Compounds and multi-word expressions in the languages of Europe

1 Introduction

This volume deals with compounds (e.g., boat house, softball) and multi-word expressions (piece of cake, dry cough) in European languages. Compounds and multi-word expressions (henceforth MWEs) are similar as they are both lexical units and complex, made up of at least two constituents. The most basic difference between compounds and MWEs seems to be that the former are the product of a morphological operation and the latter result from syntactic processes. This is, admittedly, a very vague distinction. However, as soon as one takes into account more than one specific language (or language family), it seems that this is the closest one may come to a definition that is more or less applicable to the European languages. In fact, in light of Romance examples such as French glace au chocolat, Spanish helado de chocolate ‘chocolate ice cream’ which have often been analyzed as compounds although they contain syntactic relational markers, even the morphological criterion for compoundhood seems to be questionable. Further complicating matters, whereas in many languages compounds are regarded as being opposed to MWEs, in other languages, and particularly in English, compounds are often regarded as a kind of MWE. In addition, for languages that are assumed to have an opposition between compounds and MWEs, the question arises of whether compounds and MWEs act in competition or complementation with regard to the formation of new lexical units.

Given this background, the aim of the volume is to present an overview of compounds and MWEs in a sample of European languages. Central questions that are discussed for each language concern the formal distinction between compounds and MWEs (in particular prosodic, morphological, and syntactic properties), the relation between compounding and MWE formation as well as the conclusions concerning the theory of grammar and the lexicon that follow from these observations. Although several comprehensive volumes on compounding and phraseology have appeared in recent (and not so recent) years (cf.

---

1 We would like to thank Kristel Van Goethem and Carmen Scherer for very valuable comments on an earlier version of this chapter.
the relationship between compounds and MWEs with respect to their status in lexicon and grammar has received comparatively little attention (cf. Hüning/Schlücker 2015 for an overview). For this reason, this relationship constitutes the central focus of this volume.

The aim of the present chapter is to review the language-specific properties, bring them together and compare them against German. German is well-known for its propensity for (nominal) compounding, as compared to, e.g., French. Also, there is a rather clear demarcation line between compounds and MWEs in German, in contrast to English, for instance. Taking German as a reference point may help to shed more light on some of the crucial questions with respect to the compound-MWE relationship in the various European languages such as, for instance, the potential competition between the two processes, or their demarcation line. By way of language comparison, the differences and commonalities between languages – both within language families and across these borders – become clearer, ultimately revealing that a cross-linguistically valid definition of compounds and the demarcation from MWEs may be impossible, given that languages vary greatly in their defining properties and in the number and productivity of compound and MWE subpatterns.

The volume contains chapters on English, German, Dutch, French, Italian, Spanish, Greek, Russian, Polish, Finnish, and Hungarian. Although this sample is neither complete nor representative of “the” languages of Europe, it nevertheless provides thorough analyses of a large set of central European languages. Importantly, it should be noted that the selection here is mostly due to various practical reasons, rather than an assessment of the relevance of languages. In addition to the languages mentioned, the present chapter also comprises an overview of the North Germanic languages.

The structure of this chapter is as follows: Section 2 starts with general considerations about the lexicon and the lexicon-syntax interface and discusses basic notions such as morphological vs. syntactic lexical unit, lexicalization, and the problem of correspondence. Section 3 discusses compounds and MWEs against the background of German, sorted by language families. The chapter ends with a brief conclusion in Section 4.
2 Theoretical considerations

At the outset of our overview, a short remark on the notion of MWE is in order. It is widely known that different research traditions within this field have focused on different types of MWEs, applying an extremely diverse terminology. In the early Anglo-American structuralist tradition (e.g., Weinreich 1969; Newmeyer 1974), the focus was on idioms as semantically and/or syntactically irregular MWEs. Idioms – a notorious example being kick the bucket – were mainly discussed under the assumption that they posed a problem to rule-based grammar. Traditional German phraseology, on the other hand, which is influenced by the Soviet tradition, has been investigating idioms in their own right, as a core phenomenon of the linguistic subfield of phraseology (Häusermann 1977; Fleischer 1982; Burger et al. 2007). This tradition has put much effort into issues of classification, studying not only idioms, but also other types of MWEs which need not be idiomatic, for instance collocations such as starker Raucher (lit. strong smoker, ‘heavy smoker’) or routine formulae such as Kein Problem (‘no problem’) (e.g., Burger 1998). However, under the growing influence of theories such as Construction Grammar (Fillmore/Kay/O’Connor 1988; Goldberg 2006; Hoffmann/Trousdale (eds.) 2013), and insights from applied linguistics, such as research in foreign language learning (Pawley/Syder 1983; Wray 2002), and with the advent of new technologies within quantitative linguistics and corpus linguistics (Sinclair 1991; Gries 2008), the notion of MWE has broadened dramatically in the last decades. In particular, it has become increasingly accepted that there is a large inventory of lexically partially fixed patterns in the lexicon such as \([N\ by\ N]\) (page by page, year by year, country by country, cf. Jackendoff 2008) that may or may not be fully compositional, and that may be used productively to create new instances. Under such a broad view, MWEs are “co-occurrence phenomena at the syntax-lexis interface” (Gries 2008: 8) that may be defined as syntactic patterns consisting of at least two words, the combination of which may be more or less fixed, more or less idiomatic, and more or less productive. Crucially, as idiomaticity is not a defining feature of all of these patterns, their status as stored MWEs hinges on sufficient frequency and on their function as a lexical unit; hence the term ‘phrasal lexical unit’, which is regularly employed throughout the volume and the remainder of this introduction. To decide whether or not a frequent syntactic pattern is a lexical unit, a well-defined notion of lexical unit, and of the lexicon, is required.
2.1 The notion of the lexicon

It is a widely held assumption that the lexicon is a repository of stored linguistic knowledge, in particular, a repository of words. In fact, it may seem that under the last 50 years of linguistic research, this assumption has hardly been challenged, compared to the lively and ongoing debate about what the most adequate theory of grammar is (cf. Wunderlich 2006: 1). However, it is clear that our theory of the lexicon crucially depends on our theory of grammar. For example, whether the lexicon is viewed as a repository of only words or also of affixes depends on whether morphology is conceptualized as a subcomponent of the lexicon or as part of syntax. Under a mainstream view, linguistic knowledge comprises two components:

One is a finite list of structural elements that are available to be combined. This list is traditionally called the “lexicon”, and its elements are called “lexical items”. [...] The other component is a finite set of combinatorial principles, or a grammar. (Jackendoff 2002: 39)

This view entails the idea that lexical items have to be learned, as they are not predictable. By contrast, grammar – which is often equated with syntax – is viewed as the domain of rules, or principles, that enable speakers of a language to productively generate new sentences. For example, it is an idiosyncrasy of English that the word squirrel (and not, say, the word dog, or the word hamburger) refers to the concept SQUIRREL. Speakers of English have to learn this word with its specific phonological, categorial and semantic features. However, they do not have to learn the sentence The squirrel is eating nuts, as they can productively generate it by combining the respective words according to the rules of grammar. Therefore, the dichotomy between lexicon and grammar also tends to be conceptualized as a dichotomy between words and phrases, and between idiosyncrasies and rules (Engelberg/Holler/Proost 2011: 1).

However, it has long been recognized that there are a considerable number of phenomena in the languages of the world that pose a serious problem to the view of a strict lexicon/grammar divide. Compounds and MWEs are a pertinent case in point. As to compounds, Jackendoff (2009: 108) points out that on the one hand, speakers must store thousands of lexicalized compounds, e.g., peanut butter, but on the other hand, they may build compounds “on the fly”, e.g., bike girl for a girl who left her bike in the vestibule. Thus, compounds arguably are part of the lexicon, but at the same time, compounding is a productive, and therefore rule-based process. For this reason, it is necessary to distinguish between the properties of being morphological, and of being lexical (Gaeta/Ricca 2009).

As to MWEs such as kick the bucket, it is obvious that on the one hand, they are phrasal units, often showing a fully regular syntactic behavior, but on the
other hand, they must be part of the lexicon, as their meaning is non-compositional and has to be learned (Nunberg/Sag/Wasow 1994; Gries 2008). What is more, there is ample evidence by now that not all MWEs are isolated units that have to be learned one by one, but that there must be something like MWEs “on the fly”, as well. That is, there seem to be abstract patterns in the lexicon that can be used by speakers to create new MWEs (Fillmore/Kay/O’Connor 1988). For example, speakers might newly coin the potential, but unattested phrasal simile *heavy as a truck* on the basis of the lexicalized pattern [(as) A as NP], which comprises established examples such as *strong as a horse* or *dead as a doornail* (Finkbeiner 2008).

This raises the more general question of the interrelation between the lexicon and the two “rule-based” components of grammar, morphology and syntax. If both morphology and syntax may feed the lexicon, as is evidenced by compounds and MWEs, how is the interaction of morphology and syntax with the lexicon to be represented in our theory of grammar?

### 2.2 Lexicon-syntax interface

In early conceptions of Generative Grammar, the lexicon was conceived of as a passive repository of morphemes, which would be concatenated in the transformational component of syntax. Only the later stage of lexicalism, initiated by Chomsky’s *Remarks on Nominalization* (1970), led to the recognition of the dual status of the lexicon as both a repository of words and an active component of the grammar (Giegerich 2009). Thus, in a lexicalist theory, morphology is acknowledged as an autonomous component of grammar that is part of the lexicon.² However, there is still a sharp dividing line between the lexicon, including morphology, and syntax. This divide is captured by the principle of lexical integrity, which says that syntactic processes can manipulate members of lexical categories, but not their morphological components (Di Sciullo/Williams 1987; Scalise/Guevara 2005). Behind this is the idea that the lexicon (including morphology) is a ‘pre-syntactic’ component that feeds syntax, but not vice versa. Thus, lexical items, with or without internal morphological structure, are taken from the lexicon and inserted into a syntactic tree. The resulting syntactic structures are later ‘spelled out’ in phonology as well as in semantics.

---

² A weaker form of lexicalism assumes that inflectional morphology is more closely related to syntax, while word-formation is more closely related to the lexicon (e.g., Anderson 1982; cf. also Giegerich 2009).
Under such a conception, one may account for the fact that compounds are part of the lexicon, while at the same time being the output of a productive morphological component. However, the model does not account for the difference between listed compounds and novel ones, as morphosyntactically they look exactly the same. Even more importantly, lexicalism predicts a lexicon free of syntactic phrases. Thus, not only do MWEs such as idioms and collocations pose a serious problem to lexicalism, but also phenomena like phrasal compounds, i.e. compounds with a phrasal modifier constituent (e.g., Pafel 2015; Trips/Kornfilt 2015), and particle verbs, i.e. verbs with a separable particle (e.g., Lüdeling 2001; Zeller 2001).

The linear view of the lexicon/syntax relation is abandoned in Jackendoff’s (1997) Parallel Architecture. At the heart of this approach is the hypothesis of representational modularity, which states that grammar is organized into three autonomous and generative components: viz. phonological structure, syntactic structure, and conceptual structure. Each domain generates representations of its own. The interaction between the components is established by separate interface modules between the systems that contain correspondence rules. In this model, a lexical entry is exactly such a (small-scale) correspondence rule. It links a small chunk of phonology with a small chunk of syntax and a small chunk of semantics. Instead of lexical insertion, there is lexical licensing, in that a lexical item licenses its chunks of information as the result of three independent processes. As Jackendoff (2009) puts it:

A word therefore is to be thought of not as a passive unit to be pushed around in a derivation, but as a part of the interface components. It is a long-term memory linkage of a piece of phonology, a piece of syntax, and a piece of semantics, stipulating that these three pieces can be correlated as part of a well-formed sentence. (ibid.: 107)

The crucial point is that this model allows for including into the lexicon all kinds of units, not only simplex and complex words, but also phrases of different kinds. That is, MWEs can be listed in the lexicon as correspondence rules like every other lexical item. The only difference is that in an MWE such as *kick the bucket*, the three syntactic words are associated with three phonological words, but only with one element in semantics (‘to die’). Complex words, such as compounds, are treated as instantiations of more abstract morphosyntactic schemata that contain variables at the three representational levels. Thus, morphology is not a separate component in Jackendoff’s model. There is no difference between words and rules, but both are conceived of as declarative schemata that have the status of (more or less abstract, and more or less productive) lexical units.

The Parallel Architecture has much in common with Construction Grammar and Construction Morphology (Booij 2010). In a way, one can say that Construc-
tion Grammar, or at least certain variants of it, are realizations of the Parallel Architecture. At the heart of Construction Grammar is the insight that linguistic knowledge largely consists of stored knowledge of constructional schemata, from morphological schemata via lexical, phrasal, and even discourse schemata. Both the Parallel Architecture and Construction Grammar thus argue for a continuity between lexicon and grammar. In Construction Grammar, this continuum view culminates in the notion of the ‘constructicon’, which replaces older views of a lexicon/grammar dichotomy. The constructicon is conceived of as a large structured inventory of constructions of all levels of abstraction. Under this approach, compounds and MWEs can easily be treated as on a par with each other, both being complex constructions sharing certain conceptual or functional features.

2.3 Lexicalization

The continuum view of the syntax/lexicon relationship may lay the ground for an integrated and systematic treatment of both compounds and MWEs as the output of productive or semi-productive schemata localized in the lexicon. Still, it does not say anything about the differences in the lexical status between, e.g., the compounds grass frog vs. grass slug, or the VPs hit the road vs. hit the dog. While grass frog is a lexicalized compound, grass slug is not, and while hit the road is a lexicalized MWE, hit the dog is not. Obviously, some outputs of schemata, or rules, have the status of established lexical items listed in the lexicon, while others have not (Hohenhaus 2005; Bauer 2006; Gaeta/Ricca 2009). In order to account for these differences, one needs a concept of lexicalization.

According to Hohenhaus (2005: 356), the term lexicalization denotes both the process of listing and the state of listedness, that is, the property of some element to be a lexical item of a language. The main rationale behind the joint investigation of compounds and MWEs is precisely their common status as complex lexical items. In order to delimitate the field of investigation, it is therefore crucial to

3 One difference between the Parallel Architecture approach and Construction Grammar lies in the conceptualization of productivity. While Jackendoff (2009, 2013) clearly differentiates between productive and semi-productive phenomena, Construction Grammar is somewhat less explicit in this respect, assuming a flexible continuum of productivity of constructions. Another difference lies in the conceptualization of the contents of constructions. While in a homogeneous approach (e.g., Goldberg 1995, 2006), all linguistic units are taken to be meaningful constructions – there being no autonomous syntactic principles – a heterogeneous approach takes meaningful constructions as only one kind of stored structure, assuming that the grammar can also contain independent principles of syntactic form or semantic structure (Jackendoff 2013: 78f.).
properly define the notion of ‘lexical item’. That is, while we want to include grass frog and hit the road into our field of investigation, we would like to exclude grass slug and hit the dog. In particular, the following two criteria seem to be crucial in this respect.

Firstly, a lexical item functions as a semantic, or conceptual unit. For example, grass frog refers to a unitary concept, a certain species, and hit the road refers to a specific kind of activity. Both are concepts that speakers of the language have stored together with the respective items. By contrast, while speakers of English will be able to assign an interpretation to grass slug, they do not have stored it as a unit together with a certain conventional concept, or stable referent. Similarly, speakers will be able to interpret the phrase hit the dog, but they do so on compositional grounds, and not because they have learned this phrase together with a certain concept.

Secondly, for an element to have the status of a lexical item, it must occur with significant frequency in the language. This criterion has received increasing attention with the growing influence of usage-based approaches and rapidly developing quantitative methods in corpus linguistics. It is closely related to the first criterion, because high frequency makes it more likely that an item is becoming listed with a certain meaning. For example, if during a rainy summer a plague of slugs that eat all the grass in people’s gardens were to sweep over a country, and everybody started talking about the nasty grass slugs, it might be that after a while, this compound would get stored in the English lexicon as a label of this specific concept (‘certain kind of nasty grass-eating slug’).

2.4 Compounds and multi-word expressions in the lexicon

The criterion of lexicalization, i.e. the property of being a (complex) lexical unit, thus allows us – at least, theoretically – to distinguish between those instances of morpho-syntactic schemata that are listed in the lexicon, and those that are not. However, we also need a good criterion to distinguish, within the class of complex lexical units, between compounds, on the one hand, and MWEs, on the other. This criterion, obviously, must be found in their internal structure.

Compounds are the output of morphology, while MWEs are the output of syntax. Accordingly, Gaeta/Ricca (2009: 38) suggest a quadripartite typology which is based on the idea that one has to strictly distinguish between the properties of being morphological, and of being lexical. The property of being morphological implies that an item is the output of some morphological schema or rule, which is different from a syntactic schema or rule. The property of being lexical implies that an item is lexicalized in the above-mentioned sense, i.e., that it refers to a
stable concept and occurs with sufficient frequency in the language. Cross-classifying the two properties results in the following matrix (ibid.):

(a) [+morphological], [+lexical]  
(b) [+morphological], [−lexical]  
(c) [−morphological], [+lexical]  
(d) [−morphological], [−lexical]

Of these four options, (a) represents the prototypical instance of a lexicalized compound, i.e., an item that is the output of a morphological process and that is listed in the lexicon with a stable meaning, e.g., grass frog, play list, or milkshake. Option (b), by contrast, represents an item that is the output of a morphological process, but is not listed, e.g., bike girl, grass slug, or Trump problem. Option (c) is represented by MWEs, that is, phrasal, not morphological items for which it is plausible to assume listedness, either because of semantic idiomaticity or sufficient frequency, or both, e.g., hit the road, heavy smoker, or by and large. Finally, option (d) represents the prototypical syntactic phrase, i.e., a VP such as hit the dog that is formed according to a syntactic rule, or schema, and whose meaning is compositional, therefore not requiring separate storage in the lexicon. The quadripartite typology makes it very clear that, contrary to traditional views, morphological units do not need to be lexical units, while syntactic units may be lexical units.

Against this background, we may now attempt to pin down the defining criteria of compounds, and of MWEs. In this we do not aim for more than a rough approximation, as it is clear that the respective criteria are not only in part language-specific, but also a matter of controversial theoretical debate. Generally, we take it for granted that compounds have the features [+morph], [+lex], whereas MWEs have the features [−morph], [+lex]. Compounds may be defined, following Bauer (2009a), on both phonological, morphological, and syntactic grounds (cf. also Lieber/Štekauer (eds.) 2009a; Giegerich 2015; Bauer 2017). First, compounds

---

4 These items are also called occasionalisms. They may become listed in the lexicon at a later stage, but not all of them will. Hohenhaus (2005) discusses the question whether there are occasionalisms that are not listable (non-lexicalizable) in principle.

5 Booij (2010: 190) uses the term “lexical phrasal constructions” to refer to these units.

6 While Gaeta/Ricca (2009) focus on the delimitation between compounds and MWEs, i.e., complex lexical units, it is clear that the feature combination [−morph], [+lex] also applies to established simplex words, such as grass. Likewise, the combination [−morph], [−lex] also applies to inexistent simplex words, such as the nonce verb to gorp from a textbook sentence on language acquisition (“The duck is gorpng the bunny”), cf. Saxton (2010).
usually behave like single words phonologically. For example, the stress pattern in English compounds is more like the stress pattern in single words than the stress pattern in phrases, e.g., *gréen card* (compound; ‘residence permit’) vs. *green cárd* (phrase; ‘green card’, e.g., in a game of cards).

Second, compounds are marked as word-like units morphologically. While the prototypical case is that a compound is made up of two unmarked lexemes, in languages with inflection, the non-head may carry an inflection-like element (e.g., the element -s in German *Liebe+s+brief* ‘love letter’). Crucially, though, this inflection-like element does not vary as a function of the compound’s role in the matrix sentence (Bauer 2009a: 346). What carries the inflection for the compound as a whole, according to its role in the matrix sentence, is the head (ibid.). For example, the linking element -s in the German compound *Liebe+s+brief* is carried by the non-head, while inflection according to the compound’s role in a matrix sentence goes to the end of the head, e.g., in den *Liebe+s+brief+en* (‘in the love letters, _DAT.PL_’).

Third, compounds can be defined according to syntactic criteria, most importantly syntactic inseparability and an inability to modify the non-head. For example, one cannot insert an element in between the two constituents of the German compound *Alt+bau* ‘old building’, cf. *dieser Alt teure Bau* (lit. this old expensive building), and the non-head (the first constituent) cannot be modified: *dieser sehr Alt+bau* (lit. this very old building).

As for MWEs, scholars like Nunberg/Sag/ Wasow (1994) and Gries (2008) make use of syntactic, semantic, and frequency criteria to arrive at a definition. As outlined in the beginning of this chapter, in modern phraseological research, most scholars hold a rather broad view of the notion of MWEs, including many different types of phrasal units. Syntactically, MWEs are required to consist of more than two syntactic elements, which may be of different natures. For example, the collocation *heavy smoker* consists of two words. In other MWEs, a word tends to co-occur with a particular grammatical pattern, for instance, the verb *to hem* tends to co-occur with the passive. In this case, the MWE consists of a word and a syntactic frame (Gries 2008: 5). MWEs often are syntactically more or less fixed, but there are also fully flexible MWEs. For instance, the MWE *by and large* is completely fixed (e.g., the reverse order *large and by* would be ungrammatical), while *run amok* is rather flexible (e.g., it allows for different tenses).

Semantically, it is usually required that MWEs be semantic units, i.e. that they have a meaning just like a single word or morpheme. For example, *hit the road* roughly means ‘leave’. While many MWEs tend to have a non-compositional semantics, non-compositionality is not a necessary criterion. For example, while *kick the bucket* is semantically non-compositional, *too much to ask* is fully compo-
sitional. Both can be regarded as semantic units, however. As to the frequency criterion, for something to count as an MWE, it is required that the observed frequency of the joint occurrence of the constituents be larger than the expected frequency of joint occurrence. More generally, the degree of frequency of an MWE can be related to its degree of cognitive fixedness, or “entrenchment”. Naturally, the frequency criterion can only be employed on empirical grounds.

2.5 Problem of correspondence

While MWEs such as kick the bucket and compounds such as blackbird do not seem to have much in common except their being complex lexical units, it has been pointed out repeatedly in the literature that there are certain subsets of compound words and MWEs that closely correspond to each other. For example, in German, as in many other languages, there are adjectival compounds, e.g., butter+weich ‘butter soft’, that have corresponding phrasal similes, e.g., weich wie Butter ‘as soft as butter’. These expressions share lexical material and have a very similar meaning. Another case in point are A+N combinations such as schwarzer Tee vs. Schwarz+tee ‘black tea’ (cf. Schlücker 2014; Hüning/Schlücker 2015). As both the morphological and the syntactic pattern are stored lexical units, they pose a problem to the principle of synonymy blocking in the lexicon, suggesting that this principle might not be as strong as often assumed. For such cases, potential tasks for the researcher are to find out how much the two competing processes overlap, if the overlap is systematic or only applies to a subset of the respective patterns, whether one is dealing with real doublets, or whether there are more specific differences in meaning or usage (cf. Masini, this volume; Schlücker, this volume). For example, Hüning/Schlücker (2015) point out that the morphological and the phrasal pattern in similes such as butter+weich/weich wie Butter are competitive only with regard to a relatively small subset of all possible similes. This can be shown by pairs such as *brot+dumm/dumm wie Brot (lit. dumb as bread, ‘very dumb’), where one of the two patterns is ruled out. Theoretically, the interesting question is what underlying principles guide the choice of strategy that is employed in a given language, or in a given context. For German A+N sequences, for instance, the choice between the morphological and the phrasal pattern seems to be sensitive to type frequency effects (cf. Schlücker/Plag 2011).

While all contributions to this volume discuss the compound-MWE relationship, some of them focus explicitly on corresponding patterns, while others look at the issue from a broader perspective. What can be said more generally for the different languages and language families of Europe is that the potential corre-
spondence between compounds and MWEs cannot be described in a uniform way, since it is multifaceted and manifests itself in very different ways.

An interesting aspect, from a semantic point of view, is the observation that in German, compounds such as *Rot+kraut* ‘red cabbage’, in contrast to their phrasal counterparts (*rotes Kraut*), seem to be more inclined to adopt a kind reading. Thus, *Rot+kraut* denotes a specific kind of cabbage, and not just a cabbage that is red. Härtl (2016) argues that this semantic specialization of compounds is not, as is often assumed, an effect of lexicalization, but can also be observed with novel compounds such as *Rot+dach* (‘specific kind of roof’) vs. *rotes Dach* (‘red roof’), and is therefore “somehow active ‘right from the beginning’ in the life of a compound” (Härtl 2016: 66; cf. also Lipka 1977). From a contrastive point of view, an interesting question is whether this presupposition of kind reference is true for compounds in other languages as well. Furthermore, given that Romance languages employ compounding to a far lesser extent than Germanic languages (cf. Section 3.3), one may ask whether similar effects in French are connected to the difference between the ubiquitous, determinerless [N de N] pattern and the ‘regular’ pattern with definite article [N du/de la N]. Similarly, for Swedish, one might speculate that the systematic difference between the ‘regular’ pattern with double determination on the one hand (*det röda kors+et* ‘the red cross’), and the reduced pattern with single determination, i.e. with suffixed determiner only (*röda kors+et* ‘the Red Cross’) (cf. Section 3.2) on the other hand, might be functional. If this were the case, then one would expect that a novel combination with double determination such as *den stora mur+en* ‘the big wall’ would be less inclined to adopt a kind reading or a naming function when compared to the combination with single determination, *stora mur+en* (which should be inclined to denote a specific type of wall, e.g., the prospective wall between the United States and Mexico).

In the next section, we are going to take a more detailed contrastive look into the compound/MWE relationship in the different languages and language families of Europe as compared with German.

---

7 This is not to say that German phrasal patterns cannot adopt a kind reading, which is clearly not the case (e.g., *schwarzes Brett* ‘bulletin board’). The point in Härtl (2016: 66) is that “right from the beginning”, a compound is semantically more specialized, or more restricted than its corresponding phrase, which may, but must not adopt a kind reading. Potential counterexamples to this hypothesis are pairs such as *Warmwasser* vs. *warmes Wasser* (‘warm water’), or *Blondhaar* vs. *blondes Haar* (‘blond hair’), where the compound does not seem to be semantically more restricted than the phrase; cf. Schlücker (2014).
3 A contrastive overview

The second part of this chapter is devoted to the comparison of German with the West Germanic, North Germanic, Romance, Slavic, Greek, and Finno-Ugric language families in terms of the relationship between compounds and MWEs. It strives to illustrate the similarities and differences between these languages and to sketch some more general tendencies of the respective language families with respect to this relationship. The languages discussed in the following overview are restricted to those represented in the various chapters of the volume, except the North Germanic languages, which lack their own chapter and which have been added to this overview to complete the picture. When relevant, compound boundaries are marked by “+” in the following.

3.1 West Germanic languages and German

As German is a West Germanic language, more similarities than differences with other West Germanic languages are to be expected. In fact, German and the other major members, English and Dutch, are characterized by several common properties. First of all, there is no doubt whatsoever about the existence and productivity of the morphological pattern of compounding in these languages. Second, these languages have both nominal and adjectival compounding, with the former being unanimously regarded as the most frequent and productive subpattern and N+N compounding particularly apparent. Verbal compounding, on the other hand, is regarded as either scarce or non-existent. Regarding English, Bauer (this volume) and Bauer (2017: 136–140) provide sporadic examples of verbal compounding such as dry-burn or mock-whisper. Similarly, there are a few coordinate V+V compounds in German, such as brenn+härten ‘flame-harden’, press+ polieren ‘press-polish’. They are, however, very rare and mainly belong to technical terminology. In general, it seems clear that most forms that look like verbal compounds on the surface are in fact the result of either back-formation or conversion (e.g., German frühstücken ‘to have breakfast’, < Früh+stück, lit. early piece, ‘breakfast’). Then again there are also separable complex verbs, such as particle verbs (e.g., English drink up) and quasi-noun incorporation (e.g., Dutch piano spelen ‘play the piano’ (cf. Booij, this volume)) whose morphological/compound status is highly problematic given the fact that they are separable. Thirdly, English, Dutch and German all have MWEs, both those that in principle correspond to compounds and those that do not, such as proverbs or routine formulas. MWEs corresponding to compounds are those that share the basic naming function of compounds and possibly also share lexical material. For instance, there are various
kinds of nominal phrasal constructions with a naming function and which are therefore on a par with nominal compounds. Thus, they are lexical noun phrases, sometimes also termed ‘phrasal nouns’. Patterns of lexical noun phrases are easily found in all these languages, e.g., close apposition (German Prinzip Hoffnung ‘principle of hope’), genitive (or possessive) constructions (English baby’s chair, German Ei des Kolumbus ‘egg of columbus’), constructions with prepositional phrases (Dutch restaurant met tuin ‘garden café’), binomials (English fish and chips), or A+N phrases, often with a relational adjective (Dutch stalen zenuwen ‘nerves of steel’).

However, in addition to these similarities, there are also differences within West Germanic. In particular, there is one fundamental contrast that distinguishes English from the other two. Overall, in German and Dutch, compounds can be very clearly distinguished from phrasal constructions on the basis of formal criteria, primarily stress and inflection. This distinction is reflected in spelling, with compounds displaying solid spelling and MWEs being written in two (or more) orthographic words. There are only very few patterns that resist a clear classification as either morphological or phrasal, at least at first view, such as phrasal (particle) verbs. In fact, German and Dutch seem to pattern very much alike with regard to the (number of) types of compound and MWE patterns that exist in both languages.

Leaving aside various minor differences and specific characteristics of each language, the major difference between German and Dutch seems to lie in the often noted observation that – at least in the nominal domain – Dutch seems to use phrasal patterns more often than German, which in contrast opts for compounding more frequently, although both patterns are in principle available in both languages, e.g., German Tag+es+gespräch, Dutch gesprek van de dag (lit. talk of the day, ‘nine days’ wonder’), German Stumm+film, Dutch stomme film (‘silent film’) (cf. van Haeringen 1956; De Caluwe 1990; Booij 2002; Hüning 2010; Hünig/Schlücker 2010, among others).

In English, on the other hand, the formal distinction between (nominal) compounds and phrases is notoriously difficult. First of all, the criterion of inflection is inapplicable in English. Secondly, the (formerly often invoked) criterion of stress has been shown in a number of works (cf., for instance, Plag 2006; Kunter 2011) to be incapable for drawing this distinction because although the vast majority of (NN) compounds have forestress, as predicted, there are also numerous exceptions, as can be seen from classical examples such as ‘apple

---

8 Stalen is a relational adjective derived from the noun staal ‘steel’.
cake vs. apple 'pie, 'Madison Street vs. Madison 'Avenue. Thirdly, the distinctive force of other tests which refer to the idea that compounds, being words, should be subject to lexical integrity (contrary to phrases), such as the pro-one test, internal modification, or coordination, have been proven weak in works such as Bauer (1998), Giegerich (2015) and Bauer (to appear). Also, the forms that evolve as either morphological or phrasal on basis of the stress criterion do not necessarily coincide with the outcomes of the other tests. For this reason, very divergent opinions on the definition of compounds in English and the demarcation from phrases can be found in the literature. A literature survey is beyond the scope of the present paper (but see, for instance, Olsen 2000; Lieber/Štekauer 2009b). Generally speaking, in addition to uniform analyses that assume that the constructions in question are either all morphological, and thus compounds, or all syntactic, and thus phrases, it has also been suggested that some of them are morphological whereas others are syntactic, depending on how the above-mentioned criteria are weighted (e.g., Giegerich 2004). Finally, it has been advocated that the inconclusive data are an indication of the fact that the compound-phrase distinction does not exist and that there is either a continuum or an overlap between syntax and the lexicon (e.g., Giegerich 2015; Bauer, this volume). Another problematic case is the ‘descriptive’ or ‘classifying’ genitives, e.g., lawyer’s fee, mother’s milk. Regardless of their obvious phrasal form, they are alike compounds in that the genitive dependent has a classifying rather than a determinative function, that it is immediately adjacent to the head noun, and that the constituents cannot be separated, e.g., by another modifier. For this reason, they have often been treated as compounds in the literature (cf. Rosenbach 2006: 82–89 for a literature survey).

In sum, the major difference between German (and Dutch) on the one hand and English on the other is that in English, due to the apparent impossibility of distinguishing clearly between morphological and syntactic N+N and A+N sequences, compounds are often regarded as just one kind of MWE, cf., for instance, Ramisch (2015), Bauer (this volume), whereas in German and Dutch, compounds and MWEs are clearly opposed and there are only few patterns that elude immediate classifications as either compound or MWE. Apart from that, the West Germanic languages pattern very much alike regarding the existence of various specific subtypes of compounds and MWEs. This similarity becomes particularly obvious when German is compared to other languages and language families.

9 Moon (2015), on the other hand, explicitly excludes compounds from the set of MWEs.
3.2 North Germanic languages and German

The North Germanic languages comprise the continental Scandinavian languages Swedish, Danish, and Norwegian, as well as the insular Scandinavian languages Icelandic and Faroese. As there is no separate chapter on complex lexical units in a North Germanic language in this volume, a short general description of the language family is in order. Generally, North Germanic languages are very similar to West Germanic languages in many respects. Distinctive features common to all North Germanic languages that are lacking in West Germanic languages include the suffixed definite article; the agreement of the adjective in gender and number not only in attributive, but also in predicative position; and the existence of a synthetic passive (termed s-passive or medio-passive, cf. Torp 2002). As to word order, North Germanic languages share with Dutch and German V2 in declarative sentences, where English dominantly has SV. On the other hand, North Germanic shares with English the predominant VO-pattern, where Dutch and German have OV. Within the North Germanic languages, the insular Scandinavian languages differ from the continental Scandinavian languages most notably in their rich inflectional morphology. While Swedish, Norwegian and Danish have a rather reduced inflectional morphology, Icelandic has, of all modern Germanic languages, the most differentiated inflection in the nominal, adjectival and verbal domain (Braunmüller 2007: 248), comparable to that of Ancient Greek or Latin, but with additional combinatorial phonological changes.

Compounding is a highly productive morphological process in all North Germanic languages, as in German. Generally, compounds in North Germanic languages, as in German, are right-headed, with inflectional endings attaching to the word-final element. Also, compounding in North Germanic languages is recur-

---

10 In this overview, we will concentrate on examples from Swedish, Danish, and Icelandic.
11 E.g., Swedish bil + en (lit. car + the, ‘the car’).
12 E.g., Swedish en stor bil (common gender, ‘a big car’), ett stort hus (neuter, ‘a big house’); bilen är stor (common gender, ‘The car is big’), huset är stort (neuter, ‘The house is big’).
13 E.g., Swedish dörren öppnade-s ‘the door was opened’, with the -s-suffix marking passive.
14 Cf. Swedish Där kommer hon, German Da kommt sie (Adv V S) (both lit. there comes she), but English There she comes (Adv S V).
15 This is reflected not only in subordinate clauses, but also in main clauses if one takes into account the position of infinite verbal parts. Cf. for main clauses Swedish Hon har sett huset, English She has seen the house (Vfin Vinf O) (both ‘She has seen the house’), but German Sie hat das Haus gesehen (Vfin O Vinf) (lit. she has the house seen); for subordinate clauses Swedish [Jag vet att] hon har sett huset, English [I know that] she has seen the house (Vfin Vinf O) (both ‘I know that she has seen the house’), but German [Ich weiß, dass] sie das Haus gesehen hat (O Vinf Vfin) (lit. I know that she the house seen has).
Compo unds and multi-word expressions in the languages of Europe

sive (e.g., Danish Kilde+skatte+direktorat+et ‘internal revenue service’ (lit. source tax directorate), cf. Haberland 1994, Icelandic Norð+austur+atlant+s+haf+s+fisk+veiði+nefndin ‘The North East Atlantic Ocean Fisheries (lit. Fish-Catching) Committee’, cf. Bjarnadóttir 2017). North Germanic compounds normally display solid spelling and carry stress on the first constituent. Nominal compounding (N+N, A+N, V+N) is by far the most common process, with N+N being the most productive pattern, approximately as in German (cf. Thráinsson 1994; Teleman 2005; Bauer 2009b). Some examples are Swedish ång+båt, Danish damp+skib, Icelandic gufu+bátur ‘steam boat’ (N+N); Swedish lille+finger, Danish lille+finger, Icelandic litli+fingur ‘little finger’, ‘pinkie’ (A+N); Swedish skriv+bord, Danish skrive+bord, Icelandic skrif+bord, lit. write table, ‘desk’ (V+N).

One difference concerning V+N compounding in the three languages is that V+N compounds in Swedish and Icelandic use the verbal stem as first constituent (skriv-, skrif-), while Danish V+N uses the infinitive of the verb ([at] skrive). This feature of Danish V+N compounds is distinct from German, which is also interesting from a theoretical point of view. If one takes infinitival endings as inflectional endings, the question arises whether Danish V+N compounds should be regarded as cases of compound-internal inflection. However, it is clear that infinitival non-heads are to be distinguished from cases where a non-head exhibits agreement features with the head. Only the latter case may pose a serious problem to the delimitation between compounds and syntactic phrases, since in cases with compound-internal agreement there is a potential overlap between compound and syntactic phrase.

A highly particular feature of Icelandic compounds, in contrast with all other Germanic languages, is that they systematically exhibit compound-internal inflection (cf. Bjarnadóttir 2017). This pertains both to a subclass of Icelandic N+N compounds, i.e. those with a genitive (or sometimes also a dative) non-head, as well as to all A+N compounds. As to N+N compounds, Bjarnadóttir (ibid.: 18) distinguishes between compounds with a stem as non-head (e.g., fjár+hús ‘sheep house’); compounds with a genitive as non-head (e.g., vegar+endi, ’end of road’, with vegar being one of two possible genitive forms of the noun vegur ‘way’); and a very small class of compounds with a special stem form or a linking element.

---

*Note that compounds in North Germanic languages do not exhibit regular capitalization (in contrast to German). The examples Kildeskattedirektoratet and Norðausturatlantshafsfiskveiðinefndin exhibit upper case because they function as proper names.

*Note that internal inflection, more generally, pertains to all nouns with a suffixed definite article in Icelandic. Thus, in definite nouns, the noun and suffixed definite article both inflect, e.g., hestur ‘horse’, hestur-inn ‘theNOM horseNOM’, hesti-num ‘theDAT horseDAT’.*
also Thráinsson 1994). While she acknowledges that the nature of the genitives in Icelandic compounds and the question of whether these are true inflectional forms or linking elements are matters of debate, Bjarnadóttir (2017: 19) argues for a genitive/inflectional analysis of these forms. One of her arguments in favor of this analysis is that the inflected forms of the non-head nouns are always the “correct” genitives, in spite of the complexity of the inflectional patterns. This stands in contrast with German, where forms such as *Liebesbrief (‘love letter’) are paradigmatically incorrect, the expected genitive feminine being Liebe, not Liebe+s+brief. Internal inflection is also found in the adjectival non-heads of A+N compounds, where agreement of gender, case, and number “is exactly the same within the compounds as in syntax” (ibid.: 28f.). For example, in litli+fingur ‘pinkie’ (lit. small finger), the ending -i in litli ‘small’ is a marker for masculine, singular, nominative, definite. In the accusative case, the compound form would be litla+fingur, with the ending -a in litla marking masculine, singular, accusative, definite. Thus, on purely inflectional grounds, it is not possible in Icelandic to differentiate between a definite noun phrase litli fingurinn ‘the small finger’ and a compound word in definite form, litli+fingurinn ‘the pinkie’. This distinction can be made only with the help of word stress (and spelling, though this is not a very robust criterion), with compound words carrying primary stress on the first constituent, and secondary stress on the second constituent in a binary compound.

In Swedish and Danish, on the other hand, compounds can be distinguished from phrasal constructions based on prosodic, morphological, and syntactic criteria. Swedish and Danish compounds, as in Icelandic, carry primary stress on the first constituent and secondary stress on the last constituent (cf. Teleman 2005; Bauer 2009b). According to the Swedish tonal system, which differentiates between accent 1 (“acute”) and accent 2 (“grave”), compounds carry accent 2, which is characteristic for polysyllabic words with primary stress on the first syllable (2-sport+bil ‘sports car’, 2-läs+glas+ögon, lit. read glass eyes, ‘reading glasses’).18 The difference between accent 1 and accent 2 is distinctive in pairs such as 2ande+n (‘spirit+definite’) and 1and+en (‘duck+plural’). For these pairs, accent 2 is a lexical accent differentiating lexical words from inflected word forms. This is specific for Swedish and contrasts with German. In German, there is a difference between lexical stress and phrasal stress, but not between lexical stress in words vs. word forms.

Moreover, in Swedish and Danish, compounds may be distinguished from phrases on formal grounds. Generally, in contrast to Icelandic, Swedish and Dan-

---

18 The exponent 1/2 replaces the primary stress sign.
ish compounds do not exhibit internal inflection. While in A+N phrases, the adjective must carry inflection (e.g., Danish et stort køb ‘a big purchase’, det store køb ‘the big purchase’), in A+N compounds, it is uninflected (e.g., Danish et stor+køb, stor+køb+et ‘a wholesale’). Syntactically, while the adjective in an A+N phrase may be modified (e.g., et meget stort køb ‘a very big purchase’, det største køb ‘the biggest purchase’), in an A+N compound, it may not (e.g., *meget stor+køb, lit. very big purchase, *størst+køb, lit. biggest purchase). Further evidence for the compoundhood of A+N compounds comes from definiteness inflection on the noun. While a single noun in Swedish and Danish takes a postposed definite article (hus+et ‘the house’), a premodified noun takes a preposed definite article (cf. Swedish det stora hus+et, Danish det store hus ‘the big house’). Thus, the correct definite form of the Danish compound hvid+vin ‘white wine’ is hvid+vin+en, but not *den hvid+vin, as would be expected of a phrase (Bauer 2009b).

In Swedish and Danish N+N compounds, the non-head may be changed morphologically in various ways. However, these forms are normally regarded not as inflection, but rather as linking elements, as in German (Niemi, S. 2009; Bauer 2009b). Swedish compounds may display vowel deletion (flicka > flick+skola ‘girl school’), vowel addition (tjänst > tjänst+e+man ‘service man’, ‘clerk’), or the addition of -s (stol > stol+s+ben ‘chair leg’) (cf. Josefsson 1997; Teleman 2005). Danish compounds may display an s-link (træning+s+bane ‘training ground’), an e-link (jul+e+dag ‘christmas day’), an er-link (blomst+er+bed ‘flower bed’) or an (e)n-link (rose+n+gaard ‘rose garden’). In general, this picture is consistent with West Germanic languages such as German and Dutch (but not English, which lacks linking elements).

Apart from compounds on the one hand and regular syntactic phrases on the other, a large stock of MWEs can be found in North Germanic languages, both those that in principle correspond to compounds and those that do not. For example, in Swedish, there are A+N phrases with a naming function such as röda hund ‘measles’ and hög hatt ‘top hat’; collocations such as ymnig grönska ‘lush greenery’ and duka bordet ‘lay the table’; complex verbs incorporating a non-referential noun such as knipa käft (lit. shut mouth) ‘keep one’s trap shut’ and vålla storm, lit. cause storm, ‘to cause a great stir’; idioms such as tala i skägget ‘to express oneself in an obscure way’; and speech act formulae such as Tack för senast ‘thanks for the other day’. As Koptjevskaja-Tamm (2009: 134) observes, lexicalized A+N phrases in Swedish may contain both indefinite (hög hatt ‘top hat’) and definite adjectives (röda hund, lit. red dog, ‘measles’), with definite adjectives combining with either unmarked nouns (röda hund ‘measles’) or nouns with the suffixed definite article (röda korset ‘the Red Cross’). However, what is avoided, according to Koptjevskaja-Tamm (2009), are lexicalizations of the nor-
mal pattern with preposed determiners, definite adjectives and nouns with suffixed article (as in *den gula hatten* ‘the yellow hat’) (cf. Section 2.5). A specific feature of Swedish MWEs is their connective prosody (cf. Anward/Linell 1976), whereby all stressed syllables in the MWE become deaccentuated, except for the last one. This can be taken as a distinctive feature for telling apart phrasal lexical units from phrasal syntactic units. In this respect, Swedish clearly differs from German, which does not distinguish lexical phrases from non-lexicalized phrases on prosodic grounds.

Generally speaking, the North Germanic MWE systems are very similar to the MWE system of German. Thus, there are overall commonalities both as to the number and the types of MWEs, including rather specific idioms such as German *auf keinen grünen Zweig kommen*, which directly corresponds to Swedish *ej komma på grön kvist* (lit. to not come onto a green branch, ‘to get nowhere’). However, there are also many language-specific differences in lexicalization which can be easily demonstrated, e.g., for the case of collocations. For example, in Swedish, there are several collocations with the verb *torka* ‘to dry (sth.)’, e.g., *torka bordet* (‘wipe the table’), *torka golvet* (‘wipe/clean the floor’), *torka disken* (‘dry the dishes’). While German has a direct verbal equivalent, *trocknen* ‘to dry (sth.)’, it uses three different verbs in combination with the respective nouns: *den Tisch abwischen/*trocknen ‘wipe the table’, *den Boden wischen/*trocknen ‘wipe/clean the floor’, *das Geschirr abtrocknen/*trocknen ‘dry the dishes’.

An interesting question is whether there are any tendencies in the North Germanic languages as to the use of compounds compared to their corresponding MWEs. It is well-known that Dutch, relative to German, tends to prefer MWEs over compounds, while German, relative to Dutch, tends to prefer compounds over MWEs (cf. Section 3.1). As to North Germanic languages, as far as we can see, comprehensive studies on this issue are lacking. There is some evidence, though, that Swedish tends to use compounds more frequently than corresponding MWEs compared to other languages. For example, Dura/Gawronska (2007), in a parallel corpus study on novel expressions, found that legislative concepts such as ‘quality control’ were realized in the Swedish corpus as compound nouns (*kvalitet+s+kontroll*), whereas the Polish parallel corpora used nominal phrases (*kontrola jakośki*). Combinations with ‘animal food’ were realized as compound nouns (*djur+foder* ‘animal food’, *fisk+foder* ‘fish food’) in the Swedish corpus, but as lexical noun phrases containing prepositional phrases (*karma dla zwierząt* ‘animal food’, *karma dla ryb* ‘fish food’) in the Polish corpus. Inghult (1991), in an investigation of the principles of lexical innovations in German and Swedish, found that only 3% of all new formations in dictionaries of neologisms were phrases, while 97% were word formations. Moreover, he found that Swedish often has compounds where German has MWEs, for instance, German *kupferne*
**Kanne** vs. Swedish *koppar+kanna*, ‘copper pot’. However, these somewhat outdated results from dictionaries should be treated with caution and are in need of confirmation by corpus-driven studies.

Comparing German and Danish, Farø (2015) finds that Danish tends to have MWEs where German has compounds, e.g., Danish *råget laks* vs. German *Räucher+lachs* (‘smoked salmon’), Danish *stor begivenhed* vs. German *Groß+ereignis* (‘major event’). However, there are also reverse pairs such as Danish *spanskrør* vs. German *Spanisches Rohr* (‘cane’). For a comparison of Dutch and Danish, Haberland (1994: 347) remarks that where Dutch would use derivational processes, Danish would use compounds, cf. Danish *vel+smagende* ‘well tasting’, ‘tasty’ vs. Dutch *smakelijk*. While more comprehensive studies on this issue are lacking, these observations suggest, overall, that the North Germanic languages tend to pattern with German with respect to the utilization of the two competing processes.

An interesting commonality between the North Germanic languages, German, and Dutch, which clearly sets them apart from English, is put forward by Klinge (2006). Klinge investigates the [N de N] construction, which is well-known from French (e.g., *prisonnier de guerre*). Interestingly, this is also a productive pattern of formation in English (e.g., *prisoner of war*), yet not or only marginally in other West Germanic languages (German, Dutch) or indeed in North Germanic languages such as Danish or Icelandic. Thus, where English has *bird of prey*, German has *Raub+vogel*, Danish *rov+fugl*, and Icelandic *rán+fugl*. The hypothesis put forward by Klinge is that this may be explained as a language contact phenomenon. Thus, the originally Romance [N de N] pattern was adopted in English from Norman French. This would explain why it does exist in English, but not in Dutch, German, Danish, or Icelandic. Importantly, Klinge argues that MWEs such as *weapons of mass destruction* in English are not the result of some isolated lexicalization of a syntactic phrase, but instead reflect the presence of a lexical formation pattern [N de N] in English which instantiates such structures directly as lexical units.

In sum, one can say that the North Germanic languages largely pattern with German with respect to the availability and utilization of the processes of compounding and MWE formation. The most significant differences between North Germanic and German are to be found in the Icelandic possibility of compound-internal inflection, which makes Icelandic compounds look more “syntactic” than German compounds. However, in many other respects, the commonalities outweigh the differences.
3.3 Romance languages and German

In Romance, morphological compounding is much more restricted than in German. Verbal compounding does not exist (or is very marginal) just as in German, but the number and productivity of nominal and adjectival compound patterns is lower than in German. In general, the notion of compound has often been used also to include phrases, and thus MWEs, for instance nominal constructions containing a preposition or an inflected adjective, e.g., French *moulin à vent* ‘wind mill’, Italian *macchina da scrivere* (lit. machine to write) ‘typewriter’, Spanish *casa de campo* ‘country house’, French *guerre froid-e* ‘cold war’, Spanish *mal-a suerte* ‘bad luck’. Obviously, the key reason for classifying such forms as compounds is their semantic-functional property of serving as a conventional naming entity for a unitary concept. It seems safe to say that in comparison to German such “syntagmatic/syntactic/improper compounds” (as they are often termed in the literature) are much more frequent in Romance. This can also be illustrated by the fact that the German counterparts of all of the above-mentioned examples are compounds, except for the last two, which are a lexical phrase (*kaler Krieg*) and a simplex word (*Pech*). Just as in German, these MWEs either have a fully regular syntactic structure (e.g., French *homme de la rue*, lit. man from the street, ‘average person’) or are syntactically deficient, for example in that the determiner is missing, e.g., French *château d’eau* (lit. palace of water, ‘water tower’) (e.g., Gunkel/Zifonun 2011; Gunkel et al. 2017: 1625).

Turning to “proper”, morphological compounds, it is striking that for each of the languages under discussion there is no general agreement in the literature as to precisely which constructions should be classified as such. Obviously, the main reason for this is the difficulty in providing generally valid properties of morphological compounding. This problem is illustrated by the definition given in Fradin (2009: 417): “Compounds may not be built by syntax (they are morphological constructs).” Thus, compounds are defined only negatively as non-syntactic, yet this leaves open the exact nature of morphological constructs. The problem is that many of the criteria that can be positively established for compounds in other languages, and in particular in German, are not available in Romance. The first one is the absence of a unitary compound stress rule in Romance (Rainer/Varela 1992; Arnaud 2015; Fernández-Domínguez, this volume). Thus, compounds and MWEs are basically stressed in the same way, contrary to German where compounds can clearly be distinguished from phrases on the basis of stress (modifier vs. head stress). (Native) linking elements, another common property of German (N+N and V+N) compounds, do not exist in French and Italian. However, the native linking vowel *-i* is found regularly in some adjectival and nominal compound patterns of Spanish, e.g., *roj+i+blanco* (lit. red white,
‘red and white’). Regarding headedness, French and Italian compounds are generally left-headed, e.g., French *stylo-bille* (lit. pen ball, ‘ball pen’), Italian *pesce-spada* (lit. fish sword, ‘sword fish’). However, Spanish, in addition to left-headed compounds, e.g., *célula madre* (lit. cell mother, ‘stem cell’), also has some right-headed compound patterns (cf., e.g., Guevara 2012; Rainer 2016), both adjectival and nominal ones (cf. Fernández-Domínguez, this volume), e.g., *drog+adicto* (lit. drug addict, ‘addicted to drugs’). For Italian, on the other hand, Masini/Scalise (2012) argue that the existence of right-headed compounds does not provide evidence against the assumption that Italian compounding is generally left-headed because these cases are either neoclassical formations, Latin relics, or English calques, such as *scuolabus* ‘school bus’. Another frequently mentioned property of compounds, which is again particularly valid for compounding in German (although not for all compound subpatterns) is recursivity. In general, compounding is not considered to be recursive in the Romance languages under discussion (cf., for instance, Scalise 1992 on Italian), with the exception of coordinate (or: copulative) compounds (e.g., Arnaud 2015 on French). Also, solid spelling – which is often said to be indicative of the compound (versus phrase) status in German – is often found with morphological compounds, as well as hyphenated spelling. At the same time, however, there are also compounds with an unstable spelling (cf. Fernández-Domínguez, this volume) as well as MWEs written as one word (e.g., Van Goethem 2009; Van Goethem/Amiot, this volume).

So far, this brief overview has shown that in contrast to German it seems much more difficult to provide clear criteria for morphological compounds as opposed to MWEs in French, Spanish, and Italian. However, two important criteria are still missing. They are among those that have been established by Lieber/Štekauer (2009b: 8) as more general, cross-linguistic criteria of compounding, namely (in addition to stress) (a) syntactic impenetrability, inseparability, and unalterability, and (b) inflection. The first criterion is difficult to assess. On the one hand, it is a basic criterion for distinguishing compounds from phrases (cf., for instance, Fernández-Domínguez, this volume, Van Goethem/Amiot, this volume). On the other hand, however, it is well-known that it also applies to some, though not all kinds of lexicalized phrases (cf., for instance, Gunkel/Zifonun

---

19 In addition, Latinate and Greek linking elements are found in neoclassical compounding of all three languages, cf., for instance, Villoing (2012) on French.
20 It goes without saying that spelling is subject to conventional norms and possible changes of normative rules and, for these reasons, cannot be regarded as evidence for the grammatical status of forms. However, in particular non-normative writing tendencies might be indicative of the writer’s assessment of a form as a conceptual unit.
Thus, syntactic impenetrability, inseparability, and unalterability can be regarded as a necessary criterion of morphological compounds but not as a sufficient one that distinguishes compounds from MWEs. The second criterion, inflection, very clearly distinguishes compounds from phrases in German, as compounds contain only stems and not inflected constituents. This criterion is highly problematic for the Romance languages which has, among other things, to do with the fact that Romance compounds are generally left-headed. Thus, if in a nominal compound the plural is marked on the head, this results in word-internal inflection, e.g., French *poisson*<sub>SING</sub>*-scie*<sub>SING</sub> – *poissons*<sub>PL</sub>*-scies*<sub>PL</sub> (lit. fish<sub>SING/PL</sub> saw<sub>SING/PL</sub>, ‘sawfish’). It seems that the three languages have both constructions with word-internal inflection and those without. Thus, plural, for instance, is sometimes marked only on one (usually the left) constituent and on both constituents in other cases. In particular, coordinate compounds seem to regularly inflect for plural on both constituents (e.g., French *auteurs*<sub>PL</sub>*-compositeurs*<sub>PL</sub> ‘songwriters-composers’). The question then is which conclusions can be drawn from these observations. In other words: how much value is attached to this criterion regarding the definition of compound? As expected, different positions can be found in the literature: Whereas some scholars quite naturally accept inflected forms as word-internal building-forms of compounds (e.g., Scalise 1992; Guevara 2012; Masini/Scalise 2012; Arnaud 2015), others are more restrictive (e.g., Villoing 2012, on French).

In sum, it is obvious that although in all three languages at hand there are constructions that are clearly morphological (and thus compounds) and others that are clearly syntactic (and thus MWEs) it is very difficult to draw a clear border between them. In this connection, proposals have been made in the context of constructionist frameworks which do away with the idea of a clear-cut borderline between syntax and the lexicon (cf. Masini 2009; Van Goethem/Amiot, this volume; Masini, this volume). If we compare German and the Romance languages with regard to compounding and MWEs, three differences can be noted: firstly, in contrast to the Romance languages, German does allow (with very few exceptions, cf. Schlücker, this volume) a clear-cut distinction between compounds and MWEs. Secondly, although empirical evidence cannot be provided here, it seems that both the number of clearly morphological compound patterns as well as the specific forms instantiated from these patterns are much rarer in Romance languages than in German, and for this reason, MWEs prevail in Romance. Thirdly, if we compare the morphological compound patterns of the Romance languages and German, two Romance patterns stand out from a German (or Germanic) perspective. The first one is V+N compounding, a productive pattern of exocentric compounding, consisting of a verb and a noun which functions as the direct object of that verb, e.g., French *abat-jour* (lit. weaken light, ‘lampshade’),
Italian *porta+bagagli* (lit. carry luggage, ‘trunk’), Spanish *cubre+cama* (lit. cover bed, ‘bedspread’). The pattern is regarded as typical for the Romance languages and it does rarely exist in other Indo-European languages, and not at all in German nor in Dutch (there are sporadic English examples such as *turncoat, killjoy*). Although there have been debates concerning the stem form (and thus the morphological nature) of the left, verbal constituent, these constructions are relatively uniformly regarded as morphological compounds in contemporary works (see the literature cited in this section as well as Ricca 2015, who also discusses the interlinguistic differences of V+N compounds within Romance).

The second pattern are coordinate compounds (A+A, N+N) which can be said to be fairly regular and productive in French, Italian, and Spanish (though with some restrictions regarding specific subpatterns in the individual languages). In comparison to German, they are interesting for two reasons: first, with regard to form, coordinate compounds often show inflectional marking on both heads and thus word-internal marking, which is impossible in German. One could therefore argue that the pattern is more morphological in German than in Romance. Second, the existence of N+N coordinate compounds has been widely discussed in the literature on German (in contrast to A+A coordinate compounds, whose existence has not been questioned). The main argument is that in many cases of alleged N+N coordinate compounds it seems hard to establish a semantic coordinate relationship and thus two semantic heads; instead, a determinative interpretation is available in equal measure or even preferred. There are only very few clearly nominal coordinate compounds in German (with an additive meaning) such as toponyms, for instance the names of federal states that consist of two regions, e.g., *Nordrhein-Westfalen* ‘North Rhine-Westphalia’, or technical terms such as *Sprecherschreiber* ‘speaker-writer’ which is however restricted to linguistic terminology. Thus, although in general German seems to be much more prone to morphological compounding than the Romance languages which in contrast make much more use of MWEs, there are at least these two patterns of compounding that constitute an exception from this general distribution of use of forms.

### 3.4 Modern Greek and German

Regarding compounds and MWEs, Modern Greek and German display many similarities. Compounding in Greek is, just as in German, a very productive device of word-formation, and both languages have various MWE patterns. As in German, compounds can be distinguished clearly from syntactic phrases (both MWEs and common ones) in Greek.
Starting with compounding proper, it is remarkable that virtually all properties that have been identified for German compounds can also be found in Greek (for comprehensive descriptions cf. Ralli 1992, 2009, 2013a, 2013b, 2016). The vast majority of Greek compounds is endocentric and right-headed, e.g., domat+o+saláta ‘tomato salad’. Also, Greek compounds have lexical stress. More precisely, they are single-stressed and therefore form one phonological word, contrary to phrases (e.g., compound stress on the antepenultimate syllable in kapnoxórafo ‘tobacco field’ < kapn(ós) ‘tobacco’ xoráf(i) ‘field’). In contrast to German compounds, however, which (in simple compounds) always have stress on the first constituent, compound stress in Greek compounds is more variable, depending largely on the phonological properties of the second constituent. Thus, there are several single-stressed compound patterns (e.g., Ralli 2013b: 186f.). Greek compounds consist of either stem or word constituents (most frequently, the left constituent is a stem with the right one either a stem or a word). In any case, they clearly do not have word-internal inflection. Another important point relates to linking elements. There is only one linking element, -o-, e.g., kapn+o+xórafo ‘tobacco field’, which is almost compulsory in Greek compounds (there are only a few phonologically conditioned exceptions). For this reason, Ralli (2008) treats linking elements in Greek and in general as compound markers. The occurrence of Greek -o- is much more systematic compared to linking elements in German compounds, which are restricted to particular compound subpatterns and display a broad variety of forms, including the zero form. Finally, Greek compounds display solid spelling, contrary to phrases, just as in German.

As to the differences, it seems that recursiveness – which is usually considered a typical property of German N+N compounds – is possible in Greek, too (cf. Ralli 2009), but much rarer (cf. Koliopoulou, this volume). More importantly, while German does not have verbal compounding, it is a productive pattern in Greek, with either verbs, nouns or adverbs as left constituents, e.g., N+V: xaropalévo ‘fight (with) death’ (xár(os) ‘death’, palévo ‘fight’), Adv+V: kakopernó ‘live badly’ (kak(á) ‘badly’, pernó ‘pass, live’) (e.g., Ralli 1992). Meanwhile phrasal compounds, that is, compounds with a phrasal modifier constituent, do not exist in Greek, in contrast to German.

Greek compounds can clearly be distinguished from phrases on the basis of stress, the linking element -o- and the absence of inflectional markers. Furthermore, morphological compounds are subject to lexical integrity and thus the usual tests (as known from the literature on English and other languages) can be applied: inseparability and the inability to modify the non-head constituent and refer pronominally to the individual constituents (a comprehensive overview of the diagnostics for the compound-phrase distinction is given in Bağrıaçık/Ralli 2015, for instance).
As for MWEs, two particularly interesting constructions that relate to the present issue have been discussed in detail in the literature on Greek, namely [A N] and [N N\text{GEN}] sequences. Classical examples are *psixrós pólemos* ‘Cold War’ ([A N]) and *zóni asfalías* (lit. belt safety, ‘safety belt’), i.e. an [N N\text{GEN}] sequence with the second, non-head constituent assigned genitive case. These [A N] and [N N\text{GEN}] sequences are lexical units with a stable conventional meaning, many of them being scientific terms. They are phrasal, and thus syntactic entities, sharing some features with (morphological) compounds and are inaccessible for the syntactic operations that phrases normally allow. Thus, they are hybrid constructions and have, for this reason, been termed phrasal compounds (not to be confused with compounds containing a phrasal modifier constituent), syntactic compounds or loose multi-word compounds (cf., e.g., Ralli 1992; Ralli/Stavrou 1998; Bağrıaçık/Ralli 2015; Ralli 2016; Koliopoulou, this volume). They are phrasal in that they exhibit full inflectional marking as well as phrasal stress, thus they have two distinct prosodic domains. Also, the [N N\text{GEN}] sequences are left-headed. On the other hand, they behave unlike syntactic phrases, and like morphological compounds in that they are inseparable and do not allow modification of the non-head constituent, e.g., *métria psixrós pólemos* (lit. moderately cold war). Also, the [A N] sequences do not allow doubling of the definite article (and neither do A+N compounds), which is a usual constellation in common A N phrases, e.g., *o psixrós o pólemos* (lit. the cold the war, ‘the Cold War’), but *o meýálos o pólemos* (lit. the big the war, ‘the big war’) (Ralli 2016: 3147). Also, many of these lexical phrases do not have a compositional meaning, just like compounds (e.g., Ralli 1992). In addition to these two patterns that have been described in detail in the works of Ralli (and colleagues), there are also several other lexical syntactic patterns, consisting of two inflected nouns, cf. Gavriilidou (2013), Ralli (2013a), Koliopoulou (this volume).

In sum, these sequences are clearly both lexical units and syntactic entities, and thus MWEs. For this reason, they pose a challenge as to their exact grammatical status, given the fact that they combine syntactic and morphological properties (which is obviously not necessarily the case for all MWEs). This is particularly important because they are not individual instances of lexicalization but rather the result of productive patterns for creating new lexical units, just as with compounding.\footnote{Contrary to compounding, though, they seem to be a rather recent pattern. According to Ralli (2013a), they have been observed only in the last two centuries and have most probably emerged under the influence of French and English. Also, they are almost always restricted to specific registers.} In the light of this, Booij (2009, 2010) offers a formal analysis for
Greek lexical [A N] sequences as syntactic compounds (N₁) within the framework of Construction Morphology; similarly, a constructional analysis for various lexical [N N] sequences is proposed in Gavrilidou (2013).

3.5 Slavic languages and German

Slavic languages, exemplified here by Russian and Polish, differ clearly from German with respect to the formation of new lexical items and in particular compounding. Although compounding, and in particular nominal compounding, is a productive word-formation process in both languages, it is a less important means for expanding the lexicon than it is in German (and other languages, such as English), particularly since derivation is highly productive (Uluhanov 2016).

As in German, nominal compounds, in particular N+N compounds, are the predominant compound type both in Russian and Polish, e.g., Polish gwiazd+zbiór (lit. star+set, ‘constellation’), Russian gaz+snabżenie (‘gas supply’), followed by adjectival compounds, e.g., Polish ciemn+niebieski (‘darkblue’), Russian tëmn+sinij (‘darkblue’). Verbal compounding is considered unproductive in Polish (cf. Szymane 2009) and only marginally productive in Russian (cf. Benigni/Masini 2009), although both languages have a rather small inventory of (older) verbal compounds. Generally, there are neither compounds with verbal modifiers (V+X) (cf. Ohnheiser 2015: 761) nor phrasal modifiers (XP+X) (cf. Bağrıaçık/Ralli 2015: 344; Szymane 2017) in Slavic, in contrast to German. Compounding is mostly right-headed, although there are also some (minor) left-headed subpatterns. Compounds proper have a linking element, mostly -o-, as in the above-mentioned examples or, less frequently, -e-, -i-, -u-, and they are written in one word (or with a hyphen). Compounds in Polish display lexical stress on the penultimate syllable which clearly sets them apart from phrases. Finally, Polish and Russian compounds are hardly recursive; compounds with more than two constituents are only found with adjectival coordinate compounds, e.g., Polish polsko-rosyjsko-ukraińskie (‘Polish-Russian-Ukrainian’).

---

22 In the following, only the most important and basic properties of compounding in Russian and Polish are described. For further details, such as the difference between proper compounds and solid compounds, the various kinds of input elements, neoclassical compounding, the gender class shift etc. the reader is referred to the contributions by Ohnheiser (on Russian) and Cetnarowska (on Polish) in this volume, as well as Szymane (2009), Benigni/Masini (2009), Ohnheiser (2015), Uluhanov (2016), Nagórko (2016).
In addition to the formation of endocentric and coordinate compounds of various kinds which are familiar from the German(ic) perspective, Polish and Russian (and Slavic in general) have another frequent and productive type of compounding, namely synthetic (or parasynthetic) compounding (cf., e.g., Benigni/Masini 2009; Melloni/Bisetto 2010; Ohnheiser 2015). This pattern is rare in German(ic), e.g., English *blue-eyed*, German *blauäugig*. In this case, a suffix is added to the compound simultaneously with the combination of the two constituents, e.g., Polish *nos+o+roż+ec* (‘rhinoceros’), with *nos* ‘nose’, *róg* ‘horn’, the linking element -o- and the nominal suffix -ec, or Russian *rabot+o+da+tel* (‘employer’), with *rabota* ‘work’, *dat* ‘give’, the linking element -o- and the nominal suffix -tel’. Importantly, the linking element and the suffix necessarily co-occur and enter the structure at the same time. For this reason, they are referred to as co-formatives (cf., e.g., Szymanek 2009; Nagórko 2016). Another interesting recent phenomenon in Russian are N+N compounds that are modelled on Germanic N+N compounds and which contain stems borrowed from English or German, e.g., *press-diskussija* (‘press discussion’), *eskort-uslugi* (‘escort service’), cf. Kapatsinski/Vakareliyska (2013). The authors suggest that they are not instances of borrowing of individual lexemes but rather a specific compound pattern that has been developed on the basis of the individual forms.

The observation that Russian and Polish make only limited use of compounding compared to German and English has often been attributed to the fact that – in addition to the high productivity of derivational processes – these languages have various productive MWE patterns, in particular nominal ones. For instance, Szymanek (2009: 465f.) notes that the equivalents of English N+N compounds such as *telephone number*, *toothpaste*, and *computer paper* are realized in Polish either as a noun phrase with an inflected noun modifier, usually in the genitive (e.g., *numer telefonu* ‘telephone number’), a noun phrase with a PP modifier (e.g., *pasta do zębów* ‘toothpaste’), or a noun phrase with a relational adjective as modifier (e.g., *papier komputerowy* ‘computer paper’). Other patterns for the formation of lexical noun phrases (or: phrasal nouns) mentioned in the literature are N+N sequences with a noun modifier case other than genitive, N conj N (binomials), and A+N patterns. Thus, there are both N+A and A+N patterns, the adjective being often but not necessarily relational (cf. Masini/Benigni 2012; Cetnarowska 2015; Nagórko 2016; Cetnarowska 2018; Cetnarowska, this volume; Ohnheiser, this volume). Masini/Benigni (2012) stress that of all these patterns the A+N pattern is by far the most productive one in Russian.

In a similar way as has been discussed for the various phrasal lexical units in other languages in the preceding sections, Russian and Polish phrasal lexical units, or more specifically, lexical noun phrases, can be distinguished both from free syntactic phrases and from compounds on formal grounds. At the same time,
they also share properties with free syntactic phrases and compounds. For instance, these lexical noun phrases are inseparable. That is, they cannot be interrupted by intervening material, e.g., Russian *sotovyi telefon* (‘mobile phone’), but *sotovyi služebnyi telefon* (lit. cellular official telephone). Also, the individual constituents cannot be modified internally, e.g., Russian *posobie po bezrabotice* ‘unemployment benefit’, but *posobie po ženskoj bezrabotice* (lit. benefit by female unemployment). These are properties typical of morphological entities and unlike free syntactic phrases; also, the function of lexical noun phrases as lexical naming unit equals that of compounds. On the other hand, lexical noun phrases display inflectional markers, like free syntactic phrases and unlike compounds, and some patterns contain relational elements, thus prepositions and conjunctions (as po in the last example), again like free syntactic phrases and unlike compounds (for a more detailed discussion of the tests employed including (apparent) counterexamples cf. Masini/Benigni 2012; Cetnarowska 2015; Ohnheiser 2015; Cetnarowska 2018; Cetnarowska, this volume; Ohnheiser, this volume). Thus, again it can be shown that these lexical noun phrases are lexical entities on the interface of syntax and the lexicon, i.e. lexical entities that are created in syntax. Building on works by Booij (2009, 2010) on A+N phrases in Dutch and Greek, among others, constructionist analyses have been proposed for these Russian and Polish lexical noun phrases in Masini/Benigni (2012), Cetnarowska (2018), Cetnarowska (this volume).

MWEs, and lexical noun phrases in particular, are also known from German, although it seems likely that these (or comparable patterns) are less productive in German than they are in Slavic, given the predominance of compounding in German. There is, however, another process for the formation of lexical items which stands in a close relationship to MWEs and MWE formation. This process is specific for Slavic and without a real equivalent in German. It is a process of shortening phrasal items to a single morphological lexeme.23 More precisely, there are several shortening processes, among them ellipsis, truncation, clipping and de-suffixation (cf. Masini/Benigni 2012 on Russian; Martincová 2015 on Slavic in general), e.g., Polish *rzut karny* (‘penalty throw’) > *karny* (lit. penal), Russian *mineral’naja voda* (‘mineral water’) > *mineralka*. These processes are referred to either as shortening, condensation or univerbation, although the latter term is somewhat misleading as univerbation is often understood elsewhere.

---

23 There are, obviously, also shortening processes such as clipping and contamination in German. They are much less systematic in nature than the shortening processes in Slavic however. Also, they occur only sporadically and are much less frequent. Finally, they do not require MWEs as input structures.
Compositions and multi-word expressions in the languages of Europe

(i.e. in the non-Slavic literature) as the fixation of a phrasal item as a single word, without any shortening or change of the form. These shortenings produce forms that are synonymous to the input phrases. However, they belong to a different register as they are usually considered to be much more colloquial, expressive and informal than the corresponding phrases. They are considered to be very productive which, according to Martincová (2015), might also be related to the lower productivity of compounding in Slavic. Importantly, it is usually assumed that the input of these shortenings are not phrases in general but rather MWEs (for discussion on this point cf. Masini/Benigni 2012; Martincová 2015). If this is the case, then the productivity of shortenings presupposes productivity of MWE formation processes and ultimately, MWE formation not only creates phrasal lexical units but also systematically underlies the formation of non-phrasal lexemes in Slavic.

3.6 Finno-Ugric languages and German

Compounding is a productive word-formation pattern both in Finnish and Hungarian and can, according to Niemi, J. (2009) and Pitkänen-Heikkilä (2016), even be considered the most productive word-formation device in Finnish. The output classes are nominal and adjectival compounds, with N+N compounds being particularly productive, e.g., Hungarian vér+nyomás ‘blood pressure’, Finnish tee+kuppi ‘tea cup’. According to Kiefer (2016: 3310), the productivity of nominal compounding in Hungarian has been considerably increased through loan-translations of thousands of German nominal compounds at the beginning of the 19th century. There are very few (apparent) verbal compounds (less than 1% in Finnish according to Kolehmainen/Savolainen 2007) and these forms are regarded either as backformations, univerbations from verbal phrases, or derivates rather than as compounds proper (cf. Kolehmainen/Savolainen 2007; Kiefer 2009, 2016; Pitkänen-Heikkilä 2016). Compounding in Finnish and Hungarian is also similar to German(ic) in that the vast majority of compounds is endocentric and right-headed. Furthermore, Finnish and Hungarian compounds display lexical stress which distinguishes them from phrases. (N+N) compounds are recursive, e.g., Hungarian [(vér+nyomás]+ mérő] ‘blood-pressure measuring’, [[[vér+nyomás]+ mérő]+ készülék] ‘blood-pressure measuring apparatus’ (Kiefer 2009). Finally, compounds are written as one word.

However, there are also clear differences between compounding in German and in Finnish and Hungarian. The first one is that compounds in Finnish and Hungarian never have linking elements. The second, and more important one, is that in addition to uninflected adjectival and nominal modifiers, e.g., Hungarian
feketeszoftver (A+N) ‘black/illegal software’, Finnish kylmävarasto (A+N) ‘cool storage’,24 Finnish and Hungarian compounds also include inflected modifier constituents, thus, word-internal inflection, e.g., Hungarian bolond-ok-ház-a (N+N) ‘mad house’, with the plural suffix -ok (and the possessive suffix -a), Finnish käde-sija (N+N) (lit. hand’s place) ‘handle’, with the genitive suffix -n. Regarding Hungarian, Kiefer (2009: 539) argues that they are not productive and morphologically formed compound patterns but rather univerbations from verb phrases or possessive noun phrases. Accordingly, there are no compounds with word-internal inflection in Hungarian. In Finnish, on the other hand, sequences with an inflected adjectival or nominal modifier constituent (mostly in the genitive, but also in other cases) are considered compounds proper (e.g., Niemi, J. 2009; Karlsson 2015; Pitkänen-Heikkilä 2016; Hyvärinen, this volume), similar to Icelandic (cf. Section 3.2). They form a considerable part of all cases: according to a corpus study by Niemi, J. (2009), about 14% of all nominal modifiers are inflected, about 20% of all adjectival modifiers and 22% of the verbal modifiers. Interestingly, compounds with a genitive modifier often have a possessive interpretation, e.g., tuoli-jalka (lit. chair’s leg) ‘leg of a chair’ (cf. Pitkänen-Heikkilä 2016: 3214), which seems to suggest that these sequences are univerbated possessive phrases rather than compounds. However, possessive relationships are also found with non-genitive modifiers and genitive modifiers may also express other meaning relations (cf. Hyvärinen, this volume). Niemi, J. (2009: 239f.) claims that with respect to syntactic islandhood and lexical integrity, respectively, which underlie the debarment of word-internal inflection, Finnish compounds differ from other Standard European languages. Although from a morphological perspective, the sequences in question are phrases rather than compounds, they are regarded as compounds due to their lexical stress pattern which distinguishes them clearly from phrases.25 Thus, the prosodic structure is regarded as decisive (and more important than the morphological one) for the classification as compound in the Finnish literature; again as in Icelandic. Recall from the previous sections that, in contrast, sequences classified as lexical noun phrases (phrasal nouns) in various languages retain phrasal stress, which is one property that distinguishes them from morphological compounds.

---

24 More precisely, in the Finnish literature, this is regarded as the nominative case, as the nominative equals the base form without any inflectional suffixes (cf., e.g., Hyvärinen, this volume).

25 An additional, morphosyntactic criterion is that possessive suffixes and clitics that can be added to Finnish nouns are not allowed inside compounds (Niemi, J. 2009: 241f.).
Lexical noun phrases, their morphosyntactic properties and the question of whether there are productive patterns of the formation of lexical noun phrases have to our knowledge not yet been discussed in the literature, at least not in the non-Finnish and non-Hungarian speaking one. However, lexical noun phrases obviously exist, e.g., Hungarian nyári szünet (AREL N) ‘summer holidays’, állatok világa (NPL NPOSS) ‘animal kingdom’, Finnish valkoinen valhe (A N) ‘white lie’, also in terminology, e.g., Finnish jätteiden poltto (N PL GEN N) ‘waste combustion’, batalainen vyöhyke (AREL N) ‘bathyal zone’ (cf. Liimatainen 2008).

In the verbal domain, meanwhile, several interesting phenomena with respect to the lexicon-syntax interface have been discussed (cf. Hyvärinen, this volume). For Hungarian, Kiefer (1990, 1992, 2009) and Kiefer/Németh (this volume) describe the phenomenon of quasi-noun incorporation, that is, combinations of a bare noun and a verb, e.g., levelet ír (lit. letter write) ‘to do letter writing’, zenét hallgat (lit. music listen) ‘to do music listening’, in contrast to ‘writing a letter’ or ‘listening to a (particular) piece of music’. These complex verbs always denote institutionalized activities. They are similar to compounds in that they exhibit compound stress. Also, the non-head cannot be modified, pluralized and is non-referential, just as a compound modifier. On the other hand, the noun and the verb can be separated, e.g., by the negative particle nem ‘not’ (cf. Kiefer 1992: 76) which indicates their phrasal nature. Thus, these complex verbs can be regarded as verbal MWEs. Quasi noun-incorporation also exists in German and Dutch (as well as in Danish, Norwegian, and Swedish, cf. Section 3.2). For a constructionist analysis for quasi-noun incorporation in Dutch, see Booij (this volume).

3.7 Discussion

This section is devoted to a discussion and summary of the preceding sections. The first, very simple observation is that all languages examined here have morphological compounds. However, it turned out that the compounds in these languages do not all share the same defining properties. While lexical (compound) stress, headedness (either right or left), inseparability and debarment of word-internal inflection, recursiveness, and linking elements are generally considered essential criteria for the definition of compound, in particular from a German(ic)

26 There are, however, numerous studies on MWEs in Finnish and Hungarian in the more traditional sense, written in German or English, in particular on verbal idioms, including various Hungarian-German and Finnish-German contrastive studies. For an overview on Finnish cf. Hyvärinen (2007).
perspective, all of them also emerged as problematic in at least one language, or as non-existent.\textsuperscript{27} Thus, it seems that there is no universal definition of compound. Rather, as pointed out by Ralli (2013b: 184):

> What makes a compound morphological should be defined on a language-specific basis, since languages vary with respect to the realization of their morphological features and the use of morphologically-proper units.

Although it is ultimately impossible to weigh the various criteria against each other, it seems that compounding in German is – hardly surprisingly given the genetical relation – particularly similar to Dutch as well as the continental North Germanic languages, but also to Greek.

In addition to the defining criteria, also the number of compound subpatterns and the productivity of these patterns vary considerably between the languages. Verbal compounding, for instance, is regarded as either unproductive or only marginally productive in most languages, in contrast however to Greek which has several productive verbal compound patterns. What all languages discussed here have in common is that nominal compounding, and in particular N+N compounding, is considered the most frequent and probably also the most productive compound type (cf. likewise Guevara/Scalise 2009 for a much larger language sample).

A second observation is that all languages under discussion have MWEs and in particular MWEs that correspond or equal functionally to compounds.\textsuperscript{28} Notably, it has been observed that all languages have various productive patterns that instantiate these phrasal lexical units.\textsuperscript{29}

In the literature, the existence and productivity of MWE patterns is usually explained in relation to the existence and productivity of corresponding compound patterns and other word-formation processes, in particular derivation. Thus, for instance, compounding has been deemed comparatively limited in

\textsuperscript{27} Guevara/Scalise (2009) correctly point out that defining criteria of compounding such as those mentioned here usually reflect the Germanic perspective, given the huge amount of studies on compounding in Germanic, but cannot do justice to compounding from a broader perspective.

\textsuperscript{28} In this overview, more attention has been given to nominal MWEs than to verbal ones. This is due to reasons of space as well as to the fact that the starting point of this study is German, and that German compounding is predominantly nominal. Verbal lexical phrasal units are studied in detail in the chapters on Dutch, Finnish, and Hungarian (this volume).

\textsuperscript{29} As noted in Section 3.6, as far as we are aware there is no English or German speaking literature on lexical noun phrases and, in particular, on the respective lexical patterns in Finnish and Hungarian so far. There are, however, studies on complex verbal lexical units.
Slavic due to the productivity of both derivation and MWE formation, whereas the high productivity of nominal compounding in German has often been used as an explanation for the fact that the number and productivity of nominal German MWEs seem to be lower than in other languages.

Comparing the MWE patterns, it turned out that all languages have, among others, productive patterns for the formation of \([A \text{ N}]\) phrasal units (or \([\text{ N A}]\), in left-headed configurations). Among these, units with a relational adjective play an important role. In addition, some languages (among which German, Dutch, Danish, Swedish, Polish, and Greek) also have morphological \(A+N\) compounds, which raises – for each language – the question of synonymy and synonymy blocking.

Another phrasal pattern that can be observed cross-linguistically are so-called phrasal similes, i.e. comparative adjectival phrases of the type \([\text{ (as) A as NP}]\), e.g., \textit{as red as blood} (cf. Section 2.5). Phrasal similes are attested in the West and North Germanic languages as well as in Finnish and Italian, \(^{30}\) e.g., Swedish \textit{mjuk som silke}, German \textit{weich wie Seide}, both ‘soft as silk’, Italian \textit{rosso come il sangue} ‘red as blood’ (note that not all comparisons make sense in their literal meaning, e.g., Danish \textit{dum som en dør} ‘as stupid as a door’). They are particularly interesting with respect to the question of synonymy and synonymy blocking since all these languages also have an equivalent \(A+N\) compound pattern with a comparative meaning, e.g., \textit{blood-red}, Swedish \textit{silkesmjuk} ‘silky smooth’. It can be observed that in some cases the existence of a phrasal or morphological form blocks the other (e.g., Swedish \textit{mjuk som smör} ‘soft as butter’, but \textit{*smörmjuk} (lit. butter-soft)), but in other cases the phrasal and morphological form co-exist (e.g., Danish \textit{dum som snot} (lit. stupid as snot, ‘very stupid’), \textit{snotdum} (id.)). The principles that underlie the (non-)blocking in the various cases, both within single languages and cross-linguistically, are however not yet fully understood. While both phrasal \(A+N\) units and phrasal similes seem to arise quite naturally from the usual syntactic patterns of the various languages, it is very interesting to see that phrasal patterns that are more specific in that they violate the syntactic rules can also be found in various languages. A case in point is the \([an \text{ N}_1 \text{ of } an \text{ N}_2]\) pattern (e.g., \textit{a hell of a guy}), again a comparative pattern. It expresses a comparison of \(N_2\) to the reference value provided by \(N_1\). Hence, there is a mismatch between the semantic head of the construction (\(N_1\)) and the syntactic one (\(N_2\)), referred to as ‘dependency reversal’ in Rijkhoff (2009: 76). This pattern exists not only in Germanic, as for instance in English (\textit{a hell of a guy}), German (\textit{ein Idiot von (einem) Arzt} ‘an idiot of (a) doctor’), Danish (\textit{en klovn av en statsråd} ‘a clown of privy council’), Swedish (\textit{en kretin till

\(^{30}\) More detailed studies on phrasal similes can be found in this volume in the chapters on German, Dutch, and Italian.
polisprefekt ‘an imbecile of police chief’) and Dutch (where it is well-known in the linguistic literature in connection with the famous example schat van een kind (lit. sweetheart of a child, ‘very sweet child’), cf. Paardekoper 1956), but also in Italian, French, and Spanish (e.g., esta maravilla de niño ‘this wonder of a child’) (cf. Gun- kel et al. 2017: 1627ff.).

One has to add that while in the present context attention is given only to the formal side, i.e. the morphosyntactic and possibly phonological properties of patterns such as [(as) A as NP] and [an N\textsubscript{1} of an N\textsubscript{2}], cross-linguistic similarities (and differences) have also been studied with respect to the semantic side, in particular themes and images that feed imagery and metaphors in phrasal patterns and that re-occur cross-linguistically, due to cultural links and other factors (cf. Piirainen 2012, among many others). Thus, from this perspective, it is not unexpected that similar patterns occur cross-culturally in different, even genetically unrelated languages.

4 Summary

In this chapter, we sought to present an introductory overview of compound and MWE formation in a sample of European languages. We started with some general considerations about the notion of complex lexical unit, the lexicon, and the lexicon-syntax interface, and provided some preliminary criteria for the distinction between compounds and MWEs. In the second part of the chapter, we reviewed the language-specific properties of compounds and MWEs in West Germanic, North Germanic, Romance, Greek, Slavic, and Finno-Ugric languages, comparing them to German. Central questions that were discussed for each language family included the formal distinction between compounds and MWEs (in particular prosodic, morphological, and syntactic properties), the relationship between compounding and MWE formation as well as the conclusions concerning the theory of grammar and the lexicon following from these observations. One major finding is that while there are great similarities as well as differences regarding compound and MWE formation in the languages of Europe, a cross-linguistically valid definition of compounds and MWEs is hard to establish, because the languages differ greatly with respect to both the compound criteria that can be relevantly applied to them, and the relevant types of compound and MWE patterns and their degree of productivity. The various chapters of this volume provide in-depth analyses of the situation in the respective languages and language families, also discussing in more detail the relevant implications for the theory of the lexicon-grammar interface.
References

Anderson, Stephen R. (1982): Where's morphology? In: Linguistic Inquiry 13. 571–612.
Anward, Jan/Linell, Per (1976): Om lexikaliserade fraser i svenskan. In: Nysvenska studier 55/56. 76–119.
Arnau, Pierre J. L. (2015): Noun-noun compounds in French. In: Müller, Peter O. et al. (eds.). 673–687.
Bağrıaçık, Metin/Ralli, Angela (2015): Phrasal vs. morphological compounds: Insights from Modern Greek and Turkish. In: STUF – Language Typology and Universals 68, 3. 323–357.
Bandle, Oscar et al. (eds.) (2005): The Nordic Languages. An International Handbook of the History of the North Germanic Languages. Vol. 2. (= Handbooks of Linguistics and Communication Science (HSK) 22.2). Berlin/New York: De Gruyter.
Bauer, Laurie (1998): When is a sequence of two nouns a compound in English? In: English Language and Linguistics 2. 65–86.
Bauer, Laurie (2006): Morphological productivity. (= Cambridge Studies in Linguistics 95). Cambridge, UK: Cambridge University Press.
Bauer, Laurie (2009a): Typology of compounds. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 343–356.
Bauer, Laurie (2009b): IE, Germanic: Danish. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 400–416.
Bauer, Laurie (2017): Compounds and Compounding. (= Cambridge Studies in Linguistics 155). Cambridge, UK: Cambridge University Press.
Bauer, Laurie (to appear): Compounds. In: Aarts, Bas/Bowie, Jillian/Popova, Geri (eds.): Oxford Handbook of English Grammar. Oxford: Oxford University Press.
Benigni, Valentina/Masini, Francesca (2009): Compounds in Russian. In: Lingue e Linguaggio 8, 2. 171–193.
Bjarnadóttir, Kristín (2017): Phrasal compounds in Modern Icelandic with reference to Icelandic word formation in general. In: Trips, Carola/Kornfilt, Jaklin (eds.). 13–48.
Booij, Geert (2002): The morphology of Dutch. Oxford: Oxford University Press.
Booij, Geert (2009): Phrasal names. A constructionist analysis. In: Word Structure 2, 2. 219–240.
Booij, Geert (2010): Construction morphology. Oxford/New York: Oxford University Press.
Braunmüller, Kurt (2007): Die skandinavischen Sprachen im Überblick. 3rd ed. Tübingen/Basel: Francke.
Burger, Harald (1998): Phraseologie. Eine Einführung am Beispiel des Deutschen. Berlin: Schmidt.
Burger, Harald et al. (eds.) (2007): Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research. (= Handbooks of Linguistics and Communication Science (HSK) 28). Berlin/New York: De Gruyter.
Cetnarowska, Bożena (2015): The lexical/phrasal status of Polish Noun+Adjective or Noun+Noun combinations and the relevance of coordination as a diagnostic test. In: SKASE Journal of Theoretical Linguistics 12, 3. 142–170.
Cetnarowska, Bożena (2018): Phrasal names in Polish: A+N, N+A and N+N units. In: Booij, Geert (ed.): The Construction of Words. Advances in Construction Morphology. (= Studies in Morphology Series 4). Cham: Springer. 287–313.
Chomsky, Noam (1970): Remarks on Nominalization. In: Jacobs, Roderick A./Rosenbaum, Peter S. (eds.): Readings in English Transformational Grammar. Waltham, MA: Ginn & Co. 184–221.

De Caluwe, Johan (1990): Complementariteit tussen morfologische en in oorsprong syntactische benoemingsprocedéés. In: De Caluwe, Johan (ed.): Betekenis en productiviteit: Gentse bijdragen tot de studie van de Nederlandse woordvorming. Gent: Seminaria voor Duitse Taalkunde. 9–23.

Di Sciullo, Anna-Maria/Williams, Edwin (1987): On the definition of word. Cambridge, MA: MIT Press.

Dura, Elsbieta/Gawronska, Barbara (2007): Novelty extraction from special and parallel corpora. In: Vetulani, Zygmunt/Uszkoreit, Hans (eds.): Third Language and Technology Conference (LTC 2007), Poznan, Poland, October 5–7, 2007. (= Lecture Notes in Artificial Intelligence 5603). Berlin/Heidelberg: Springer. 291–302.

Engelberg, Stefan/Holler, Anke/Proost, Kristel (2011): Zwischenräume. Phänomene, Methoden und Modellierung im Bereich zwischen Lexikon und Grammatik. In: Engelberg, Stefan/Holler, Anke/Proost, Kristel (eds.): Sprachliches Wissen zwischen Lexikon und Grammatik. Berlin/Boston: De Gruyter. 1–35.

Farø, Ken (2015): Feste Wortgruppen/Phraseologie II: Phraseme. In: Haß, Ulrike/Storjohann, Petra (eds.): Handbuch Wort und Wortschatz. Berlin/Boston: De Gruyter. 226–247.

Fillmore, Charles J./Kay, Paul/O’Connor, Mary C. (1988): Regularity and idiomaticity in grammatical constructions: The case of let alone. In: Language 64. 501–538.

Finkbeiner, Rita (2008): Zur Produktivität idiomatischer Konstruktionsmuster. Interpretierbarkeit und Produzierbarkeit idiomatischer Sätze im Test. In: Linguistische Berichte 216. 391–430.

Fleischer, Wolfgang (1982): Phraseologie der deutschen Gegenwartssprache. Leipzig: VEB Bibliographisches Institut.

Fradin, Bernhard (2009): IE, Romance: French. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 417–435.

Gaeta, Livio/Grossmann, Maria (eds.) (2009): Compounds between syntax and lexicon. Special Issue of Italian Journal of Linguistics/Rivista di Linguistica 21, 1.

Gaeta, Livio/Ricca, Davide (2009): Composita solvantur: Compounds as lexical units or morphological objects? In: Gaeta, Livio/Grossmann, Maria (eds.). 35–70.

Gaeta, Livio/Schlücker, Barbara (eds.) (2012): Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte. (= Linguistik – Impulse & Tendenzen 46). Berlin/New York: De Gruyter.

Gavriilidou, Zoe (2013): NN Combinations in Greek. In: Journal of Greek Linguistics 13, 1. 5–29.

Giegerich, Heinz J. (2004): Compound or phrase? English noun-plus-noun constructions and the stress criterion. In: English Language and Linguistics 8, 1. 1–24.

Giegerich, Heinz J. (2009): Compounding and Lexicalism. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 178–200.

Giegerich, Heinz J. (2015): Lexical structures: compounding and the modules of grammar. (= Edinburgh Studies in Theoretical Linguistics 1). Edinburgh: Edinburgh University Press.

Goldberg, Adele E. (1995): Constructions. A construction grammar approach to argument structure. Oxford: Oxford University Press.

Goldberg, Adele E. (2006): Constructions at Work. The nature of generalization in language. Oxford: Oxford University Press.
Gries, Stefan Th. (2008): Phraseology and linguistic theory: a brief survey. In: Granger, Sylviane/Meunier, Fanny (eds.): Phraseology. An interdisciplinary perspective. Amsterdam: Benjamins. 3–25.

Guevara, Emiliano/Scalise, Sergio (2009): Searching for Universals in Compounding. In: Scalise, Sergio/Bisetto, Antonietta (eds.): Universals of Language Today. Dordrecht: Springer. 101–128.

Guevara, Emiliano R. (2012): Spanish compounds. In: Probus 24, 1. 175–195.

Gunkel, Lutz/Zifonun, Gisela (2011): Klassifikatorische Modifikation im Deutschen und Französischen. In: Lavric, Eva/Pöckl, Wolfgang/Schallhart, Florian (eds.): Comparatio delectat: Akten der VI. Internationalen Arbeitstagung zum romanisch-deutschen und innerromanischen Sprachvergleich, Innsbruck, 3.–5. September 2008. Frankfurt a.M.: Lang. 549–562.

Gunkel, Lutz et al. (2017): Grammatik des Deutschen im europäischen Vergleich: das Nominal. Vol. 2: Nominalflexion, Nominale Syntagmen. (= Schriften des Instituts für Deutsche Sprache 14). Berlin/Boston: De Gruyter.

Haberland, Hartmut (1994): Danish. In: König, Ekkehard/Auweera, Johan van der (eds.). 313–348.

Haeringen, Coenraad Bernardus van (1956): Nederlands tussen Duits en Engels. 2de druk. Den Haag: Servire.

Härtl, Holden (2016): Normality at the boundary between word-formation and syntax. In: d’Avis, Franz/Lohnstein, Horst (eds.): Normalität in der Sprache. In: Linguistische Berichte Sonderheft 22. Hamburg: Buske. 65–92.

Häusermann, Jürg (1977): Phraseologie. Hauptprobleme der deutschen Phraseologie auf Basis sowjetischer Forschungsergebnisse. Tübingen: Niemeyer.

Hoffmann, Thomas/Trousdale, Graeme (eds.) (2013): The Oxford Handbook of Construction Grammar. Oxford: Oxford University Press.

Hohenhaus, Peter (2005): Lexicalization and institutionalization. In: Štekauer, Pavol/Lieber, Rochelle (eds.). 353–373.

Hüning, Matthias (2010): Adjective + Noun constructions between syntax and word formation in Dutch and German. In: Onysko, Alexander/Michel, Sascha (eds.): Cognitive Perspectives on Word Formation. (= Trends in Linguistics. Studies and Monographs 221). Berlin/New York: De Gruyter. 195–215.

Hüning, Matthias/Schlücker, Barbara (2010): Konvergenz und Divergenz in der Wortbildung. Komposition im Niederländischen und im Deutschen. In: Dammel, Antje/Kürschnler, Sebastian/Nübling, Damaris (eds.): Kontrastive Germanistische Linguistik (= Germanistische Linguistik 206–209). Vol. 2. Hildesheim i.a.: Olms. 783–825.

Hüning, Matthias/Schlücker, Barbara (2015): Multi-word expressions. In: Müller, Peter O. et al. (eds.). 450–467.

Hyvärinen, Irma (2007): Phraseologie des Finnischen. In: Burger, Harald et al. (eds.). Vol. 2. 737–752.

Inghult, Göran (1991): Lexikalische Innovationen in Wortgruppenform. Zu einer Untersuchung über die Erweiterung des Lexembestandes im Deutschen und Schwedischen. In: Palm, Christine (ed.): Europhrasas 90. Akten der internationalen Tagung zur germanistischen Phraseologieforschung Aske/Schweden, 12.–15. Juni 1990. Uppsala: Almqvist & Wiksells. 101–114.

Jackendoff, Ray (1997): The architecture of the language faculty. Cambridge, MA: MIT Press.

Jackendoff, Ray (2002): Foundations of language. Oxford: Oxford University Press.
Jackendoff, Ray (2008): Construction after construction and its theoretical challenges. In: Language 84. 8–28.
Jackendoff, Ray (2009): Compounding in the Parallel Architecture and Conceptual Semantics. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 105–128.
Jackendoff, Ray (2013): Constructions in the Parallel Architecture. In: Hoffmann, Thomas/Trousdale, Graeme (eds.). 70–92.
Josefsson, Gunlog (1997): On the principles of word formation in Swedish. Lundastudier i nordisk språkvetenskap 51. Lund: Lund University Press.
Kapatsinski, Vsevolod/Vakareliyska, Cynthia (2013): [N[N]] compounds in Russian: A growing family of constructions. In: Constructions and Frames 5, 1. 69–87.
Karlsson, Fred (2015): Finnish: an essential grammar. 3rd edition. Milton Park i.a.: Routledge.
[Trans. by Andrew Chesterman].
Kiefer, Ferenc (1990): Noun incorporation in Hungarian. In: Acta Linguistica Hungarica 40, 1–2. 149–177.
Kiefer, Ferenc (1992): Compounding in Hungarian. Rivista di Linguistica 4. 61–78.
Kiefer, Ferenc (2009): Uralic, Finno-Ugric: Hungarian. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 527–541.
Kiefer, Ferenc (2016): Hungarian. In: Müller, Peter O. et al. (eds.). 3308–3326.
Klinge, Alex (2006): The Origin of Weapons of Mass Destruction. Investigating Traces of Lexical Formation Patterns in the (Linguistic) History of Europe. In: Nölke, Henning (ed.): Grammatica: Festschrift in honour of Michael Herslund/Hommage à Michael Herslund. Frankfurt a.M./New York: Lang. 233–248.
Kolehmainen, Leena/Savolainen, Tiina (2007): Deverbale Verbbildung im Deutschen und im Finnischen: ein Überblick. Würzburg. (Internet: https://opus.bibliothek.uni-wuerzburg.de/opus4-wuerzburg/frontdoor/index/index/docId/1937, last access: 30.5.2018).
König, Ekkehard/Auwera, Johan van der (eds.) (1994): The Germanic Languages. London/New York: Routledge.
Koptjevskaja-Tamm, Maria (2009): Proper-name nominal compounds in Swedish between syntax and lexicon. In: Rivista di Linguistica 21, 1. 119–148.
Kunter, Gero (2011): Compound stress in English: the phonetics and phonology of prosodic prominence. (= Linguistische Arbeiten 539). Berlin: De Gruyter.
Lieber, Rochelle/Štekauer, Pavol (eds.) (2009a): The Oxford handbook of compounding. Oxford: Oxford University Press.
Lieber, Rochelle/Štekauer, Pavol (2009b): Introduction: Status and definition of compounding. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 3–18.
Liimatainen, Annikki (2008): Untersuchungen zur Fachsprache der Ökologie und des Umweltschutzes im Deutschen und Finnischen: Bezeichnungsvarianten unter einem geschichtlichen, lexikografischen, morphologischen und linguistisch-pragmatischen Aspekt. (= Finnische Beiträge zur Germanistik 22). Frankfurt a.M.: Lang.
Lipka, Leonhard (1977): Lexikalisisierung, Idiomatisierung und Hypostasierung als Probleme einer synchronischen Wortbildungsllehre. In: Brekle, Herbert E./Kastovsky, Dieter (eds.): Perspektiven der Wortbildungsforschung. Beiträge zum Wuppertaler Wortbildungskolloquium vom 9.–10. Juli 1976. Bonn: Bouvier. 155–164.
Lüdeling, Anke (2001): On particle verbs and similar constructions in German. Stanford: CSLI Publications.
Martinčová, Řího (2015): Multi-word expressions and univerbation in Slavic. In: Müller, Peter O. et al. (eds.). 742–757.
Masini, Francesca (2009): Phrasal lexemes, compounds and phrases: A constructionist perspective. Word Structure 2, 2. 254–271.
Masini, Francesca/Benigni, Valentina (2012): Phrasal lexemes and shortening strategies in Russian: the case for constructions. In: Morphology 22, 3. 417–451.
Masini, Francesca/Scalise, Sergio (2012): Italian compounds. In: Probus 24, 1. 61–91.
Melloni, Chiara/Bisetto, Antonietta (2010): Parasynthetic compounds. Data and theory. In: Scalise, Sergio/Vogel, Irene (eds.): Cross-Disciplinary Issues in Compounding. Amsterdam/Philadelphia: Benjamins. 199–217.
Moon, Rosamund (2015): Multi-word items. In: Taylor, John R. (ed.): The Oxford Handbook of The Word. 1st ed. Oxford: Oxford University Press. 120–140.
Müller, Peter O. et al. (eds.) (2015–2016): Word-formation. An international handbook of the languages of Europe. (= Handbooks of Linguistics and Communication Science (HSK) 40). Berlin/Boston: De Gruyter.

Nagórkó, Alicja (2016): Polish. In: Müller, Peter O. et al. (eds.). 2831–2852.
Newmeyer, Frederick J. (1974): The regularity of idiom behavior. In: Lingua 34. 327–42.
Niemi, Jussi (2009): Compounds in Finnish. In: Lingue e Linguaggio 8, 2. 237–256.
Niemi, Sinikka (2009): Compounds in Swedish. In: Lingue e Linguaggio 8, 2. 257–269.
Nunberg, Geoffrey/Sag, Ivan/Wasow, Thomas (1994): Idioms. In: Language 70, 3. 491–538.
Ohnheiser, Ingeborg (2015): Compounds and multi-word expressions in Slavic. In: Müller, Peter O. et al. (eds.). 757–779.
Olsen, Susan (2000): Composition. In: Booij, Geert/Lehmann, Christian/Mugdan, Joachim (eds.): Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung/Morphology. An international handbook on inflection and word-formation. Berlin: De Gruyter. 897–916.
Paardekoper, Petrus C. (1956): Een schat van een kind. In: De Nieuwe Taalgids 49. 93–99.
Pafel, Jürgen (2015): Phrasal compounds are compatible with lexical integrity. In: Language typology and universals (STUF) 68, 3. 263–280.
Pawley, Andrew/Syder, Frances H. (1983): Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In: Richards, Jack C./Schmidt, Richard W. (eds.): Language and Communication. London/New York: Routledge. 191–226.
Piirainen, Elissabeth (2012): Widespread idioms in Europe and beyond: toward a lexicon of common figurative units. (= International Folkloristics 5). New York i.a.: Lang.
Pitkänen-Heikkilä, Kaarina (2016): Finnish. In: Müller, Peter O. et al. (eds.). 3209–3228.
Plag, Ingo (2006): The variability of compound stress in English: structural, semantic, and analogical factors. In: English Language and Linguistics 10, 1. 143–172.
Rainer, Franz (2016): Spanish. In: Müller, Peter O. et al. (eds.). 2620–2640.
Rainer, Franz/Varela, Soledad (1992): Compounding in Spanish. In: Rivista di Linguistica 4, 1. 117–142.
Ralli, Angela (1992): Compounds in Modern Greek. In: Rivista di Linguistica 4. 143–174.
Ralli, Angela (2008): Compound markers and parametric variation. In: STUF – Language Typology and Universals 61, 1. 19–38.
Ralli, Angela (2009): IE, Hellenic: Modern Greek. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 453–463.
Ralli, Angela (2013a): Compounding in modern Greek. (= Studies in Morphology 2). Dordrecht: Springer.
Ralli, Angela (2013b): Compounding and its locus of realization: Evidence from Greek and Turkish. In: Word Structure 6, 2. 181–200.
Ralli, Angela (2016): Greek. In: Müller, Peter O. et al. (eds.). 3138–3156.
Ralli, Angela/Stavrou, Melita (1998): Morphology-syntax interface: A-N compounds vs. A-N constructs in Modern Greek. In: Booij, Geert E./Marle, Jaap Van (eds.): Yearbook of Morphology 1997. Dordrecht: Springer. 243–264.
Ramisch, Carlos (2015): Multiword expressions acquisition. A generic and open framework. New York i. a.: Springer.
Ricca, Davide (2015): Verb-noun compounds in Romance. In: Müller, Peter O. et al. (eds.). 688–707.
Rijkhoff, Jan (2009): On the co-variation between form and function of adnominal possessive modifiers in Dutch and English. In: McGregor, William (ed.): The expression of possession. Berlin/New York: De Gruyter. 51–106.
Rosenbach, Anette (2006): Descriptive genitives in English. A case study on constructional gradience. In: English Language and Linguistics 10. 77–118.
Saxton, Matthew (2010): Child Language. Acquisition and Development. London: SAGE.
Scalise, Sergio (1992): Compounding in Italian. In: Rivista di Linguistica 4. 175–199.
Scalise, Sergio (ed.) (1992): The Morphology of Compounding. Special issue of Rivista di Linguistica 4, 1.
Scalise, Sergio/Guevara, Emiliano (2005): The lexicalist approach to word formation and the notion of the lexicon. In: Štekauer, Pavol/Lieber, Rochelle (eds.). 147–187.
Scalise, Sergio/Vogel, Irene (eds.) (2010): Cross-disciplinary issues in compounding. Amsterdam/Philadelphia: Benjamins.
Schlücker, Barbara (2014): Grammatik im Lexikon. Adjektiv+Nomen-Verbindungen im Deutschen und Niederländischen. (= Linguistische Arbeiten 553). Berlin/Boston: De Gruyter.
Schlücker, Barbara/Plag, Ingo (2011): Compound or Phrase? Analogy in Naming. In: Lingu 121. 1539–1551.
Sinclair, John (1991): Corpus, Concordance, Collocation. Oxford: Oxford University Press.
Štekauer, Pavol/Lieber, Rochelle (eds.) (2005): Handbook of Word-Formation. Dordrecht: Springer.
Szymanek, Bogdan (2009): IE, Slavonic: Polish. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 464–477.
Szymanek, Bogdan (2017): Compounding in Polish and the absence of phrasal compounding. In: Trips, Carola/Kornfilt, Jaklin (eds.). 49–79.
Telemann, Ulf (2005): The Standard languages and their systems in the 20th century: Swedish. In: Bandle, Oscar et al. (eds.). 1603–1626.
Thráinsson, Höskuldur (1994): Icelandic. In: König, Ekkehard/Auwera, Johan van der (eds.). 142–190.
Torp, Arne (2002): The Nordic Languages in a Germanic Perspective. In: Bandle, Oscar et al. (eds.). 13–24.
Trips, Carola/Kornfilt, Jaklin (2015): Typological aspects of phrasal compounds in German, English, Turkish and Turkic. In: Language typology and universals (STUF) 68, 3. 281–321.
Trips, Carola/Kornfilt, Jaklin (eds.) (2017): Further investigations into the nature of phrasal compounding. (= Morphological Investigations 1). Berlin: Language Science Press.
Uluhanov, Igor’ S. (2016): Russian. In: Müller, Peter O. et al. (eds.). 2953–2978.
Van Goethem, Kristel (2009): Choosing between A+N compounds and lexicalized A+N phrases: The position of French in comparison to Germanic languages. In: Word Structure 2, 2. 241–253.
Villoing, Florence (2012): French compounds. In: Probus 24, 1. 29–60.
Weinreich, Uriel (1969): Problems in the analysis of idioms. In: Puhvel, Jaan (ed.): Substance and Structure of Language. Berkeley: University of California Press. 23–76.
Wray, Alice (2002): Formulaic Language and the Lexicon. Cambridge, UK: Cambridge University Press.
Wunderlich, Dieter (2006): Introduction: What the theory of the lexicon is about. In: Wunderlich, Dieter (ed.): Advances in the Theory of the Lexicon. Berlin/New York: De Gruyter. 1–26.
Zeller, Jochen (2001): Particle verbs and local domains. Amsterdam: Benjamins.
