Kristin Kopf

The role of syntax in the productivity of German N+N compounds. A diachronic corpus study[1]

Abstract: This paper studies the morphological productivity of German N+N compounding patterns from a diachronic perspective. It argues that the productivity of compounds increases due to syntactic influence from genitive constructions ("improper compounds") in Early New High German. Both quantitative and qualitative productivity measures are adapted from derivational morphology and tested on compound data from the Mainz Corpus of (Early) New High German (1500–1710).

1. Introduction

While N+N compounding in present-day German is subject to very few restrictions, the situation was different in Old High German (OHG, 500–1050 AD) and Middle High German (MHG, 1050–1350 AD). Obviously, the productivity of compounding has changed between MHG and today. This development warrants a more detailed investigation.

The paper is structured as follows. I will first review some fundamental aspects of compounding in modern German and its history, and describe my data sources and analysis (Section 2). I will then turn to productivity in word-formation which can be measured in different ways: First, it can be calculated from the numbers of types, tokens and hapax legomena. This is what I will refer to as quantitative measures. In its most basic sense, productivity is the ability of a word-formation pattern to form new units (operationalized as potential productivity; cf. Baayen 2009: 902–904) and to actually do so (operationalized as realized productivity; cf. Baayen 2009: 901–902). The question of to what extent productivity in word-formation can be measured has been hotly debated, albeit almost exclusively in the context of derivation. Compounding is considered to follow similar principles (e.g. "all the discussion here should equally apply to compounds"; Bauer 2005: 316), although I am unaware of case studies. I will apply quantitative productivity measures to corpus data from the second half of Early New High German (ENHG, 1350–1650) and from the beginning New High German (NHG, since 1650) in Section 3, and discuss their benefits and problems. Second, a word-formation pattern can lose semantic, morphological or phonological

---

1   I'd like to thank the two anonymous reviewers, Susanne Flach and Celeste Brennecka for their helpful comments.

restrictions, thereby opening up to formations that were unavailable before. While this is reflected in the quantitative measures, it warrants an in-depth discussion (Section 4). I will refer to this approach as qualitative, although the loss of restrictions will also be measured in numbers – strictly speaking, it is also largely quantitative, but applied to selected phenomena within compounding, not to compounding as such. I will show that the two perspectives complement each other and that it would not be wise to forego one for the other. In Section 5, I will combine the results and observations, offering an explanation for the considerable productivity of compounding in present-day German. I will show that it is likely rooted in changes in ENHG compounding patterns that were facilitated by syntactic constructions.

2. Fundamentals

2.1. Compounds in New High German

Modern German shows a strong affinity for compounding as a means to expand its lexicon. Munske (2009: 227) considers compounding "the greatest laboratory of 'integration' for lexical elements of different origin into new word structures" (own translation).[2] 83% of all new word-formation products in Harlass & Vater's (1974: 91) corpus data are compounds – compared to single-digit percentages for derivation (9%), simplexes (3%) and abbreviations (2%).[3] They conclude: "The extension of the lexicon thus relies mostly on compounding."[4]

However, in present-day German compounding, input mostly consists of nouns and adjectives for both constituents, with N+N compounds being by far the most common type (Ortner et al. 1991: 37; Pümpel-Mader et al. 1992: 19). Many other types often subsumed under compounding are in fact cases of univerbation with a change in the part of speech: A combination of two prepositions used as an adverb, such as *durch|aus* 'definitely (verb. through|out)', should therefore not be analyzed as compounding.[5]

---

2   „[D]as größte Labor der ‚Integration' lexikalischer Elemente verschiedener Herkunft in neuen Wortkonstruktionen."

3   The material also contains phrasal constructions like *sowjetische Besatzungszone* 'Soviet occupation zone', i.e. mostly names in the form of an NP. These were not included here. In total, the corpus comprises 20,000 types from newspapers published in the early 1960s. The types denote concepts/objects that were either not known before 1945 or had not been recorded in dictionaries before. All types were recorded on punch cards and analyzed automatically.

4   „Eine Erweiterung des Wortbestandes findet also vor allem durch Komposition statt."

5   A few notes on formal conventions: The internal structure of compounds is marked by <|> instead of <-> to avoid confusion with actual spelling, e.g.

Not only are compounds a productive means of word formation in German, the process also lacks phonological, morphological or semantic restrictions apart from preference for certain parts of speech. The widely cited *Donau|dampf|schiff|fahrt|s|gesellschaft|s|kapitän* 'captain of the Danube Steamboat Shipping Company'[6] is no anecdotal exception: The length of a compound is merely restricted by the speaker's working memory, much as syntactic units are. The longest compound (by letters) that appears at least 40 times in the Duden corpus consists of seven nouns, many of them containing derivational affixes.[7] As long as there is a communicative need to further modify a concept, speakers form compounds (cf. Meineke 1991: 41). Even when presented with compounds that do not seem to make much sense (i.e. *Brücke|n|brücke* ,bridge bridge', Günther 1981), they are willing to construct possible contexts (a general presupposition of existence, see Klos 2011). The main semantic relation between the two constituents of a German compound is determinative, i.e. the first constituent modifies the second, which forms the semantic core. Other semantic relations do exist but will only be discussed in passing.[8]

Anfahrt|s|beschreibung 'directions (verb. approach|ʟᴇ|description)', standard spelling: <Anfahrtsbeschreibung >. Linking elements are always separated and marked as ʟᴇ. Historical corpus data often contains hyphenation in the original (using <=> or <->). In these cases, <|> is only used to additionally segment linking elements, while the original hyphen is kept, e.g. *Jnfection|s=zeiten* 'times of infection (verb. infection|ʟᴇ|times)'. Likewise, if a compound is not written as one orthographic word, the original space is preserved, e.g. *Vermanung|s red* 'admonishing speech (verb. admonishment|ʟᴇ|speech)'. If internal compound structure is relevant, the embedded structure is enclosed in square brackets, e.g. [*Weih|rauch*]*essig* 'vinegar of frankincense (verb. consecrate|smoke|vinegar)'. If a compound is not glossed, its English translation is accompanied by a verbatim equivalent in parentheses where the two differ, this is marked by 'verb.'. Glosses follow the Leipzig Glossing Rules.

6   Verb. Danube|steam|ship|drive|ʟᴇ|company|ʟᴇ|captain.
7   *Grund|stück|s|verkehr|s|genehmigung|s|zuständigkeit|s|übertragung|s|verordnung* 'Regulation on the delegation of authority concerning land conveyance permissions (verb. ground|piece|ʟᴇ|traffic|ʟᴇ|permission|ʟᴇ|jurisdiction|ʟᴇ|transfer|ʟᴇ|decree)', see http://www.duden.de/sprachwissen/sprachratgeber/die-laengsten-woerter-im-dudenkorpus (8/9/2017).
8   Compounding terminology is notoriously heterogeneous (for an overview, see Scalise & Bisetto 2009). I follow Olsen (2001), distinguishing determinative, copulative and possessive compounds. Determinative compounds subsume both synthetic compounds like *Arbeit|nehmer* 'employee (verb. work|taker)' and root compounds like *Arbeit|s|zeit* 'working time (verb. work|ʟᴇ|time)'. The choice is motivated by the different genesis of these types: Only determinative compounds could arise from syntactic constructions in ENHG (see below) while copulative and possessive compounds were always combined directly by a word-formation pattern (cf. Kopf 2018b: 5–8).

2.2. Compounding types: Etymology and form

N+N compounds can be divided into two groups from a synchronic and a diachronic perspective, respectively. Synchronically, we distinguish compounds with linking elements (LE, "linked compounds") and compounds without linking elements ("unlinked compounds"). Diachronically, there are proper compounds (formed after the IE pattern) and improper compounds (reanalyzed from genitive phrases). These four groups are not independent, but there is no straightforward correlation between linked/improper compounds and unlinked/proper compounds, respectively. We will now consider all four types in turn.

IE nominal morphology consists of three elements: The root combines with a thematic vowel (stem-building element, SBE)[9] to form a stem that can then be combined with a second noun or adjective (for compounds, c.f. 1a), a derivational suffix or an inflectional suffix (Krahe & Meid 1967). Only a few so-called root nouns act as stems without further additions (c.f. 1b).

(1) Gothic
   a. mar|i|saiws 'lake'
      sea|SBE|lake
   b. man|leika 'figure/image of a man'
      man|image

This pattern has been called "proper compounding" ("eigentliche Komposition") by Grimm (1826: 407): It is older than improper compounding and its compounds are formed by a morphological process. In OHG and MHG, most thematic vowels in compounds lose their systematicity, and are then lost completely (cf. Gröger 1911; Demske 2001: 297: Nübling & Szczepaniak 2013: 69–70). The only residuals can be found in a small group of short-stemmed nouns, e.g. MHD *tag|e|lôn* 'daily wage (verb. day|LE|wage)'.[10] Thus, the new default case is a direct combination of roots without any intervening elements.

Improper compounds are a later addition to the system. They develop from univerbation of prenominal genitive attributes (cf. Pavlov 1983; Nitta 1987; Demske 1999, Nübling & Szczepaniak 2008; 2013; Solling 2012; Kopf 2017; 2018a; 2018b). Isolated cases are attested early on in many IE languages (Krahe & Meid 1967: 17–18):

---

9 Or a combination of vowel and consonant; these are, however, irrelevant for German compounds (cf. Nübling & Szczepaniak 2013, Szczepaniak 2016).

10 Vowels following former long-stemmed nouns must, on the other hand, have originated in improper compounds (e.g. *Schwein e futter* 'pig LE food', cf. Szczepaniak 2016: 329).

(2) a. Greek   diós|kouroi
                    Zeus.GEN-son.PL
     b. Latin   aquae|ductus
                    water.GEN-pipeline
     c. Sanskrit   jás|pati-
                    house/family.GEN-lord-
     d. Gothic   baurg|s|waddjus
                    castle-GEN-wall

However, the pattern only becomes more common in ENHG. Nouns with overt genitive markers carry these with them when they undergo univerbation, hence giving rise to compounds with linking elements. This can be seen in (3) for linking-*s* (univerbation from a to b) and in (4) for linking-*n* (Kopf 2018a):

(3) ENHG: genesis of linking-*s*
     a.   wegen   [des   Leib|s]   Erbe|n   >
          because the.GEN.SG   body(M)-GEN.SG   heir(M)-GEN.SG
     b.   wegen   des   Leib|s|erbe|n
          because the.GEN.SG   body-LE-heir(M)-GEN.SG
          'because of the heir of the body'

(4) ENHG: genesis of linking-*n*
     a.   [der   Sonne|n]   Schein   >
          the.GEN.SG   sun(F)-GEN.SG   shine(M).NOM.SG
          'the shine of the sun'
     b.   der   Sonne|n|schein
          the.NOM.SG   sun-LE-shine(M).NOM.SG
          'the sunshine'

The results resemble compounds, but did not arise by compounding. At some point, they are reanalyzed as part of the word-formation system. Linking elements are then productively inserted into new compounds that do not go back to genitive constructions. Such cases are still called "improper compounds" by Grimm, even though they are "properly" (i.e. directly) compounded.

Nouns unmarked for genitive case in the singular – i.e. feminine nouns from the strong *i*-declension (5a) – and strong masculine/neuter nouns without plural markers (5b) are formally identical with the old compounding type (for a detailed heuristic cf. Kopf 2018b, in preparation). It is thus impossible to tell how they came about:

(5) ENHG

a. *Nacht*:   *Nacht|zeit*   < *Nacht + Zeit*
                      <   *in*   *der*   *Nacht*   *Zeit*
   night (F)   night-time   in   the   night:GEN.SG time

b. *Engel*:   *Engel|schar*   < *Engel + Schar*
                      <   *mit*   *der*   *Engel*   *Schar*
   angel (M)   angel-group   with   the   angel:GEN.PL group

The distinction of proper and improper compounds thus poses two problems for empirical studies: First, in many cases without a linking element, it is impossible to tell if the compound is "proper" or not. Second, in many cases with a linking element, it is impossible to tell if the linking element is due to univerbation or part of a productive word-formation pattern.

In studies of present-day German, the division between proper and improper compounds is irrelevant: The direct connection with syntactic structures has been lost due to postposing of genitive attributes and the spread of determiner use (especially articles, cf. Pavlov 1983, Demske 2001, Kopf 2018b). Hence, no new univerbations occur. All compounds are now understood to belong to word-formation, even if their individual histories might not bear this out. Linked compounds amount for 27.2%to 42.4% of all compound types (Wellmann et al. 1974: 365, Kürschner 2003) and 29.5% of compound tokens (Donalies & Bubenhofer 2011: 32). The distribution of linking elements has been widely discussed (e.g. Fuhrhop 1996, Aronoff & Fuhrhop 2002, Kürschner 2003, Nübling & Szczepaniak 2008, Kopf 2018b), and although interested laypersons often concentrate on dubious cases, there are large areas in which the use of a linking element can be predicted with high certainty.

Almost all linked compounds go back to former improper compounds or were formed analogically, in line with their pattern. Only very few cases still contain an old linking vowel (like *Tag|e|lohn*, see Tab. 1, upper row). Most unlinked compounds go back to former proper compounds and the pattern they provided. There is, however, a sizeable number of potential exceptions: In cases in which the genitive singular or plural did not have a suffix in MHG/ENHG (such as *(der) Nacht* 'night's (GEN SG)' or *(der) Engel* 'angels' (GEN PL)'), we cannot tell whether compounds containing them are due to reanalysis/analogy or the old compounding pattern (see Tab. 1, lower row).

Tab. 1: Difficulties when distinguishing proper and improper compounds by linking elements.

	proper compounds	improper compounds
linked compounds	*Tag|e|lohn* (< OHG *tag|allōn* 'daily wage (verb. day|LE|wage')	from univerbation: *Tag|es|zeit* 'time of day (verb. day|LE|time)' (< OHG *zuo tag|es zīt* 'at day|GEN time') from compounding: *Religion|s=Anfechtungen* 'contestations of religion (verb. religion|LE|contestation|PL'
unlinked compounds	*Nacht|Ø|zeit* 'nighttime' (< OHG *zuo naht zīt* 'at night. GEN time' OR OHG *naht|Ø|zīt* 'night-time')[11]	

2.3. Corpus & data analysis

Most of the empirical data for this paper was collected using the following corpora:

time	corpus	composition	tokens	annotation	data extracted	description
1500–1710	Mainz Corpus of (E)NHG	8 periods, 2 genres, 5 regions	320,000	none	all Genitive constructions and NN compounds with common nouns	Kopf (2018b)
1650–1800	GerManC (subcorpora SCIE, SERM, NEWS)	3 periods, 3 genres, 5 regions	90,000	POS	all NN compounds (manual check of all tokens POS-tagged NN with 6+ characters)	Durrell et al. (2007)
1850, 1900	excerpts from Mannheim Historical Newspapers	2 periods, 1 genre	60,000	none	all NN compounds (manual check of all tokens with initial upper case and 6+ characters)	http://hdl.handle.net/10932/00-01B8-AE41-41A4-DC01-5

---

11 The empty set symbol ($\emptyset$) is used where no linking element occurs to mark the contrast. It does, however, not assume some kind of invisible element.

The Mainz Corpus of (E)NHG, based on Bergmann & Nerius (1998) but heavily modified for present purposes (cf. Kopf 2018b), has been analyzed most thoroughly. It contains 8 periods, 2 genres (non-fiction and theological texts) and 5 regions (East Upper German, West Upper German, North Upper German, East Middle German, West Middle German). The GerManC subcorpora SCIE and SERM (scientific texts and sermons) were chosen because they are similar in text composition. From 1650 to 1900, newspaper texts of the same size can also be compared (3 periods of NEWS in GerManC and 2 added periods from the Mannheim Corpus of Historical Newspapers). Each corpus was balanced for token number per period and, where applicable, genre and region.

Quantitative productivity measures were used for the Mainz Corpus, qualitative data was analyzed for all three corpora. In one case, additional material from the German Text Archive (DTA) was used; in another case, the Corpus of the MHG Grammar (MiGraKo) was consulted.

Identification of compounds in historical sources poses serious problems, especially in the Mainz Corpus, as they partially overlap with syntactic constructions (cf. Pavlov 1983, Nitta 1987, Reagan 1981, Solling 2011, 2012). Due to morphosyntactic changes in ENHG (e.g. positional change of the genitive attribute, obligatorization of definite and indefinite article), this is no longer the case. Today, morphosyntactic, phonological and graphematic properties clearly distinguish compounds from phrases. In Kopf (2018b, in preparation), I develop a detailed heuristic based on Pavlov (1983) for distinguishing unambiguous compounds, unambiguous genitive constructions and three types of ambiguous intermediate constructions. One of these ambiguous constructions, the so-called "framing constructions" (Rahmenkonstruktionen), is often taken for compounding (e.g. Kopf in print) as it has a determiner or modifier that precedes the whole construction and agrees with the second noun (see 4a). I have included this construction when measuring quantitative productivity. As shown in Kopf (2018b: 243–252), there is, however, little difference with a scenario in which they are excluded.
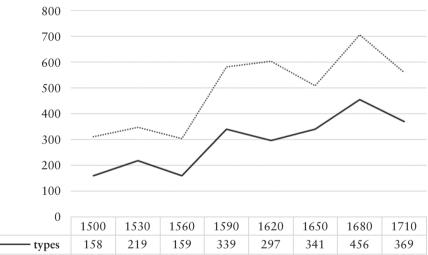
3. Quantitative measures of productivity

3.1. Compounding in general

It is indisputable that compounding has become more and more frequent when comparing OHG and NHG inventory and usage. While numbers for types and tokens are lacking in earlier periods, they have been determined for the Mainz Corpus, so part of the gap between OHG and NHG can now be filled with data. This allows us to apply established measures of productivity and compare the outcome with findings from a more quantitative approach.

The numbers for tokens and types, given in Fig. 1, show a strong increase from 1590 onwards: Compounds were used to a greater extent than before, and people used more different compounds than previously – e.g. compounds that were not possible in OHG or MHG due to morphological restrictions (see Section 4). The type frequency corresponds to the realized productivity of the pattern,[12] i.e. it shows how productive the pattern has been in the past (Baayen 2009: 901–902). As it is rather low for the years 1500 to 1560, its productivity must have expanded immensely shortly before 1590.

Fig. 1: Types (n=2,338) and tokens (n=3,927) of compounds in the Mainz Corpus. The increase is statistically highly significant for both types and tokens (Kendall's Tau, τ=1; T=28; p<0.001).

	1500	1530	1560	1590	1620	1650	1680	1710
types	158	219	159	339	297	341	456	369
tokens	312	347	304	582	605	510	708	559

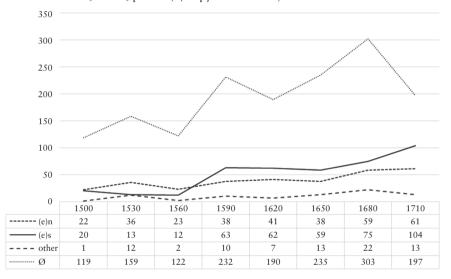## 3.2. Compounding patterns: linked vs. unlinked compounds

As has been discussed in Section 2.2, most compounds with linking elements (apart from a few exceptions with linking-*e*) point either to reanalysis from genitive phrases or to a new, purely morphological pattern of compounding. Compounds without linking elements, on the other hand, do not point to a particular genesis (but see Kopf 2018b: 228–230). Changes in their productivity cannot be clearly attributed to either univerbation or compounding. I will therefore focus on linked compounds for the remainder of this section.

---

12 In a corpus of invariant size, realized productivity is simply the number of types (while usually it would be the number of types divided by the number of text words).

3.2.1. Realized productivity

The realized productivity of compounds in general has been shown to increase over time (3.1). If we look at different compounding patterns in turn, large differences arise: The unlinked pattern increases, but shows a great variation, probably due to the heterogeneous genesis of its members. The linked patterns increase in a more straightforward way (*(e)n*, *(e)s*, other in Fig. 2).

Fig. 2: *Realized productivity by compounding pattern. The increase is statistically significant for the three major patterns (Kendall's Tau, (e) s: τ=0.571; T=22; p<0.05), (e)n: τ=0.815; z=2.764;[13] p<0.01, no LE: τ=0.643; T=23; p<0.05) (Kopf 2018b: 245).*

	1500	1530	1560	1590	1620	1650	1680	1710
------- (e)n	22	36	23	38	41	38	59	61
——— (e)s	20	13	12	63	62	59	75	104
- - - - other	1	12	2	10	7	13	22	13
·········· Ø	119	159	122	232	190	235	303	197

As for the overall development, compounds with linking-*(e)s* show a drastic increase in 1590, and again in 1710. We cannot tell for sure from the numbers if this is due to reanalysis or due to the establishment of a new, purely morphological pattern. I would, however, argue for the latter, as most constructions that form potential input for reanalysis have disappeared by this time (due to advanced positional change of genitive attributes, see Kopf 2018: 86–110).

3.2.2. Potential productivity

Potential productivity is a measure that is supposed to determine not how productive a pattern has been in the past, but what potential it has to form new words in the present (Baayen 2009: 901–902). This makes a lot of

---

13 T could not be calculated due to multiple identical values.

sense when looking at derivational morphology: If a suffix appears on many words, but morphological or semantic restrictions hinder its further spread, the pattern has done all it can already. The existing types will be used, but as they are firmly established, they will surely occur several times in a large collection of texts, not just once. If, however, the pattern can still be expanded (e.g. by shedding restrictions), speakers will coin words that are not (yet) part of the lexicon, and those will appear seldom at first. Potential productivity is therefore estimated by dividing hapax legomena by tokens of the pattern under investigation. The reasoning goes as follows: If a word appears only once in a good-sized corpus, it is very likely a new formation. We can then determine the ratio of such new formations to all formations. The closer the resulting number is to 1, the higher the potential productivity.

By drawing on hapax legomena, potential productivity hinges on corpus size: The smaller the corpus, the higher the number of compounds that appear only once, and the higher the probability that these are not actually new formations. The relationship between hapax legomena, corpus size and number of tokens relevant to the pattern is not linear (see Kopf 2018b: 246–250). While the Mainz Corpus accounts for invariable corpus size, comparison between patterns of different frequency is notoriously difficult. It has been shown that the overall number of compounds increases over time (3.1), thus lowering the probability of hapax legomena in later periods. When separating the groups of linking elements as done here, the problem increases: As unlinked compounds show much higher numbers than linked compounds (and thus lower probabilities for hapax legomena), their development cannot be compared to each other. This renders the concept of potential productivity highly problematic.
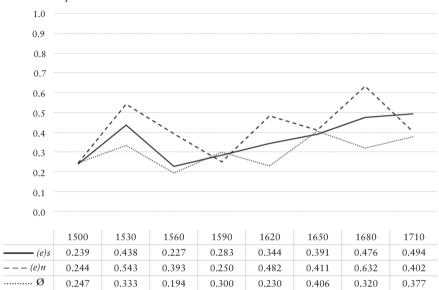
Fig. 3: *Potential productivity of different compounding patterns in the Mainz Corpus.*

	1500	1530	1560	1590	1620	1650	1680	1710
——— (e)s	0.239	0.438	0.227	0.283	0.344	0.391	0.476	0.494
– – – (e)n	0.244	0.543	0.393	0.250	0.482	0.411	0.632	0.402
·········· Ø	0.247	0.333	0.194	0.300	0.230	0.406	0.320	0.377

The findings are therefore limited to determining if there is an increase of productivity in one of the patterns. This is not the case; Kendall's Tau yields no significant results for any of them. This seems peculiar when compared to realized productivity: How can the potential of a pattern be unchanged if the number of different types it produces does in fact increase? The most likely explanation is that the corpus is not suited for measuring potential productivity: The number of hapax legomena is artificially high for two reasons: First, the corpus size is rather small (40,000 tokens per period). Second, the compounds are not evenly distributed over time. The number of compound tokens is low in the first three periods (between 300 and 350) and then almost doubles in 1590 (see Fig. 1). This means that the number of hapax legomena will be higher than expected from 1500 to 1560 compared to the following years. That might lead to the impression that potential productivity does not change over time. We cannot tell for sure whether realized or potential productivity is a better indicator of what is happening in the corpus, which is why it is instructive to have a closer look at qualitative measures of productivity.

4. Qualitative measures of productivity: Loss of restrictions, extension to new contexts

4.1. General remarks

The use of new word-formation patterns can be limited by pragmatic or structural restrictions (e.g. Rainer 2005, Schmid 2011: 115–119). The main prerequisite for the use of a pattern is of course the need to designate a concept or object. This need may not arise if the referent is non-existent (and not imaginable, e.g. *carpet opener*), if the pattern would make use of obvious properties (**eyed man*) or if there is no "nameable" concept (Schmid 2011: 116). Preexisting synonyms[14] or homonyms (**liver* 'living person') might also be hindering an otherwise regular word-formation process. All these "restrictions" pertain to individual words; they do not hinder a word-formation pattern as such, and thus they tell us more about human imagination than about word-formation (see also Schmid 2011: 116, who calls them "not very helpful").[15]

The case is different for semantic, morphological or phonological restrictions: A pattern might only be applicable to certain parts of speech (i.e. present-day German *-er* denoting agentive nouns can only be used with a verbal base) or to certain parts of the lexicon (i.e. native vs. non-native bases,[16] complex vs. simple bases). It might be hindered when certain sound patterns would arise

---

14 Much-used examples like *Stehler* (*stehl-* 'steal' + *-er* 'agentive suffix') which is supposedly blocked by *Dieb* 'thief' lose much of their strength by a diachronic perspective: When *er*-formations arose in OHG, they met with quite a number of synonyms, e.g. *wartil* (*-il*) 'warden' or *becko* (*-o*) 'baker'. Still, they were finally replaced by *Wärter* and *Bäcker*. Cases like *Dieb* are marginal at best and hold no influence over the word-formation pattern as such; they only hinder individual words. Kempf (2016) shows partially synonymous adjectival suffixes (e.g. *-lich* and *-bar*) competing for bases which lead to functional differentiation (via temporarily coexisting forms).

15 It is also not true that speakers would generally feel no need to coin such terms: They might want to promote a new tool used to pry a carpet away from the floor to remove it (*carpet opener*), or they might want to distinguish a person employed with unfolding a rug from a roll from other workers. Children might speak about the only doll left that still has its eyes as *eyed man* and a science-fiction author may combine a seemingly disparate array of actions and objects into a single verb to denote something aliens do.

16 This restriction is considered etymological by Plag (1999: 58). I would argue that – at least for German – the synchronic structure of the base is the deciding factor. Of course suffixes like *-ier* '*-ize*' (restricted to latinate bases) can also be found with roots that do not look like loans, but in fact are (e.g. *kassieren* 'to cash in' < *Kasse* 'cash register', which looks like native *Gasse* 'alley'). However, most of these cases are remnants from earlier times – *kassieren* hails from the

(i.e. German diminutive suffix *-chen* which doesn't combine with bases ending in a palatal fricative – in these cases, synonymous *-lein* is preferred, so *Büchlein* 'small book' instead of *Büchchen*) or favored when a certain prosody arises.

4.2. What to measure in compounding

While most derivatives and compounds show a binary structure, the functions of the two morphemes are not identical. In derivational morphology, the analysis of restrictions is usually framed in one of two perspectives: Either the base accepts only certain affixes, or the affix selects only certain bases. This is clearly not the case in compounding. Apart from the fact that there should be a semantically motivated reason to combine the two elements, the specific constituents to not restrict each other. However, there are restrictions on the whole compounding pattern – or at least there were in earlier periods of German. In OHG, compounds with a verbal stem as first constituent are not yet in use (for their development, see Gröger 1911), tripartite compounds are uncommon and mostly found in translations and glosses (Carr 1939:197), and derivationally complex first constituents are almost nonexistent. The default case for the first constituent is a monosyllabic stem (Wilmanns 1896: 388).[17] The case is different for second constituents: They show no restrictions in older periods (Wilmanns 1896: 388). Even though it cannot be said that the second constituent formally restricts the first, OHG speakers were clearly hesitant to make use of morphologically complex elements while NHG speakers are not. In this section, I will focus on the change the compounding pattern underwent in ENHG. First, the use of morphologically complex input over time will be analyzed with data from grammars and the Mainz Corpus. Then, a second aspect must be considered: Most OHG and MHG compounds adhere to the old compounding pattern, although there are a few cases of linking elements from genitive reanalysis (Grimm 1826: 409). From ENHG onwards, the new compounding pattern plays an important role. I will show that there is an overarching explanation for both observations.

4.2.1. *Number of roots: tripartite compounds*

Tripartite compounds with three nominal elements are not attested in Gothic or other minor Germanic sources (Carr 1939: 197). Even if "particle compounds" like Gothic [*fra\bauhta*]*boka* 'deeds of sale' or *ufar*[*himina\kunda*] 'heavenly'

17th century when its base still was an obvious loan from Italian (*Cassa*). It seems more adequate to classify the input restriction for *-ier* as structural.

17   There are different opinions on how common compounding was in Germanic (Gröger 1911: 1, Salus 1963: 47, Carr 1939), but we can assume tentatively that it was similar to early OHG.

are included, cases are rare and the embedded compounds are "no longer felt as compounds" (Wilmanns 1896: 388).[18]

In early OHG, tripartite compounds are still rare (Carr 1939: 197), although more numerous than in Gothic (Wilmanns 1896: 388). They are mostly found as glosses and often contain a lexicalized constituent. In many cases, the second constituent is undergoing grammaticalization (6d, ending in today's derivational suffix -*schaft* '-ship').

(6) a. [*pooh|stap*]*zila* 'row of letters'
         book-staff-row
         (Samanunga glossary, Vienna, Cod. 162)
     b. [*uuerolt|reht*]*uuiso|n* 'scholars of the law of the world'
         world-right-scholar-PL
         (Muspilli, Clm 14098)
     c. [*hasel|nuze*]*chern|en* 'hazelnut kernel'
         hazel-nut-pit-DAT.SG
         (Notker: Martianus Capella, St Gall, Cod. 872)
     d. [*heri|ganoz*]*scaf* 'military division'
         army-comrade-form
         (Monsee fragments, Munich, Clm 14747)

Starting from the later OHG period, tripartite compounds are "not uncommon" (Carr 1939: 197), although most seem to be a result of translation:[19] Language users rely on them to solve a communicative problem. Their usage is telling: As compounding is almost nonexistent in Latin, tripartite compounds are a Germanic structure through and through. Even though sparingly used in indigenous texts, the pattern is available. Such structures are most likely not (yet) a part of the general compounding pattern, but this pattern's extension comes naturally when expressing complex concepts.

We can assume that the use of the pattern rose gradually, though not steeply from OHG to MHG and then ENHG. Reliable data is available from 1500 onwards: Compounds with three roots occur from the beginning in the Mainz corpus, albeit at a low level (Kopf 2018b: 268–270):[20] They account for 2.6% of all compound types and are used even less (1.8% of all compound

---

18 I count such cases towards prefixation (see below).
19 "The whole type is, in fact, mainly restricted to glosses, and is not found in the OHG poetry, although there is one example from the L[ow]G[erman] Heliand." (Carr 1939: 198)
20 The data includes compounds with three roots. The embedded compound may have one of the structures V+N, A+N or N+N.

tokens). Almost half of the types contain a lexicalized constituent that is already attested in OHG (e.g. [*Weih|rauch*]*essig* 'vinegar of frankincense (verb. consecrate|smoke|vinegar'). The GerManC data (1650–1800) shows no difference in types (2.7%) but a higher usage frequency (2.9%), which is mostly due to the inclusion of newspaper texts (they contain 93 of 155 tripartite tokens).

In present-day German, compounds with three or more lexical roots are more common, but not widely used: While accounting for 11.8% of Ortner et al.'s (1991: 13) nominal compound types (X+N), their usage varies between 1% and 8% in newspaper texts. If we consider only new formations instead of inventory, tripartite compounds play a bigger role: In an early corpus study on German neologisms, Harlass & Vater (1974: 94) show 36% of all compounds to be made up of more than two lexical roots. This points to a stronger productivity of the pattern, probably genre-related.

4.2.2. Derivationally complex constituents

This group contains cases of suffixation (e.g. *Beziehung|s|ratgeber* 'relationship advisor'), prefixation (e.g. **Umwelt|schutz** 'environmental protection (verb. environment|protection)', but see below) and circumfixation (**Geschmeide|kunst** 'jewelry (verb. jewelry|art')). Such derivationally complex first constituents occur regularly in present-day German. They are, however, uncommon in OHG, which limits itself to complex second constituents (Wilmanns 1896: 388).

Only when a **suffix** has become intransparent in OHG (e.g. in *kuning* 'king', *jugund* 'youth'), its word-formation product may be used at the beginning of a compound. Transparent formations in *-āri*, *-ida*, *-unga*, *-nissi* cannot be found in OHG compounds (Tatian, Otfrid, Notker, cf. Wilmanns 1896: 388).[21] These restrictions still hold for MHG, first constituents in *-er(e)* (< OHG *-āri*) being the only exception (Carr 1939: 222):

(7) a. *dörper|diet* 'farmers (verb. farmer|people)'
    b. *jeger|meister* 'hunter (verb. hunter|master)'
    c. *rihter|stuol* 'bench (verb. judge|chair)'

The group denotes agent nouns, as opposed to most other OHG/MHG suffixes, which bring about abstract or collective semantics. This makes *er*-derivations especially common in possessive constructions and therefore in many compounds. These nouns are structurally identical with simplex

---

21 Carr (1939: 219) names two early exceptions, *hamalunc-stat* 'Skull Hill' and *gesmīde-ziereda*.'jewel ornaments' (from the Germanic circumfix *ga-*X-*ja*), but concedes that the former is likely due to OE influence Carr (1939: 221).

nouns in *-er* (e.g. MHG *kevere* 'bug', *hamer* 'hammer'), which might have helped as well.

The situation is very different in the (E)NHG data of the Mainz Corpus. I considered all first constituents ending in suffixes or suffix-like endings of Latin origin ("exogenous suffixes", cf. Fleischer & Barz 2012: 239) and mapped their first appearance in a compound (Fig. 4). While first appearance in a corpus should not be mistaken as first use or even first attested use, the relative chronology is still instructive.

Fig. 4: *First appearance of a suffixed first constituent in the Mainz Corpus (Kopf 2018b: 261). Suffixes that appeared less than ten times on either a first constituent of a compound or a genitive attribute were not considered. Real suffixes are set in boldface to set them apart from suffix-like endings.*

1500	1530	1560	1590	1620	1650	1680	1710
-er (*Burger|meister* 'mayor')							
-ent/ant (*conuent|bruder* 'friar', *Proviand|heuser* 'storehouse for food')							
			-ur (*Fractur=Feder* 'quill for black letter')				
			-ung (*Vermanung|s red* 'admonishing speech')				
			-ion (*Jnfection|s=zeiten* 'times of infection')				
				-nis (*Vermächtnuß|sachen* 'inheritance')			
				-tät (*Majestät|Briefen* 'Letters of Majesty')			
				-ei (*Cantzley|buch* 'registry')			
				-lein (*Mährlein|schreiber* 'writer of (fairy) tales')			
					-schaft (*Vormundschafft|s Zweck* 'purpose of a guardianship')		
							-at (*Canonicat=Stifft* 'monastry chapter')
							-enz (*Residentz=Stadt* 'royal seat')
1500	1530	1560	1590	1620	1650	1680	1710

The data suggest that the restriction on derivationally complex first constituents started to crumble in the late 16th century: Derivatives that look much like simplexes occur from the beginning (with *-er* even going back to MHG). However, suffixes that are formally set apart first appear in 1590 and grow in number. The first suffixes to be included are phonologically reanalyzed due to their vocalic onset, and thus merged with the base. Suffixes with consonantal onset that form their own phonological word appear from 1620 on. Some suffixes, while present outside of compounds in the Mainz Corpus (e.g. *-tum*, *-heit/keit*), do not appear yet. A closer look at the most frequent of these suffixes, *-ung*, a

cognate of engl. *-ing*, adds to the finding that complex first constituents only became part of the German compounding pattern in ENHG. Grimm (1826: 937) remarks on Martin Luther's German bible translation (first half of the 16th century):

> As in earlier use, [Luther] never compounds the frequent formations in *-ung*. He says neither *nahrung-sorge*, *versönung-geld* nor *nahrungs-sorge*, *versönungs-geld*, but *sorgen der nahrung* ['worries about food/sustenance'] Luke 21, 34, *geld der versönung* ['atonement money'] Exod. 30, 16. […] or if he wants to compound, he uses a verb, e.g. *versön-tag* ['day of atonement'] Levit. 23, 28. (own translation and formatting)[22]

This ties in nicely with the corpus data reported above, and more specifically with the usage of *-ung* in first constituents of compounds: The first use in the Mainz corpus dates from 1590. An increase can only be seen in 1680/1710 (11 tokens, 10 different *ung*-types, cf. Kopf 2018b: 258–259). One could suspect that growing numbers of *ung*-compounds simply reflect growing productivity of *ung*-derivation. However, this is not the case: The productivity of *-ung* remains diachronically stable in the Mainz Corpus (Hartmann 2016: 171). Growing use in compounds therefore shows loss of morphological restrictions. We will come back to the fact that Grimm's (1826: 937) potential compound suggestions show both unlinked and linked compounding (*nahrung-sorge* vs. *nahrung-s-sorge*) – this is no coincidence.

The topic of nominal **prefixes** is somewhat complicated: Although they could be mistaken for prefixes on first sight, preposition-like elements in many complex words are actually due to conversion (8a) or derivation of verbal stems (8b):

(8) a. *stēn* 'stand' >   *fir+stantan* 'realize, understand'
                conversion: *firstant* 'understanding, intellect'
     b. *waltan* 'govern' >   *anawalt(an)* 'to be powerful'
                derivation: *anawalt+o* 'administrator, ruler'

(9) a. *lust* 'desire, pleasure' > *un+lust* 'dislike, displeasure'
     b. *richten* 'to judge' >   *ge+riht+i* 'court of law'

---

22 German original: „die häufigen bildungen mit -ung setzt er [Luther, KK], gleich der früheren sprache, […] nie zusammen. Er sagt weder nahrung-sorge, versönung-geld noch nahrungs-sorge, versönungs-geld, sondern sorgen der nahrung Luc. 21, 34, geld der versönung Exod. 30, 16. […] oder wenn er componieren will, thut ers verbal, z. B. versön-tag Levit. 23, 28."

Nouns that have carried over verbal prefixes or particles are rare as first constituents in OHG (Wilmanns 1896: 139). Nouns with true nominal prefixes do not occur (9). This also holds true for derivatives with a *gi*-X-*i* circumfix (9b),[23] the single exception being *gesmīde|ziereda* 'jewel ornaments' (Carr 1939: 219). A spot-check in MiGraKo yields no clear examples of MHG compounds with former verbal prefixes or particles.[24] A few examples with circumfixed forms (e.g. [ge|rihte]*hûs* 'courthouse', [ge|burt]*muoter* 'birth mother') are attested in MHG dictionaries. In the (E)NHG corpus data, the pattern is present, independent of its origin: 33 tokens (9 different first constituents) are based on former prefixed verb stems (e.g. 10a)[25], and 61 tokens (11 different first constituents) bear a nominal prefix (e.g. 10b) – albeit almost all of them examples of the former circumfix.

(10) a. *beyschlaff weiber* 'common-law wives' (1557), *Ablaß|krämer* 'seller of indulgences' (1588), *Geburt|s|zeit* 'time of birth' (1617), *Anfang|s|búchstaben* 'initial letters' (1648), *Andacht|s=Blick* 'devoted gaze' (1684), *Besitz=Nehmung* 'appropriation, seizure' (1708)

     b. *gericht|ß|costung* 'court fees' (1507), *Gesang-Buch* 'hymn-book' (1706), *Geblüt=Folge* 'succession line' (1708), *Gebett|s=Puncten* 'parts of a prayer' (1708), *Unglück|s|stiffterin* '(female) causer of misfortune' (1708)

The pattern is ubiquitous in present-day German. It was probably furthered by the growing usage frequency of stem conversions in the 17th and 18th century and some new back-formations, both due to prescriptivist abhorrence of *ung*-derivatives (Hartmann 2016: 172).

4.2.3. Growing complexity: summary

Overall, the analysis of restrictions on structural complexity of compound constituents yields the following picture:

---

23 The closing part of the circumfix is later dropped in many cases, so the form resembles a prefix *Ge-* (OHG *girihti* > NHG *Gericht* 'court of law').

24 However, cases which might be syntactic phrases do occur: *antlaz(e) tag* 'Maundy Thursday', *vor dem auffert. tag* 'before (the) Ascension Day'.

25 For a potential connection with simplex stem conversions, see Kopf (2018b: Kap. 11.7.3).

Tab. 2: Restrictions on morphological complexity of compound constituents in OHG, MHG, and NHG.

	OHG	MHG	NHG
N1 = compound	✓ (mostly in glosses)	✓	✓
N2 = compound	✓ (but rare)	✓	✓
N1 = suffixed	✗ no transparent/ productive suffixes possible	✗ no transparent/ productive suffixes possible (exception: *-er(e)*)	✓
N1 = prefixed	✗	✗ (only very few cases of *ge-*)	✓
N2 = prefixed/suffixed	✓	✓	✓

Tripartite compounds have gradually expanded, but they were always possible. The increase in both types and tokens is most likely due to the growing need for a more specialized vocabulary as the use of German becomes more common in all kinds of written texts. The case is different for compounds that contain a derivative: While derivatives have always been used as second constituents, their use as first constituents only came about in ENHG and has expanded since. The input restriction on derivatives as first constituents in OHG and MHG was rather abruptly dropped in ENHG. Specialization in vocabulary might also play a role, but due to the speed of the change, it is unlikely to be the only reason. I have also shown that the change is mostly independent from increasing productivity of the respective derivational pattern. I will now offer a different explanation for the change. It hinges on the observation that the loss of this restriction coincides with the genesis of a new compounding pattern derived from syntax. I assume that in this case, the correlation is due to a causal relation between the two.

4.2.4. Compounding patterns

Productivity of the new compounding type must be further differentiated. The pure increase in linked compounds cannot tell us whether these are cases of univerbation stemming from genitive constructions or cases of direct word-formation. While quantitative productivity can be computed for the former, this is not truly productivity in a morphological sense: The numbers do not prove the existence of a compounding pattern. They simply show lexical growth in a special part of the vocabulary. There are, however, a few systematic cases that can be taken as proof of an independent word-formation pattern; they could theoretically be used to determine the terminus ante quem for morphological productivity. All are either mismatches between semantics

and form of the compounding stem form (first constituent and LE, cf. Fuhrhop 1998), or formal mismatches between inflection and word-formation. While they are often cited in literature (e.g. Demske 2001, Meibauer et al. 2015), their usefulness in actual corpus texts has not been tested before.

There are two kinds of mismatches between semantics and form: The compounding stem form in (11) matches a genitive in the original paradigm (cf. Kopf 2018b: 140–141).

(11)   *Bischof|s|konferenz* 'episcopal conference (verb. bishop-LE-conference)'

However, its number is semantically wrong: If *Bischof|s|konferenz* were a case of univerbation, the preceding genitive construction would have been *Bischofs Konferenz* 'bishop's conference'. This is impossible, as the referent is a meeting of several bishops. Speakers using a genitive construction to convey this sense would have said *(der) Bischöfe Konferenz* '(the.GEN) bishops' conference', resulting in a hypothetical univerbation *Bischöf|e|konferenz*, not *Bischof|s|konferenz*. The mismatch between form and semantics shows that the linking-*s* in *Bischof|s|konferenz* must be due to a word-formation process. The phenomenon is not very frequent, and no such cases are attested in the Mainz Corpus. A non-exhaustive search in DTA,[26] however, yields the following examples for my time period:

(12) a. 1614: *häring|s|fang* 'catch of herrings (verb. herring|LE|catch)'
      b. 1626: *Wolff|s|jagt* 'wolf hunt (verb. wolf|LE|hunt)'
      c. 1627: *Dieb|s|hauffen* 'pile of thieves (verb. thief|LE|pile)'
      d. 1640: *Strömling|s-fang* 'catch of Baltic herrings (verb. Baltic herring|LE|catch)'
      e. 1659: *Mann|es|hauffe* 'pile of men (verb. man|LE|pile)'
      f. 1660: *Feind|es-Hauffen* 'pile of enemies (verb. enemy|LE|pile)'
      g. 1682: *Hunds-Zucht* 'dog breeding (verb. dog|LE|breeding)'

All these cases show linking elements that do not fall in line with the semantics of the first constituent: A *Diebshauffen* is a pile of thieves, but contains linking-*s*, which is derived from a genitive singular. It is therefore impossible that the compound is a case of univerbation from a genitive construction; a plural (e.g. *der Dieb|e Hauffen*) would have been used. The linking-*s* must be part of a word-formation process. We can thus conclude that a productive

---

26 These data were collected by Andreas Klein (Mainz), mostly by searching for collective second constituents like *Haufen* 'pile' preceded by a linking element. Only the first attestation for each first constituent are taken into account.

compounding pattern with linking elements is already in place by the early 17th century.

The second mismatch between semantics and form is based on the case relation between the two constituents. It is usually argued that copulative compounds cannot be based on a genitive construction because a genitive relation between the two constituents is not conceivable: A genitive relation is one of subordination while a copulative compound shows coordination (e.g. Demske 2001: 311). If copulative compounds like NHG *Hosenrock* 'pant skirt, pair of culottes (verb. trouser|LE|skirt)' contain linking elements, this must therefore be due to morphological productivity. However, almost all historical examples are problematic. Demske's (2001: 311) cases like *Baur|s|man* 'farmer (verb. farmer|LE|man)' should not be used because the relation is one between hyponym and hypernym. As is shown in Kopf (2018b: 178–180), parallel genitive constructions do actually exist at the time. In the only linked example found in the Mainz Corpus, *Schalk|s|narren* 'fools (verb. joker|LE|fool|PL)', the first constituent might as well be based on the abstract meaning of *Schalk*, 'waggishness'.[27]

The last and most useful case to be considered here is that of the so-called "non-paradigmatic linking-*s*" as in (13):[28]

(13)   *Vormundschafft|s Zweck* 'purpose of a guardianship (verb. guardianship|LE|purpose)'

Compounds with feminine first constituents using linking-*s* are non-paradigmatic because the *s*-genitive is (and was) exclusively used for neutral und masculine nouns. These cases can therefore not be based on univerbation; the linking element must have been included when a regular compounding pattern was applied. Non-paradigmatic linking-*s* in NHG is mostly found with derivatives in -*heit/keit*, -*ion*, -*(i)tät*, -*sal*, -*schaft* and -*ung* and there are

---

27 For the sake of completion, I will shortly point to another possible indicator of morphological productivity that is irrelevant for my corpus data. These are first constituents which only appear after all possible reanalysis contexts disappeared due to the positional change of the genitive attribute, i.e. in later NHG, usually loans. One of the (very few) examples is *Training|s|anzug* 'tracksuit (verb. training|LE|suit)'. As univerbation may still be ongoing during the period of my investigation, linked compounds with recent loans can by definition not be used in my Corpus data to determine productivity. They are, however, useful for addressing present-day productivity questions (see Kopf 2018b: Kap. 10.9.2).

28 Other cases of non-paradigmatic linking elements came about because the simplex changed its inflectional paradigm while the compound kept the old, fossilized element. They can, naturally, not be used to determine productivity (see also Kopf 2018b: Kap. 5.1.1).

almost no cases in which these suffixes are not followed by linking-*s*. On the other hand, only 4.3% of all non-paradigmatically linked compound types (n=443) from the Wortwarte Corpus are simplexes (Kopf 2018b: Kap.12.1). Their first constituents form a small, closed group. The non-paradigmatic linking-*s* is thus closely tied to certain suffixes or lexemes; when it is used in new compounds, it is usually kept. However, it does not spread to new first constituents. This shows, unsurprisingly, that linked compounds in NHG follow a morphological pattern, although it points to low productivity of the linked pattern.

It is more instructive to analyze the phenomenon in the Mainz Corpus when it was still new and spreading. The first examples can be found in 1590, and the number of different types (i.e. the realized productivity of the pattern) increases strongly in 1710 (Fig. 5):
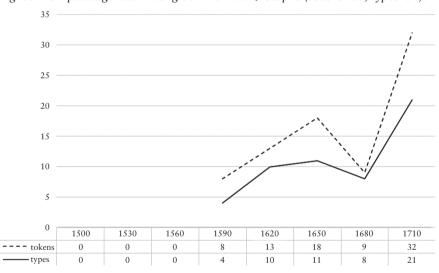
Fig. 5: Non-paradigmatic linking-s in the Mainz Corpus (tokens=80, types=54).

	1500	1530	1560	1590	1620	1650	1680	1710
---- tokens	0	0	0	8	13	18	9	32
—— types	0	0	0	4	10	11	8	21

The non-paradigmatic linking-*s* is closely connected to a number of suffixes and lexemes from the beginning (see below), but still shows some variation with unlinked forms. How and why this changed will be addressed in the next section.

Linking-*(e)n*, the only other linking element of any consequence, shows almost no innovative use (but see Klein 2015). It is firmly tied to the respective inflectional paradigm. It is the default linking element with mixed feminines and weak masculines – the latter consisting almost exclusively of a small group of animate nouns (for details, see Kopf 2018b: 43). Linking-*(e)n* is therefore not relevant to determine the productivity of the linked compounding pattern.

We can conclude that a morphologically productive compounding pattern with linking elements was present from 1590 in the Mainz Corpus. (For an overview of earlier attestations in other sources, see Kopf 2018b: Kap. 12.2.) This is mostly shown by the non-paradigmatic linking-*s*, but supported by the occurrence of semantically "wrong" compounding stem forms by the beginning of the 17th century.

5. Compounding productivity: An explanation

Thus far, the following has been shown:

1. Transparent derivatives were not used as first constituents in OHG and MHG compounds but are attested in ENHG compounds, where they increase from 1590. This happened rather abruptly, compared with the slow extension of tripartite compounds that were attested from the beginning. Derivatives are completely unremarkable in NHG compounds.
2. In ENHG, a new compounding pattern with linking elements arose due to univerbation and reanalysis of prenominal genitive constructions. This pattern is still in use today, while its source, univerbation, has run dry due to the positional change of the genitive attribute in (E)NHG.
3. One of these linking elements, -*s*-, was extended to non-paradigmatic use in combination with feminine nouns. This can be taken as proof of a word-formation pattern independent from syntax. The non-paradigmatic linking-*s* is strongly associated with a number of suffixes in NHG.

In addition, it is important to note that NHG paradigmatic linking-*s* (used with masculine and neuter nouns) also shows a remarkable affinity for complex first constituents: Apart from a few suffixes (-*tum*, -*ling*), it also combines with many prefixed nouns, although there is somevariation (Kopf 2018b: 367, see also Augst 1975: 113–120). In total, 81.3% of prefixed masculine and neuter compound types with a monosyllabic basis tested in a comprehensive corpus study showed linking-*s* (Kopf 2017, Kopf 2018b: Kap. 14.2.2).[29] The Mainz Corpus contains only 26 prefixed cases for all grammatical genders; however, 19 of these use linking-*s* (73%).

These observations lead to a plausible explanation for the increased productivity of compounding in general, which is mostly due to loss of restrictions: The traditional pattern did not include derivatives as first constituents.

---

29 Feminine nouns, for which -*s*- is non-paradigmatic, still used it in 53.8% of the prefixed cases.
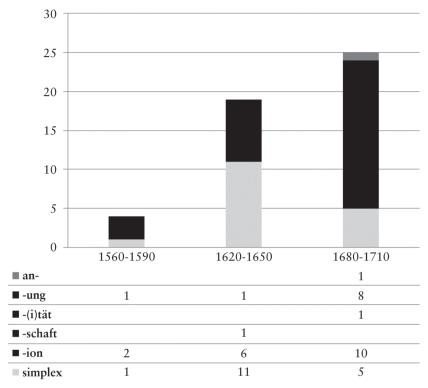
When speakers started to reanalyze genitive constructions as compounds in ENHG,[30] this changed. As genitive constructions, being syntactic units, did not show formal restrictions for their genitive attribute, univerbation "smuggled" structures into morphology that had not been used up to this point, especially derivatives. The morphological restriction on derivatives in compounds fell.

In many cases, these derivatives brought a former genitive suffix with them, which then became a paradigmatic linking element. In most cases, this was -*s*-. Of course, this not only happened with derivatives but also with simplexes, e.g. *des leib\s erben > die Leib\s\erben*. The important difference is that morphologically simple first constituents were already possible in the old, unlinked compounding pattern. The linking-*s* therefore appeared much more regularly with derivatives than with simplexes. This quickly turned the linking-*s* into a marker for internal complexity: Where it appeared, a complex first constituent was very likely. This might have been a useful segmentation aid for speakers that were now confronted with much more complicated compounds.

This marker was quickly applied to compounds with derivative feminine first constituents. As most of them did not show an inflectional suffix in the genitive singular, complex constituents would have looked like old compounds. A closer look at the internal structure of feminine first constituents in the Mainz Corpus (Fig. 6) shows, however, that they make up a considerable part of the non-paradigmatic cases from the beginning.

---

30 Several explanations have been brought forward as to why univerbation becomes so common in ENHG, e.g. by Pavlov (1983) and Demske (2001). See Kopf (2018b: Kap. 8.6) for an evaluation and some new aspects.

Fig. 6: *First attestations of compounds with non-paradigmatic linking-s by*
morphological complexity in the Mainz Corpus (n=48). For ease of
comparison, all suffixed first constituents have been assigned the same
color; individual numbers can be seen in the table.

	1560-1590	1620-1650	1680-1710
■ an-			1
■ -ung	1	1	8
■ -(i)tät			1
■ -schaft		1	
■ -ion	2	6	10
▨ simplex	1	11	5

The non-paradigmatic linking-*s* is not used as invariably as today when en-
countering suffixes. However, Fig. 7 shows a remarkably early affinity: Non-
paradigmatic cases dominate from the beginning for first constituents in *-ung*,
-(i)tät, *-schaft* and *-ion*, while cases without linking-*s* persist for the whole
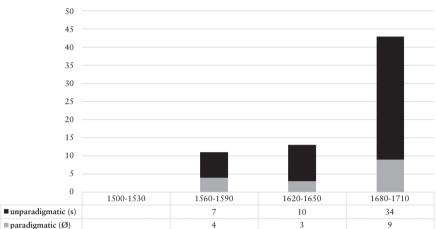period of investigation, but to a much smaller degree.

Fig. 7: *Compounding patterns of first constituents in -ung, -(i)tät, -schaft and -ion in the Mainz Corpus (tokens, n=67), the other suffixes are not yet attested. Data combined from Kopf (2018b: 313).*

	1500-1530	1560-1590	1620-1650	1680-1710
■ unparadigmatic (s)		7	10	34
■ paradigmatic (Ø)		4	3	9

Thus, it seems reasonable to assume that the spread of the non-paradigmatic linking-*s* was motivated by the unconscious will to mark derivatives in compounds, segmenting them for listeners and readers (on (E)NHG compound segmentation, see also Kopf 2017). The linking element is an even better signal here than with paradigmatic cases; as it appears only in compounds, confusion with syntactic structures is completely impossible.

The spread of the non-paradigmatic linking-*s* shows a temporary functionalization of linking-*s* in general as a complexity marker (for factors that might have helped it along additionally, see Kopf 2018b: Kap. 12.3). I consider this as a case of exaptation (Kopf 2018b: Kap. 14.4). When the new compounding pattern was firmly integrated into word-formation and the use of derivatives was no longer restricted, the linking-*s* was no longer needed. The exaptation process was interrupted and linking-*s* stopped expanding to new lexemes. It fossilized in fixed compounding stem forms.

6. Conclusion

The present paper has shown why and how productivity in German compounds can and should be measured: The status quo would hardly raise any interest in the topic, as compounds are obviously **the** means of word-formation in NHG. However, OHG and MHG show a different picture. Compounds were formed and used, but the morphological complexity of the first constituent was heavily restricted, especially in the case of derivatives. As conditions have obviously changed since then, this warrants a detailed investigation.

Quantitative measurements of productivity that are regularly employed in derivational morphology have been extended to compounding. Realized productivity shows a heavy increase in overall compounding as well as in patterns differentiated by linking pattern. Potential productivity shows no significant change, but it is shown that this is most likely due to corpus size and fluctuations in token numbers. It has become clear that the quantitative measures, while yielding interesting results, cannot stand for themselves and can hardly be used to generate any further hypotheses. This can be done much more elegantly by a qualitatively motivated approach. Comparing OHG and NHG yields morphological complexity as a suspect. It can in fact be shown that derivationally complex first constituents became possible and quickly more frequent in the 16th century. Tripartite compounds, on the other hand, have always existed. Their usage frequency rose constantly over the centuries, but is still only moderately high today. Concentrating on derivatives, it can be shown that their rise in compounds coincides with univerbation and reanalysis of compounds from genitive constructions. Syntax was a back door that enabled derivatives to enter morphology, removing a long-standing restriction on first constituents. As masculine and neuter derivatives usually brought along a former genitive-*s*, the new, more complex pattern was overtly marked. The linking-*s* was partly functionalized as a marker of internal derivational complexity, even spreading to feminine nouns where it was non-paradigmatic. It was, however, only needed as a temporary crutch: When the new compounding type was firmly integrated into word-formation, the spread of the linking-*s* was halted. Since the beginning of the 20th century, very few isolated cases of analogy are attested (e.g. *Training|s|hose* 'tracksuit (verb. training|LE|suit)').

References

Aronoff, Mark & Nanna Fuhrhop. 2002. Restricting suffix combinations in German and English: Closing suffixes and the monosuffix Constraint. *Natural Language & Linguistic Theory* 20. 451–490.

Augst, Gerhard. 1975. *Untersuchungen zum Morpheminventar der deutschen Gegenwartssprache*. Tübingen: Narr.

Baayen, R. H. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics* (Handbücher zur Sprach- und Kommunikationswissenschaft 2), 900–919. Berlin & New York: de Gruyter.

Bauer, Laurie. 2005. Productivity: Theories. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation* (Studies in natural language and linguistic theory 64), 315–334. Dordrecht: Springer.

Bergmann, Rolf & Dieter Nerius. 1998. *Die Entwicklung der Großschreibung im Deutschen von 1500 bis 1700*. Heidelberg: Universitätsverlag Winter.

Carr, Charles T. 1939. *Nominal compounds in Germanic*. London: Oxford University Press.

Demske, Ulrike. 1999. Case Compounds in the history of German. In Matthias Butt & Nanna Fuhrhop (eds.), *Variation und Stabilität in der Wortstruktur. Untersuchungen zur Entwicklung, Erwerb und Varietäten des Deutschen und anderer Sprachen*, 150–176. Hildesheim: Olms.

Demske, Ulrike. 2001. *Merkmale und Relationen. Diachrone Studien zur Nominalphrase des Deutschen*. Berlin & New York: de Gruyter.

Donalies, Elke & Noah Bubenhofer. 2011. *Tagtraum, Tageslicht, Tagedieb*: Ein korpuslinguistisches Experiment zu variierenden Wortformen und Fugenelementen in zusammengesetzten Substantiven (Arbeitspapiere und Materialien zur deutschen Sprache 42). Mannheim: Institut für Dt. Sprache.

Durrell, Martin, Astrid Ensslin & Paul Bennett. 2007. The GerManC project. *Sprache und Datenverarbeitung* 31. 71–80.

Fleischer, Wolfgang & Irmhild Barz. 2012. *Wortbildung der deutschen Gegenwartssprache* (de Gruyter Studium), 4th edn. Tübingen: Niemeyer.

Fuhrhop, Nanna. 1996. Fugenelemente. In Ewald Lang & Gisela Zifonun (eds.), *Deutsch – typologisch*, 525–550. Berlin & New York: de Gruyter.

Fuhrhop, Nanna. 1998. *Grenzfälle morphologischer Einheiten*. Tübingen: Stauffenburg.

Grimm, Jacob. 1826. *Deutsche Grammatik. Zweiter Theil* (2). Göttingen: Dieterich.

Gröger, Otto. 1911. *Die althochdeutsche und altsächsische Kompositionsfuge. Mit Verzeichnis der althochdeutschen und altsächsischen Composita*. Zürich: Zürcher und Furrer.

Günther, Hartmut. 1981. N+N: Untersuchungen zur Produktivität eines deutschen Wortbildungstyps. In Leonhard Lipka & Hartmut Günther (eds.), *Wortbildung*, 258–280. Darmstadt: Wissenschaftliche Buchgesellschaft.

Harlass, Gertrude & Heinz Vater. 1974. *Zum aktuellen deutschen Wortschatz*. Tübingen: Narr.

Hartmann, Stefan. 2016. *Wortbildungswandel. Eine diachrone Studie zu deutschen Nominalisierungsmustern*. Berlin & New York: de Gruyter.

Kempf, Luise. 2016. *Adjektivsuffixe in Konkurrenz. Wortbildungswandel vom Frühneuhochdeutschen zum Neuhochdeutschen* (Studia Linguistica Germanica 126). Berlin & New York: de Gruyter.

Klein, Andreas. 2015. *Instrumentenklänge und Temperamentszustände – Fälle zwischen -en und -s*: Zur unparadigmischen en-Fuge in Lehnwortkomposita: Genese und Distribution. Mainz: Johannes Gutenberg-Universität Hausarbeit.

Klos, Verena. 2011. *Komposition und Kompositionalität*: Möglichkeiten und Grenzen der semantischen Dekodierung von Substantivkomposita (Reihe Germanistische Linguistik 292). Zugl.: Siegen, Univ., Diss., 2009. Berlin & New York: de Gruyter.

Kopf, Kristin. in preparation. Erkennung von Komposita im Frühneuhochdeutschen. In Christian Schwarz & Christine Ganslmayer (eds.), *Historische Wortbildung. Theorie - Methoden - Perspektiven*. Hildesheim: Olms.

Kopf, Kristin. 2017. Fugenelement und Bindestrich in der Compositions-Fuge. Zur Herausbildung phonologischer und graphematischer Grenzmarkierungen in (früh)neuhochdeutschen N+N-Komposita. In Renata Szczepaniak, Nanna Fuhrhop &

Karsten Schmidt (eds.), *Sichtbare und hörbare Morphologie*, 177–204. Berlin & New York: de Gruyter.

Kopf, Kristin. 2018a. 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 Horst J. Simon, Christian Zimmer & Tanja Ackermann (eds.), *Germanic Genitives*. Amsterdam & Philadelphia: Benjamins.

Kopf, Kristin. 2018b. *Fugenelemente diachron. Eine Korpusuntersuchung zu Entstehung und Ausbreitung der verfugenden N+N-Komposita*. Berlin & Boston: de Gruyter.

Krahe, Hans & Wolfgang Meid. 1967. *Germanische Sprachwissenschaft* (Sammlung Göschen 1218). Berlin: de Gruyter.

Kürschner, Sebastian. 2003. *Von Volk-s-musik und Sport-Ø-geist im Lemming-Ø-land – af folk-e-musik og sport-s-ånd i lemming-e-landet: Fugenelemente im Deutschen und Dänischen – eine kontrastive Studie zu einem Grenzfall der Morphologie*. Freiburg: Albert-Ludwigs-Universität Magisterarbeit.

Meibauer, Jörg, Ulrike Demske, Jochen Geilfuß-Wolfgang, Jürgen Pafel, Karl-Heinz Ramers, Monika Rothweiler & Markus Steinbach. 2015. *Einführung in die germanistische Linguistik*, 3rd edn. Stuttgart: Metzler.

Meineke, Eckhard. 1991. Springlebendige Tradition. Kern und Grenzen des Kompositums. *Sprachwissenschaft* 16(1). 27–88.

Munske, Horst H. 2009. Was sind eigentlich ,hybride' Wortbildungen? In Peter O. Müller (ed.), *Studien zur Fremdwortbildung* (Germanistische Linguistik 197/198), 223–260. Hildesheim: Olms.

Nitta, Haruo. 1987. Zur Forschung der ,uneigentlichen' Zusammensetzung im Frühneuhochdeutschen. *Zeitschrift für deutsche Philologie* 106(3). 400–416.

Nübling, Damaris & Renata Szczepaniak. 2008. On the way from morphology to phonology: German linking elements and the role of the phonological word. *Morphology* 18. 1–25.

Nübling, Damaris & Renata Szczepaniak. 2013. Linking elements in German. Origin, Change, Functionalization. *Morphology* 23(1). 67–89.

Olsen, Susan. 2001. Copulative compounds: A closer look at the interface between syntax and morphology. In Geert Booij, Jaap van Marle, Stephen Anderson, Mark Aronoff, Mark Baker, Laurie Bauer, Rudie Botha, Joan Bybee, Andrew Carstairs-McCarthy, Greville Corbett, Wolfgang Dressler, Martin Haspelmath, Jack Hoeksema, Rochelle Lieber, Peter Matthews, Franz Rainer, Sergio Scalise, Henk Schultink & Andrew Spencer (eds.), *Yearbook of Morphology 2000* (Yearbook of Morphology), 279–320. Dordrecht: Springer Netherlands.

Ortner, Lorelies, Elgin Müller-Bollhagen, Hanspeter Ortner, Hans Wellmann, Maria Pümpel-Mader & Hildegard Gärtner. 1991. *Deutsche Wortbildung: Typen und Tendenzen in der Gegenwartssprache. Eine Bestandsaufnahme des Instituts für deutsche Sprache, Forschungsstelle Innsbruck* (4 Substantivkomposita (Komposita und kompositionsähnliche Strukturen 1)). Berlin & New York: de Gruyter.

Pavlov, Vladimir M. 1983. *Von der Wortgruppe zur substantivischen Zusammensetzung* (Zur Ausbildung der Norm der deutschen Literatursprache (1470–1730) 4). Berlin: Akademie-Verlag.

Plag, Ingo. 1999. *Morphological Productivity*: *Structural Constraints in English Derivation* (Topics in English Linguistics [TiEL] 28). Berlin & New York: de Gruyter.

Pümpel-Mader, Maria, Elsbeth Gassner-Koch & Hans Wellmann. 1992. *Adjektivkomposita und Partizipialbildungen*: *Komposita and kompositionsähnliche Strukturen 2* (Sprache der Gegenwart 80). Berlin & New York: de Gruyter.

Rainer, Franz. 2005. Constraints on productivity. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation* (Studies in natural language and linguistic theory 64), 335–352. Dordrecht: Springer.

Reagan, Sherman C. 1981. *Compound nouns in the Luther Bible and some of its printed predecessors*. Madison, Wisconsin: University of Wisconsin-Madison dissertation.

Salus, P. H. 1963. *Nominal composition in the Indo-European languages*. New York Dissertation.

Scalise, Sergio & Antonietta Bisetto. 2009. The Classification of Compounds. In Rochelle Lieber & Pavol Stekauer (eds.), *The Oxford Handbook of Compounding*, 49–82. Oxford: Oxford University Press.

Schmid, Hans-Jörg. 2011. *English morphology and word-formation*: *An introduction* (Grundlagen der Anglistik und Amerikanistik 25), 2nd edn. Berlin: Schmidt.

Solling, Daniel. 2011. Zur Problematik der Unterscheidung zwischen pränominalem Genitivattribut und getrennt geschriebenem Substantivkompositum im Frühneuhochdeutschen. In Hans U. Schmid & Arne Ziegler (eds.), *Historische Semantik* (Jahrbuch für germanistische Sprachgeschichte 2), 294–311. Berlin & Boston: de Gruyter.

Solling, Daniel. 2012. *Zur Getrennt-, Zusammen- und Bindestrichschreibung von Substantivkomposita im Deutschen (1550–1710)* (Acta Universitatis Upsaliensis: Studia Germanistica Upsaliensia 57). Zugl.: Uppsala, Univ., Diss., 2012.

Szczepaniak, Renata. 2016. Is the development of linking elements in German a case of exaptation? In Muriel Norde & Freek van de Velde (eds.), *Exaptation and language change* (Current Issues In Linguistic Theory Volume 336), 317–340. Amsterdam, Philadelphia: John Benjamins.

Wellmann, Hans, Nikolaus Reindli & Annemarie Fahrmeier. 1974. Zur morphologischen Regelung der Substantivkomposition im heutigen Deutschen. *Zeitschrift für deutsche Philologie* 93. 358–378.

Wilmanns, Wilhelm. 1896. *Deutsche Grammatik. Gotisch, Alt-, Mittel- und Neuhochdeutsch. Bd. 2: Wortbildung*. Straßburg: Trübner.

Dr. Kristin Kopf
Westfälische Wilhelms-Universität Münster
Germanistisches Institut
Schlossplatz 34
48143 Münster
kristin.kopf@uni-muenster.de