	"sell me your shares or shoot yourselves in the foot"	type
1	"sell cheap, the future looks bright"	technique
1	"gee-whiz, would-you-believe-it?"	fashion
1	"gee whiz, gosh, women are now making it"	syndrome

Taking into account the structure of PCs, a correlation between a higher frequency of occurrence and the non-verbal type, and a lower frequency of occurrence and the verbal type can be assumed. The 50 cases with the highest frequencies are predominantly of the non-verbal type including patterns like *Noun-and-Noun* or *Noun-prep-Noun*. The majority of hapaxes , however, are of the verbal type that are expressed by different types of complete sentences (see Tables 4 and 5 again). All of these hapaxes are not listed in any dictionary which supports the claim that they are a good indicator of the productivity of this type of PC. Thus, the quantitative study presented in this section has shown that the most productive type of PC is the verbal type, which is, as I have stated above, also most interesting from a theoretical perspective. It is that perspective we will turn to in the following.

4. A new approach

After having set the empirical basis in the previous section, in this section, I am going to sketch an analysis of PCs based on the model of Parallel Architecture (e.g. Jackendoff 1997, 2002, Culicover & Jackendoff 2005) for verbal PCs.

In the introduction, the motivation to provide a more satisfying analysis of these types of PCs was spelled out, and it was said that the hitherto proposed analyses have all run into problems because they are based on the following properties defining the classical generative framework: (i) they are syntactocentric, (ii) all derivational processes always start from syntax, (iii) there is a strict division between lexicon and grammar. Regardless of whether these analyses reflect a strictly lexicalist, word syntactic or mixed-model point

338

of view, all of them have run into serious problems explaining the "peculiar" properties of PCs because they have to allow the integration of a syntactic phrase into a morphological phrase, i.e. into word structure, although the generative system based on properties (i) to (iii) does not. From my point of view, these proposals have dealt with PCs in a one-sided way by only looking at the formal properties and neglecting the semantic aspects. I am going to discuss a different approach and will hopefully provide a more balanced analysis of this phenomenon by doing justice to both the formal and semantic properties of PCs.

So far we have seen that about half of the PCs found in the corpus are of the verbal type, and that almost all cases of that type are hapaxes being produced on the fly and thus non-lexicalised. Since this type of PC poses more problems for a formal analysis than the non-verbal type, because it really integrates a transparent syntactic structure into a word structure which cannot be said to be a whole, lexicalised, unit, in the following we will predominantly deal with this type. From the examples of the verbal type presented in Tables 4 and 5 we see that a wide variety of verbal phrases occurs, i.e., all kinds of verbs with all possible morphological inflections, with arguments and adjuncts, the elision of the verb, verbs in declarative main clauses as well as in questions, and sentences introduced by interjections as in spontaneous, authentic speech. In Table 15 the nominal heads of these PCs are classified in terms of their conceptual-semantic properties following Jackendoff 1995 and Meibauer 2003:

Table 15: Conceptual semantic classification
of the nominal head of verbal PCs (non-exhaustive)

INDIVIDUAL	variety, category, brigade, community, people, writer, attorney, team, guardian, author, searcher, teenager, theorists, Greek-Cypriot, type, group, man, wife, whiner, watchdog, starfish, sir, searcher, psychologist, prisoner, player, person, junior, guru, gang, foe, fan, expert, crew, corporation, coalition, candidate, campaigner, party
PROPERTY	image, quality, style, look, smirk, style, nature, feeling, touch, sensuality, quality, power, face
CONCEPTUAL ENTITY	idea, approach, regime, experience, theory, basis, principle
ATTITUDE	philosophy, attitude, line, position, policy, ideology, syndrome, viewpoint, vein, standpoint, statement
ACTION	series, act, routine, tactics, strategy, scheme, campaign, smokescreen, action, activity, event, exhibition, programme, conference, lunch
UTTERANCE, MEDIUM CONVEYING UTTERANCE	argument, message, gesture, story, speech, song, phrase, sound, chant, response, record, slogan, comment, report, refrain, proverb, sign, sticker, postcard, newspaper, banner, button, reader, book, letter, prospectus, chapter, section, album, LP, column, T-shirt, magazine, leaflet, guide, command, card, rhetoric, riddle, question, error, appeal, compilation, tag, box
TIME	heyday, holiday, day, session, time, era, moment, episode, phase, week(end), period, stage, situation,
THING	jacket, Oscar, machine, sweetener

In (21) a number of examples for each of the semantic concepts of the nominal heads are given:

(21) a.   INDIVIDUAL

We are left with the fun loving (overgrown kids) and the *"I am not going to   miss out on the fun" brigade*. (HP6 1079)

b.   PROPERTY

Martinho was watching, with that *"I've got nothing to do with this" look* that he put on when he'd fucked things up good. (H9N 1983)

c.   CONCEPTUAL ENTITY

Please try to avoid *the "does he take sugar?" approach*, ask the person in the chair directly "Would you like a push?" rather than ask their companion if they have one. (CHK 1298)

d.   ATTITUDE

He claimed that he was sick of *this "Steffi is Great" attitude* and he accused you of showing favour towards Steffi. (A0V 485)

e.   ACTION

They can't fool me with *that "powdering my nose" act*. (A0D 1728)

f.   UTTERANCE

If you are being pressurised by someone, use this tactic; it's *the "I'm just looking, thank you"* or *the "I'll go away and think about it" response* to the pushy salesperson. (CEF 1025)

g.   MEDIUM CONVEYING UTTERANCE

FREED from a lift in his Harare hotel, the Bearded Wonder sends us a *"Weather hot, cricket wonderful" postcard* from Zimbabwe. (K52 2291)

h.   TIME

Radio brought the main news from the outside world; nuclear tests in the Pacific, civil rights marches in America, the coronation of Queen Elizabeth and *the "never had it so good" era* in Britain. (H7E 1024)

i.   THING

Most eventually got honorary Lifetime Achievement Awards -- alias *the "Whoops, sorry, we forgot you" Oscars*, or even "Whoops, sorry, we did n't know you were still around", as happened to Sophia Loren in January, thirty years after she won Best Actress for Two Women .(ABS 2601)

We see that the non-heads of all these examples are sentences, which contain a proposition that is based on truth values. This applies to all the examples found in the different semantic categories, also to those where the copula verb be has been elided as in *a "Weather hot, cricket wonderful" postcard* (also note that the example in f. contains gapping which is generally possible in compounds, e.g. *word and sentence structure)*. Interestingly, only few cases occur where the head noun denotes a concrete thing as in (21) i. (the other examples found are *a "cut and sew" jacket*, *the "Try your strength" machine*, and *the "We're not going heavy" sweetener*), the preponderance of occurrences must be classified as MEDIUM CONVEYING UTTERANCE as can be seen from the table above. I assume that this observation can be attributed to the nature of verbal PCs (see below).

If we try to apply one of the basic functions that can fill out the function *F* in NNCs as proposed by Jackendoff (2009, 2010a), the one that seems to come closest to the semantics of these PCs is BE (Y,X) meaning 'Y is (also) an X' since it is based on a predication relation. Jackendoff provides the following examples of NNCs for this relation:

(22)   a.   *boy king* (dvandva compound)

         b.   *witch doctor* (objects that are a mixture of N1 and N2)

         c.   *tractor-trailer* (objects composed of N1 and N2) (Jackendoff 2010a: 437f)

However, if the data in (22) are compared with verbal PCs in (21), we find that the relation is not the same. For example the PC *the "Steffi is Great" attitude* (21 d.) does not denote

"'Steffi is Great" is (also) an attitude'. The same holds for the basic function KIND (X,Y), 'an X of kind Y', denoting a relation among kinds. So whereas it is true that 'a puppy is a kind of dog' (puppy dog) this relation does not underlie the PC *"Steffi is Great" attitude*: it is not true that "'Steffi is Great" is a kind of attitude'. If these basic functions were applied to all the examples in (21) the result would be that none of these underlie the PCs. To understand why this is not possible let us compare an NNC with a PC in more general terms: the former type is based on the function $F(X_1,Y_2)$ yielding the meaning of $[N_1 \ N_2]$. In the literature, it has been extensively discussed that in isolation this relation is very hard to determine due to the semantic underspecification of compounds (see e.g. Fanselow, 1981, Meyer, 1993). Thus, a *puppy dog* could be interpreted according to the KIND relation but it could also be interpreted as e.g. "a dog who eats puppies". Although PCs are also compounds, they do not share this property with non-phrasal compounds: in our example, *the "Steffi is Great" attitude*, the relationship between the phrasal non-head and the non-phrasal head is much less underspecified, i.e., more clearly defined, namely that the utterance *"Steffi is Great"* expresses an attitude. This assumption is corroborated by Meibauer's experiment we have dealt with above, since he took the semantic underspecification of compounds into account and investigated PCs in isolation and in context. What he found was that his informants evaluated PCs as equally informative, understandable and witty. Thus, NNCs are much more underspecified and context-dependent than verbal PCs. What I then claim for the interpretation of verbal PCs is that this type of compound is based on the IS-A relation because they contain a proposition:

(23)   [State IS-INSTANCE-OF ([x; TOKEN], [y; TYPE])] (Jackendoff 2010a:13)

Under this assumption, it is possible to provide a general account for the semantic interpretation of all PCs of this type (which also includes their form as I will show below), and to explain why NNCs differ from CPs semantically.

341

Based on this conceptual structure, I further claim that two types of verbal PCs must be distinguished: a) the type where the utterance refers to the concept of THING, and b) the type where the utterance refers to the concept of EVENTUALITY (cf. Varzi, 2002). Figure 1 graphically represents this classification:
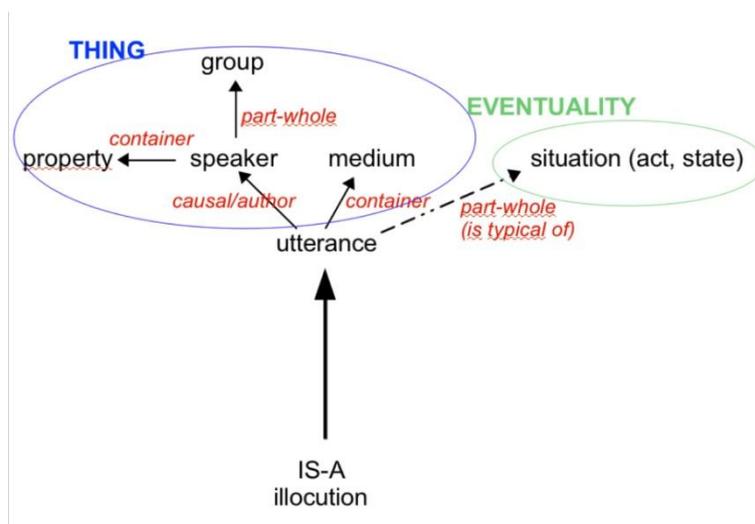
Figure 1: Conceptual semantic relations of verbal PCs

As can be seen from Figure 1 metonymic coercion, indicated by the arrows and metonymic shifts in italics, plays a crucial role in defining the properties of verbal PCs. I assume that

metonymy is a conceptual phenomenon which can be defined as proposed by Panther & Thornburg (2007: 242):

(24)   a. Conceptual metonymy is a cognitive process where a source content provides access to a target content within one cognitive domain.
    b. The relation between source content and target content is contingent (conceptually nonnecessary), i.e., in principle defeasible.
    c. The target content is foregrounded, and the source content is back-grounded.
    d. The strength of the metonymic link between source and target content may vary depending, among other things, on the conceptual distance between source and target and the salience of the metonymic source.

Type a) PCs differ from type b) PCs in that the former type refers to the cognitive domain of THINGS and almost always shows type mismatches which are resolved by metonymic coercion (for the notion of type shift or coercion see Pustejovsky, 1995, ch. 7). The only exception here are heads of the type UTTERANCE like *response*, *argument*, etc. (see Table 15) where the relation between source and target content is direct: [State IS-INSTANCE-OF ([x;I'LL GO AWAY AND THINK ABOUT IT], [y;RESPONSE])]. In all other cases of this type the relation between source and target content is indirect, for example, in the case of MEDIUM CONVEYING UTTERANCE, an example of which would be *a "Weather hot, cricket wonderful" postcard,* a type shift from UTTERANCE to MEDIUM CONVEYING UTTERANCE is assumed :

(25)   "Weather hot, cricket wonderful"$_1$ postcard$_2$ = [state IS-INSTANCE-OF (UTTERANCE$^\alpha$ [WEATHER HOT, CRICKET WONDERFUL]$_1$, MEDIUM CONVEYING UTTERANCE$^\alpha$ [POSTCARD]$_2$)]

342

In this case, the content of the coerced function *F* is filled out by material from the proper function of *postcard*. In Jackendoff's framework, the process of cocomposition (i.e. the coercion of extra functions into the structure by filling them with either basic functions or with internal semantic structure from the first and second part of a NNC; cf. Jackendoff 2009: 120) is invoked here.

    Type b) PCs including the concepts of ACTION and STATE are instances of processes of the concept of EVENT(UALITIES). Here, only one metonymic coercion occurs, *is typical of* (*part-whole*), which shows a strict association between UTTERANCE and EVENT(UALITIES). No such restriction is found for type a) PCs because they lack strict associations between speech acts and nominal heads denoting things. For illustration, let us take a closer look at the interpretation of the following two PCs: *this "powdering my nose" act* and *this "Steffi is Great" attitude*. The former can be paraphrased as 'the utterance "(I am) powdering my nose" is typical of an act', thus it includes the IS-A relation (the strict association between the utterance and the concept denoted by the nominal head is indicated by the concept of ACTION[7]):

(26)   "powdering my nose"$_1$ act$_2$ = [state IS-INSTANCE-OF (ACTION$^\alpha$ [POWDERING MY NOSE]$_1$, ACTION$^\alpha$ [ACT]$_2$)]

In this case and in other cases, it is not just an act which is clearly defined by the phrasal non-head but actually an act seen as a stereotype, i.e., the phrase is used as periphrasis to refer to a salient, conventionalised piece of information in one cognitive domain, which may even lead to using it as a euphemism:

---

[7] In the following, all relevant concepts (ACTION, PSYCHOLOGICAL STATE, etc.) are defined according to the classification of WordNet-3.1.

(27)   I'll use your bathroom. To powder my nose, as nice girls say.
        (L. P. Davies *What did I do Tomorrow?* 1972, p. 72; OED online)

In the latter case, the relation can be paraphrased as 'the utterance "Steffi is Great" is typical of an attitude', again the IS-A relation holds and the strict association between the utterance and the concept of attitude (state) is indicated in the conceptual structure as follows:

(28)   "Steffi is Great"$_1$ attitude$_2$ = [state IS-INSTANCE-OF (PSYCHOLOGICAL FEATURE$^\alpha$ [STEFFI IS GREAT]$_1$, PSYCHOLOGICAL FEATURE$^\alpha$ [ATTITUDE]$_2$)]

Whereas for type a) PCs a number of metonymic shifts are possible, for type b) PCs only one shift is. Note however, that generally these types of shifts are constrained, since a shift from source to target content is dependent on its cognitive domain. In the case of *the "I am not going to miss out on the fun" brigade,* the utterance "I am not going to miss out on the fun" conveys an attitude which is attributed to a group of people (brigade is used here to denote a social group). Here we find a more indirect metonymic shift than in the other cases above: first, a shift from the utterance to an individual (*causal/author*), and second, a shift from individual to group (of individuals) via the *part/whole* relation. Since an attitude is conveyed, the structure of the PC thus resembles the one proposed for the PC *"Steffi is Great" attitude*.

(29)   "I am not going to miss out on the fun"$_1$ brigade$_2$ = [state IS-INSTANCE-OF (PSYCHOLOGICAL FEATURE$^\alpha$ [I AM NOT GOING TO MISS OUT ON THE FUN]$_1$, PSYCHOLOGICAL FEATURE$^\alpha$ [BRIGADE]$_2$)]

Generally, it could be concluded that the difference between the interpretation of NNCs and PCs lies in the fact in the former case, the semantic relation between the non-head and head is much more underspecified because it is based on a one-to-many relation. PCs on the other hand do not share this property because they are based on the IS-A relation and rather specified instances of metonymic shifts or coercions, which have also been called 'rules of construal' (Nunberg 1979).

We have further seen that metonymic coercion plays a crucial role in explaining the nature of verbal PCs. Based on the observations made above, it could be assumed that the expressive flavour of PCs discussed in section 3 results from the strength of the metonymic link between source and target content. To put it simpler, the more indirect a metonymic link is between the source and target content, the wittier a PC is. Cases where no metonymic coercion occurs would therefore be perceived as being less witty. This seems to be borne out for type a) PCs:

(31)   a.   If you are being pressurised by someone, use this tactic; it's *the "I'm just looking, thank you" or the "I'll go away and think about it" response* to the pushy salesperson. (CEF 1025)
                => no metonymic coercion, less witty
        b.   Most eventually got honorary Lifetime Achievement Awards -- alias *the "Whoops, sorry, we forgot you" Oscars*, or even "Whoops, sorry, we did n't know you were still around", as happened to Sophia Loren in January, thirty years after she won Best Actress for Two Women .(ABS 2601)
                => metonymic coercion, more witty

For type b) PCs where a strict association between source and target content was assumed this explanation does not hold. Rather, I would assume that in this case the observation

that an act is seen as a stereotype, is relevant (apart from the fact that the unexpected occurrence of a sentence within a word must also have an effect). According to Levinson, lexical items have the potential to implicate stereotypical default readings, so "What is expressed simply is stereotypically exemplified" (2000:37). It is the interplay of these rules that can account for the properties of PCs.

In section 3, Meibauer's definition of the expressivity of PCs was discussed. Meibauer claimed that enrichment and informativity (based on a conflict between Levinson's I and Q principle) on the structural level are critical for explaining this property. But perhaps it is not the structural level but the conceptual level which plays the decisive role. Papafragou (1996) defines two communicative reasons for using metonymies: on the one hand metonymies cause extra processing effort which is "levelled out" by a gain in contextual effects (additional implicatures). On the other hand, the processing effort may be smaller than that for a literal expression of the metonymic sense. If we applied the latter communicative reason to Meibauer's assumptions and to the production of PCs, we could say that producing this type of compound leads to enrichment via metonymic coercion, i.e, to additional contextual effects, which is not possible with NNCs. What we automatically gain is a maximum of informativeness, so from this point of view, it is not a conflict that arises. Although the cognitive effort is greater, it is still the most economical way to get to enriched conceptual information, thus a PC will be preferred over an NNC because it is wittier (distance between target and source content) and more understandable (more enriched, more transparent).

5. Conclusion

This paper has dealt with phrasal compounds from an empirical and a theoretical point of view. The motivation to study this phenomenon in depth was on the one hand the lack of a qualitative and quantitative empirical study and on the other hand a satisfying analysis. In sections 2 and 3 the empirical corpus study based on the BNC was discussed, and the PCs found were classified into two main categories, verbal and non-verbal PCs. The verbal type occurred about as frequent as the non-verbal type, and it was shown that although a number of patterns do determine the occurrence of phrasal non-heads (e.g. *Nom-prep-Nom*), this mainly applies to the non-verbal type. For the verbal type, it was shown that all kinds of sentences are allowed, simple as well as complex. It was assumed that the latter type is the more interesting type since (i) it is the more productive type, (ii) the sentential non-head includes a number of entailments, (iii) a proposition based on truth values is always included, (iv) as an utterance the phrasal non-head includes illocutions. Further, apart from the qualitative analysis in section 3 a quantitative analysis in terms of the distribution of PCs across textual and demographic features was provided. The main results were that PCs are a phenomenon of written speech (significant correlation), more precisely informative prose, and that they predominantly occur in newspapers and periodicals. It was also shown that the distribution across age categories and differences between male and female producers of PCs were significant. It was said that one explanation for these results could be explained with the same factor, namely that more men between the age 25 to 34 write newspaper articles because they are more often hired than women.

Concerning the question of why PCs are produced at all, Meibauer's interesting observations regarding PCs in German were discussed and applied to the English data presented here. It was concluded that his definitions of understandability and wittiness could explain the findings, and in more general terms, the morphopragmatic character of PCs.

Based on this empirical study, in section 4 a sketch of an analysis along the lines of Jackendoff's Parallel Architecture was provided. The fact that verbal PCs contain

344

propositions invited a conceptual-semantic analysis based on the IS-A predication relation which was applied to a number of PCs to show that it generally holds, and that metonymic coercion are needed to account for the facts. Thus, I hope to have shown that an analysis based on Conceptual Semantics is quite promising and might be able to account better for the phenomenon than any other analysis that has been proposed so far in syntactocentric, derivational models of generative grammar.

Bibliography

Ackema, P. and Neeleman, A. 2004. *Beyond Morphology*. Oxford: Oxford University Press.

Baayen, H. 1993. "On frequency, transparency and productivity", In G. Booij and J. van Marle *Yearbook of Morphology 1993*, 181-208. Dordrecht: Kluwer.

Bauer, L. 1997. "Evaluative morphology: in search of universals". *Studies in Language* (21): 533–75.

Bauer, L. 2002. *English Word-Formation*. Cambridge: Cambridge University Press.

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman Addison Wesley.

Borer, H. 1988. "On the morphological parallelism between compounds and constructs". In G. Booij and J. van Marle (eds), *Yearbook of Morphology* 45–65. Dordrecht: Kluwer.

Botha, R. 1981. "A base rule theory of Afrikaans synthetic compounds". In M. Moortgat, H. van der Hulst and T. Hoekstra (eds), *The scope of lexical rules,* 1–77. Dordrecht: Foris.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambrige Mass.: MIT.

Chomsky, N. 1981. *Lectures on Government and Binding*. Studies in generative grammar 9. Dordrecht: Foris.

Culicover, P. and Jackendoff, R. 2005. *Simpler Syntax*. Oxford: Oxford University Press.

Fanselow, G. 1981. *Zur Syntax und Semantik der Nominalkomposition: ein Versuch praktischer Anwendung der Montague-Grammatik auf die Wortbildung im Deutschen*. Linguistische Arbeiten 107. Tübingen: Niemeyer.

Gallmann, P. 1990. *Kategoriell komplexe Wortformen*. Tübingen: Niemeyer.

Jackendoff, R. 1975. *"Morphological and semantic regularities in the lexicon". Language* (51): 639–71.

Hoffmann, S., S. Evert, N. Smith, D. Lee, Y. Berglund Prytz. 2008. *Corpus Linguistics with BNCweb - a Practical Guide*. Frankfurt: Peter Lang.

Jackendoff, R. 1983. Semantics and Cognition. Cambrige Mass.: MIT Press.

Jackendoff, R. 1990. Semantic Structures. Cambrige Mass.: MIT Press.

Jackendoff, R. 1995. "The Boundaries of the Lexicon". In M. Everaert, E.-J. van der Linden, A. Schenk and R. Schreuder (eds), *Idioms. Structural and Psychological Perspectives*, 133–65. Hillsdale, N.J.: Erlbaum.

Jackendoff, R. 1997. The Architecture of the Language Faculty. Cambrige Mass.: MIT Press.

Jackendoff, R. 2002. Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford: Oxford University Press.

Jackendoff, R. 2007. "A parallel architecture perspective on language processing". Brain research (1146): 2–22.

Jackendoff, R. 2008. "Alternative minimalist visions of language". In Chicago Linguistic Society 41: The Panels, R. Edwards, P. Midtlyng, K. Stensrud and C. Sprague (eds), 189–226. Chicago: Chicago Linguistic Society.

Jackendoff, R. 2009. "Compounding in the Parallel Architecture and Conceptual Semantics". In R. Lieber and P. Štekauer (eds), *The Oxford Handbook of Compounding*, 105–129. Oxford: Oxford University Press.

Jackendoff, R. 2010a. Meaning and the Lexicon. Oxford: Oxford University Press.

Jackendoff, R. 2010b. "The parallel architecture and its place in cognitive science". In B. Heine and H. Narrog (eds), *The Oxford handbook of linguistic analysis*, chapter 23, 583–605. Oxford University Press.

Lapointe, S. G. 1980. *A theory of grammatical agreement*. Ph.D. thesis, University of Massachusetts, Amherst.

Lawrenz, B. 1996. "Der Zwischen-den-Mahlzeiten-Imbiß un der Herren-der-Welt-Größenwahn: Aspekte der Struktur und Bildungsweise von Phrasenkomposita im Deutschen". In: *Zeitschrift für Germanistische Linguistik* 24, 1-15.

Levinson, S. C. 2000. *Presumptive Meanings*. The Theory of Generalized Conversational Implicature. Cambridge, Mass.: MIT Press.

Lieber, R. 1988. "Phrasal compounds and the morphology-syntax interface". Chicago Linguistic Society II Parasession on agreement in grammatical the-ory (24): 202–222.

Lieber, R. 1992. Deconstructing Morphology. Chicago: The University of Chicago Press.

Meibauer, J. 2003. "Phrasenkomposita zwischen Wortsyntax und Lexikon". *Zeitschrift für Sprachwissenschaft* 22 (2): 153–188.

Meibauer, J. 2007. "How marginal are phrasal compounds? Generalized insertion, expressivity, and I/Q-interaction". *Morphology* (17): 233–259.

Meyer, R. 1993. *Compound Comprehension in isolation and context*. Linguistische Arbeiten 299.

Nunberg, G. 1979. "The nonuniqueness of semantic solutions: Polysemy". *Linguistics and Philosophy* 3, 143-184.

Panther, K.-U. and Thornburg, L. 2007. "Metonymy". In D. Geeraerts and H. Cuyckens (eds), The Oxford Handbook of Cognitive Linguistics, 236-264. Oxford: Oxford University Press.

Papafragou, Anna. 1996. "Figurative language and the semantics-pragmatics distinction." *Language and Literature* 5: 179–93.

Plag, I. 2003. *Word-Formation in English*. Cambridge: Cambridge University Press.

Pustejovsky, J. 1995. The Generative Lexicon. A Theory of Computational Lexical Semantics. Cambrige Mass.: MIT Press.

Štekauer, P. and Lieber, R., (eds). 2009. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press.

Trips, C. 2009. Lexical Semantics and diachronic morphology: the development of the derivational suffixes -hood, -dom and -ship in the history of English. Tübingen: Niemeyer.

Trips, C. Submitted. "The relevance of phrasal compounds for the architecture of grammar".

Varzi, A. C. 2002. "Events, Truth and Indeterminacy". *The Dialogue* 2: 241–264.

Wiese, R. 1996. "Phrasal compounds and the theory of word syntax". *Linguistic Inquiry* (27): 183–193. Zwicky, A. and Pullum, G. 1987. "Plain morphology and expressive morphology". In J. Aske et al. (eds), *Proceedings of the thirteenth annual Meeting of the Berkeley linguistic society*, 330–340. Berkeley: Berkeley Linguistic Society.

346