



# The morphologization of German noun-participle combinations. A diachronic case study

Carlotta J. Hübener<sup>1</sup> 

Received: 5 October 2021 / Accepted: 24 April 2023 / Published online: 29 May 2023  
© The Author(s) 2023

## Abstract

The present paper focuses on the recent history of German noun-participle combinations in which the noun saturates an argument of the base verb. These structures are hybrids between phrases and words, yielding variation in spelling and in the form of the nominal constituent. For instance, the combinations *Mitleid erregend* vs. *mitleidØerregend* vs. *mitleidserregend* ‘pitiful’, lit. ‘pity-arousing’ (where Ø represents a zero morpheme) exemplify such variation, which is a (preliminary) result of (ongoing) language change. Therefore, this paper studies how the noun-participle pattern has diachronically evolved between syntax and word formation since the 18th century. Spelling and nominal forms are used as central indicators. For the latter, two approaches based on verb-valency are introduced. Data from the German Text Archive and DWDS core corpus show that over the past 300 years, noun-participle combinations have undergone a process of morphologization: they are increasingly written as a graphemic unit and take nominal forms that are typical of nominal root compounds. Moreover, this study shows that morphologization is partly supported by high token frequencies of individual types. It is argued that the phenomenon is best characterized by assuming a gradient distinction between syntax and word formation.

**Keywords** Synthetic compounds · Noun-participle combinations · Linking elements · Spelling · Language change · History of German

## 1 Introduction

This paper addresses the recent history of German noun-participle combinations, that is, complex structures combining a noun and a present participle. The focus is on combinations in which the noun saturates an argument of the base verb (e.g., *ekelerregend/Ekel erregend* ‘nauseating’, lit. ‘disgust-arousing’), a highly frequent pattern

---

✉ C.J. Hübener  
[huebenec@hu-berlin.de](mailto:huebenec@hu-berlin.de)

<sup>1</sup> Department of German Studies and Linguistics, Humboldt University of Berlin, Berlin, Germany

in present-day German (cf. Pümpel-Mader et al., 1992: 4, 288).<sup>1</sup> Noun-participle combinations have an ambivalent status between phrases and words, which causes variation in form. This can be seen in written language where both spaced and concatenated spellings occur diachronically. While spacing between the noun and the present participle is a feature typical of phrases (cf. (1)), noun-participle combinations may also be written as a graphemic unit, which is a feature typical of words (cf. (2)).

- (1) *Vertrauen erweckend* ‘instilling confidence’
- (2) *vertrauenerweckend* ‘instilling confidence’

For the latter type, the form of the nominal constituent also shows variation due to the ambivalent status of noun-participle combinations.<sup>2</sup> The choice of nominal forms either corresponds to the nominal form of the underlying verb phrase or to the nominal form of root compounds with the same first constituent.

- (3) [*zeitung*∅]*lesend* ‘newspaper-reading’
- (4) [*zeitungs*]*lesend* ‘newspaper-reading’

In (3), the nominal form is homophonous with the covertly inflected form of the noun in the underlying verb phrase: [*Zeitung*∅] *lesen* ‘to read (a) newspaper’. Such noun-participle combinations are close to verb phrases since the latter serve as a source of analogical extension for the nominal forms. In (4), the nominal form cannot be analyzed as an inflected form (\*[*Zeitungs*] *lesen*). Instead, the noun-participle combination takes a linking *-s-* analogously to root compounds with the same first constituent (such as [*Zeitungs*]*artikel* ‘newspaper article’, [*Zeitungs*]*junge* ‘paperboy’). These two principles compete with each other and lead to varying nominal forms as can be generally observed for synthetic compounds (Nübling & Szczepaniak, 2011: 59–62), cf. for instance [*Beitrags*]*zahler*/[*Beitrag*∅]*zahler* ‘contributor’, lit. ‘contribution payer’. Fuhrhop (1996: 546) notes that, diachronically, there is an increasing tendency for the nominal form in noun-participle combinations to align with that in root compounds, although this process has not yet progressed very far.

A large group of noun-participle combinations exhibit features that are typical of phrases, of words, or of a combination of both. Looking at the varying spellings and nominal forms, the question arises of whether noun-participle combinations are phrases (syntax) or compounds (word formation). The relationship between syntax and word formation has been extensively discussed in the literature (cf. Schlücker, 2020 for an overview) and is mostly approached from a synchronic perspective. However, the question of the relationship between syntax and word formation is ultimately a diachronic one. The variation of phrasal and morphological features observed in noun-participle combinations such as *nervenaufreibend*/*Nerven aufreibend* ‘nerve-racking’ indicates that there is ongoing language change (Nübling & Szczepaniak,

<sup>1</sup> According to corpus data from Pümpel-Mader et al. (1992: 4, 288), noun-participle combinations with an internal argument structure make up 11.71% of complex adjectives. Apparently, only combinations with concatenated spelling were included in this study.

<sup>2</sup> Regarding notation: Nominal forms are highlighted by square brackets []. The ∅ symbol represents the zero morpheme, that is, no additional phonetic material.

2011). Studying the historical development of noun-participle combinations thus allows us to draw conclusions about how the relationship between syntax and morphology is organized (diachronically). Diachronic studies on synthetic compounds in German are scarce; examples of studies on nominal synthetic compounds are Joeres (1995), Meibauer (1998), Werner (2017), and Werner et al. (2020). Noun-participle combinations are especially underrepresented in prior research. As will be shown in Sect. 2.2, they are more closely related to the syntax than nominal synthetic compounds are (e.g., *Feuerlöscher* ‘fire-extinguisher’). Due to this, their characteristics can shed some light on the relationship between phrases and words, particularly from a diachronic perspective. This paper aims to explore how noun-participle combinations have developed since the 18th century with regard to their phrasal and morphological features. Based on prior research, it is expected that noun-participle combinations have diachronically evolved away from the syntax and have taken on morphological features, supported by lexicalization. To examine this assumption, a diachronic corpus study of newspaper texts will be conducted.

This paper is organized as follows: Sect. 2 gives a brief overview of prior research on the relationship of word formation and syntax, highlighting the need for a diachronic approach (2.1). Subsequently, the morphological and syntactic features of noun-participle combinations are introduced in more detail (2.2), with special attention to the form of the nominal constituents (2.3). The design of the corpus study will be presented in Sect. 3. Sect. 4 introduces the empirical results. The data confirm that noun-participle combinations have become more morphological over time. However, combinations of low frequency have tended to keep spaced spellings. Moreover, high-frequency combinations have tended to preserve the nominal forms of corresponding verb phrases. In Sect. 5 it is argued that the phenomenon is best characterized by assuming a gradient distinction between syntax and word formation.

## 2 Between word formation and syntax

### 2.1 The diachronic relationship of word formation and syntax

Many languages provide evidence for linguistic units exhibiting both morphological and phrasal features (e.g., Bauer, 2019 on compounds and multi-word expressions in English; Booij, 2002a on separable complex verbs in Dutch; Hüning, 2010 on Dutch and German adjective-noun combinations; see Schlücker, 2020 for crosslinguistic discussion). Such units cannot be clearly classified as words or phrases. Hence, the relationship between word formation and syntax has been controversial in the literature. In this regard, there are two key issues: First, how can words and phrases be distinguished from each other? Second, do word formation and syntax form distinct linguistic domains or not? Works from different theoretical backgrounds have given divergent answers to the latter. Generative approaches assume a modular structure and view morphology and syntax either as separate domains (Ackema & Neeleman, 2004), as overlapping domains (Giegerich, 2015: 122–123), or as regulated by a single system (Lieber, 1992: 21). In contrast, works in the constructionist framework view word formation and syntax as poles belonging to a continuum (e.g., Booij &

Audring, 2017; Nenonen & Penttilä, 2014; Schlücker, 2020). Most of these pieces of work take a synchronic perspective. However, structures between words and phrases are often the (preliminary) result of (ongoing) language change. Diachronic data can account for many of the intermediate phenomena and furthermore shed light on present-day linguistic variation (cf. Nübling & Szczepaniak, 2010 on linking elements in German; Sanchez-Stockhammer, 2018: 3, 210–226 on the spelling of English compounds). In this respect, the question of the relationship between word formation and syntax is not least a matter of diachrony. In addition to the diachronic view, the relationship of word formation and syntax crucially depends on the language under consideration (Schlücker, 2020: 26). In some languages, morphological and phrasal entities can be clearly distinguished on the basis of their forms. Consider the contrast between the German examples *Schwärztee* and *schwarzer Tee* ‘black tea’. The former is a compound, exhibiting concatenated spelling, an uninflected first constituent, and main stress on the non-head. The latter is a phrase, exhibiting spaced spelling, an inflected adjective, and main stress on the head. Spelling, inflection, and stress help to distinguish nominal compounds and phrases also in Dutch and Danish (Hüning & Schlücker, 2015: 454). Other languages such as English (cf. Bauer, 1998) lack a systematic marking of words or phrases, making a distinction much more difficult to define. Given that German has a rich inventory of morphological features, it is considered a “prime example of a clear compound-phrase distinction” (Schlücker, 2020: 26). Nevertheless, a clear demarcation of the categories ‘word’ and ‘phrase’ in German is by no means trivial since certain linguistic patterns combine morphological and phrasal features and/or show variation in this respect. Examples of such borderline cases are noun-participle combinations such as *händeschüttelnd* ‘shaking hands’ and *Vertrauen erweckend/vertrauenerweckend* ‘instilling confidence’. Their ambivalent characteristics will be outlined below in Sect. 2.2.

In order to examine the relationship between word formation and syntax, it is necessary to clarify what distinguishes words from phrases. There is a long history of researchers attempting to define the notion of ‘word’ (e.g., Bloomfield, 1984: 178; Krámský, 1969: 67; Lyons, 1968: 200). Due to space considerations, the discussion cannot be presented here in its entirety. Haspelmath (2011) convincingly argues that a universal definition of ‘word’ is impossible (also see Lyons, 1968: 181; Matthews, 1972: 147–156). This is especially the case when taking a diachronic perspective as the notion of ‘word’ can change. Nonetheless, it is possible to approach the notion of ‘word’ by identifying features that are typical of words. In this respect, German is a promising language of investigation due to its aforementioned morphological features. The present study focuses on two word-typical features, namely concatenated spelling and linking elements. In the history of German, concatenated spelling is considered to go hand in hand with word status (e.g., Solms, 1999: 133, fn. 14; Kopf, 2018a: 151–154). As in English (e.g., Bauer, 1998: 68–69), the reverse does not apply: throughout the history of German, spaced spelling is not a clear indicator of non-word status (Solling, 2012; Jacobs, 2005: 6, 168, fn. 13). Hyphenation can be considered as a “compromise” (Bauer, 1998: 69) between spaced and concatenated spelling, which is especially interesting with regard to structures between syntax and word formation. In the context of this study, concatenated spelling is not conceived

of as a feature strictly implying word status.<sup>3</sup> Instead, it is considered a prototypical feature of words in that it occurs almost exclusively with words diachronically. Linking elements are the second word-typical feature considered here. They occur between the constituents of complex words (Fleischer & Barz, 2012: 185), as for example *-s-* in *Geburstag* (= *Geburt* + *s* + *tag*) ‘birthday’. It is widely assumed that linking elements have lost their initial morphosyntactic meaning (Ortner et al., 1991: 51; Hennig, 2016: 337). Consequently, many linking elements cannot be analyzed as inflectional morphemes, making them features of word formation. In the case of *lebensrettend* ‘life-saving’, for example, the base verb *retten* ‘to save’ requires a complement in the accusative case. *Leben* ‘life’ is covertly marked for accusative: (*ein*) [*Leben*Ø] *retten* ‘to save a life’. The linking *-s-* in [*Lebens*] *rettend* cannot be analyzed as an inflectional morpheme because the form *Lebens* is ill-formed in the phrase \* [*Lebens*] *retten* ‘to save (a) life’. In such cases, linking elements are clearly categorial indicators of word formation. However, there are many ambiguous cases where linking elements and inflectional morphemes correspond in form. Sect. 2.3 therefore argues that it is useful to avoid the distinction between linking elements and inflectional morphemes when investigating the relationship of words and phrases. As an alternative, an approach via nominal forms is proposed. Moreover, Sect. 2.3 introduces three approaches for categorizing nominal forms as indicators of word formation or syntax.

In view of structures combining morphological and phrasal features (e.g., in terms of spelling and nominal forms: *Leben rettend/lebenrettend/lebensrettend*), the question arises as to which direction this development is taking: Are formerly syntactic structures becoming more morphological or are morphological structures taking on syntactic characteristics? Both directions are possible, but with regard to word-formation morphology, the former is apparently the standard case.<sup>4</sup> The diachronic view holds countless examples of multi-morphemic entities taking on morphological features or abandoning syntactic features. According to the literature, this development is supported by lexicalization. For example, Plag et al. (2008) show that more lexicalized compounds (with frequency of use and spelling as indicators of lexicalization) tend toward more stress on the first constituent in English, a stress pattern typical of compounds. Similar observations have been made for the concatenated spelling of English compounds (Kuperman & Bertram, 2013: 946; Sanchez-Stockhammer, 2018: 134). Booij (2002b: 315) gives examples of lexicalized adjective-noun sequences exhibiting concatenated spelling despite the inflectional schwa of the adjective such as *blindedarm* ‘appendix’ and *jongeman* ‘young man’. Other lexicalized adjective-noun sequences in Dutch such as *de wetenschappelijkØ directeur* ‘the scientific director’ allow for deletion of the inflectional schwa despite spaced spelling, making them more similar to compounds, which have no internal inflection either (Booij, 2002b: 315–316; Hüning, 2010: 208–209).

<sup>3</sup>Particularly in earlier stages of German, syntactic sequences written as one graphemic unit also occurred (cf. Tophinke, 2000; Bredel, 2006; Hübener, 2021). Although this phenomenon is rare, it shows that diachronically, concatenated spelling does not necessarily imply word status.

<sup>4</sup>An example of a pattern that became more syntactic is adjective-noun combinations in Dutch. While they were often realized as compounds in Middle Dutch, they are mostly expressed by phrases today, consider for example *soetwijn* > *zoete wijn* ‘sweet wine’ (Hüning, 2010: 205).

Fuhrhop (2000: 201) refers to this process in which linguistic units abandon structural features that are typical of the syntax and instead adopt typically morphological characteristics as *morphologization*. She suggests that noun-participle combinations have undergone morphologization as well and that this development is supported by lexicalization (Fuhrhop, 2000: 211; also see Kopf, 2018a: 183, fn. 164). With regard to the relationship between word formation and syntax, it is very insightful to trace this development on the basis of diachronic data. To date, however, there has been little empirical research on the history of German noun-participle combinations. The same holds for synthetic compounds in German in general, of which noun-participle combinations form a subgroup. Synthetic compounds are defined here as structures of the form noun + verbal stem + affix with the noun saturating an argument of the base verb (e.g., *Feuerlöscher* ‘fire-extinguisher’).<sup>5</sup> This means that in the corresponding verb phrase, the noun functions as an object of the verb: *Feuerlöscher* – *Feuer löschen* ‘to extinguish fire’. In recent years there has been growing interest in German synthetic compounds in empirical work with consideration given to their exceptional status between words and phrases (e.g., Gaeta & Zeldes, 2012, 2017; Jores, 1995; Meibauer, 1998; Werner 2017, 2020; Werner et al., 2020; cf. Neef, 2015 for an overview). However, large quantitative studies on the diachronic development of synthetic compounds still present a scientific lacuna. It is the aim of this study to fill this research gap by investigating the recent history of noun-participle combinations in German. The following sections introduce the ambivalent characteristics of noun-participle combinations.

## 2.2 Syntactic and morphological characteristics of noun-participle combinations

The structural ambiguity of noun-participle combinations between words and phrases is triggered by the present participle, which has adjectival and verbal features (Fuhrhop, 2007: 129; also see Fuhrhop & Teuber, 2000).

Unlike in English, present participles in German do not occur in analytic verb forms (*He is walking* – *\*Er ist gehend*).<sup>6</sup> Present participles occur in attributive position, agreeing with the head noun for case, number, and gender (*die scharrenden Hühner* ‘the scratching chickens’), and in adverbial position (*er sah lächelnd auf* ‘he looked up smiling’). In some cases, present participles are gradable (*das bedeutendste Werk Goethes* ‘the most important work of Goethe’). Due to their distributional traits, present participles are considered as adjectives here. However, they are peripheral adjectives (Sommerfeldt, 1988: 225; also see Charitonova, 1977: 30) since some characteristics of their verbal bases are preserved. Central to this study is that they take on complements of the underlying verb, highlighted in square brackets in (5)–(8). In (5-a), *das Fahrrad* is an object of the verb *reparieren*. In the corresponding participle phrase (5-b), *das Fahrrad* can be seen as a secondary object of the participle

<sup>5</sup>This paper focuses on synthetic compounds with a verbal base. In general, the governing constituent can also be an adjective, as for example in *liebenswert* ‘kind’, lit. “worthy of love”.

<sup>6</sup>According to Blatz (1900: 603), the present participle frequently occurs in syntactic constructions such as the following: *Ich sah den Reiter über die Brücke sprenkend* ‘I saw the rider jumping over the bridge’. Fuhrhop and Teuber (2000: 180), on the other hand, evaluate such constructions as ungrammatical. Obviously, this is a case of syntactic change (cf. Zifonun et al., 1997: 1418).

*reparierend*.<sup>7</sup> In fact, present participles can convert verb phrases of any complexity into participial phrases (functioning as adjective phrases) with almost no restrictions. The ability of present participles to form phrases demonstrates their strong linkage to the syntax.<sup>8</sup> This paper focuses on complements with nominal heads. The nouns of participial complements can have a determiner (5-b), can be quantified (6), modified (7), or occur autonomously (8).

- (5) a. *Die Frau repariert [das Fahrrad]* ‘the woman is fixing the bicycle’  
 b. *Die [das Fahrrad] reparierende Frau* ‘the woman fixing the bicycle’, lit. “the bicycle fixing woman”
- (6) *eine [viele Bücher] umfassende Sammlung* ‘a collection including many books’
- (7) *die [getrocknete Tomaten] essenden Mitarbeiter* ‘the employees eating dried tomatoes’
- (8) *die [Kaugummi] kauenden/[kaugummi]kauenden Kinder* ‘the children chewing gum’

Cases (5) to (7) are clearly syntactic because the nouns occur with a determiner or a modifier. However, noun-participle combinations<sup>9</sup> with autonomous nouns such as (8) are structurally ambiguous (Fuhrhop, 2007: 134). On the one hand, they can be analyzed as participle phrases, analogously to phrases (5-b)–(7). On the other hand, they can be seen as adjectival compounds.<sup>10</sup> This hybrid status is reflected in spelling. Noun-participle combinations with autonomous nouns can occur both in concatenated and spaced spellings (Dudenredaktion, 2020: 56), often yielding variation in individual types (*Mitleid erregend* vs. *mitleiderregend* ‘pitiful’, *kräftesparend* vs. *Kräfte sparend* ‘effort-saving’). Concatenated spelling, however, is more common (Hennig, 2016: 411–412). Concatenated spelling is a word-typical feature, whereas spaced spelling suggests a stronger link of a noun-participle combination to the syntax. Several studies report that lexicalization or high token frequencies give rise to concatenated spelling (e.g., Weidman, 1941: 97–98 on Middle High German compounds; Kuperman & Bertram, 2013: 946; Sanchez-Stockhammer, 2018: 134 on En-

<sup>7</sup>The term *secondary object* is used here for illustrative reasons. See Imo (2016: 172–173) for a critical discussion of this analysis.

<sup>8</sup>In this respect, participle structures are more closely related to the syntax than patterns of nominal synthetic compounding are. The latter can only form complex words, not phrases: [*das Fahrrad*] *reparierend* ‘fixing [the bicycle]’ vs. \*[*das Fahrrad*] *Reparierer* ‘[the bicycle] repairer’.

<sup>9</sup>In the remainder of this study, the term *noun-participle combination* refers to complex structures combining an autonomous noun and a present participle with the noun saturating an argument of the base verb.

<sup>10</sup>Other than nominal compounds including modification such as *Heißwasserspeicher* ‘hot water tank’ or *deutsche Sprachwissenschaft* ‘German linguistics’, lit. “German language science” (cf. Bergmann, 1980), noun-participle combinations refrain from integrating complex nominal phrases morphologically (?*getrocknete-tomaten-essend* ‘eating dried tomatoes’). Apparently, due to the syntactic linkage of present participles, it is preferred to realize modified nouns as noun phrases rather than as complex first constituents. Therefore, cases like (7) are not part of the ambivalent construction that is the scope of this study.

glish compounds). This could also be the case for noun-participle combinations as lexicalization could enforce them to be perceived as lexical units.

In terms of their syntactic distribution, noun-participle combinations are adjectives (or form adjective phrases), parallel to simple participles. For example, they occur in the attributive position, inflectionally agreeing with the head noun: *ein aufsehenerregender/Aufsehen erregender Hut* ‘a sensational hat’. In some instances, noun-participle combinations are gradable (e.g., *am erfolgversprechendsten* ‘the most promising’, lit. “the most success-promising”). Linking elements, or more generally, the form of the nominal constituents, are another feature that reveals the ambivalence of noun-participle combinations. They will be presented in the following section.

### 2.3 Nominal forms as a toolkit for investigating morphologization

While spelling is a straightforward variable for operationalizing morphologization, linking elements require more detailed explanations. Linking elements have mostly been considered in the context of root compounds<sup>11</sup> (e.g., Nübling & Szczepaniak, 2008; Nübling & Szczepaniak, 2010; Kopf, 2018a,b) and rarely in the context of synthetic compounds. However, due to their relation to verb phrases, synthetic compounds differ from root compounds in terms of linking elements. To grasp this phenomenon, new theoretical and methodological approaches are needed. This section starts by arguing that it is more useful to analyze the nominal form rather than considering linking elements in isolation. In succession, three approaches are presented that use nominal forms as a toolkit for investigating the morphologization of synthetic compounds.

In the remainder of this paper, the form in which a nominal constituent occurs is referred to as the nominal form and is highlighted in square brackets, as for example in [*hände*]ringend ‘hand-wringing’. It includes possible linking elements and inflectional morphemes. This approach is closely related to the concept of stem paradigms as proposed by Fuhrhop (1998). However, nominal forms are first and foremost a descriptive tool that does not require separation of compounding and inflection.<sup>12</sup>

The reason why nominal forms provide information on morphologization becomes immediately evident when comparing the nominal forms of synthetic compounds, verb phrases and root compounds. In present-day German, many nouns have a fixed nominal form in which they occur as first constituents in nominal root compounds (e.g., Kopf, 2018a: 284–285). Nominal forms of nouns ending in the suffix *-ung* such as [*Heizungs*]rohr ‘heating pipe’ and [*Meinungs*]freiheit ‘freedom of opinion’ usually contain a linking *-s-*. In contrast, synthetic compounds and noun-participle combinations in particular often lack the linking *-s-* although it is typical of several first constituents (Hennig, 2016: 339), as for example in [*achtung*Ø]gebietend ‘imposing’

<sup>11</sup>Synthetic compounds need to be distinguished from root compounds. Throughout this paper, the term *root compound* is used to refer to compounds without a word-internal argument relation (e.g., *Teetasse* ‘tea cup’, *Regenwald* ‘rain forest’). Root compounds may, nevertheless, have deverbal second constituents (e.g., *silberglänzend* ‘silvery’, lit. “silver-shining”).

<sup>12</sup>Fuhrhop’s (1998, 2007) approach focuses on comparing synthetic compounds with root compounds. For instance, she refers to nominal forms corresponding to verb phrases as compounding stem forms (e.g., *achtunggebietend*), although inflectional stem forms could also be assumed here.



as opposed to [*Achtungs*]erfolg ‘respectable achievement’. In such cases, the nominal forms correspond to the nominal forms of the underlying verb phrases: [*Achtung*Ø] gebieten ‘to command respect’ – [*achtung*Ø]gebietend. The analogy to verb phrases is also a frequently used gateway for plural nominal forms to occur in noun-participle combinations (Fleischer & Barz, 2012: 321), as for example in [*Kräfte*] sparen ‘to save efforts’ – [*kräfte*]sparend (Archiv der Gegenwart, 2001 [1944], DWDS) ‘effort-saving’. In contrast, for corresponding nominal forms in root compounds, singular forms as in [*Kraft*Ø]werk ‘power station’ are the default (cf. Kopf, 2018a: 277–278 on (Early) New High German), although counterexamples such as [*Kräfte*]gleichgewicht ‘balance of forces’ exist. In such cases, it is hardly possible to speak of mere linking elements (Fleischer & Barz, 2012: 321, 331) since they have a grammatical function indicating plurality.<sup>13</sup> However, in analogy to root compounds, plural forms as in [*kräfte*]sparend ‘effort-saving/saving efforts’ can be neutralized as in [*kraft*Ø]sparend ‘effort-saving’ (Fleischer & Barz, 2012: 321).

Many authors agree that nominal forms within synthetic compounds and particularly within noun-participle combinations correspond to nominal forms in the underlying verb phrases (Augst, 1975: 121; Hennig, 2016: 339; Fuhrhop, 2000: 211; Nübling & Szczepaniak, 2011: 59–62). This is not necessarily the case. Consider the example [*Lebens*]verlängernd ‘life-prolonging’. The base verb *verlängern* requires a complement in the accusative case, which is covertly marked for *Leben*: (*das*) [*Leben*Ø] verlängern ‘to prolong life’. However, the nominal form of [*Lebens*]verlängernd has an additional *-s-*. Therefore, it does not correspond to the verb phrase [*Leben*Ø] verlängern. Instead, the nominal form corresponds to the nominal form of root compounds with the same first constituent. *Leben* usually takes a linking *-s-* when occurring as a first constituent in root compounds (e.g., [*Lebens*]lauf ‘curriculum vitae’). Apparently, synthetic compounds can be influenced by the model of root compounds (Fuhrhop, 2007: 30; cf. Krott et al., 2007 on analogical effects on linking elements in German) and evolve away from their underlying verb phrases. Thus, nominal forms that are influenced by root compounds are an indicator of morphologization (cf. Fuhrhop, 2000).

Relatively little is known about how nominal forms of synthetic compounds in German are distributed. Diachronically, synthetic compounds often show variation in terms of nominal forms, as for example in [*kriegs*]führend/[*krieg*Ø]führend ‘warring’. Kopf (2018a: 183–184) finds that the linking *-s-* occurs considerably less frequently in synthetic compounds than in root compounds (16.3% vs. 44.3% linking *-s-*) in the Mainz (Early) New High German corpus (1500 to 1710). It competes with the zero morpheme, which occurs most likely by analogy to corresponding verb phrases in Kopf’s data ([*Urtheil*Ø]sprecher ‘judge’, lit. “judgment speaker” – ein [*Urteil*Ø] sprechen ‘to pronounce a verdict’). On the basis of data extracted by means of a Google search, Nübling and Szczepaniak (2011: 59–60) show that linking *-s-* and the zero morpheme alternate greatly for numerous synthetic compounds in present-day German. Alternating nominal forms are due in part to ongoing language change (Nübling & Szczepaniak, 2011). In the case of synthetic compounds, the analogical extension of nominal forms (cf. Krott et al., 2007) has undergone change: there is

<sup>13</sup>This, again, highlights that it is useful to consider nominal forms instead of linking elements.

**Table 1** Competing analogical sources for nominal forms in synthetic compounds

Verb phrase	Synthetic compound	Root compound
[ <i>Ordnung</i> ∅] <i>lieben</i> 'to love order'	[ <i>ordnung</i> ∅/s] <i>liebend</i> 'order-loving'	[ <i>Ordnungs</i> ] <i>amt</i> 'public order office', ...
( <i>eine</i> ) [ <i>Brille</i> ∅] <i>tragen</i> 'to wear glasses'	[ <i>brille</i> ∅/n] <i>tragend</i> 'wearing glasses'	[ <i>Brillen</i> ] <i>bügel</i> 'eyeglass temple', ...
[ <i>Leben</i> ∅] <i>retten</i> 'to save lives'	[ <i>leben</i> ∅/s] <i>rettend</i> 'life-saving'	[ <i>Lebens</i> ] <i>ziel</i> 'life goal', ...

competition between the analogical source of verb phrases and the analogical source of root compounds, which is illustrated by examples in Table 1. In general, the current trend is towards a more morphological and less phrasal character for synthetic compounds (Fuhrhop, 2000: 211; Fuhrhop, 1998: 217–218; also see Werner, 2017), which means that root compounds gain ground as analogical sources for nominal forms. Fuhrhop (2000: 211) states that the more a noun-participle combination is lexicalized, the more likely it is to occur with the nominal form that is commonly used for root compounds with the same first constituent, as can be observed in [*richtung*∅] *weisend* > [*richtungs*] *weisend* 'trendsetting' (also see Kopf, 2018a: 183, fn. 164).<sup>14</sup> This would be in line with the universal tendency of lexicalization to support morphologization (cf. Sect. 2.1).

To sum up, nominal forms are an important tool for investigating morphologization as they can either correspond to verb phrases or to root compounds. Three approaches are possible to analyze how morphologized individual synthetic compounds are. Their nominal forms can either be compared to nouns in corresponding verb phrases or to root compounds with the same first constituent. However, as we will see in Sect. 4.3, 62.25% to 67.8% of the nominal forms are ambiguous between nominal forms of verb phrases and root compounds. Due to the many ambiguous cases, a third approach combines the former two approaches and treats ambiguity as a separate category. In what follows, the three approaches will be discussed in more detail.

**Inflectional compatibility** A classification of linking elements often considered in the literature is based on the inflectional paradigms of the first constituents (cf. Wellmann et al., 1974: 359, fn. 5). Wellmann et al. (1974: 359, fn. 5) distinguish between paradigmatic and nonparadigmatic linking elements. A paradigmatic linking element is identical in form to an inflectional suffix of the noun. For example, the linking *-n* in [*Wolken*] *formation* 'cloud formation' is paradigmatic because *Wolken* 'clouds' is the plural form of *Wolke*. A nonparadigmatic linking element does not occur in the inflectional paradigm of the noun. For instance, the *-s*- in [*Geburts*] *tag* 'birthday' is nonparadigmatic since \**Geburts* does not correspond to an inflectional form of *Geburt* 'birth'. This distinction is useful for root compounds but insufficient for determining whether a synthetic compound is syntactically or morphologically shaped.

<sup>14</sup>In the present paper, it is assumed that linguistic patterns are lexicalized or conventionalized with increasing token frequency (e.g., Zeldes, 2013: 243).

For example, the *-s-* in [*lebens*]verlängernd ‘life prolonging’ would be paradigmatic according to this approach (*Lebens* ‘life.GEN.SG’). However, the fact that the verb *verlängern* normally requires a complement in the accusative case (*Leben*∅) is disregarded ([*Leben*∅] *verlängern* ‘to prolong life’, but \*[*Lebens*] *verlängern*).

This shows that a comprehensive analysis of synthetic compounds has to consider the valency of the governing constituent. Hence, this study takes into account the case required by the base verb of the noun-participle combination. Nominal forms like [*lebens*] in [*lebens*]verlängernd will be referred to as inflectionally incompatible. In contrast, consider [*ordnung*∅]liebend ‘order-loving’. Here, the nominal form [*ordnung*∅] is consistent with an accusative form as required by the base verb *lieben*: [*Ordnung*∅] *lieben* ‘to love order’. Therefore, this nominal form can be described as inflectionally compatible with this combination of base verb and noun.

**Comparison to root compounds** Fuhrhop (2007: 141–142) examines whether synthetic compounds take the linking elements that would be expected to occur after corresponding first constituents in root compounds. Since root compounds are so dominant and ubiquitous in German (e.g., Ortner et al., 1991: 3, 112; Schlücker, 2012: 2), the distribution of their linking elements can be seen as a benchmark for other patterns that involve linking elements. Fuhrhop (2007) thus refers to linking elements in root compounds as conventional. For example, in [*produktions*]anregend (Berliner Tageblatt, morning issue, 02-13-1902, DWDS) ‘production-stimulating’ the linking *-s-* is conventional because it regularly occurs after derivatives ending in *-ion* (cf. Nübling & Szczepaniak, 2008: 4), as for example in [*Produktions*]kosten ‘manufacturing costs’. In contrast, the conventional *-s-* does not occur in [*religion*∅]bildend (Völkischer Beobachter, Berlin issue, 03-04-1934, DWDS) ‘religion-forming’. Since [*religion*∅]bildend ‘religion-forming’ is unlinked, this combination apparently is not modelled analogously to root compounds but rather corresponds to the verb phrase [*Religion*∅] bilden ‘to form religion’.

According to Fuhrhop’s (2007) examination of 420 noun-participle combinations from a contemporary newspaper corpus, most combinations take the same linking elements as root compounds with corresponding first constituents (e.g., [*entzündungs*]hemmend ‘anti-inflammatory’, [*arbeits*]fördernd ‘work-promoting’) (Fuhrhop, 2007: 141).

**Source of analogical extension** The third approach combines the parameters presented above and adds an ambiguous category: whenever a nominal form is inflectionally compatible in the above-defined sense, it can be considered to be motivated by a verb phrase (e.g., [*vertrauen*∅]erweckend ‘instilling confidence’, cf. [*Vertrauen*∅] *erwecken* ‘to instill confidence’). Whenever a nominal form is conventional as defined above, it can be analyzed as motivated by root compounds (e.g., [*arbeits*]fördernd ‘life-saving’, cf. [*Arbeits*]tag ‘working day’).

One weakness in Fuhrhop’s (2007) approach and in the approach based on inflectional compatibility is that linking elements and inflectional suffixes are often identical in form. For example, root compounds with *Blume* ‘flower’ as a first constituent usually take a linking *-n-* (e.g., [*Blumen*]strauß ‘bouquet of flowers’). The nominal form in [*blumen*]pflückend ‘flower-picking’ is thus conventional. However, it is also

identical to the inflected nominal form in the corresponding verb phrase [*Blumen*] *pflücken* ‘to pick flowers’. Therefore, it cannot be decided whether the nominal form [*blumen*] is motivated by root compounds or by a verb phrase. As a result, a third category for ambiguous cases like [*blumen*] *pflückend* needs to be established. Note that the zero morpheme can be ambiguously motivated as well: [*gold*Ø] *erzeugend* ‘gold-producing’ – [*Gold*Ø] *erzeugen* ‘to produce gold’ – [*Gold*Ø] *barren* ‘gold bar’. Given that most root compounds are unlinked (e.g., Krott et al., 2007: 29; such as [*Gold*Ø] *barren*), it is not surprising that ambiguously motivated cases form the largest class (cf. Sect. 4.3).

Morphological indicators such as concatenated spelling and linking elements had largely been established for root compounds by the early New High German period (by around 1700; e.g., Solling, 2012; Kopf, 2018a). Therefore, this paper aims to study how noun-participle combinations have diachronically developed between word formation and syntax from the early 18th century up until today. For a thorough empirical analysis of noun-participle combinations, a diachronic corpus of newspaper texts will be examined. Spellings and nominal forms will be used as central indicators for locating noun-participle combinations between phrases and words. Based on the literature, the following hypotheses will be tested:

1. Noun-participle combinations diachronically undergo morphologization. In this case, we would expect them to be increasingly written as one graphemic unit and to take the same nominal forms as root compounds with corresponding first constituents.
2. The higher their token frequency, the more prone noun-participle combinations are to undergo morphologization. This should be reflected in concatenated spellings and in the occurrence of nominal forms that match those of root compounds with corresponding first constituents.

The methodology—including data sampling and annotation—is presented in the following section.

### 3 Corpus study: data and methodology

#### 3.1 Sampling

The data for this study were taken from two newspaper corpora extending over the period from the 18th century to the 20th. Newspaper language is expected to have properties that are typical of the standard German language (Eisenberg, 2007: 217). The focus of this study is on the standard variety of German as spoken in Germany. Newspaper articles from the German Text Archive (DTA) were used to cover the 18th and 19th centuries.<sup>15</sup> Since the DTA is continuously extended, it is a dynamic corpus.<sup>16</sup> The DTA newspaper corpus contained a total of 13,280,605 tokens when the data were collected in December 2019. Since the DTA mostly consists of texts

<sup>15</sup><https://www.dwds.de/d/k-referenz#dta>.

<sup>16</sup>A snapshot of the DTA as of October 16, 2019, is accessible via <https://kaskade.dwds.de/dstar/dta2019/>.

of trans-regional importance,<sup>17</sup> it is expected to represent the early stages of German standard language. It must be noted that the DTA is not balanced over time. This means that earlier decades in the corpus comprise significantly fewer tokens than later decades. The 1700s, 1760s, 1820s, and 1860s are not represented in the newspaper corpus at all. This was taken into account in the analyses, which focus primarily on the period from 1840 onwards. From then on, the size of the corpus allows for more reliable statements and for better comparability between the decades. To compensate for remaining differences between the decades, absolute numbers and relative frequencies will be reported.

In order to study the years 1900 to 1999, the core corpus of the Digital Dictionary of the German Language (DWDS) was used and was also limited to newspaper articles.<sup>18</sup> For copyright reasons, not all parts of this corpus are accessible to the public, especially for texts dating from 1980 onwards. These texts cannot be queried, so only the publicly available part of the corpus is used here. The number of publicly available tokens in the DWDS newspaper corpus is 30,154,346.

The DTA and DWDS core corpus are lemmatized and tagged for parts of speech with the same tag set and according to similar guidelines, thus identical queries can be used. To retrieve noun-participle combinations, the interfaces were queried for adjectives in any syntactic position ( $\$p=ADJ^*$ ) with lemmata ending in *-end*, *-elnd* or *-ernnd* ( $\$l=/e[lr]?nd\$/$ ), compare (9). Since present participles are tagged as adjectives in both corpora and end in the grapheme sequences mentioned, this query retrieves present participles or noun-participle-combinations with concatenated spelling or hyphenation. Additionally, appellative nouns ( $\$p=NN$ ) followed by such adjectives were queried to retrieve combinations with spaced spelling, compare (10). Hits that were already retrieved by the query in (9) were manually removed.

(9)  $\$l=/e[lr]?nd\$/$  WITH  $\$p=ADJ^*$

(10)  $\$p=NN$   $\$l=/e[lr]?nd\$/$  WITH  $\$p=ADJ^*$

Both queries ensure high recall. However, the queries incur low precision and require forms such as *tausend* ‘thousand’ and single participles to be manually excluded. The noun had to have the interpretation as an object of the underlying verb, thus hits such as *schweißglänzend* ‘shining with sweat’ were excluded. Additionally, noun-participle combinations with determiners, quantifiers, or modifiers were excluded (*den Redner beleidigend* ‘insulting the speaker’, Berliner Tageblatt (morning issue), 02-12-1902, DWDS). These structures are clearly phrasal and thus not part of the continuum between syntax and morphology that is the scope of this study (cf. Sect. 2.2).

The syntactic context was used to clarify whether a noun occurs autonomously. In (11), for example, the noun *Butter* ‘butter’ is preceded by the determiner *der* ‘the’. This determiner, however, refers to *Frauen* ‘women’, as can be seen from the inflection. Hence, *Butter* occurs autonomously, so the hit is included in the study. By contrast, the noun *Umweltschutz* ‘environmental protection’ in (12) has a determiner

<sup>17</sup><http://www.deutschestextarchiv.de/doku/textauswahl>.

<sup>18</sup><https://www.dwds.de/d/k-referenz#kern>.

and was excluded. Coordinated nouns were considered if none of the nouns were modified or part of a prepositional phrase.

- (11) *die Abwesenheit der Butter einkaufenden Frauen* (DWDS)<sup>19</sup>  
 the.absence.the.GEN.PL butter.ACC.SG buy.PTCP woman.GEN.PL  
 ‘the absence of the butter-buying women’
- (12) *ihre den Umweltschutz betreffenden Maßnahmen* (DWDS)<sup>20</sup>  
 their.ACC.PL the.ACC.SG environmental.protection.ACC.SG concern.PTCP  
 measure.ACC.PL  
 ‘their measures relating to environmental protection’

Twenty-two cases were ambiguous between autonomous and modified nouns.<sup>21</sup> Considering the broader syntactic context of these cases, the possibility that the nouns are intended to occur autonomously could not be ruled out. Therefore, these cases were included in the study. Finally, noun-participle combinations that are used as nouns were excluded (e.g., *Vorstandsvorsitzende* ‘chairwoman’).

Descriptive statistics were performed with *R* (version 4.1.3, *R* Core Team, 2022). In order to investigate which factors determine the choice for spellings and nominal forms, binomial logistic regressions were performed using the package *lme4* (Bates et al., 2022). Model selection was based on the *AIC* values and on the results of an ANOVA. Marginal and conditional  $R^2$  values were calculated using the function *r.squaredGLMM* (method: *delta*) from the package *MuMIn* (Bartón, 2022). The full dataset and *R* code are available at <https://osf.io/ey54w/>.

### 3.2 Annotation

Noun-participle combinations were annotated for spelling, valency of the base verb, token frequency of individual types, lemma, and features of nominal forms. In order to simplify the statistical analysis, token frequencies were determined over the entire period of investigation, that is, diachronic developments of frequencies were not considered. Each noun-participle combination was lemmatized, with variants of spelling and nominal forms (e.g., *Erfolg versprechend*, *erfolgversprechend*, *erfolgsversprechend* ‘promising’) put together as one type. Nominal forms were analyzed according to the three approaches introduced in Sect. 2.3: inflectional compatibility, comparison to root compounds (Fuhrhop, 2007), and source of analogical extension. Inflectionally incompatible nominal forms corresponding to root compounds are not expected to occur in noun-participle combinations with spaced spelling in New High German (*\*Erfolgs versprechend* ‘promising’; e.g., Kopf, 2018a: 342–343), as confirmed by the present study. Hence, the annotation of nominal forms was limited to noun-participle combinations with concatenated spelling or hyphenation.

<sup>19</sup>Vossische Zeitung (evening edition), 03-03-1916, provided by the DWDS core corpus.

<sup>20</sup>Archiv der Gegenwart, 2001 [1979], provided by the DWDS core corpus.

<sup>21</sup>An example of this is *frohe Hoffnung erweckende Nachrichten* (Vossische Zeitung (morning edition), 03-02-1912, provided by the DWDS core corpus). This phrase can be analyzed either as *[[frohe Hoffnung] erweckende] Nachrichten* ‘news that awakens joyful hope’ or as *[frohe] [Hoffnung erweckende] Nachrichten* ‘happy hopeful news’.

Regarding the comparison to root compounds (Fuhrhop, 2007), it would not have been feasible within the scope of this study to determine the conventional nominal form for each noun on an empirical basis, especially since diachronic developments would have to be considered as well. To operationalize the concept of conventional and unconventional nominal forms, the annotation is limited to noun-participle combinations which have derivatives ending in *-ung* or nominalized infinitives as first constituents, as for example [*achtungs/Ø*]*gebietend* ‘awe-inspiring’, [*vertrauens/Ø*]*erweckend* ‘instilling confidence’. These two patterns show a very clear preference as first constituents in root compounds: in contemporary German, their nominal forms regularly contain a linking *-s-* (Nübling & Szczepaniak, 2008: 4; Nübling & Szczepaniak, 2011: 58) (e.g., [*Achtungs*]*applaus* ‘polite applause’, [*Vertrauens*]*person* ‘person of trust’). This was already true by the 18th century (Kopf, 2018a: 264–265), which means that *-s-* is conventional throughout the entire period of investigation, whereas a zero morpheme occurs most likely by analogy to covertly inflected nouns in verb phrases. For both derivatives ending in *-ung* and nominalized infinitives, root compounds must be the analogical source of nominal forms since the *-s-* cannot be explained by the syntax. In the case of derivatives ending in *-ung*, the linking *-s-* cannot be analyzed as an inflectional suffix (*\*Leistungs*, *\*Zeitungs*); it is nonparadigmatic (cf. Sect. 2.3). For nominalized infinitives, a genitive singular interpretation of *-s-* (*des Vertrauens*) is impossible in this context as nominalized infinitives do not occur with the rare cases of verbs requiring genitive objects in the corpus data (cf. Sect. 4.1). Hence, derivatives ending in *-ung* and nominalized infinitives are a promising test case for conventional or unconventional nominal forms.

For annotating the source of analogical extension of nominal forms, a random sample of 50 tokens was annotated for each span of 20 years between 1840 and 1999. Note that this sample is not limited to a specific morphological structure of first constituents. It was not feasible to annotate larger samples within the scope of this study. The random sample is likely to contain mainly noun-participle combinations with high token frequencies. In order to consider low-frequency types as well, a further random sample of 50 tokens per decade was taken from hapax legomena originating from the 20th century.

## 4 Results of the corpus study

### 4.1 Frequencies

The procedure described above retrieved 14,933 tokens of noun-participle combinations and 2,050 types in the period from the early 18th century to 1999.<sup>22</sup>

Noun-participle combinations have undergone a remarkable increase in token frequency over the past three centuries. Figure 1 illustrates this development. The number of tokens per one million words increased from 68.74 in the 1710s to 626.45 in the 1970s. The outlier from the 1800s comes from individual texts in which certain

<sup>22</sup>Variants of spellings and nominal forms are put together as one type. For instance, *Vertrauen erweckend*, *vertrauenerweckend*, and *vertrauenserweckend* ‘instilling confidence’ all belong to the same type.

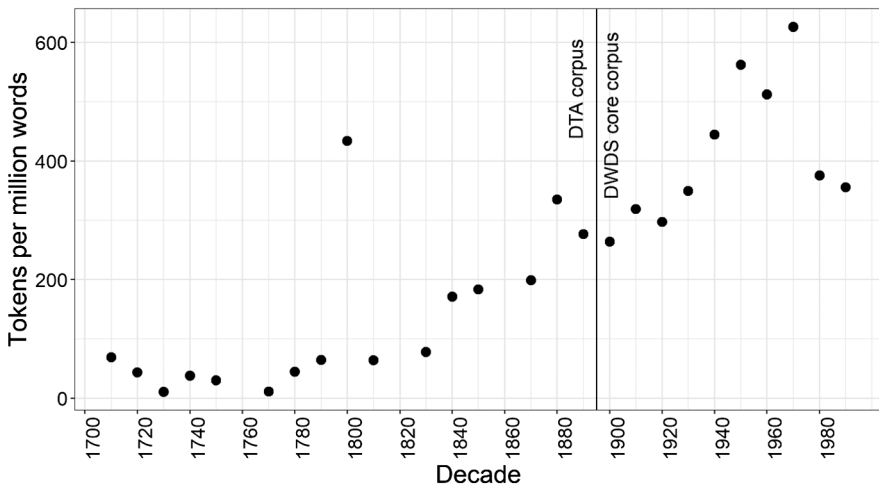


Fig. 1 Token frequencies in the DTA and DWDS core corpus projected to one million words

types are frequently repeated. The sudden drop to 375.66 (1980s) or 355.97 tokens per million words (1990s) at the end of the 20th century is apparently due to the composition of the corpus. Until the 1970s, the DWDS newspaper corpus largely consists of texts from the source *Archiv der Gegenwart* (newspaper articles with a focus on day-to-day politics) with a share of up to 91.68% (1970s), which provides most of the noun-participle combinations up until that point. This share drops to 33.62% in the 1980s, while at the same time 60.17% of tokens of noun-participle combinations still come from the *Archiv der Gegenwart*. In the 1990s, 13.84% of tokens in the corpus come from the *Archiv der Gegenwart* providing 31.43% of tokens of noun-participle combinations in that decade. Thus, the share of a source that generally provides many noun-participle combinations declines in the composition of the corpus in the late 20th century.

The relative increase of tokens of noun-participle combinations plus the increasing corpus size over the decades lead to an unequal distribution of tokens along the time axis. There are only 31 records of noun-participle combinations from the 18th century. 2,286 combinations date from the 19th century. 12,616 tokens originate from the 20th century.

Absolute frequencies of specific types suggest that several noun-participle combinations are highly lexicalized. The most frequent types are *stellvertretend* ‘deputy’, lit. “filling in for a position” (2,587 tokens), *grundlegend* ‘basic’, lit. “ground-laying” (1,440 tokens) and *maßgebend* ‘decisive’, lit. “measure-giving” (1,083 tokens). Note that *stellvertretend* has idiosyncratic characteristics. It has a subtractive nominal form (*Stelle* > *stell*) and an isolated, lexicalized meaning differing from the underlying verb phrase (*die Stelle vertreten* ‘to fill in the position’). Nevertheless, even lexicalized noun-participle combinations as such are generally not completely demotivated (Pümpel-Mader et al., 1992: 256).



Among the 2,050 types are 1,482 hapax legomena, suggesting a high level of productivity of the noun-participle pattern (cf. Baayen, 1994, 2001: 203–205).<sup>23</sup> Among the 1,482 hapax legomena, there are 905 types of nouns and 521 types of participles. The most frequent participle occurring in hapax legomena is *suchend* ‘searching’ with 42 occurrences, followed by *erzeugend* ‘producing’ (29 occurrences) and *betreffend* ‘concerning’ (28 occurrences). Top nouns are *Leben* ‘life’, *Welt* ‘world’, and *Herz* ‘heart’, occurring 22, 21, and 18 times, respectively.

99.04% of noun-participle combinations contain base verbs that require the noun to be in the accusative case in the underlying verb phrase (14,790 instances, e.g., *erfolgversprechend* ‘promising’). Combinations with dative verbs (0.74%, 111 instances, e.g., *gottvertrauend* ‘trusting in God’) and genitive verbs (0.21%, 32 instances, e.g., *dienstenthebend* ‘relieving of duty’) are considerably less frequent. This result is consistent with the literature on noun-participle combinations, according to which accusative combinations are most frequent (Fuhrhop, 2007: 135; Wilss, 1983: 230; Müller & Müller, 1961: 72; Lohde, 2006: 167).

## 4.2 Spelling: concatenated spellings spread diachronically

Most of the evidence occurs in concatenated spelling (93.15%, 13,910 tokens). 6.7% of the noun-participle combinations (1,001 tokens) exhibit spaced spelling, while only 0.15% of the combinations (22 tokens) are hyphenated.<sup>24</sup> Figure 2 illustrates how the ratio of concatenated to spaced spelling has developed in diachrony. Concatenated spellings clearly predominated in most of the decades;<sup>25</sup> their share even increased over the course of the 19th and 20th centuries. 80.97% of the combinations were written as one graphemic unit between 1840 and 1849. This percentage rose and reached a peak of 96.93% in the 1960s. The ratio between the spelling variants has stabilized over time. Although noun-participle combinations increasingly tended to be written as a graphemic unit, combinations with spaced spellings have persisted.<sup>26</sup>

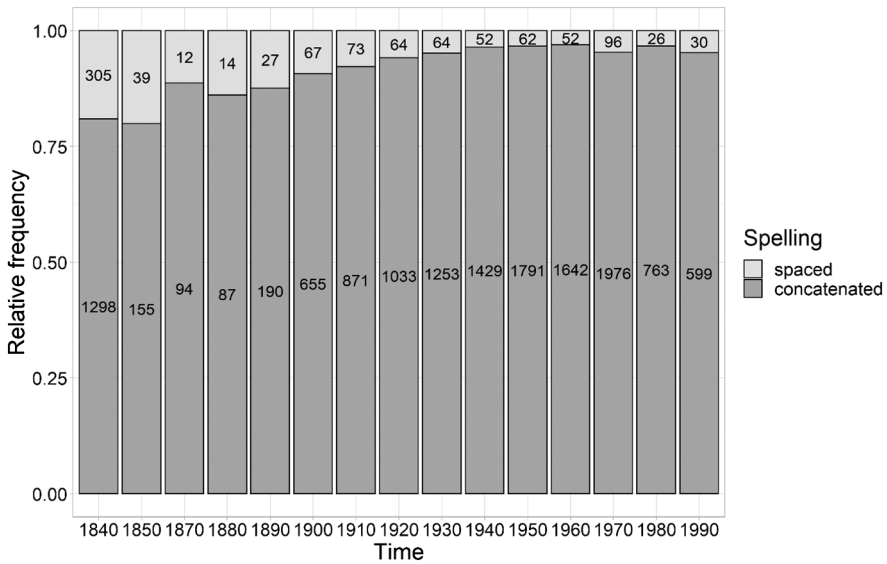
The development of types is illustrated in Fig. 3. Whenever a noun-participle combination displayed both spaced and concatenated spelling, it was treated as two types (e.g., *Öl fördernd* vs. *ölfördernd* ‘oil-producing’). As for tokens, there is a diachronic increase of concatenated spellings for types, although the proportions of noun-participle combinations written as a graphemic unit are generally lower than for tokens. The peak was reached between 1980 and 1989 with 89.96% of the types being written as a graphemic unit. Note that for the 1980s, the average number of

<sup>23</sup>Note that the number of hapax legomena largely depends on the corpus size (cf. Baayen, 2001). This is one of the many reasons why hapax legomena are not necessarily neologisms (Hilpert, 2018: 95).

<sup>24</sup>Since hyphenations are highly infrequent, they will not be taken into account in the following analyses.

<sup>25</sup>Due to the scarcity of the data, the time sections from 1710 to 1839 (93 records in total) are not included in the figure. Up to 1839, the data show more variability of spelling than in the subsequent decades, though concatenated spelling is generally favored.

<sup>26</sup>Note that spelling in Germany is determined by state regulations, which are also implemented in the extra-governmental sphere (Dürscheid, 2016: 166). However, orthographic regulation is unlikely to play a role in the data given that before 1996, the spelling of noun-participle combinations was not officially regulated or variation was tolerated, respectively. The prescriptive rule from 2004 banning concatenated spelling of single nouns and participles was removed in 2006 after having been heavily criticized (Dürscheid, 2016: 191, 201–202).

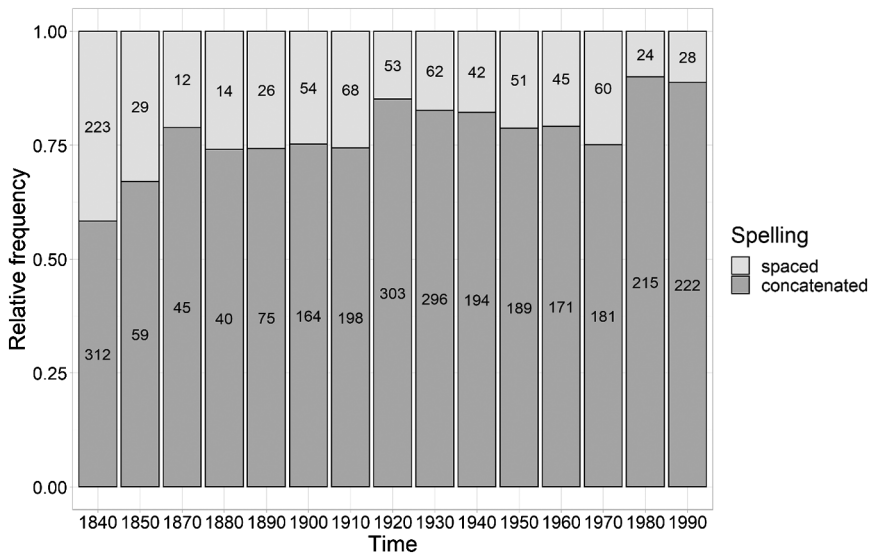


**Fig. 2** Diachronic development of the spelling of noun-participle combinations for tokens, including absolute numbers and relative frequencies ( $n = 14,819$ )

tokens per type is 3.55 (763:215) for concatenated spelling, but only 1.08 (26:24) for spaced spelling. Similar results can be found for the other decades. This is a first hint in the data that high token frequencies of individual types support concatenated spellings.

To find out which factors determine the choice between concatenated and spaced spelling, a binomial logistic regression was performed using the package *lme4* (Bates et al., 2022) in *R*.<sup>27</sup> Tokens that are hyphenated or that originate from the period before 1840 were excluded from this analysis due to scarcity of data. In the maximum model, the variables *DECADE* and *FREQUENCY* were considered as fixed effects. Since token frequencies were determined based on the entire period of investigation (cf. Sect. 3.2), no interaction between the two fixed effects was modeled here. To control for the influence of individual newspapers, the variable *NEWSPAPER* was set as a random effect. This was done to prevent any editorial guidelines of individual newspapers from distorting the results. Furthermore, the variable *LEMMA* was set as random effect in order to control for possible spelling preferences of individual types. According to the AIC values, the maximum model is the best-fit model. The ANOVA shows that simplified models result in a significant loss of information (comparison between the maximum model and a model without *FREQUENCY*:  $\chi^2 = 41.815$ ,  $df = 1$ ,  $\Pr(> \chi^2) = 1.004e - 10^{***}$ ; comparison between the maximum model and a model without *DECADE*:  $\chi^2 = 35.681$ ,  $df = 1$ ,  $\Pr(> \chi^2) = 2.325e - 09^{***}$ ). Looking only at the data from 1840 until 1999 and ignoring hyphenation, this analysis shows that *DECADE* proves to be a significant predictor for spelling. The FRE-

<sup>27</sup>Model formula for the maximum model in *R*: `glmer(Spelling ~ Decade + Frequency + (1 | Newspaper) + (1 | Lemma), data = df.1840.nohyph, nAGQ = 0, family = 'binomial')`.



**Fig. 3** Diachronic development of the spelling of noun-participle combinations for types, including absolute numbers and relative frequencies ( $n = 3,455$ )

**Table 2** Binomial logistic regression with all factors as predictors (final model,  $n = 14,819$ , marginal  $R^2 = 0.9094$ , conditional  $R^2 = 0.9719$ )

Fixed effects				
	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	-33.404863	6.452229	-5.177	2.25e-07***
Decade	0.018176	0.003383	5.373	7.75e-08***
Frequency	0.011933	0.002155	5.536	3.09e-08***
Random effects				
	Variance	Std. Dev.		
Lemma	8.0167	2.8314		
Newspaper	0.9289	0.9638		
Goodness of Fit				
	AIC	BIC	log likelihood	deviance
	4441.1	4479.1	-2215.5	4431.1

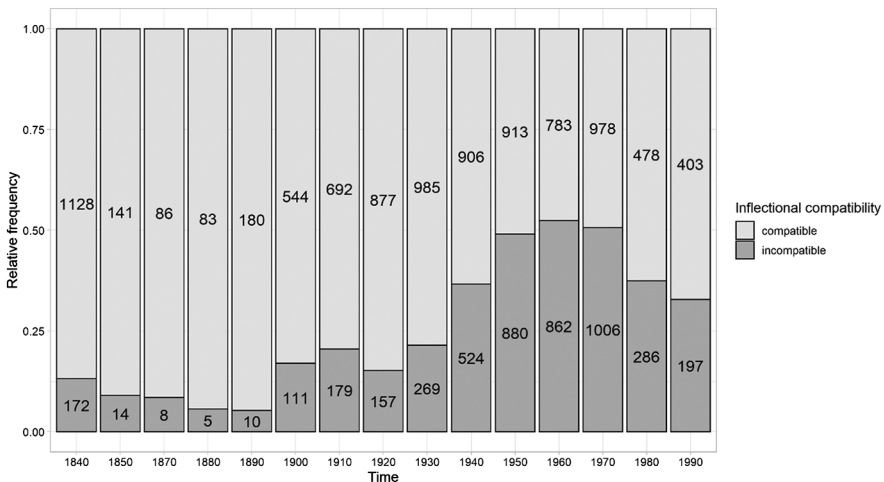
QUENCY of individual types also turns out to be a significant predictor for spelling. The estimates of the final model are given in Table 2.<sup>28</sup>

<sup>28</sup>For the binomial logistic regression, the characteristic value *concatenated spelling* was set to 1, *spaced spelling* was set to 0.

### 4.3 Nominal forms: root compounds as a model

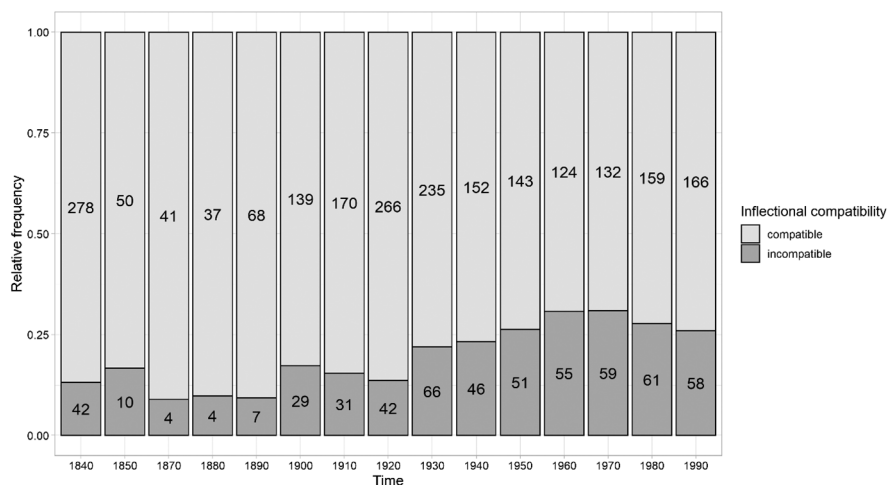
As is shown in Sect. 2.3, nominal forms indicate the degree of morphologization (Fuhrhop, 2000) since they can either correspond to nominal forms in verb phrases ([*leben*Ø]*bedrohend*, Berliner Tageblatt (evening issue), 03-01-1918, DWDS, cf. [*Leben*Ø] *bedrohen* ‘threaten life’) or to nominal forms in root compounds ([*lebens*]bejahend ‘life-affirming’, Berliner Tageblatt (morning issue), 03-02-1915, DWDS, cf. [*Lebens*]ziel ‘life goal’). Several approaches to classify nominal forms were introduced in Sect. 2.3. In the following, the results of the analyses will be presented. First, we will look at inflectional compatibility. Then we will turn to conventional and unconventional nominal forms (Fuhrhop, 2007), focusing on derivatives ending in *-ung* and nominalized infinitives as first constituents. Subsequently, the results on the source of analogical extension of nominal forms will be presented.

Figure 4 shows how the inflectional (in)compatibility of nominal forms has developed from 1840 up to present-day German.<sup>29</sup> The share of incompatible nominal forms has increased over time. In the 19th century, incompatible nominal forms were rather rare with a share of at most 13.23%. During the 20th century, they spread rapidly, reaching a peak of 52.4% in the 1960s. At the end of the 20th century, the share of incompatible nominal forms dropped to 37.43% in the 1980s and to 32.83% in the 1990s. This sudden decrease in frequency of incompatible nominal forms is likely due to changing token frequencies of individual types. For example, the frequency of the type *stellvertretend* (incompatible nominal form) dropped from 36.25% of all tokens in the 1970s to 8.57% in the 1990s. A diachronic increase of incompatible nominal forms is also evident when types are considered (cf. Fig. 5).



**Fig. 4** Diachronic development of the inflectional compatibility of nominal forms (tokens), including absolute numbers and relative frequencies ( $n = 13,857$ )

<sup>29</sup>Before 1840, too, inflectionally compatible nominal forms clearly prevailed in most decades.



**Fig. 5** Diachronic development of the inflectional compatibility of nominal forms (types), including absolute numbers and relative frequencies ( $n = 2,725$ )

To find out which factors determine inflectional (in)compatibility, a binomial logistic regression was performed.<sup>30</sup> In the maximum model, the variables DECADE, FREQUENCY, and NEWSPAPER were included as in the spelling analysis. Furthermore, the variable NOUN was specified as a random effect. This was done to control for individual first constituents that prefer a particular nominal form or that do not allow for variation between nominal forms.

An ANOVA shows that FREQUENCY is not a significant predictor for inflectional (in)compatibility and can be neglected in the final model. The model without FREQUENCY emerged as the model with the lowest AIC value. DECADE proves to be a significant predictor for the choice of nominal forms. The ANOVA shows that neglecting DECADE results in a significant loss of information (comparison between the maximum model and a model without DECADE:  $\chi^2 = 11.492$ ,  $df = 1$ ,  $\text{Pr}( > \chi^2 ) = 0.000699^{***}$ ). The estimates of the final model are given in Table 3.<sup>31</sup>

As discussed above, investigating inflectional compatibility sheds light on the extent to which nominal forms can be explained by analogy to verb phrases. From a diachronic perspective, this approach is the most conservative one because the compatible category by definition includes cases that are ambiguous between verb phrases and root compounds—while the incompatible category is on the rise diachronically. To gain deeper insight into cases that are ambiguous between verb phrases and root compounds, it is useful to additionally consider the other two approaches introduced in Sect. 2. In the following, the results of these analyses are presented.

Fuhrhop's (2007) approach determines the extent to which nominal forms can be explained by analogy to root compounds. For root compounds with derivatives

<sup>30</sup>Model formula for the maximum model in R: `glmer(Compatibility ~ Decade + Frequency + (1 | Newspaper) + (1 | Noun), data = df.1840.nosep, nAGQ = 0, family = 'binomial')`.

<sup>31</sup>For the binomial logistic regression, the characteristic value *inflectionally incompatible* was set to 1, *inflectionally compatible* was set to 0.

**Table 3** Binomial logistic regression with DECADE as predictor (final model,  $n = 13,857$ , marginal  $R^2 = 0.0006$ , conditional  $R^2 = 0.1354$ )

	Fixed effects			
	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	-21.550092	5.746247	-3.750	0.000177***
Decade	0.009464	0.00299	3.165	0.001549**
	Random effects			
	Variance	Std. Dev.		
Noun	33.3251	5.7728		
Newspaper	0.3222	0.5676		
Goodness of Fit				
	AIC	BIC	log likelihood	deviance
	2626.7	2656.9	-1309.4	2618.7

**Table 4** Diachronic development of linking elements after derivatives ending in *-ung* (tokens), including absolute numbers and relative frequencies ( $n = 471$ )

Decade	1840	1850	1860	1870	1880	1890	1900	1910
<i>-ung+Ø</i>	45 (0.67)	5 (0.71)	-	4 (1)	1 (0.5)	0	9 (0.56)	9 (0.6)
<i>-ung+s</i>	22 (0.33)	2 (0.29)	-	0 (0)	1 (0.5)	0	7 (0.44)	6 (0.4)
Decade	1920	1930	1940	1950	1960	1970	1980	1990
<i>-ung+Ø</i>	16 (0.46)	35 (0.67)	44 (0.71)	47 (0.55)	25 (0.69)	23 (0.48)	15 (0.71)	7 (0.35)
<i>-ung+s</i>	19 (0.54)	17 (0.33)	18 (0.29)	39 (0.45)	11 (0.31)	25 (0.52)	6 (0.29)	13 (0.65)

ending in *-ung* and nominalized infinitives as first constituents, *-s-* is the conventional linking element throughout the period of investigation. Therefore, noun-participle combinations with corresponding first constituents are a good test case. The data suggest that noun-participle combinations do not show the same clear preference for the linking *-s-*. As can be seen in Tables 4 and 5, the spread of the linking *-s-* in noun-participle combinations is clearly lagging behind.

471 noun-participle combinations exhibiting concatenated spelling or hyphenation and having a derivative first constituent ending in *-ung* were identified. Of these, 285 are unlinked (60.51%, [*verfassungØ*]gebend) and 186 take a linking *-s-* (39.49%, [*verfassung*s]ändernd).<sup>32</sup> Taken together, the data show that there is great variation between the zero morpheme and linking *-s-* during the entire period of investigation. Unlike root compounds (Kopf, 2018a), there seems to be no clear trend towards the (conventional) linking *-s-* for noun-participle combinations.

Similar results can be found for nominalized infinitives. The study retrieved 252 noun-participle combinations with concatenated spellings or hyphenation that have a nominalized infinitive as first constituent. 151 of these take a zero morpheme

<sup>32</sup>This difference is statistically significant ( $\chi^2 = 20.81$ ,  $df = 1$ ,  $p < 0.001$ ).

**Table 5** Diachronic development of linking elements after nominalized infinitives (tokens), including absolute numbers and relative frequencies ( $n = 252$ )

Decade	1840	1850	1860	1870	1880	1890	1900	1910
infinitive+∅	21 (1)	4 (1)	-	1 (1)	1 (1)	2 (1)	12 (1)	13 (0.72)
infinitive+s	0 (0)	0 (0)	-	0 (0)	0 (0)	0 (0)	0 (0)	5 (0.28)
infinitive+subtraction	0 (0)	0 (0)	-	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Decade	1920	1930	1940	1950	1960	1970	1980	1990
infinitive+∅	30 (0.94)	19 (0.79)	7 (0.64)	9 (0.9)	8 (0.67)	9 (0.21)	7 (0.2)	8 (0.3)
infinitive+s	2 (0.06)	5 (0.21)	4 (0.36)	0 (0)	4 (0.33)	33 (0.79)	28 (0.8)	19 (0.7)
infinitive+subtraction	0 (0)	0 (0)	0 (0)	1 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)

(59.92%), 100 tokens occur with a linking *-s-* (39.68%) and one exhibits a subtractive nominal form<sup>33</sup> (0.4%, *Rennen + entscheidend > renn\_entscheidend* ‘race-determining’).<sup>34</sup> For the nominalized infinitives there is a clear and rapid development towards the conventional linking *-s-*, which begins in the early 20th century. In the 1980s, linking *-s-* already occurs with a relative frequency of 0.8. However, it had not yet completely spread and at that time still competed with the zero morpheme.

As is shown in Sect. 2.3, the analogical source of nominal forms can be ambiguous. The first two approaches presented here are more conservative and rely on binary categorizations (compatible vs. incompatible; conventional vs. unconventional), whereas the approach of analogical sources reflects ambiguous cases in considering three categories: verb phrases as analogical source, root compounds as analogical source, and ambiguity of the analogical source. In order to study the source of analogical extension of nominal forms diachronically, the period from 1840 to 1999 was divided into sections of twenty years. From each section, 50 random records were selected, so that a total of 400 noun-participle combinations were analyzed. For 249 noun-participle combinations the analogical source was ambiguous (62.25%, e.g., [*holz*∅]*verarbeitend* ‘woodworking’). 92 combinations exhibit nominal forms motivated by root compounds (23%, e.g., [*staats*]*erhaltend* ‘state-preserving’) and 59 nominal forms are motivated by verb phrases (14.75%, e.g., [*vertrag*∅]*schließend* ‘contracting’, lit. ‘contract-concluding’). The diachronic distribution is illustrated in Fig. 6.

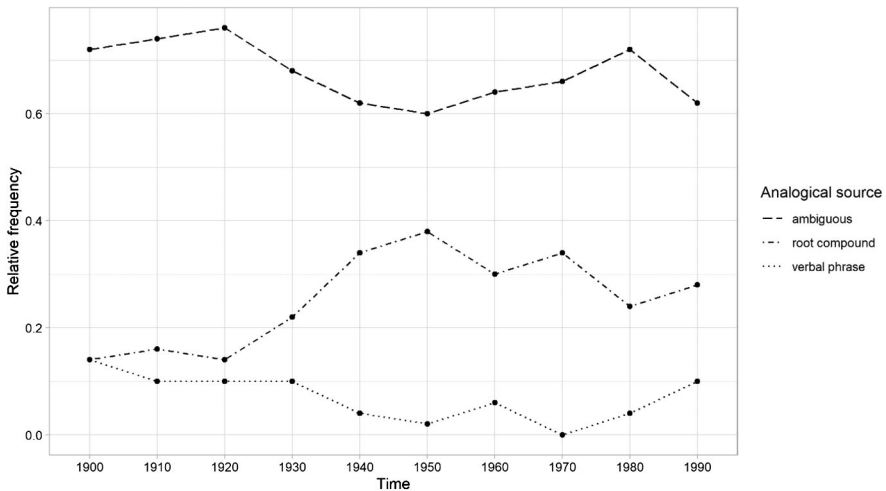
It is particularly striking how quickly nominal forms motivated by root compounds spread over the course of the 20th century. In the second half of the 19th century, only two or three out of fifty nominal forms were clearly motivated by root compounds. Between 1960 and 1979, 28 of 50 nominal forms were motivated by root compounds. The change over time is only significant for nominal forms motivated by root compounds (Kendall’s tau:  $\tau = 0.691$ ,  $z = 2.369$ ,  $p = 0.018$ ; cf. Hilpert & Gries, 2009;

<sup>33</sup>The subtractive linking element is most likely motivated by the model of root compounds as well. Though the nominalized infinitive *Rennen* ‘race’ is a rather unusual first constituent for root compounds (cf. Kopf, 2018a: 266), the corresponding verbal stem does occur as a first constituent in root compounds and apparently served as a model here (e.g., *Renn\_stall* ‘racing stable’). Furthermore, avoidance of hapology could play a role here due to the repeating *en*: *Rennen*, *entscheidend*.

<sup>34</sup>These differences are statistically significant ( $\chi^2 = 138.5$ ,  $df = 2$ ,  $p < 0.001$ ).



**Fig. 6** Diachronic development of the source of analogical extensions of nominal forms, relative frequencies (tokens,  $n = 400$ )



**Fig. 7** Diachronic development of the sources of analogical extension of nominal forms in hapax legomena, relative frequencies ( $n = 500$ )

Kopf, 2018a: 244–246 for the methodology). However, the picture is distorted by some highly frequent types.

To balance potential frequency effects which may have arisen because of the presence of some highly frequent types in Fig. 6, 50 hapax legomena per decade between 1900 and 1999<sup>35</sup> were randomly chosen, making for a total of 500 tokens. The results are illustrated in Fig. 7. According to a Kendall’s tau test, the diachronic develop-

<sup>35</sup>Only in the 20th century is there a sufficiently large number of hapax legomena.



ment is not significant for any of the three groups. Nominal forms with an ambiguous source of analogical extension were predominant in all decades (relative frequencies between 0.6 and 0.76), followed by nominal forms motivated by root compounds (relative frequencies between 0.14 and 0.38) and nominal forms motivated by verb phrases (relative frequencies between 0 and 0.14).

## 5 Discussion and conclusion

This study was the first to examine the diachronic change of noun-participle combinations in German on a large-scale empirical basis. In particular, this study has investigated how the combinations have evolved between syntax and word formation over the past 300 years. Based on a diachronic newspaper corpus, two hypotheses were tested: (1) Noun-participle combinations diachronically undergo morphologization and (2) the morphologization of noun-participle combinations is supported by lexicalization. Spelling and nominal forms were used as central features to determine the degree of morphologization; token frequencies were used as an indicator of lexicalization. Both hypotheses could be essentially confirmed: as time progresses, concatenated spellings and nominal forms that are distinct from corresponding verb phrases are more likely to occur (*Aufsehen erregend* > *aufsehenerregend* ‘sensational’; [*leben*Ø]*bedrohend* > [*Lebens*]*bedrohend* ‘life-threatening’ – \**[Lebens]bedrohen* ‘to threaten (a) life’). As expected, the proportion of concatenated spellings is rising with an increasing degree of lexicalization of the individual types. However, lexicalization is not a significant predictor for the shape of nominal forms. Contrary to expectations, the probability of an inflectionally incompatible nominal form to occur in a noun-participle combination does not increase the more this combination is lexicalized.

Note that the observed spread of concatenated spellings between 1700 and 1999 is not a general trend for compounds. For nominal root compounds, concatenated spelling (or hyphenation) became the norm much earlier (cf. Solling, 2012: 103–121; Kopf, 2018a: 342–354). The same largely applies to nominal forms of root compounds, which have mostly been established by the beginning of the New High German period (cf. Kopf, 2018a: 233–287; Kopf, 2018b).

Overall, the results suggest that noun-participle combinations undergo morphologization (cf. Fuhrhop, 2000): they take on features that are typical of words (concatenated spelling, linking elements) and abandon characteristics that are typical of phrases (spaced spelling, inflected nominal constituents). In terms of spelling, this process is supported by lexicalization. To date, the morphologization of noun-participle combinations has not been completed since there is still variation, as for example in [*achtung*Ø]*gebietend*/*achtungs**gebietend* ‘awe-inspiring’. Additionally, the spread of concatenated spellings seems to have come to a standstill. It could be shown that the share of spaced spellings has stabilized at around 3.07% (1960s) to 4.77% (1990s) in the second half of the 20th century. This is due to the strong syntactic linkage of noun-participle combinations. Their correspondence to more complex participle phrases (e.g., *den frisch gemähten Rasen düngend* ‘fertilizing the freshly mowed lawn’, lit. “the-ACC freshly mowed-ACC lawn-ACC fertilizing”) keeps a subset of the combinations tied to the syntax. By analogy with more complex participle

phrases (cf. Sect. 2.2), they can still be analyzed as phrases themselves, which is expressed by means of spaced spelling. This prevents concatenated spellings from becoming fully established, even though there is a clear trend toward morphologization.

The results are essentially consistent with Fuhrhop (2007), who expects noun-participle combinations to become more morphological over time. The results also fit into the crosslinguistic trend that multi-morphemic units with two or more constituents carrying lexical meaning take on features diachronically that are typical of words or occur (almost) exclusively in words. Here, too, this development is supported by lexicalization, as has been shown by Plag et al. (2008), Kuperman and Bertram (2013), and Sanchez-Stockhammer (2018) for English compounds, for example.

This raises the question of why lexicalization is not a significant predictor for inflectional compatibility of nominal forms. In the statistical analysis, no significant effect of FREQUENCY on COMPATIBILITY was detected. Contrary to expectations, the maximum model (cf. Table 6, Appendix) even yields a negative correlation between the two variables (which cannot be generalized beyond the sample). That is, in the data, the share of incompatible nominal forms decreases with increasing token frequency. This result runs counter to Fuhrhop (2000: 211) and to Kopf (2018a: 183) who state that conventional linking elements increasingly occur with increasing lexicalization.<sup>36</sup> Note that the statistical analysis was performed on the comparison to verb phrases (inflectional compatibility), whereas Fuhrhop refers to the comparison to root compounds. However, the different approaches do not sufficiently account for the inconsistency between Fuhrhop (2000) and the present data. Reconsidering first constituents suffixed by *-ung*, we still see that the share of the conventional linking *-s-* decreases with increasing token frequency. Thus, the most frequent noun-participle combination containing an *-ung*-derivative is [*verfassung*∅]*gebend*/[*verfassung*s]*gebend* ‘constitutional’, lit. “constitution-giving”, with 255 occurrences, of which 194 (76.08%) are unlinked and 61 (23.92%) contain a linking *-s-*. In the case of the 45 hapax legomena containing a first constituent suffixed by *-ung*, these ratios are reversed: 10 of them (relative frequency: 0.22) are unlinked, while 35 (relative frequency: 0.78) contain a linking *-s-*. Similar results can be found for noun-participle combinations with nominalized infinitives as first constituents. Note that these statements rely on small samples. However, Nübling and Szczepaniak’s (2011) Google search suggests similar conclusions. Their data suggest that highly frequent synthetic compounds are more likely to preserve the unlinked nominal form corresponding to verb phrases (e.g., [*Stellung*∅]*nahme* ‘statement’ – [*Stellung*∅] *nehmen* ‘to comment’, Nübling & Szczepaniak, 2011: 60).<sup>37</sup> This indicates that even if we consider the comparison to root compounds in the sense of Fuhrhop, there seems to be a negative correlation between conventional linking el-

<sup>36</sup>Note that Fuhrhop’s (2000) statement is not obviously based on empirical data. Kopf’s (2018a: 183) statement is based on a corpus query for only two types (*achtung(s)gebietend*, *richtung(s)weisend*).

<sup>37</sup>Nübling and Szczepaniak (2011) consider three text types: books, forums, and blogs. Although the share of unlinked nouns differs between the text types, there is a general tendency that highly frequent types remain unlinked. Due to this, it seems unlikely that the text type newspaper could play a role in this respect.

ements and token frequency. A possible explanation for this is that high token frequency blocks regularization (cf. Bybee, 2006). Since the vast majority of compounds are root compounds (cf. Ortner et al., 1991: 112; Gaeta & Zeldes, 2012: 203–204), we can view the nominal forms of root compounds (e.g., [*Erfolgs*]geschichte ‘success story’) as the benchmark for other compounding patterns. Therefore, the adoption of their nominal forms is a case of regularization or analogical reformation (e.g., [*erfolg*Ø]versprechend > [*erfolgs*]versprechend ‘promising’). Bybee (2006) argues that frequency strengthens the cognitive representation of words or phrases, making them more likely to be accessed as a whole and hence more resistant to regularization. This study’s results support this assumption and furthermore suggest that the conserving effect is graded. The higher their token frequencies, the more likely it is that individual types of noun-participle combinations block the alignment with root compounds. Future research should investigate this trend using larger data sets. Furthermore, conservatism of scientific jargon can delay analogical developments (cf. Kopf, 2018a: 183, fn. 163), as is the case for [*verfassung*Ø]gebend ‘constitutional’, for example. Apparently, the conserving effect only affects nominal forms but not the level of spelling since spaced spelling is not conserved for highly frequent types. A reason for this could be that noun-participle combinations which are accessed as a whole are more likely to be perceived as lexical units. This could in turn give rise to concatenated spelling. In this respect, spelling may reflect the strength of cognitive representation.

Following from that, the question arises as to why hyphenations are exceedingly rare in the data. Hyphens would be an appropriate graphemic device to express categorial ambivalence of noun-participle combinations. In the data, however, they account for only 0.15% of tokens. The results of the present study reflect a crosslinguistic trend in which graphemic integration does not generally proceed from spaced spelling via hyphenation to concatenated spelling. When linked root compounds evolved from genitive constructions in Early New High German (1350–1650), they directly changed from spaced to concatenated spelling (Kopf, 2018a: 342–343). Only after that, a “century of hyphenation” (Kopf, 2017: 177) occurred from 1650 to 1750, with more than half of the root compounds being hyphenated (Kopf, 2018a: 342–343; Solling, 2012: 103–125). Similar results have been found for spelling change in English compounds, which is generally not mediated by hyphenation (Kuperman & Bertram, 2013: 945–947; Sanchez-Stockhammer, 2018: 219–225). With regard to noun-participle combinations, hyphenation is used to highlight categorically striking nouns. 9 of 22 hyphens in the sample occur after nouns denoting languages, as in *spanisch-sprechend* (Der Spiegel, 08.11.1982, DWDS) ‘Spanish-speaking’. In these cases, the nouns are converted from adjectives. The hyphen is used here to indicate the categorial markedness of the first constituents.<sup>38</sup>

Let us now turn to the two key questions posed in Sect. 2.1, that is, how words and phrases can be distinguished from each other, and whether or not word formation and syntax form distinct linguistic domains. The present study has shown that there

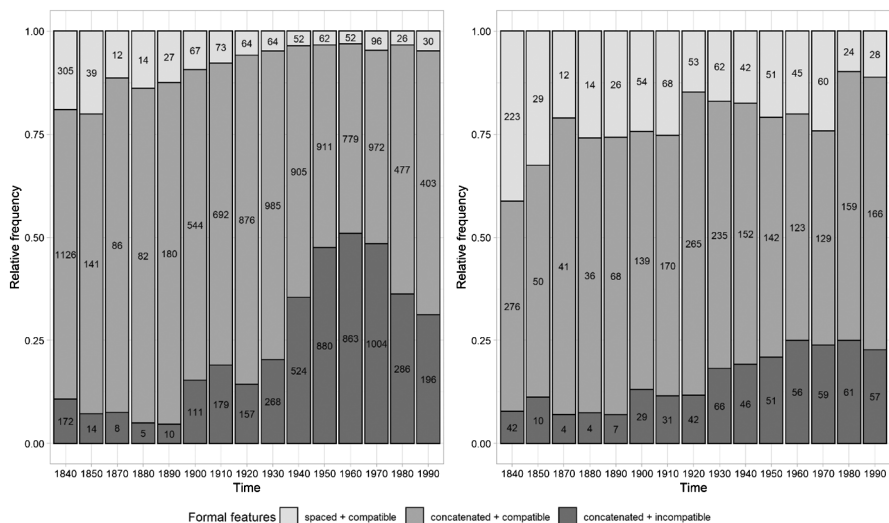
<sup>38</sup>Note that in New High German, nominalization is usually marked by capitalization (*spanisch* > *Spanisch* ‘Spanish’). Because concatenated noun-participle combinations are adjectives, they are normally written in lower case. Thus, the nominalization of the first constituents cannot be expressed by capitalization. Therefore, hyphens are used to express the categorial change (*spanisch-sprechend* ‘Spanish speaking’).

is a general trend toward morphologization of noun-participle combinations. However, frequency effects work against this process in at least two ways. On the one hand, infrequent types are more likely to retain spaced spelling. On the other hand, data from Nübling and Szczepaniak (2011) suggest that highly frequent types of synthetic compounds are more likely to retain nominal forms corresponding to inflected nouns in verb phrases. This study supports this tendency, although there is a lack of statistical significance. As a result of the interplay of morphologization and lexicalization, noun-participle combinations have become far more structurally diverse over the past 300 years. Some combinations have moved close to the morphological pole (cf. [*vertrauens*]*erweckend*, Archiv der Gegenwart, 2001 [1974], DWDS ‘instilling confidence’), whilst others have stayed close to the syntax (cf. *Abschied nehmend* ‘farewell’), and many combinations exhibit ambiguously motivated nominal forms and are therefore in between (cf. [*holz*Ø]*liefernd*, Vossische Zeitung (morning issue), 03-05-1903, DWDS ‘wood-supplying’). Still other combinations combine concatenated spelling and nominal forms corresponding to verb phrases, as for example [*krieg*Ø]*führend* ‘warring’ (cf. [*Krieg*Ø]*führen* ‘to wage war’ vs. [*Kriegs*]*recht* ‘martial law’). Moreover, single types occur in different shapes (cf. *Achtung gebietend* vs. [*achtung*Ø]*gebietend* vs. [*achtungs*]*gebietend*).

What does this imply for the cognitive representation of language? A modular demarcation of word formation and syntax, as proposed by Ackema and Neeleman (2004) and other works from the generative framework, cannot account for the different degrees of morphologization exhibited by individual types and the noun-participle pattern as a whole (cf. Schlücker, 2020: 66–67 on Giegerich’s 2015 proposal). If there was a modular organization of language, we would expect noun-participle combinations to exhibit morphological indicators on all linguistic levels as soon as they cross the boundary between words and phrases. This in turn means that the development of spelling and nominal forms should correlate. However, the present study has shown that this is not the case (also cf. Kopf, 2018b: 111). This can be seen in Fig. 8, which combines the results for the diachrony of spelling and compatibility of nominal forms for tokens and types. Concatenated spelling already dominated in the 18th century, and its share even increased over the 19th century (cf. the dark and medium shaded bars in Fig. 8, as compared to the light shaded bars). Contrary to that, nominal forms that are distinct in form from nominal forms in corresponding verb phrases were rare during the 18th and 19th centuries and mainly spread over the course of the 20th century (cf. the dark shaded bars in Fig. 8). This result makes a modular demarcation of word formation and syntax implausible.

Due to morphologization on different linguistic levels and due to the great structural variation still evident today, it is reasonable to assume a multi-dimensional continuum (Aikhenvald, 2002: 43) between word formation and syntax, reflected in spelling and nominal forms. Additional linguistic levels that provide information about this continuum, but were not considered here, include intonation and inflectional behavior (e.g., gradability). The poles of this continuum are instances that fulfill the prototypical features of the respective domains.

Following from this, another question we are left to consider is that of whether a delimitation of ‘word’ is possible after all. Even Booij (2010, 2012), who is a prominent representative of the idea that there is no sharp boundary between word



**Fig. 8** Structural change of noun-participle combinations (spelling + inflectional compatibility) for tokens (left,  $n = 14,819$ ) and types (right,  $n = 3,507$ ), absolute numbers and relative frequencies

formation and syntax, assumes so. He defines words based on the principle of Lexical Integrity (e.g., Booij, 2012: 188) as formulated by Anderson (1992: 84). This principle implies that syntactic rules cannot operate on the constituents of complex words. However, noun-participle combinations show that this approach is problematic. Highly lexicalized combinations like *verfassunggebend* ‘constitutional’, lit. “constitution-giving”, with nominal forms corresponding to verb phrases would be syntactic according to the principle of Lexical Integrity. However, the fact that these combinations exhibit word-typical features like concatenated spelling and (potentially) conventionalized meaning is completely disregarded in this view. For further criticism of the hypothesis of Lexical Integrity, see Haspelmath (2011: 67–69) and references therein.

Booij’s approach shows that taking individual linguistic levels as decisive thresholds for word status neglects the fact that there can still be more or less wordhood on the other levels, as has been confirmed by the present study. Therefore, it has to be assumed that the word is a non-discrete and multi-level unit. Unlike Booij (2010; 2012), the present study proposes that there is fluidity between words and phrases (e.g., Bloomfield, 1984: 179–181; Lyons, 1968: 204; Schlücker, 2020: 67; Fuhrhop, 2007: 7). Nevertheless, it is necessary to clarify whether there is a hierarchy between the levels of ‘word’. Crosslinguistically and diachronically, this is a promising direction for future research. With regard to German, a hierarchy between spelling and nominal forms has evolved. The occurrence of linking elements requires concatenated spelling or at least hyphenation (Dudenredaktion, 2020: 57), which is also evident in the data presented here, as there is no noun-participle combination in the corpora that exhibits both an inflectionally incompatible nominal form and spaced spelling. The reverse is not the case: noun-participle combinations that are written as one graphemic unit can contain clearly inflected nominal forms (e.g., [*arbeitsplätze*]*schaffend* ‘creating

jobs'). From this, we can conclude that the level of nominal forms is more crucial for wordhood than the level of spelling. That said, it is not the defining feature of words as argued above.

Assuming that there is a continuum between word formation and syntax, it would not make sense to classify multi-morphemic lexical units as either words or phrases, as also argued by Schlücker (2020) and Bauer (2019), for example. It would not be possible to draw a line at one or more linguistic levels without violating other levels as there are many patterns with mixed and/or ambiguous features (e.g., Schlücker, 2020: 60–61; Bloomfield, 1984: 181). Nonetheless, it is possible to determine the respective degree of morphologization or wordhood based on these morphological or phrasal features. The present study has exemplified this approach for German noun-participle combinations. Applying this approach to other ambivalent patterns in other languages as well, both from a diachronic and a synchronic perspective, is promising for future research.

A final point to be addressed is the diachronic development of the continuum between word formation and syntax. The rich inventory of morphological indicators in German is a result of language change. Historically, morphological indicators such as concatenated spellings and, more recently, linking elements are linguistic innovations (cf. Solling, 2012; Kopf, 2018a). In New High German they are well-established features, while they were less common in Old High German, for instance. Given this emergence of morphological indicators, the question arises whether the postulated continuum between words and phrases is diachronically dissolving. Categorical indicators could increasingly disambiguate structures that lie between words and phrases so that the intermediate area between the poles disappears. However, it is to be expected that there should be no dismantling of the continuum. Since the inventory of categorial indicators in German yields patterns combining morphological and phrasal features (cf. Schlücker, 2020), the indicators should consolidate the continuum instead of dismantling it. This is supported by the present study, which has shown that morphologization affects noun-participle combinations at different linguistic levels and to varying degrees, and interacts with lexicalization in complex ways. As a result, noun-participle combinations have become far more structurally diverse. Moreover, the respective indicators should strengthen the morphological and the phrasal pole. Consider that in present-day German root compounds, morphological indicators on different linguistic levels are largely interdependent. As was already mentioned, linking elements and concatenated spellings usually co-occur. This is a hint that the morphological pole has become more strongly contoured. This study supports this assumption since there is a diachronic spread of noun-participle combinations that exhibit both concatenated spellings and incompatible nominal forms (cf. Fig. 8). However, for a more empirically sound investigation of this question, a broader data sample is needed, involving greater diachronic depth and a larger variety of indicators studied. This would be a promising direction for future research looking not only to focus on the recent history of German, but also to take into account its earlier stages and to compare it to other languages.

## Appendix: binomial logistic regression

**Table 6** Binomial logistic regression with DECADE and FREQUENCY as predictors for INFLECTIONAL COMPATIBILITY as the dependent variable (maximum model,  $n = 13,857$ , marginal  $R^2 = 0.0006$ , conditional  $R^2 = 0.1354$ ); tokens that exhibit spaced spelling or that originate from the period before 1840 were excluded from this analysis

	Fixed effects			
	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.100e+01	5.774e+00	-3.638	0.000275 ***
Decade	9.192e-03	3.004e-03	3.060	0.002214 **
Frequency	-1.197e-03	8.091e-04	-1.480	0.138889
	Random effects			
	Variance	Std. Dev.		
Noun	33.4696	5.7853		
Newspaper	0.3259	0.5709		
	Goodness of Fit			
	AIC	BIC	log likelihood	deviance
	2629.1	2666.7	-1309.5	2619.1

**Acknowledgements** I would like to dedicate this paper to Melitta Gillmann, who has encouraged my interest in linguistics from the beginning. Thank you, Melitta, for your helpful and patient nature as well as for your support and advice, especially while I wrote this paper. Furthermore, I thank the participants of the corpus linguistics colloquium at HU Berlin for discussions as well as Hagen Hirschmann, Julia Lukassek, and two anonymous reviewers for helpful comments on this article.

**Funding** Open Access funding enabled and organized by Projekt DEAL. No funding was received for conducting this study.

## Declarations

**Competing Interests** The author has no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ackema, P., & Neeleman, A. (2004). *Beyond morphology. Interface conditions on word formation*. Oxford: Oxford University Press.
- Aikhenvald, A. Y. (2002). Typological parameters for the study of clitics, with special reference to Tarina. In R. M. W. Dixon, & A. Y. Aikhenvald (Eds.), *Word. A cross-linguistic typology* (pp. 42–78). Cambridge: Cambridge University Press.

- Anderson, S. R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.
- Augst, G. (1975). *Untersuchungen zum Morpheminventar der deutschen Gegenwartssprache*. Tübingen: Narr.
- Baayen, R. H. (1994). Derivational productivity and text typology. *Journal of Quantitative Linguistics*, 1, 16–34.
- Baayen, R. H. (2001). *Word frequency distributions*. Dordrecht: Kluwer.
- Bartón, K. (2022). MuMIn. Multi-model inference. R-package, version 1.46.0.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2022). lme4. Linear mixed-effects models using 'eigen' and S4. R-package, version 1.1–28.
- Bauer, L. (1998). When is a sequence of two nouns a compound in English? *English Language and Linguistics*, 2(1), 65–86.
- Bauer, L. (2019). Compounds and multi-word expressions in English. In B. Schlücker (Ed.), *Complex lexical units. Compounds and multi-word expressions* (pp. 45–68). Berlin: de Gruyter.
- BBAW (online) = Deutsches Textarchiv. Die Textauswahl. Berlin-Brandenburgische Akademie der Wissenschaften. <http://www.deutschestextarchiv.de/doku/textauswahl>. Accessed May 20, 2020.
- Bergmann, R. (1980). Verregnete Feriengefahr und Deutsche Sprachwissenschaft. Zum Verhältnis von Substantivkompositum und Adjektivattribut. *Sprachwissenschaft*, 5, 234–265.
- Digitales Wörterbuch der deutschen Sprache. Berlin-Brandenburgische Akademie der Wissenschaften. Core corpus. <https://www.dwds.de/d/k-referenz#kern>. Accessed December 2019.
- Blatz, F. (1900). *Neuhochdeutsche Grammatik mit Berücksichtigung der historischen Entwicklung der Deutschen Sprache*. Karlsruhe: Lang.
- Bloomfield, L. (1984). *Language*. Chicago: University of Chicago Press. Original work published 1933.
- Booij, G. (2002a). Separable complex verbs in Dutch. A case of periphrastic word formation. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Verb-particle explorations* (pp. 21–42). Berlin: de Gruyter.
- Booij, G. (2002b). Constructional idioms, morphology, and the Dutch lexicon. *Journal of Germanic Linguistics*, 14(4), 301–329.
- Booij, G. (2010). *Construction morphology*. New York: Oxford University Press.
- Booij, G. (2012). *The grammar of words. An introduction to linguistic morphology*. Oxford: Oxford University Press.
- Booij, G., & Audring, J. (2017). Construction morphology and the parallel architecture of grammar. *Cognitive Science*, 41(2), 277–302.
- Bredel, U. (2006). Die Herausbildung des syntaktischen Prinzips in der Historiogenese und in der Ontogenese der Schrift. In U. Bredel, & H. Günther (Eds.), *Orthographietheorie und Rechtschreibunterricht* (pp. 139–163). Tübingen: Niemeyer.
- Bybee, J. (2006). From usage to grammar. The mind's response to repetition. *Language*, 82(4), 711–733.
- Charitonova, I. J. (1977). Zur Frage von Zentrum und Peripherie einer Wortart im Deutschen. In G. Helbig (Ed.), *Beiträge zur Klassifizierung der Wortarten* (pp. 28–38). Leipzig: VEB Verlag Enzyklopädie Leipzig.
- Dudenredaktion (Ed.) (2020). *Duden: Vol. 1. Die deutsche Rechtschreibung*. Berlin: Dudenverlag.
- Dürscheid, C. (2016). *Einführung in die Schriftlinguistik*. Göttingen: Vandenhoeck & Ruprecht.
- Eisenberg, P. (2007). Sprachliches Wissen im Wörterbuch der Zweifelsfälle. Über die Rekonstruktion einer Gebrauchsnorm. *Aptum. Zeitschrift für Sprachkritik und Sprachkultur*, 3, 209–228.
- Fleischer, W., & Barz, I. (2012). *Wortbildung der deutschen Gegenwartssprache*. Berlin: de Gruyter.
- Fuhrhop, N. (1996). Fugenelemente. In E. Lang, & G. Zifonun (Eds.), *Deutsch – typologisch* (pp. 525–550). Berlin: de Gruyter.
- Fuhrhop, N. (1998). *Grenzfälle morphologischer Einheiten*. Tübingen: Stauffenburg.
- Fuhrhop, N. (2000). Zeigen Fugenelemente die Morphologisierung von Komposita an? In R. Thieroff, M. Tamrat, N. Fuhrhop, & O. Teuber (Eds.), *Deutsche Grammatik in Theorie und Praxis* (pp. 201–213). Tübingen: Niemeyer.
- Fuhrhop, N. (2007). *Zwischen Wort und Syntagma. Zur grammatischen Fundierung der Getrennt- und Zusammenschreibung*. Tübingen: Niemeyer.
- Fuhrhop, N., & Teuber, O. (2000). Das Partizip 1 als adjektivischer Infinitiv. In A. Bittner, D. Bittner, & K.-M. Köpcke (Eds.), *Angemessene Strukturen. Systemorganisation in Phonologie, Morphologie und Syntax* (pp. 173–190). Hildesheim, Zürich, New York: Olms.
- Gaeta, L., & Zeldes, A. (2012). Deutsche Komposita zwischen Syntax und Morphologie. Ein korpusbasierter Ansatz. In L. Gaeta, & B. Schlücker (Eds.), *Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte* (pp. 197–217). Berlin, Boston: de Gruyter.



- Gaeta, L., & Zeldes, A. (2017). Between VP and NN. On the constructional types of German *-er* compounds. *Constructions and Frames*, 9(1), 1–40.
- Giegerich, H. J. (2015). *Lexicalist structures. Compounding and the modules of grammar*. Edinburgh: Edinburgh University Press.
- Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Berlin-Brandenburgische Akademie der Wissenschaften. <http://www.deutschestextarchiv.de/>. Accessed December 2019.
- Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica*, 45(1), 31–80.
- Hennig, M. (Ed.) (2016). *Duden: Vol. 9. Das Wörterbuch der sprachlichen Zweifelsfälle. Richtiges und gutes Deutsch*. Berlin: Dudenverlag.
- Hilpert, M. (2018). Wie viele Konstruktionen stecken in einem Wortbildungsmuster? Eine Problematisierung des Produktivitätsbegriffs aus konstruktionsgrammatischer Sicht. In S. Engelberg, H. Lobin, K. Steyer, & S. Wolfer (Eds.), *Wortschätze. Dynamik, Muster, Komplexität* (pp. 91–105). Berlin: de Gruyter.
- Hilpert, M., & Gries, S. T. (2009). Assessing frequency changes in multistage diachronic corpora. Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing*, 24(4), 385–401.
- Hübener, C. J. (2021). What does solid spelling reveal about cognition? Evidence from Middle Low German. In S. Hartmann (Ed.), *Yearbook of the German Cognitive Linguistics Association 2021* (pp. 117–134). Berlin: de Gruyter.
- Hüning, M. (2010). Adjective + Noun constructions between syntax and word formation in Dutch and German. In S. Michel, & M. Onysko (Eds.), *Cognitive perspectives on word formation* (pp. 195–215). Berlin: de Gruyter.
- Hüning, M., & Schlücker, B. (2015). 24. Multi-word expressions. In P. O. Müller, I. Ohnheiser, S. Olsen, & F. Rainer (Eds.), *Word-formation. An international handbook of the languages of Europe* (Vol. 1, pp. 450–467). Berlin: de Gruyter.
- Imo, W. (2016). *Grammatik. Eine Einführung*. Stuttgart: Metzler.
- Jacobs, J. (2005). *Spatien. Zum System der Getrennt- und Zusammenschreibung im heutigen Deutsch*. Berlin: de Gruyter.
- Joeres, R. (1995). *Wortbildungen mit -macher im Althochdeutschen, Mittelhochdeutschen und Neuhochdeutschen*. Heidelberg: Winter.
- Kopf, K. (2017). Fugenelement und Bindestrich in der Compositions-Fuge. Zur Herausbildung phonologischer und graphematischer Grenzmarkierungen in (früh)neuhochdeutschen N+N-Komposita. In N. Fuhrhop, R. Szczeplaniak, & K. Schmidt (Eds.), *Sichtbare und hörbare Morphologie* (pp. 177–204). Berlin: de Gruyter.
- Kopf, K. (2018a). *Fugenelemente diachron. Eine Korpusuntersuchung zu Entstehung und Ausbreitung der verfügbaren N+N-Komposita*. Berlin: de Gruyter.
- Kopf, K. (2018b). From genitive suffix to linking element. A corpus study on the genesis and productivity of a new compounding pattern in (Early) New High German. In T. Ackermann, H. J. Simon, & C. Zimmer (Eds.), *Germanic genitives* (pp. 91–114). Amsterdam: Benjamins.
- Krámský, J. (1969). *The word as a linguistic unit*. The Hague, Paris: Mouton.
- Krott, A., Schreuder, R., Baayen, R. H., & Dressler, W. U. (2007). Analogical effects on linking elements in German compound words. *Language and Cognitive Processes*, 22(1), 25–57.
- Kuperman, V., & Bertram, R. (2013). Moving spaces. Spelling alternations in English noun-noun compounds. *Language and Cognitive Processes*, 28(7), 939–966.
- Lieber, R. (1992). *Deconstructing morphology. Word formation in syntactic theory*. Chicago: University of Chicago Press.
- Lohde, M. (2006). *Wortbildung des modernen Deutschen. Ein Lehr- und Übungsbuch*. Tübingen: Narr Francke Attempto.
- Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- Matthews, P. H. (1972). *Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.
- Meibauer, J. (1998). *kunst verrücker und kolengreber*. Zum Wortbildungswandel der N+V+er-Bildungen im Frühneuhochdeutschen. In I. Barz, & G. Öhlschläger (Eds.), *Zwischen Grammatik und Lexik* (pp. 81–101). Tübingen: Niemeyer.
- Müller, W., & Müller, E. (1961). Wortbildung. Ausdruck der Zeit. *Muttersprache*, 71(3), 65–78.

- Neef, M. (2015). Synthetic compounds in German. In P. O. Müller, I. Ohnheiser, S. Olsen, & F. Rainer (Eds.), *Word-formation. An international handbook of the languages of Europe* (Vol. 1, pp. 582–593). Berlin: de Gruyter.
- Nononen, N., & Penttilä, E. (2014). Constructional continuity. (Where) does lexicon turn into syntax? *The Mental Lexicon*, 9(2), 316–337.
- Nübling, D., & Szczepaniak, R. (2008). On the way from morphology to phonology. German linking elements and the role of the phonological word. *Morphology*, 18, 1–25.
- Nübling, D., & Szczepaniak, R. (2010). Was erklärt die Diachronie für die Synchronie der deutschen Gegenwartssprache? Am Beispiel schwankender Fugenelemente. In H. U. Schmid (Ed.), *Perspektiven der germanistischen Sprachgeschichtsforschung* (pp. 205–224). Berlin: de Gruyter.
- Nübling, D., & Szczepaniak, R. (2011). *Merkmal(s?)analyse, Seminar(s?)arbeit und Essen(s?)ausgabe*. Zweifelsfälle der Verfungung als Indikatoren für Sprachwandel. *Zeitschrift für Sprachwissenschaft*, 30(11), 45–73.
- Ortner, L., Kühnhold, I., Müller-Bollhagen, E., Ortner, H., Wellmann, H., Pümpel-Mader, M., & Gärtner, H. (1991). *Deutsche Wortbildung. Typen und Tendenzen in der Gegenwartssprache. Vierter Hauptteil. Substantivkomposita. Komposita und kompositionsähnliche Strukturen 1*. Berlin: de Gruyter.
- Plag, I., Kunter, G., Lappe, S., & Braun, M. (2008). The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language*, 84(4), 760–794.
- Pümpel-Mader, M., Gassner-Koch, E., & Wellmann, H. (1992). *Deutsche Wortbildung. Typen und Tendenzen in der Gegenwartssprache. Fünfter Hauptteil. Adjektivkomposita und Partizipialbildungen. With the collaboration of L. Ortner*. Berlin: de Gruyter.
- R Core Team (2022). *R. A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Sanchez-Stockhammer, C. (2018). *English compounds and their spelling*. Cambridge: Cambridge University Press.
- Schlücker, B. (2012). Die deutsche Kompositionsfreudigkeit. Übersicht und Einführung. In L. Gaeta, & B. Schlücker (Eds.), *Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte* (pp. 1–25). Berlin: de Gruyter.
- Schlücker, B. (2020). Between word-formation and syntax. Cross-linguistic perspectives on an ongoing debate. *Zeitschrift für Wortbildung*, 4(1), 26–74.
- Solling, D. (2012). *Zur Getrennt-, Zusammen und Bindestrichschreibung von Substantivkomposita im Deutschen (1550–1710)*. Uppsala: Acta Universitatis Upsaliensis.
- Solms, H.-J. (1999). Der Gebrauch uneigentlicher Substantivkomposita im Mittel- und Frühneuhochdeutschen als Indikator kultureller Veränderung. In A. Gardt, U. Haß-Zumkehr, & T. Roelcke (Eds.), *Sprachgeschichte als Kulturgeschichte* (pp. 225–246). Berlin: de Gruyter.
- Sommerfeldt, K.-E. (1988). Zur Wortartklassifikation des Deutschen. Untersuchungen zur Wortartzugehörigkeit partizipialer Bildungen. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 41(2), 221–229.
- Tophinke, D. (2000). Zur Wortabtrennung in den ‘Werler Statuten’ des 14. und 15. Jahrhunderts. Eine exemplarische Analyse. In M. Elmentaler (Ed.), *Regionalsprachen, Stadtsprachen und Institutionssprachen im historischen Prozess* (pp. 73–98). Wien: Edition Praesens.
- Weidman, R. H. (1941). The orthographic conflation of nominal compounds in MHG based on a study of the manesse manuscript. In A. Schirokauer, & W. Paulsen (Eds.), *Corona. Studies in celebration of the eightieth birthday of Samuel Singer, professor emeritus, University of Berne, Switzerland* (pp. 89–99). Durham: Duke University Press.
- Wellmann, H., Reindl, N., & Fahrmaier, A. (1974). Zur morphologischen Regelung der Substantivkomposition im heutigen Deutsch. *Zeitschrift für deutsche Philologie*, 93, 358–378.
- Werner, M. (2017). Zur Entwicklung der synthetischen Komposition in der Geschichte des Deutschen. *Zeitschrift für Wortbildung*, 1(1), 73–92.
- Werner, M. (2020). Three diachronic sources for the development of *-erei*-based synthetic compounds in German. *Word Structure*, 13(3), 347–370.
- Werner, M., Mattes, V., & Korecky-Kröll, K. (2020). The development of synthetic compounds in German. Relating diachrony with L1 acquisition. *Word Structure*, 13(2), 166–188.
- Wilss, W. (1983). Wortbildungstendenzen in der deutschen Gegenwartssprache, dargestellt an Syntagmen des Typs Substantiv + Partizip I (*kostendeckend*). *Muttersprache*, 93, 230–241.
- Zeldes, A. (2013). Komposition als Konstruktionsnetzwerk im fortgeschrittenen L2-Deutsch. *Zeitschrift für Germanistische Linguistik*, 41(2), 240–276.
- Zifonun, G., Hoffmann, L., & Strecker, B. (1997). *Grammatik der deutschen Sprache*. Berlin: de Gruyter.

---

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.