Chapter 1

Further insights into phrasal compounding

Carola Trips

Universität Mannheim

Jaklin Kornfilt

Syracuse University

1 Further insights into phrasal compounding from a typological and theoretical perspective

This collection of papers on phrasal compounds is part of a bigger project whose aims are twofold: First, it seeks to broaden the typological perspective by providing data for as many different languages as possible to gain a better understanding of the phenomenon itself. Second, based on these data, which clearly show interaction between syntax and morphology, it aims to discuss theoretical models which deal with this kind of interaction in different ways. For example, models like Generative Grammar assume components of grammar and a clear-cut distinction between the lexicon (often including morphology) and grammar which mostly stands for the computational system (syntax). Other models, like construction grammar do not assume such components and are rather based on a lexicon including constructs. A comparison of these models makes it then possible to assess their explanatory power.

The field of morphology and syntax started to acknowledge the existence of phrasal compounds predominantly in the context of Lexicalist theories because a number of authors realised that they are not easy to handle in models of linguistic theory which demarcate the lexicon (morphology) from syntax. Commenting on

the difference between base and derived forms Chomsky said in his "Remarks on Nominalization":

> "However, when the lexicon is separated from the categorial component of the base and its entries are analyzed in terms of contextual features, this difficulty disappears."
> (Chomsky (1970: 190))

This assumption was dubbed The Lexicalist Hypothesis and in the course of time a number of different versions surfaced. For example, Lapointe (1980: 8) put forward the Generalized Lexicalist Hypothesis which stated that "No syntactic rule can refer to elements of morphological structure." Botha (1981: 18) took the perspective from morphology and established The No Phrase Constraint which postulated that "Syntactic phrases cannot occur inside of root compounds." In 1987, Di Sciullo & Williams summarised these hypotheses and constraints in their Atomicity Thesis:

> "Words are "atomic" at the level of phrasal syntax and phrasal semantics. The words have "features" or properties, but these features have no structure, and the relation of these features to the internal composition of the word cannot be relevant in syntax – this is the thesis of the atomicity of words, or the lexical integrity hypothesis, or the strong lexicalist hypothesis (as in Lapointe 1980), or a version of the lexicalist hypothesis of Chomsky (1970), Williams (1978; 1978a), and numerous others."
> (Di Sciullo & Williams 1987:49)

Some of these authors commented on instances of phrasal compounding like Botha (2015) (who coined the term "phrasal compounds") and Savini (1984) and came to the conclusion that they constitute negative evidence for these constraints because they clearly showed interaction between syntax and morphology (see the following examples from Dutch):

(1) a. uit-die-bottel-drink   alkoholis
         from-the-bottle-drink alcoholic
         'alcoholic who drinks straight from the bottle'
         (Botha 1980:143)

     b. laat-in-die-aand   drankie
         late-in-the-evening drink
         'drink taken late in the evening'
         (Savini 1984: 39)

In the same vein, Lieber (1988; 1992) put forward examples for English and came to the conclusion that they violate these constraints, or in more general terms, the Lexical Integrity Hypothesis:

(2)   a. slept all day look

        b. a who's the boss wink
           (Lieber 1992:11)

But despite these rather sporadic discussions of the phenomenon no comprehensive study of phrasal compounds in individual languages or cross-linguistically existed.

Fortunately, with a growing interest in compounding as an interface phenomenon the situation has changed in the last five years. This can be seen by the publication of a number of volumes dedicating themselves explicitly to this type of word formation by providing detailed accounts of types of compounds across languages (see e.g. Scalise & Vogel (2010); Štekauer & Lieber (2009)), and this development brings phrasal compounds now to the fore as well.

To gain a better understanding of phrasal compounds, in 2013 a workshop with the topic "Phrasal compounds from a typological and theoretical perspective" brought together scholars who had been working on (phrasal) compounding in different languages and from different theoretical perspectives. The outcome of this fruitful workshop was a collection on the topic which was published in 2015 as a special edition of STUF (Trips & Kornfilt 2015). The languages under investigation were German, English, Italian, Turkish, some additional Turkic languages and Greek. Concerning the approaches chosen for an analysis of the phenomenon, some authors (Pafel, Göksel) analysed the phrasal non-head of phrasal compounds in terms of quotes, quotations, citations whereas authors like Meibauer and Trips favoured a semantic analysis which attributes an important role to pragmatics (Trips to some degree in the form of coercion, Meibauer even more so in terms of pragmatic enrichment). Some of the authors (Bisetto, Bağrıaçık & Angela Ralli) made a distinction between phrasal compounds that are lexical/morphological and syntactic (either within one and the same language or comparing languages) and some authors (Trips & Kornfilt) found similar semantic restrictions in diverse languages (Germanic, Turkish) but also clear structural differences.

Despite this valuable contribution to a phenomenon underrepresented in current research, it became evident quickly that to come closer to fulfilling the aims defined above it would be necessary to add further languages, on the one hand, and to deepen the theoretical discussion, on the other hand.

Concerning the typological aspect of (phrasal) compounding we wanted to include further languages which had not been investigated so far; especially interesting are, for example, Slavic languages, because they seem to exhibit compounds, but they occur less frequently than for example in the Germanic languages. Another aspect worth investigating is whether all Germanic languages behave in the same way. One very interesting example is Icelandic which has much more inflectional morphology than the other contemporary Germanic languages. Can we then expect that Icelandic behaves differently because of different morphology? Another, more general question is if languages which are of the same syntactic type (e.g. SOV) behave in the same way when it comes to PCs. Would we, for example, expect to find the same patterns we identified for German as an SOV language in another SOV language like Japanese? And what about languages in contact? Would we expect to find the borrowing of phrasal compounding from a source language to a recipient language since, after all, they are complex (under the assumption that contact generally leads to simplification)?

Concerning questions relevant for linguistic theory it would be worthwhile investigating if there is a correlation between the morphological and syntactic typology of a language. So for example is the rightheadedness in morphology (always) related to SOV? Or is a rich inflectional system a prerequisite for rightheadedness in morphology? Another interesting question is whether the distinction between PCs containing a predicate and PCs not containing a predicate made by Trips related to the property of the nominal head requiring an argument (or not) as the non-head? Focussing on the semantic relation between the non-head and the head in languages like English and German we find a tight semantic relation. The same is true for Turkish, but in addition we have selectional restrictions. In contrast, languages like Sakha (Turkic) show looser semantic relations between the non-head and head. So would we find these similarities/differences in other language pairs? And, from a more general point of view, are there theories which model the general properties of phrasal compounds more adequately than others? And if so, which properties would such a theory have?

Our interest in these questions made us open up our workshop in 2015 as well as this special issue to papers conceived in different frameworks. While we cannot answer these evaluative questions yet, we hope that this collection of case studies conducted in a variety of models will bring us closer to such answers.

Turning back to structural and semantic properties of phrasal compounds, questions about the relationship of the head and the non-head of phrasal compounds were addressed by the presentations at the workshop and continue to

be a focus in the contributions to this special issue. In many simple as well as phrasal compounds, the semantics appear to be similar to that of a predicate — argument relationship, as in Turkish and German:

(3) Turkish
dilbilim   öğrenci-si
linguistics student-CM
'linguistics student'

(4) German
Linguistikstudent
linguistics-student
'linguistics student'

However, especially with respect to quotative phrasal compounds, it is clear that much more general semantic relationships must be allowed to hold. This is shown quite clearly in the examples above, especially by those in (2).

Another issue that contributions have focused on is the overt (syntactic and/or morphological) expression of the head — non-head relationship in compounds, and in phrasal compounds in particular. As illustrated in (3), Turkish (nominal) compounds have a compound marker (CM) on their head; similar compounds in German and English don't have such a marker; Greek does, as well as Pharasiot, a variety of Asia Minor Greek influenced by Turkish. However, the compound markers of these Greek varieties differ with respect to their sources and their shapes — one of the issues discussed in one of the contributions in this volume. Does the presence versus absence of a compound marker determine other properties of a compound, whether phrasal or otherwise? This is a fascinating question whose answer has been attempted in the contribution on Pharasiot, but one which can only be answered more definitively after a good deal of further cross-linguistic research.

One property which appears to hold cross-linguistically is adjacency between the head and the non-head in compounds, setting them apart from phrases:

(5) a. (çalışkan) dilbilim   (*çalışkan) öğrenci-si (Turkish)
        (diligent) linguistics diligent   student-CM
        'diligent linguistics (*diligent) student'

     b. der (fleißige) Linguistik(*fleißige)student (German)
        the diligent   linguistics-diligent-student
        'the diligent linguistics (*diligent) student'

     c. the (diligent) linguistics (*diligent) student (English)

Thus, adjacency turns out to be a reliable diagnostic device for distinguishing compounds from phrases. This becomes particularly important when distinguishing phrasal compounds from phrases, given that in both, the non-head constituent is phrasal, making the relevant distinction less clear at first glance.

The non-head in phrasal compounds can be expressed in a variety of different ways cross-linguistically. Limiting attention to clausal non-heads in phrasal compounds, we see that in some languages, that constituent can be either identical to a root clause (and thus a "quotative"), or it can show up in the typical shape of an embedded clause in the language in question. Thus, in Turkic languages, embedded clauses typically show up as gerund-like nominalizations, and this is a pattern that shows up in Turkish phrasal (non-quotative) compounds:

(6) [en çabuk nasıl zengin ol   -un -duğ   -u] (*ilginç)
     most fast   how rich   become -PASS -FACT-NOM -3.SG (interesting)
     soru   -su
     question -CM
     'The (interesting) question (of) how one gets rich fastest'

In German, on the other hand, embedded clauses typically show up as fully finite, verb-final clauses, in contrast to root clauses which are verb-second; not surprisingly, this is a pattern that shows up in German phrasal (non-quotative) compounds:

(7) die (interessante) [wie man am schnellsten reich wird] (*interessante)
     the interesting   how one the fastest   rich gets   interesting
     Frage
     question
     'The (interesting) question (of) how one gets rich fastest'

In quotative phrasal compounds, we find the non-head exhibiting the morphosyntactic properties of the root clause; this appears to be similar cross-linguistically, as illustrated in (8a) for Turkish, German, and English:

(8) a. Turkish
         [en çabuk nasıl zengin ol   -un -ur]   (*ilginç)
         most fast   how rich   become -PASS -AORIST interesting
         soru-su
         question-CM
         'The "how does one get rich fastest" (*interesting) question'

b. German

die [wie wird   man am schnellsten reich] (*interessante) Frage
the how become one the fastest   rich   interesting   question
'The "how does one get rich fastest" (*interesting) question'

Similar semantics can be expressed by phrases rather than compounds in many instances. Often, a preposition or a postposition is involved in the equivalent phrase, heading the clause; this is illustrated in (9) for Turkish and German, respectively:

(9)   a. [en   çabuk nasıl zengin ol   -un -duğ   -u] hakkında
        most fast   how rich   become -PASS -FACT-NOM -3.SG about
        (ilginç)   soru-lar
        (interesting) question-pl
        '(interesting) questions about how one gets rich fastest'

     b. (interessante) Fragen   darüber, [wie man am schnellsten reich
        interesting   questions about   how one the fastest   rich
        wird]
        becomes
        '(interesting) questions about how one becomes rich fastest'

The possibility of non-adjacency between the phrasal (here, clausal) non-head and the head shows, for both Turkish and German, that these constructions are not compounds, but rather phrases. In addition, the fact that in the Turkish example there is no compound marker strengthens this observational claim.

We thus see that phrasal compounds exhibit similarities as well as differences cross-linguistically. Among the latter, we saw that in Turkish, clausal non-heads in phrasal compounds can be nominalized; this is not an option in German and English phrasal compounds. Furthermore, Turkish phrasal compounds exhibit a compound marker attached to the head; no such marker is ever found in German or English phrasal compounds. Future research will, we hope, show explanations for these differences, beyond those we were able to sketch in this brief overview.

To come closer to an answer to these questions, a second workshop on phrasal compounding from a typological and theoretical perspective took place in 2015 adding further languages and theoretical models. The present volume is a collection of these contributions.

Kristín Bjarnadóttir provides a description of compounding in Icelandic in general terms including phrasal compounding as a marked case. She shows that

compounds are extremely productive in Icelandic and are traditionally grouped into a class containing stems and a class containing inflected words (mainly genitive) as non-heads. Phrasal compounds are also found, and a more common type, well established in the vocabulary, can be distinguished from a more current, complex type. Interestingly, phrasal compounds may also contain a genitive non-head and then the question arises how they can be distinguished from the genitival non-phrasal compounds.

Bogdan Szymanek discusses compounding in Polish (and more generally, in Slavic). He shows that compounds exists in Polish but that they are much less productive than in German or English. Phrasal compounds do not seem to occur at all, as in all the other Slavic languages. The author identifies a number of reasons why this type of word formation is absent, for example the presence of 'multi-word units' that are frequently used to express complex nominal concepts.

Alexandra Bagasheva provides a study of phrasal compounds in Bulgarian. Despite the fact that this type of compound is said not to exist in Slavic languages she shows that they do, especially so in life style magazines. The author discusses her data in the constructionalist framework and proposes the process of "pattern" borrowing from English as an explanation of why phrasal compounds have started to emerge in Bulgarian.

Katrin Hein provides a comprehensive description of phrasal compounds in German and models the different types found in construction grammar. She prefers this model because "traditional" generative approaches do not allow for syntax in morphology and because such an approach also fails to explain why a speaker chooses to use a phrasal compound instead of a nominal compound. Based on a corpus study she shows that the types of phrasal compounds she found can all be captured as form-meaning pairings in this model and that their frequency and productivity justify defining them as constructions. In addition, she notes that the model serves well to explain why the second constituent with its semantic properties has to be seen as the main element and not the first constituent with its abstract syntactic properties.

Kunio Nishiyama describes and categorizes various types of compounds in Japanese whose non-heads are phrasal. Nishiyama proposes that the main criterion of categorization is whether noun incorporation is involved or not in the formation of a given phrasal compound in Japanese. The author is careful not to take a stand on whether an explicit Baker-type incorporation is involved or not, but the derivation he assumes is based on a head-movement approach, similar to a Baker-type noun incorporation, given that the evidence for noun incorporation having taken place is the appearance of "modifier stranding" effects, i.e. that a

"modifier" can be separated from its head only when it is stranded (as a result of incorporation). If noun incorporation has applied in the derivation of a phrasal compound, a further division is made according to whether the "predicate", i.e. the verbal noun which is the host of the incorporated noun, is of Sino-Japanese or of native origin. Nishiyama proposes that there are two licensing conditions for modifier stranding: the complement of the verbal noun, i.e. the left-hand element of the compound, should be a relational noun or a part of a cliché.

If no noun incorporation is involved, there are four subclasses, depending on the phrasal non-head: a modifying non-head, a coordinate structure as a non-head, phrasal non-heads to which prefixes (which the author is inclined to analyze as proclitics) are attached, and non-heads to which suffixes (which, again, the author suggests are enclitics in contemporary Japanese) are attached. Nishiyama further proposes that in phrasal compounds whose non-heads are modifying structures and coordinate structures, the licensing condition is again cliché.

Metin Bağrıaçık, Aslı Göksel & Angela Ralli The paper argues that compounding in Pharasiot Greek (PhG), an endangered Asia Minor Greek variety, is selectively copied from Turkish, based on differences between PhG compounds and Hellenic compounds on the one hand, and similar properties between PhG compounds and Turkish compounds, on the other: As opposed to various other Hellenic varieties, compounds in PhG are exclusively composed of two fully inflected nouns, where the non-head, the left-hand constituent, is marked with one of the two compound markers, -u and -s, whose shape is conditioned morphologically. According to the authors, these compound markers have been exapted from the genitive markers in PhG. Hellenic compounds have a compound marker, as well, located similarly between the head and the non-head, but it is quite a different marker, with a different history; it has been exapted from an Ancient Greek thematic vowel. Furthermore, in Hellenic compounds, there has to be at least one (uninflected) stem. Similarities between PhG and Turkish compounds include, in addition to certain structural common features, the provenance of the respective compound markers: in Turkish, the compound marker is identical to the third person singular possessive (agreement) marker and is placed, just like that agreement marker in possessive constructions, on the head, i.e. the rightmost nominal element. In PhG, the compound marker has the shape of a genitive marker and is placed, just like the genitive, on the non-head. A parallel is drawn by the authors between the respective sources of the compound markers in Turkish and PhG (i.e. the possessive agreement marker in Turkish, and the genitive marker in PhG), basing their view on a possible identification of the genitive in

PhG with the Turkish possessive agreement marker (rather than with the genitive in Turkish, which is placed on the non-head in Turkish possessives). The paper discusses, in addition to the similarities between PhG and Turkish compounds, also differences between them: Turkish compounds can have phrasal (and even clausal) non-heads, while PhG compounds cannot. This difference is attributed mainly to the location of the compound marker within the compound: the PhG compound marker, being a purely morphological affix, attaches to stems, similar to all affixes in the language (as well as in all Hellenic varieties). Therefore, no phrasal constituent can be hosted in the position to which the compound marker attaches. In Turkish, on the other hand, since the compound marker attaches to the head, the non-head can host phrasal constituents. This correlation is claimed to also hold in Khalkha Mongolian, an Altaic language like Turkish, in which, however, the compound marker attaches to the non-head. The authors claim that similar to PhG, but unlike Turkish, phrasal constituents cannot be hosted in the non-head position in Mongolian, thus supporting the correlation they propose between the locus of the compound marker and the availability of phrasal non-heads. Apparent counterexamples in Khalkha, they argue, involve a covert preposition which assigns genitive Case, thus imposing a phrasal, rather than a compound, structure on these counterexamples.

Jürgen Pafel takes a theoretical stance and discusses the morphology-syntax relation in modular approaches. He analyses phrasal compounds in the conversion approach in a number of languages and shows, contra the Lexical Integrity Hypothesis, that morphology and syntax are separate levels of grammar with separate structures and distinct properties. Further, the properties of phrasal compounding speak in favour of a parallel architecture framework, where general interface relations constrain their properties.

Acknowledgements

References

Botha, Rudolf P. 1981. A base rule theory of Afrikaans synthetic compounds. In Michael Moortgat, Harry van der Hulst & Teun Hoekstra (eds.), *The scope of lexical rules*, 1–77. Dordrecht: Foris.

Botha, Rudolf P. 2015. Do Romance languages have phrasal compounds? A look at Italian. *STUF–Language Typology and Universals* 68. 395–419.

Chomsky, Noam. 1970. Remarks on nominalization. In R. Jacobs & P. Rosenbaum (eds.), *Readings in english transformational grammar*. Waltham, Mass.: Ginn & Co.

DiSciullo, Anna-Maria & Edwin Williams. 1987. *On the definition of word*. 2nd edn. Cambridge, Mass.: The MIT Press.

Lapointe, S. G. 1980. *A theory of grammatical agreement*. Amherst: University of Massachusetts dissertation.

Lieber, Rochelle. 1988. Phrasal compounds and the morphology-syntax interface. *Chicago Linguistic Society* II Parasession on agreement in grammatical theory(24). 202–222.

Lieber, Rochelle. 1992. *Deconstructing morphology. Word formation in syntactic theory*. Chicago: University of Chicago Press.

Savini, Marina. 1984. Phrasal compounds in Afrikaans: A generative analysis. *Stellenbosch Papers in Linguistics* 12. 34–114.

Scalise, Sergio & Irene Vogel (eds.). 2010. *Cross-Disciplinary issues in compounding*. Amsterdam/Philadelphia: Benjamins.

Štekauer, Pavol & Rochelle Lieber (eds.). 2009. *The Oxford handbook of compounding*. Oxford: Oxford University Press.

Trips, Carola & Jaklin Kornfilt (eds.). 2015. *Phrasal compounds from a typological and theoretical perspective*. Vol. 68. Berlin: De Gruyter. Special edition of STUF.