The lexicalization of complex constructions: an analysis of adjective-noun combinations

Marcel Schlechtweg
*Universität Kassel*
marcelschlechtweg@gmail.com

Holden Härtl
*Universität Kassel*
holden.haertl@uni-kassel.de

1. Introduction

Languages can differ with respect to their use of compounds and phrases as naming units that become lexicalized. A comparison of Germanic and Romance languages exemplifies cross-linguistic variation in the choice of the preferred construction type in that the former seem to favor compounds where the latter often use phrases. In the current paper, we aim at discussing the lexicalization of compounds and phrases both from a language-specific and from a cross-linguistic angle. Specifically, we pursue the idea that compounds are naturally more appropriate to become lexicalized than phrases. For this purpose, we will reflect upon some fundamental characteristics of compounds and present psycholinguistic evidence that supports the conception of the lexicalization affinity of compounds.

The paper is structured in the following way. In Section 2, we will discuss the proposal that compounds are well suited to enter the process of lexicalization. In Section 3, we will introduce the psychological process of memorization and show how it is related to the process of lexicalization. In Section 4, we will present empirical evidence suggesting a fundamental difference between the memorization of compounds and phrases. Finally, in Section 5, we will summarize our discussion and conclude the paper.

2. The lexicalization affinity of compounds

While compounds have often been regarded as constructions that typically represent names of specific phenomena, phrases have been argued to fulfill a descriptive function in most cases (Bauer 1988: 102; Hüning 2010: 197). This difference can be easily recognized in German: Whereas the compound *Grünreiher* ‘green heron’ names the particular kind of heron also called *butorides virescens*, the phrase *grüner Reiher* ‘green heron’ can refer to any heron that is green for whatever reason. Nevertheless, we also find examples that show that the functional distinction between compounds and phrases is not a definite rule. The German phrase *gelbes Trikot* ‘yellow jersey’, taken from Schlücker (2014: 148), cannot only function as a descriptive unit but can also name a specific shirt worn by the leader of some sports competitions, e.g. the *Tour de France*.

Although counterexamples exist, the aforementioned functional separation between compounds as typical naming units and phrases, which are usually descriptions, holds in many cases. This has led researchers to claim that compounds are better equipped to fulfill the naming function than phrases. Bücking (2009, 2010) contrasts adjective-noun (AN) compounds to AN phrases in German. Crucially, he concentrates on novel constructions in order to exclude the influence of lexicalization. The author emphasizes that the variable $R_{\text{INTEGRAL}}$, which determines how the adjective and the noun are related, plays a crucial role in compounds. The phenomenon is connected to a fundamental semantic difference between German AN compounds and AN phrases: compounds, as opposed to phrases, are known for their semantically non-compositional character. Therefore, the meaning of a compound goes

---

beyond the simple sum of the meaning of the adjective and the meaning of the noun and necessitates an additional element, namely \textsc{RIntegral}, that establishes the precise relationship between a compound’s constituents. Since names in general, similar to compounds in particular, tend to lack compositional semantics, \textsc{RIntegral} makes compounds more appropriate to fulfill a naming purpose (Herbermann 1981: 334-335; Bücking 2009, 2010).

Härtl (2015) goes in a similar direction and emphasizes the consequences of the structural difference between German AN compounds and AN phrases. Since a compound structurally deviates from a phrase, which represents the default AN combination, it is also likely to differ from the phrase in semantic terms. The creation of a new compound such as \textit{Rotdach ‘red roof’} implies that its meaning does not equal the meaning of the phrase \textit{rotes Dach ‘red roof’}; instead, the compound is semantically non-compositional and ready to name a particular complex concept right from the beginning of its existence. The author connects this thought to the idea that compounds are more suited for lexicalization. The special structural and semantic status of compounds, in comparison to phrases, calls for the lexicalization of the complex construction as a name of a particular concept.

The aforementioned contributions by Bücking and Härtl suggest an interesting link between the structure, semantics, function and lexicalization of compounds and phrases. It is of utmost significance that the authors focus on the contrast between AN compounds and AN phrases in German. In this language, the two construction types can be clearly kept apart by means of inflectional agreement. Whereas the two constituents of an AN phrase such as \textit{grüner Reiher ‘green heron’} are in agreement with respect to gender, number, case and definiteness, the mere adjectival root of a German AN compound such as \textit{Grünreiher ‘green_heron’} does not agree with the noun it precedes (cf. also Booij 2012: 84). Inflectional agreement between the adjective and the noun, in the case of phrases, or the lack of it, in the case of compounds, are reflected on a structural basis through the presence or absence of a suffix. The structural divergence between compounds and phrases seems to be prone to mirror a semantic difference between the two construction types as well. As a consequence of their structural deviation from a normal phrase, compounds tend to carry a meaning that also differs from the default interpretation expressed by a syntactic phrase. The question now arises of how we can find further evidence for the idea that the peculiarities of compounds make them more suitable to become lexicalized than phrases.

3. Lexicalization and memorization

Lexicalization represents a diverse phenomenon, which has been defined in several ways in the literature (Lipka 2002: 111; Bakken 2006: 106-108). For the purpose of the current paper, we consider a construction to be lexicalized if it serves as the linguistic form, or the name, of a specific concept (Lipka 1981: 131; Blank 2001: 1596; Gaeta and Ricca 2009: 38). Moreover, the relation between lexicalization and another crucial term, namely the psychological notion of memorization, turns out to be significant in the context of the present contribution. Several authors have discussed the demarcation between the two terms. Wunderlich (1986: 231) assumes that memorization represents a kind of a pre-stage of lexicalization. Approaches taken by some other authors go well with this idea: While memorization is often considered to be a mental operation attributed to individual language users, lexicalization represents a process that takes place within society if an item enters the shared vocabulary of a language (Pawley and Syder 1983: 208-209; Schwarze and Wunderlich 1985: 16; Lüdi 1986: 226). Specifically, Lüdi (1986: 226) regards lexicalization as “collective memorization”. Relying on the proposal that memorization by language users can lead to the lexicalization of a specific expression in a language, we can use the notion of
memorization as a point of departure in order to shed more light on the lexicalization of complex constructions such as compounds and phrases.

4. The memorization of compounds

4.1 General remarks and existing evidence

A recent study by Kotowski, Böer and Härtl (2014: 195-196) compared German AN compounds and AN phrases in terms of memorization. The authors carried out a memorization study by means of an experimental design that consisted of two phases on each of three test days. In the first phase, the task of all subjects was to memorize novel AN compounds as well as AN phrases that were combined with an image showing the entity of interest. In the second phase, a lexical-decision task was conducted where participants had to decide whether a compound or a phrase appeared with the same picture as in the memorization phase. Kotowski et al.’s (2014: 195-196) analysis revealed that subjects gave slower and less accurate responses when reacting to compounds that had not occurred in the memorization phase in comparison to the responses to phrases that had not been part of the memorization phase. However, the reactions to memorized compounds and memorized phrases did not significantly differ. Therefore, the authors argue that memorization can overcome initial difficulties in the processing of compounds. Being exposed to a compound for the first time seems to come with processing difficulties as these constructions are, in contrast to phrases, marked. The process of memorization, however, pushes compounds so that the processing of memorized compounds equals the processing of memorized phrases.

The study presented in the previous paragraph shows that compounds benefit more from memorization than phrases. Although subjects reacted faster and more accurately to non-memorized phrases than to non-memorized compounds, no difference between the responses to memorized compounds and memorized phrases was detected. In other words, compounds were retained better than phrases in the course of memorization over three days. Assuming the above-mentioned relationship between memorization and lexicalization, we can say in a more general sense that compounds and phrases differ in the way they are lexicalized because differences in the process of memorization have an impact on the process of lexicalization. Since memorization represents a crucial step towards lexicalization, we aim at shedding more light on the process of memorization in the context of the demarcation of compounds and phrases in the discussion below. We hypothesize that compounds benefit more from the process of memorization than phrases. At this point, we must specify the exact nature of the expected advantage. Put differently, we have to define a potential memorization advantage and state how it becomes evident. Let us consider two definitions of the term “memorization advantage” for the present contribution (cf. also Schlechtweg and Härtl 2015). First of all, we can speak of a memorization advantage if subjects react faster/more accurately to compounds than to phrases overall, i.e. on all test days taken together. The definition is strictly unidirectional because faster/more accurate responses of phrases cannot be regarded as a memorization advantage. A potential advantage of phrases would be rooted in the fact that phrases represent the normal or standard construction type (ten Hacken 2013: 97). The proposal is connected to the phenomenon of markedness. As stated in Kotowski et al. (2014), compounds are more marked than phrases. Since unmarked items are usually more frequent than marked items and since higher frequency is known to cause faster/more accurate responses in lexical-decision tasks (e.g. Forster and Davis 1984; Bybee 1995: 237 referring to Greenberg 1966), an advantage of phrases on all test days together would not be a memorization advantage but rather an advantage triggered by the higher frequency of the phrasal pattern in general. As a consequence, it is important to take a different definition of
the notion of memorization advantage into account if the responses to phrases are faster/more accurate on all test days together or if the reactions to phrases and those to compounds are similar on all test days together: Although compounds are responded to more slowly/less accurately than phrases on the initial day, they show a memorization advantage in comparison to phrases if responses to compounds do not differ from the responses to phrases on the consecutive day(s). This proposal originates in the fact that the phrasal pattern is more common and frequent (ten Hacken 2013: 97) and, therefore, more likely to cause faster latencies and more accurate responses at an early stage of the memorization process, i.e. on the first test day. If the advantage of phrases disappears through memorization, we can speak of a memorization advantage of compounds because they improved more than phrases in the course of the study.

4.2 Cross-linguistic evidence

Having defined the notion of memorization advantage, we can now turn to the question of how to find further evidence for the idea that compounds show a memorization advantage in comparison to phrases. The study by Kotowski et al. (2014: 195-196) outlined above focused on the analysis of German AN compounds and phrases. In the current paper, we will go a step further and present empirical evidence from three languages, namely German, French and English. Similar to Kotowski et al. (2014: 195-196), we will concentrate on combinations of an adjective and a noun. These constructions represent an interesting group to work on because many potentially confounding variables can be controlled for across the languages under investigation. Remember that we defined an AN phrase as a construction where the adjective agrees with the following or preceding noun and an AN compound as a construction without internal agreement. Applying this definition to the three aforementioned languages and asking what kind of construction (AN compound or AN phrase) each of the languages prefers when naming a new complex lexical concept, we get a clear picture for German and French but not for English. While German tends to use AN compounds as naming units, French favors AN phrases (Van Goethem 2009). Relying on inflection as the decisive factor to differentiate between compounds and phrases, we cannot define AN compounds and AN phrases in English. The only thing we can do is to use our intuitions in the case of English constructions. Specifically, we can assume that English AN constructions with initial stress are compound-like constructions and English AN constructions with non-initial stress are phrase-like constructions (cf. also Schlechtweg and Härtl 2015). The idea is based on Chomsky and Halle’s (1968: 17) distinction between compound stress, i.e. initial stress, and nuclear or phrasal stress, i.e. non-initial stress. Despite the criticism of this proposal (Bell and Plag 2012: 487), it represents a good intuition. Although initial stress is not the defining criterion of AN compounds, it usually occurs in AN compounds. At this point, it is helpful to refer to German where the factor of inflection unambiguously identifies AN compounds and phrases. In this language, AN compounds are typically stressed on the adjective (Erben 2000: 43). Since both German and English are languages of Germanic origin, it is plausible to regard initial stress as a common marker of AN compounds (Pereltsvaig 2012: 10) and, thus, to call English AN constructions with initial stress compound-like constructions.

So far, we have assumed that German prefers compounds and French favors phrases to express a complex lexical concept through an adjective-noun combination. Since we have evidence from a single language suggesting a different pattern of memorization for compounds and phrases, we might now ask whether the contrast is also reflected cross-linguistically. Specifically, we can raise the question of whether German AN compounds deviate from French AN/NA phrases in terms of memorization. If languages differ in their preferences for either compounds or phrases and if compounds and phrases differ with regard
to memorization, we can expect that complex constructions from one language, namely compounds, that are composed of two specific constituents are memorized differently than complex constructions from another language, namely phrases, that contain the same constituents. In order to test this hypothesis, a study was conducted that contrasted the memorization of German AN compounds and French AN/NA phrases. In addition to items from these two languages, English complex AN constructions were examined in the same study with respect to how they were memorized. The experiments included not only English AN constructions with initial stress but also AN constructions with non-initial stress.

In the following, we will summarize the experimental study. Note that a very detailed presentation of the study is given in Schlechtweg and Härtl (2015). In this earlier contribution, we describe the participants, the material, the procedure as well as the central hypotheses in detail, present a comprehensive result section and all items under investigation.

Contrasting the memorization of German AN compounds, French AN/NA phrases, English AN constructions with initial stress and English AN constructions with non-initial stress, we aimed at investigating whether a memorization advantage of compounds/compound-like constructions can be observed from a cross-linguistic perspective. Note that we regarded the English constructions with initial stress as compound-like constructions and the constructions with non-initial stress as phrase-like constructions. Speakers of the three aforementioned languages participated in the study and were tested on complex constructions of their native language. For this purpose, they were divided into the four groups German, French, EnglishA and EnglishB. Note that two English groups were created in order to examine AN constructions stressed on the adjective in one group (EnglishA) and AN constructions stressed on the noun in the other group (EnglishB). In our study, we included different types of items. While novel AN/NA constructions such as Jungtourist/jeune touriste/YOUNG tourist/young TOURist represented the experimental items, existing nouns of the languages under investigation, e.g. Architekt/architecte/architect/architect, were used as control items (baseline). Subjects were asked to memorize both the experimental and the control items on three test days, i.e. on day one, day four and day eight. Apart from these items that had to be memorized on three days, we included filler items that did not have to be memorized. Filler items were either other AN/NA constructions or other existing nouns. When creating our set of experimental, control and filler items, we took several potentially confounding variables into account and controlled for them across the three languages examined in the study (e.g. number of syllables, duration in seconds, frequency of the constituents, control and filler nouns, lexicalization status of the complex AN/NA constructions). The experiment was conducted by using the computer program E-Prime (Psychology Software Tools, Inc. 2010). All subjects participated in the experiment on three days and in two phases on each day. In the first phase, the memorization phase, participants heard and memorized the experimental and the control items. Right after that, i.e. in the second or recall phase, subjects heard both the memorized items from the first phase as well as other, non-memorized items, i.e. the filler items. Having heard a memorized item, a participant was expected to press a “Yes”-button in the recall phase. If a subject heard a non-memorized item, however, he or she was supposed to press a “No”-button.

While analyzing the study, we examined both how fast and how accurate subjects responded to the experimental and control items they heard in the recall phase. Note that we will focus on reaction time in the following. Contrasting compounds/compound-like constructions to phrases/phrase-like constructions, we expected differences in the response latencies, i.e. we hypothesized a memorization advantage of the compounds/compound-like constructions. Crucially, however, there should be no difference between the response times of the German, French and English control items. That means, the control items served as a baseline in order to verify that it is possible to compare these languages in a psycholinguistic
The lexicalization of complex constructions: an analysis of adjective-noun combinations

study at all. Looking at the response latencies, our statistical analyses confirmed our hypothesis concerning the control items only for the comparisons German-French as well as EnglishA-EnglishB. In other words, the reaction times of the German control items did not significantly differ from the reaction times of the French control items and the response latencies of the control items of the group EnglishA did not significantly differ from the response latencies of the control items of the group EnglishB. Therefore, when examining the complex AN/NA constructions, we compared the German compounds only to the French phrases. Moreover, we contrasted the English AN constructions with initial stress only to the English AN constructions with non-initial stress. Our analyses revealed that subjects responded significantly more slowly to the French phrases than to the German compounds (Difference of means (henceforth: DM) = 65.0, t = 3.12, p = .046). We consider this effect to mirror a memorization advantage of compounds in comparison to phrases. Furthermore, the reaction times of the English AN constructions with non-initial stress were shorter than the response latencies of the English AN constructions with initial stress (DM = -65.2, t = -3.13, p = .045). The results reported so far are presented in Figure 1.

![Figure 1: Interaction of LANGUAGE x ITEM TYPE (F2)](image)

At this point, however, remember that the effect in English does not reflect a memorization advantage but an advantage that originates in the fact that non-initial stress represents the default prosodic structure in English AN constructions (cf. also Fudge 1984: 146; Zwicky 1986: 51; Bauer, Lieber and Plag 2013: 448). Schiller, Levelt and Fikkert’s (2004: 237-238) results suggest that responses are faster to the standard stress pattern than to a marked prosodic structure. Therefore, we have to apply the second definition of the term memorization advantage that we introduced above, i.e. the comparison of the response times between the AN constructions with initial stress and those with non-initial stress on the three individual test days. It turned out, however, that the latencies of the two groups did not significantly differ from each other on any of the individual days. In sum, we found a

---

2 All results of this study that are reported in the current paper refer to the item analysis (F2).
memorization advantage of the German compounds compared to the French phrases but we did not detect a memorization advantage of the English compound-like constructions in comparison to the phrase-like constructions of the same language.

When looking at the set of English AN constructions under investigation in the aforementioned study and reflecting upon the nature of compounds and phrases, we realized that our compound-like constructions were characterized by only one typical feature of compounds, namely initial stress. Therefore, we decided to conduct a second study that investigated not only the factor of stress but also the factor of semantic compositionality. Since compounds are also known to be semantically non-compositional (Downing 1977: 820), we aimed at combining two factors and examining their influence on memorization. Specifically, we redefined our compound-like as well as our phrase-like constructions and investigated whether a memorization advantage can be observed for English compound-like constructions, i.e. AN constructions that carry initial stress and are not semantically compositional (e.g. HARD shirt), in comparison to English phrase-like constructions, i.e. AN constructions that bear non-initial stress and are semantically compositional (e.g. short BRUSH). Again, as in the case of the first study, the reader is advised to consult Schlechtweg and Härtl (2015) for a very detailed description of this study. In the present work, we can only give a short summary.

The procedure used in this study resembled the procedure of the first investigation described above. However, the examination consisted of two memorization phases and one recall phase on each of the three test days. Our experimental items had to be memorized on three days and could be divided into four groups: Items like short BRUSH were stressed on the second constituent and semantically compositional, items like SHORT brush were stressed on the first constituent and semantically compositional, items like HARD shirt were stressed on the second constituent and semantically non-compositional and items like HARD shirt were stressed on the first constituent and semantically non-compositional. Our filler items were other AN constructions that had not to be memorized and had the function to trigger a “No”-response in the recall phase. When creating the complex AN constructions, we took several potentially confounding variables into consideration (e.g. number of syllables, duration in seconds, frequency of the constituents, lexicalization status of the AN constructions). The decision of whether an item was considered to be semantically compositional or not was based on the results of an online survey (SoSci, Leiner 2014).

In this study, we expected a memorization advantage – in the second sense of the term – of the compound-like constructions, i.e. of the semantically non-compositional constructions with initial stress, compared to the phrase-like constructions, i.e. the semantically compositional constructions with non-initial stress. Put differently, we hypothesized that the reactions to compound-like constructions are slower than the response latencies of phrase-like constructions on the first but not on the following days. Our statistical analyses revealed that the responses to phrase-like constructions were highly significantly faster than the responses to compound-like constructions when considering all three days together (DM$_1$ = -68.8, $t_1$ = -5.52, $p_1$ = .000; DM$_2$ = -61.3, $t_2$ = -4.40, $p_2$ = .000). Since phrase-like constructions represent the more common pattern (Liberman and Sproat 1992: 134; Giegerich 2009: 5-7), this result was expected and simply mirrored an advantage caused by the frequency of the constructions. Looking at the individual days, we found that the phrase-like constructions were responded to (highly) significantly faster than the compound-like constructions on the first but not on the second and third day (Day 1: DM$_1$ = -107.1, $t_1$ = -4.96, $p_1$ = .000; DM$_2$ = -81.1; $t_2$ = -3.36, $p_2$ = .048). These results are presented in Figure 2. We regard this effect as a memorization advantage of compound-like constructions in comparison to phrase-like constructions as the

---

$^3$ Values with the subscript “1” refer to the subject analysis (F$_1$) and values with the subscript “2” refer to the item analysis (F$_2$).
former clearly showed a bigger improvement than the latter from an early stage to a later stage in the memorization process.

![Graph showing response time over days for phrase-like and compound-like constructions.](image)

**Figure 2:** Phrase-like constructions (semantically compositional items with non-initial stress) versus compound-like constructions (semantically non-compositional items with initial stress) on the three test days ($F_1$)

### 5. Summary and conclusion

We started our paper with a discussion of the idea that compounds are more likely to become lexicalized than phrases. Then, we analyzed this proposal in more detail with the help of the notion of memorization, which is connected to the process of lexicalization in an important way, and by presenting evidence from studies that investigated the memorization of compounds and phrases.

The results presented in the analysis suggest that compounds/compound-like constructions show a memorization advantage in comparison to phrases/phrase-like constructions. Both the comparison of compounds/compound-like constructions and phrases/phrase-like constructions within a single language and the contrast of these construction types across different languages suggest a fundamental processing difference between compounds/compound-like constructions and phrases/phrase-like constructions. Since memorization represents a crucial step in lexicalization, the evidence reported in the current paper indicates that compounds/compound-like constructions and phrases/phrase-like constructions differ in the way they become a lexicalized complex construction.

### References


