

Compounding in German and English

A quantitative translation study

Thomas Berg

University of Hamburg

German is well-known for its propensity for nominal compounding. This claim is put on a firmer empirical footing by means of a bidirectional translation study between German and English. The difference between the two languages crystallizes in the competition between compounds and phrases. Two complementary asymmetries emerge: first, German compounds are more frequently translated by English phrases than English compounds by German phrases; second, English phrases are more frequently translated by German compounds than German phrases by English compounds. An extension to other word classes shows that the compounding bias in German is not restricted to nouns. It is tentatively argued that the token frequency of word classes plays a role in the emergence of compound propensity. The heavier use of nouns and adjectives in German than in English might be partly responsible for the higher rate of nominal and adjectival compounding in the former than the latter language.

Keywords: compounding, English, German, type frequency, token frequency

1. Introduction

Long is the list of works claiming that compounding, in particular nominal compounding, is more heavily made use of in German than in English (e.g. Potter 1957: 95, Marchand 1969: 26, Wandruszka 1969: 139, Bauer 1978: 16, Zimmer 1981: 252, Donalies 2004: 55, Gunkel & Zifonun 2008: 286, not to mention Mark

Twain's polemic 1880: 397).^{1, 2} What is common to these and other studies is that their claims are entirely impressionistic, being based on individual examples. No care is taken to ensure the veracity of the statements made. An extreme case is Donalies (2008) who formulates sweeping generalizations without adequate empirical support. For instance, she contends that German tends to use syntactic phrases where other Germanic languages, including English, use compounds. It is obvious that this conjecture plainly contradicts what may be regarded as the received view.

In contrast to the above, the list of reports presenting reliable data is much shorter. Stöhr (1987) carried out a picture-naming experiment designed to elicit nominal compounds. Under identical experimental conditions, German participants produced compounds in 52.3% of cases while English participants uttered them in only 34.1% of cases. Thus, German speakers used compounds 1.5 times more often than English speakers. This is a statistically significant difference ($p < 0.001$).

Among the 2,500 most frequent words in the British National Corpus (BNC) as well as the DWDS ("Digitales Wörterbuch der deutschen Sprache") Core Corpus, Sanchez (2008) found 133 compounds on the German list but only 50 on the English list. This ratio of 2.7:1 is higher than that reported by Stöhr. A breakdown of these compounds by word class is not provided. Hence, it is not clear whether this difference between English and German is carried by a single word class (e.g. nouns) or evenly by all word classes allowing compounding. It is also unknown whether this difference interacts with frequency, i.e. whether this difference is smaller or larger in the highest frequency range of the lexicon.

In a large-scale study investigating the extent of nominal-compound use in English and German as a function of compound size, Berg, Helmer, Neubauer & Lohmann (2012) found that compounds occurred twice as often in written German as in written English.³ This disparity held for both type and token frequency as

1. Schlücker (2012: 2) joins the chorus by stating that German is particularly prone to compounding compared to other languages, by which she apparently means English and Dutch. The title of the volume in which her article appears ("Das Deutsche als kompositionsfreudige Sprache") suggests the same interpretation although it is not clear in relation to which languages German can be regarded as showing a proclivity to compounding.

2. Sadeniemi, Kettunen, Lindh-Knuutila & Honkela (2008) note that German ranks highest on a scale of morphological complexity among 21 languages spoken in the European Union and suspect that German owes its top position to its penchant for compounding.

3. As orthography was ignored in the definition of compounds, this result cannot be put down to different spelling conventions in the two languages.

well as for different text genres. No comparable analyses of the spoken language in naturalistic settings have so far been published.

Given that German is more propitious to compounding than English, English can be expected to use alternative means where German uses compounds. In fact, such claims have been variously made in the relevant literature. Translational equivalents of German nominal compounds include nouns postmodified by PPs, adjective-noun phrases and non-compounds (i.e. monomorphemic and derived words) (e.g. Wandruszka 1968, Klinge 2005 and Donalies 2008).

The problem with these claims is the same as above: they are erected on intuitions or experience of linguists and translators rather than on serious quantitative analyses (but see Ermlich 2004 for a notable exception). It is therefore not known how frequent these alternative means of expression are and what a complete list of English analogues of German compounds would look like.

A further gap in our knowledge derives from the unidirectionality of previous publications (but see Djirackor 1982). The perspective of accepting German as the standard against which other languages are compared represents only one side of the coin. The reverse side needs to be given equal attention. Only if it can be demonstrated that non-compound structures in English tend to be rendered as compounds in German can a convincing case for a difference between the two languages be made.

Once the cross-language difference has been firmly established, we have to tackle the question of why German and English are differentially susceptible to compounding. The linguistic literature is noticeably silent on this score. The few explanations that have been proposed are riddled with problems. Berg et al. (2012) suggested a link between the propensity for compounding and the length of monomorphemic words. The underlying logic was that monomorphemic words create a schema that is applied to compounds. Given the truism that compounds are longer than simple items, a language with longer monomorphemic words is arguably more receptive to compounding than a language with shorter words. In support of their hypothesis, Berg et al. calculated the average length of non-compound words and indeed found that German lexical items were significantly longer than their English counterparts.

Unfortunately, Berg et al.'s argument does not hold up in its present form. It predicts that the greater the length of monomorphemic words, the larger the role of compounding. In a language with very long non-compound words, the simple juxtaposition of two such words would on average create compounds twice the size of non-compounds. Specifically, the combination of two four-syllable words would give rise to an octosyllabic compound. Clearly, such juggernauts would go against the grain of the economy of language. It may accordingly be predicted that the *smaller* the phonological size of monomorphemic items in a given language, the

greater its proclivity to compounding.⁴ This is exactly the opposite of Berg et al.'s claim.⁵ The fact that two opposite predictions can be formulated with more or less equal justification casts serious doubt on either or both. Since it is not known which (if any) of the two is the correct one, this proposal is of limited explanatory value.

Another explanation offered by Berg et al. revolves around potential interactions between compounding, derivation, and inflection. German is known to have a higher degree of syntheticity than English (Greenberg 1960). This difference is partly brought about by the stronger use of inflectional suffixes and derivational prefixes in German (Berg 2016). Supposing the syntheticity index was exclusively based on inflection and derivation, it would allow one to predict the rate of compounding. At first sight, this approach looks promising: the higher rate of compounding in German matches its higher rate of inflection and (to a lesser extent) derivation.

However, there are two weaknesses in this argument. Correlations should not be mistaken for causes. Even if this general point was ignored and we contented ourselves with establishing correlations, the strength of the correlation is currently unknown. The establishment of such a correlation would require determining propensity indices for compounding, derivation, and inflection across a good number of languages and ascertaining to what extent one index can predict the other. This has yet to be done. What we do know is that there is a certain independence between the morphological subsystems. Romance languages infrequently use compounding but are comfortable with derivation. Isolating languages such as Cantonese and Tok Pisin by definition disallow inflection but are less opposed to compounding (e.g. Matthews & Yip 1994, Mühlhäusler 1979). Thus, the relatively heavy use of inflections in German cannot be straightforwardly taken as an explanation of the compounding advantage in this language.

Confronting this state of affairs, the present study pursues a two-fold objective. It seeks to put the assumed cross-language difference in compound propensity on a firmer empirical footing by comparing German and English more directly than has hitherto been done. Moreover, the empirical picture will be extended from

4. I owe this argument to Katerina Stathi.

5. It might seem that there is at least indirect support for this hypothesis. Several studies have argued for an inverse correlation between the phonological or orthographic length of base words and compound propensity. The longer the base, the lower the number of compounds that it is a part of (e.g. Krott, Schreuder & Baayen 1999, Fan & Altmann 2007, Deng 2009). However, although there is clearly an effect of language economy, it does not necessarily contradict Berg et al.'s (2012) claim that a language with longer bases is more propitious to compounding than a language with shorter bases. It is certainly possible that the inverse correlation holds in all languages regardless of their compound propensity.

nouns to other word classes. This allows us to assess whether variable compound propensity is restricted to nouns or extends to other word classes. On a more theoretical level, it will be examined whether a link can be established between compound propensity and the frequency of particular word classes.

2. Method

In an effort to align the two languages to the greatest possible degree, a translation study was designed. As emphasized by Doherty (e.g. 1993), translations which are not only grammatically but also stylistically adequate bring into focus language-particular preferences and are therefore apt to reveal properties that are characteristic of individual languages. It is no wonder then that the translation method looms large in contrastive linguistics (see e.g. Ahlemeyer & Kohlhof 1999).

The translation method has the additional advantage of providing a *tertium comparationis* at the semantic level. This is James's (1980) and Krzeszowski's (1990) notion of translation equivalence, which is not to be equated with synonymy. The idea is that translators may be assumed to have made the best choice in their attempt to render the content of the original as faithfully as possible.

Note that the translation method may also be argued to be flawed. It goes without saying that the decisions made by the translators have to be considered final. In the event of "unexpected" decisions, it is out of the question to replace the "unexpected" by an expected translation. For example, the German (G.) word *Gitarrenspieler* was translated into English (E.) as *busher with a guitar* rather than *guitar player*. There is an obvious structural difference between the two translations. The German compound is translated by a postmodified simple noun rather than a compound. Thus, the data are coloured by the individual translator's decisions. It is impossible to determine whether these choices are a matter of personal taste or motivated by text-internal constraints (as perceived by the translator). In either case, it is indefensible to question the translators' linguistic intuition and competence.

As is only too well-known, the contrastive analysis is hampered by the fact that the "same" structures in different languages are never completely alike. We therefore have to rely on criteria which are applicable to both languages even though they do not play exactly the same role in the two languages. Following Marchand (1969), a rather useful criterion is stress which distinguishes between compounds and phrases rather reliably. As a general rule, noun-noun (NN) compounds are initially stressed and phrases finally stressed in both languages. While this generalization works well for German, it is not fully reliable in English where a minority of compounds are finally stressed (e.g. *stone circle*). When stress placement was

not clear, native speakers and standard reference works such as the Dictionary of Contemporary English and Longman's Pronunciation Dictionary were consulted. The presence or absence of inflections reliably distinguishes between adjective-noun (AN) phrases and compounds in German. This distinction is more difficult to draw in English. Again, the stress criterion proves helpful. When the main stress falls on the noun (e.g. *musical instrument*), the candidate structure is treated as a phrase; however, when stress falls on the adjective (e.g. *blueprint*), it is treated as a compound. Obviously, a binary distinction between compounds and phrases cannot do justice to the fuzziness of the boundary between these two types of units in English (Bauer 1998).

A parallel, bidirectional corpus of original texts and their translations was compiled. The corpus consists of a German original text translated into English as well as an English original text translated into German. An attempt was made to select texts that were as similar as possible in terms of genre, topic and date of publication. The choice fell on detective stories. It was felt that fiction represents a type of written language which is relatively close to what is widely regarded as everyday language. Specifically, the German text by Schirach (2011) and the English text by Rankin (2007) were selected for analysis. Bibliographical details are provided before the list of references. It goes without saying that the choice of only two texts limits the generality of the results. There is no guarantee that exactly the same patterns emerge in different genres or even different texts from the same genre.

The procedure was as follows. The originals were manually searched for nominal compounds, i.e. all compounds with a nominal head. The modifier could be any word class. In both languages, the typical modifier is a noun or an adjective, with adverbs and verbs being less frequent. The adequate categorization of modifiers was not always an easy matter. For example, the modifier in *seaside resort* could be classified as a noun or an adjective. However, little hinges on this decision in the present context because the focus of this study is on compounds with nominal heads. It is a remarkable fact that heads are less susceptible to ambiguity than modifiers. Generally speaking, in order to keep the number of categories to a minimum, an attempt was made to assimilate the ambiguous cases to the most frequent types.

The texts were not only perused for nominal compounds (and how they were translated) but also for AN phrases such as G. *chirurgische Instrumente* – E. *surgical instruments*. Wandruszka (1969) notes that German compounds can be rendered as AN phrases in English (e.g. G. *Musikinstrument* – E. *musical instrument*). The partial overlap of compounds and AN phrases across languages raises the possibility that AN phrases might also be translated as compounds. The bidirectional nature of the corpus also allows one to examine whether English AN

phrases show up as compounds in German more often than German phrases as English compounds.

The ensuing analysis is type- rather than token-based. How often a particular item in the source language is translated by a particular item in the target language is of no concern in the present study. A type was defined as one specific item that was rendered identically. In other terms, when one and the same item in the source language was translated in different ways into the target language, it was counted as a separate type. For example, G. *Waschbecken* was rendered as *washbasin* on one page but as *sink* on another. This way, the non-uniqueness of translational equivalents was taken into account.

Text perusal was terminated when 500 compound types per language had been spotted. The AN phrases that occurred along the way did not enter into this calculation. The compounds could be simple or complex, i.e. consist of two or three constituents, and they were optionally preceded by an adjective that was not part of the compound proper. A distinction was made between compounds containing or being proper nouns (e.g. G. *Humboldt-Universität*, E. *Charlock Bay*) and those consisting only of common nouns. While both types were collected, only the latter type was subjected to analysis. The former was not included in the set of 1000 German and English compounds.

3. Translating nominal compounds (and phrases)

We begin by introducing the categories into which the data were sorted. Owing to the strong similarity between English and German, it was possible to set up essentially the same categories for the two languages. Obviously, this greatly facilitates the cross-language comparison. However, one minor difference is worth pointing out. It concerns the status of <s> in the two languages. Contrast (1) and (2).

- (1) G. Bootshaus 'boat house'
- (2) E. printer's ink

Whereas the /s/ in German is a compound-internal linking element without any morphosemantic function (e.g. Fabb 1998, Dressler, Libben, Stark, Pons & Jarema 2001, Koester, Gunter, Wagner & Friederici 2004), the <'s> in English is part of an NP's N construction and signals a syntactic relationship between the two nouns. The German example was therefore classified as a compound but the English example as a phrase.

The first and foremost distinction imposed on the data was that between compounds (abbreviated as C) and phrases (abbreviated as P), i.e. between

morphological and syntactic objects. Within the group of compounds, two further variables were introduced, viz. size and word class. Basically, size refers to the contrast between two-part and three-part compounds. As the word class of the modifier is variable, compounds were divided into AN and NN subsets. This decision respects the contrast between AN compounds and AN phrases, as illustrated by such celebrated examples as *a blackbird* vs. *a black bird*. These are the four categories that were extracted from the original texts in both languages.

The renditions of these categories in the other language run the full gamut from simple noun (N), compound noun (AN/NN), AN phrase, other phrase and clause. The label N is an umbrella term for bare, inflected and derived nouns. Since derivation is not the focus of this study, it appears justified to lump these groups together. As in the source language, compounds could be bipartite or tripartite in nature. The category “other phrase” consists in the main of postmodified nouns, i.e. nouns followed by PPs. These PPs can be headed by *of* or any other preposition in English and by *von* or any other preposition in German. Unlike English nouns, German nouns were often followed by a genitive phrase. Translations in which the original noun was rendered by a VP were classified as “clause”.

There was a residue of cases which did not fit gracefully into any of the above categories. These were not further analyzed. Many of them represent liberal, if not whimsical, renditions of the translator which may be perfectly justified within the context of the story but which are too heterogeneous to be of any major linguistic interest. These miscellaneous cases make up 1.8% of the translations into English and 5.4% of the translations into German.

Table 1 summarizes the results of the empirical analysis on the basis of the categories introduced above. It is preceded by eight examples illustrating all four categories of the source language and some of the structural types of rendition in the target language. The first block of four has German as the source language while the second block of four has English as the source language. For easy reference, the numbers of the categories in Table 1 are also provided in square brackets in the examples below. Numerals stand for the source-language categories, small letters for the different structural types of rendition. Arabic numerals represent the translation from German into English, Roman numerals the translation from English into German. Examples (3)–(10) do not require special comment.

- (3) [1a] G. Bücherregale -> E. bookshelves
- (4) [2d] G. Nadelstreifenanzug -> E. pinstriped suit
- (5) [3d] G. Graugänse -> E. grey geese
- (6) [4f] G. überfahrener Fuchs -> E. fox lying dead at ...
- (7) [Ie] E. job satisfaction -> G. befriedigende Arbeit

- (8) [IIe] E. car park exit -> G. Ausfahrt des Parkhauses
 (9) [IIIc] E. sweetheart -> G. Schätzchen
 (10) [IVa] E. free speech -> G. Meinungsfreiheit

Table 1. Frequency of structural types of translational equivalence

Category		Direction German→English	Category	Direction English→German	
1. NN-C (464)	a. NN-C	223 (48.1%)	I. NN-C (451)	a. NN-C	310 (68.7%)
	b. NNN-C	14 (3.0%)		b. NNN-C	18 (4.0%)
	c. AN-C	0 (0.0%)		c. AN-C	1 (0.2%)
	d. N	79 (17.0%)		d. N	78 (17.3%)
	e. AN-P	53 (11.4%)		e. AN-P	18 (4.0%)
	f. other P	88 (19.0%)		f. other P	19 (4.2%)
	g. clause	7 (1.5%)		g. clause	7 (1.6%)
2. NNN-C (27)	a. NN-C	10 (37%)	II. NNN-C (31)	a. NN-C	13 (42%)
	b. NNN-C	3 (11%)		b. NNN-C	11 (35%)
	c. N	1 (4%)		c. N	1 (3%)
	d. AN-P	5 (19%)		d. AN-P	0 (0%)
	e. other P	8 (30%)		e. other P	5 (16%)
	f. clause	0 (0%)		f. clause	1 (3%)
3. AN-C (9)	a. NN-C	1 (11%)	III. AN-C (18)	a. NN-C	0 (0%)
	b. AN-C	0 (0%)		b. AN-C	6 (33%)
	c. N	0 (0%)		c. N	12 (67%)
	d. AN-P	7 (78%)		d. AN-P	0 (0%)
	e. other P	1 (11%)		e. other P	0 (0%)
		500			500
4. AN-P (212)	a. NN-C	15 (7.1%)	IV. AN-P (1425)	a. NN-C	160 (11.2%)
	b. AN-C	0 (0.0%)		b. AN-C	8 (0.6%)
	c. N	5 (2.4%)		c. N	95 (6.7%)
	d. AN-P	166 (78.3%)		d. AN-P	981 (68.8%)
	e. other P	15 (7.1%)		e. other P	158 (11.1)
	f. clause	11 (5.2%)		f. clause	23 (1.6%)

The analysis of Table 1 begins with a look at the three compound types in German and English (1–3, I–III). It is perhaps surprising that they are fairly equally distributed across the two languages. With more than 90%, the lion's share among the three types is held by NN compounds. NNN compounds are infrequent and almost equally so in German and English. Less than 1 instance out of 15 nominal compounds is a tripartite NNN compound. AN compounds are even less frequent but more than twice as frequent in English as in German. This latter result may be taken as evidence for the claim that English distinguishes less clearly than German

between adjectives and nouns and therefore accepts adjectives more readily in a schema which is strongly dominated by nouns.

Let us focus our attention on the predominant NN compounds (1 and I) and begin with the observation that in both languages the majority option is for a compound to be translated as such. This finding contradicts Djirackor (1982: 87–88) who reports a predominance of renditions which do not preserve the structural type of the source language. To be specific, he finds that compounds are more frequently translated as phrases than as compounds in both languages. The source of this disparity is unknown.

There are three major discrepancies between the way a German NN compound is translated into English and vice versa. The same structural type in the target language is selected significantly less commonly in English than in German ($\chi^2(1) = 43.1, p < 0.001$). This finding invites the conclusion that German is more congenial to nominal compounding than English.

This difference is almost perfectly compensated for by the more frequent selection of phrases in English as renditions of the German NN compounds. This is true for both AN and all other phrases. This makes English the more syntactic, and German the more morphological, language. It is worth emphasizing that this account captures the cross-language difference almost exhaustively. Other structural options play less of a role. It is true that German selects non-compound nouns less often than English as a target language. However, this difference is much smaller than in the case of NN compounds. Note that the rate of this structural option (Id) may be somewhat inflated in German because of the rather broad definition of simple nouns adopted in this study (see above). Also, only minor cross-language differences are observed in the frequency of complex NNN compounds and clauses.

The analysis of the translations of NNN compounds (2 and II) reveals no statistically significant contrasts between the two languages. It is likely that this is due to the low number of relevant items in this category. There are two non-significant tendencies which may or may not be validated in larger samples. One is that the structural pattern of the source language seems to be preserved in the target language more frequently in German than in English. If true, this would suggest that complex compounds are tolerated better in the former than the latter language. The other tendency ties in with the analysis of NN compounds. A larger number of phrasal solutions appear to be found for the translation of complex compounds in English than in German. This would again testify to the relatively more phrasal nature of English.

The analysis of AN compounds (3 and III) is also hampered by the low number of pertinent cases. However, some tentative observations can be made. Putting aside the by now well-known observation that English prefers phrasal solutions,

we note two differences in the data. As a target language, German likes to adhere to the English structure or favours a simple noun as a possible rendition of an AN compound. English does not resort to these options at all. The tendency towards preserving the compound pattern in German may be attributed to the compounding bias in this language.

Next, we turn to the renditions of AN phrases (4 and IV). It is noteworthy that there are almost three times as many such structures in the English than in the German sample. This is in part a simple consequence of the fact that more text had to be scanned in order to collect 500 compounds in English than in German. Note that these phrases were mainly culled as a “by-product” of the search for compounds.

The first two rows 4a, b and IVa, b are the most relevant ones for the present purposes. The critical question is complementary to the foregoing analysis. While it was previously asked whether compounds are rendered as phrases, the point now is whether phrases can also be rendered as compounds, and if so, to what extent. Table 1 shows that this option is available in both languages but that it is more commonly utilized in German (IVa, b) than in English (4a, b) ($\chi^2(1) = 8.2$, $p < 0.01$). This cements our earlier claim that compounding has a higher probability of occurring in German than in English. It can also be seen that both languages prefer to translate AN phrases as NN rather than AN compounds. This links up naturally with the above claim that NN compounds are a more natural choice than AN compounds.

Summarizing, clear differences emerge between German and English translations. German gravitates more strongly to compounding than English does. Inversely, English leans more strongly towards phrasal solutions than German does. Notably, these opposite tendencies are almost equally strong. They capture the larger part of the difference between the two languages. Little support was found for a widely different use of simple nouns in the two languages (as long as simple nouns are understood to be non-compound nouns). Similarly, no cross-language difference was observed in the complexity of nominal compounds. This demonstrates that compound propensity and compound complexity are two variables which are independently controlled by individual languages. This result is in keeping with Berg et al.’s (2012) investigation of differences in compound size between English and German.

4. Assessing the rate of non-nominal compounding

In order to properly interpret the higher propensity of nominal compounding in German than in English, it is necessary to know which of the following two options is the correct one: either the difference between the two languages is confined to nouns, or German is generally more propitious to compounding than English, irrespective of individual word classes. In the latter case, the higher propensity of nominal compounding would be just one instance of a general compounding propensity in German. It is also conceivable that we find, let us say, two word classes which are more prone to compounding in German than in English, and another two word classes where the opposite holds. However, this would seem less likely.

The next step is, then, to determine the rate of non-nominal compounding in German and English. Actually, much less research has been done on non-nominal than on nominal compounding in both languages. To the best of my knowledge, no quantitative comparison of non-nominal compounding in English and German has so far been performed. To rectify this state of affairs, a further empirical study was carried out on a database similar to that used for the analysis of nominal compounds. The ensuing investigation focuses on the frequency of occurrence of non-nominal compounds in the two languages rather than on the translation of compounds. The comparison draws on text samples of identical size. The German sample is based on the first 106 pages of Schirach (2011). The English sample is drawn from all of Christie (1984) and the first page of Baron et al. (1984). Each sample consists of 20,000 words.

The following decisions were taken in the selection of compounds. Undoubtedly, the major problem area is complex verbs in both German and English. German inseparable verbs such as *unterschreiben* 'to sign' and *durchqueren* 'to cross' were classified as derivations and therefore excluded. German separable verbs such as *einladen* 'to invite' and *aufhören* 'to stop' were treated likewise because the syntactically determined separation of the two morphemes violates one of the best criteria for compoundhood, viz. inseparability (e.g. Lieber 1992, Gaeta & Ricca 2009). This criterion also eliminates complex verbs such as *standhalten* 'to withstand' and *Recht behalten* 'to be right', for instance, whose first constituent is an open-class word.

The case of English complex verbs is similar. Inseparable verbs such as *to overdo* and *to outshine* were classed as prefix-stem combinations. Intransitive phrasal verbs such as *to stand up* and *to come along* as well as inseparable transitive phrasal verbs such as *to bring up* (in the sense of 'to rear') were also discarded because, as their name implies, they have a rather phrasal status. Finally, separable transitive phrasal verbs such as *to pick up* were ignored by virtue of their separability. It is

obvious that this filtering procedure makes compound verbs an excessively rare event in both German and English.

In the domain of non-verbs, the following decisions should be noted. English complex adverbs such as *absentmindedly* were classified as compound adjectives because the compounding process takes place at the level of the adjective *absent-minded*. The adverbial suffix is irrelevant in this context. English complex adjectives and adverbs such as *right-hand (side)* and *downstairs*, which have a noun in final position, were classified as compound adjectives and adverbs.

With respect to German, a line had to be drawn between adjective phrases and adjective compounds. There is a difference between *frisch geschnitten* 'newly mown' and *hochgestochen* 'pretentious'. Orthography is deliberately put aside here even though it does contain a clue as to the linguistic status of these units. A more reliable criterion than spelling is stress. Open-class compounds almost invariably carry stress on the first constituent. The adjective *hochgestochen* was therefore categorized as a compound whereas *frisch geschnitten* was treated as a phrase and therefore left out of consideration.

The other problem in German is what Becker (1992) refers to as improper compounds, in particular particle compounds. These form a rather heterogeneous group and encompass items with adverbs, prepositions or pronouns as their second constituents. Pertinent examples include *irgendwo* 'somewhere' and *worauf* 'what ... P'. Since verbs prefixed by *(r)auf-* were not considered compounds, it was only consistent to ignore case like *worauf* for instance. By contrast, cases such as *irgendwo* were retained because *wo* 'where' is an independent interrogative pronoun. As *irgendwo* is functionally an adverb, it was classified as a compound adverb.

A number of dvandva compounds cropped up in the text samples. These were discarded from the analysis because their status is quite different from that of hierarchical compounds.

Table 2 presents the results of the quantitative analysis of non-nominal compounds in German and English. In contrast to the study of nominal compounds, the frequency analysis of non-nominal compounds is both type- and token-based. As a general rule, the compounds were categorized according to the word class of the head. If this differed from the word class of the entire compound, precedence was given to the latter. The different compound types are illustrated in (11)–(15). Compound adjectives are shown in (11), compound adverbs in (12), compound prepositions in (13), compound pronouns in (14) and a compound verb in (15).

- (11) a. G. eiskalt 'ice-cold'
b. E. middle-aged

- (12) a. G. jedenfalls ‘in any case’
b. E. beforehand
- (13) a. G. innerhalb ‘inside’
b. E. inside
- (14) a. G. irgendetwas ‘something’
b. E. himself
- (15) a. G. –
b. E. daresay

Table 2. Type and token frequency of non-nominal compounds in German and English

Language	Word class					
	verbs	adjectives	adverbs	conjunctions	pronouns	prepositions
German: types	0	54	43	3	3	3
English: types	1	37	17	0	17	7
German: tokens	0	63	110	12	7	4
English: tokens	2	53	45	0	134	32

The major finding contained in Table 2 is that the proclivity of German towards nominal compounding extends to other word classes. It is useful at this point to distinguish between open-class words and closed-class items and to begin with the former. The fact that verbs hardly undergo compounding in Table 2 largely results from the rigorous selection criteria imposed on them (see above). However, it also confirms the oft-made claim that both English and German are hostile to compound verbs (e.g. Marchand 1969, Erben 2003, Hüning & Schlücker 2010). Not even (15b) can be regarded as an uncontroversial case.

The only other open-class category is adjectives. In terms of type frequency, German adjectives are 1.5 times more frequent than their English congeners. Remarkably, this difference resembles that reported for nominal compounds by Stöhr (1987) (see introductory section). The cross-language difference at the level of token frequency is somewhat smaller.

The results are mixed among the closed-class items. Whereas adverbs and conjunctions clearly undergo compounding more frequently in German than in English, the opposite is true of pronouns and prepositions. Not surprisingly, this latter difference is much more pronounced at the level of token than at the level of type frequency.

To conclude, the compounding propensity in German is not confined to nouns. However, the empirical picture is not entirely consistent. In terms of its compound propensity, German does not outweigh English in all word classes. As

the closed-class items are more variegated and more difficult to compare across languages, the following conclusions are limited to open-class words.

The relatively heavy use that German makes of compounding generalizes across different word classes. So does the less frequent use that English makes of compounding. By implication, the notion of compound propensity is applicable not only to individual word classes but also at a more general level. However, nouns seem to bring out the compounding bias in German more clearly than do other word classes. The claim that high compound propensity is a characteristic of the morphology of German open-class words which is not confined to nouns, does not imply that all word classes behave alike. Individual word classes may vary in their compound propensity and even defy compounding in the presence of a general proclivity to compounding, as in the case of German verbs.⁶

5. A frequency analysis of German and English word classes

It has repeatedly been stated that English is a more verbal, and German a more nominal, language (e.g. Kortmann & Meyer 1992, Rohdenburg 1998, Fischer 1999).⁷ Fischer (2007: 396) explicitly links German NPs and English VPs, contending that the structural complexity of German NPs is mirrored by the structural complexity of English VPs. The underlying assumption is that languages may differentially distribute information across syntactic phrases and thereby differentially exploit the structural possibilities offered by these constituents. If German really is more nominal than English, we might have an explanation for the higher propensity for nominal compounding in the former than the latter language. At first sight, the logic might seem trivial: the higher the extent of single-noun use, the greater the likelihood that these nouns will be combined to form compounds. However, this is a non sequitur. The proclivity for compounding is logically independent of the extent to which individual morphemes are used. On the other hand, it may very well be that a nominal language develops a higher availability of nouns and that this heightened availability encourages the use of nominal compounding.

6. Of course, if the rigorous selection criteria for verbal compounds were relaxed, we would end up with a far higher number of such compounds. Notably, such a move would bring verbs in line with other word classes and thereby make the data much more homogeneous.

7. Steiner (2012) arrives at the same result although his conclusion is based on a comparison of a set of “nominal word classes” to a set of “verbal word classes”. When the set of nouns is compared to the set of verbs, the expected cross-language difference fails to show up. His corpus analysis yields not only more verbs but also more nouns in English than in German.

The relevant literature has remained largely silent on how a nominal or verbal slant should be defined and how it could be operationalized for comparative purposes. In this study, we will draw on the notions of type and token frequency. Type frequency taps into the structure of the lexicon. It provides information about the proportions of word classes in the overall vocabulary of a language. These proportions can be directly compared across languages. If a given language possesses an elevated rate of nouns, it qualifies as a nominal language. Token frequency is closer to the literary notion of style than type frequency is. It refers to the use that is made of the lexicon in the act of speaking or writing. It measures the proportions of different word classes in texts. If a representative sample from language A contains an elevated rate of verbs (relative to a sample from language B), language A qualifies as a verbal language.

Type frequency was calculated on the basis of CELEX, an electronic dictionary of English, German and Dutch (Baayen, Piepenbrock & Gulikers 1995). If type frequency lies at the bottom of the propensity for nominal compounding in German, we would expect a larger number of nouns in the German than the English lexicon. The same expectation holds for adjectives. The prediction for verb types is less clear. If there are fewer verb types (and more noun types) in German than in English, a stronger case can be made for the link between compound propensity and the open-class lexicon; if, however, there are more verb and more noun types in German than in English, such a link cannot be easily established.

Table 3 presents the results of the CELEX analysis. Given our focus on open-class words, all closed-class items were conflated.

Table 3. Type frequency of various word classes in German and English

Language	Word class				
	nouns	verbs	adjectives	closed-class items	total
German	26,586	9,204	9,075	6,863	51,728
	51.4%	17.8%	17.5%	13.3%	100%
English	27,165	8,433	8,502	8,444	52,544
	51.7%	16.0%	16.2%	16.1%	100%

The most important result to emerge from Table 3 is the almost identical proportions of nouns in German and English. This implies that nominal compounding cannot be linked to type frequency. The higher rate of verb types in German than in English is meaningless in the present context because verbal compounding is a strongly dispreferred option in both languages. If anything, this difference argues against a link between compounding propensity and the structure of the lexicon. Viewed in isolation, the higher rate of adjective types in German than in English

might be taken as a potential explanation of the higher occurrence of adjectival compounding in the former than the latter language. However, it is indefensible to try to forge a link between compounding propensity and type frequency for one word class and at the same time reject such a link for another word class. We are therefore inclined to believe that the larger number of adjective types in German than in English is an unlikely cause of the differing rates of adjectival compounding in the two languages. Our conclusion is that type frequency cannot explain compound propensity.

The comparison of token frequencies is a tricky business because its validity stands or falls on the comparability of the corpora from which the token frequencies are derived. Token frequency is known to be relatively variable, depending on modality (spoken vs. written), text type and sundry other variables. This variability creates the additional difficulty of generalizing from particular corpora to the language at large. Even though corpora are almost by definition designed to be balanced (at least in certain respects), they cannot a priori be expected to be balanced with respect to the factors determining compound use.

Given this state of affairs, a double strategy was deployed. To assess the variability between corpora and, by implication, the validity of the conclusions to be drawn, four corpora, two from each language, were subjected to scrutiny. In addition, an effort was made to select two corpora which are highly similar in design and size and hence to the greatest possible extent comparable. As a secondary strategy, the diachronic dimension was introduced.

To be specific, the following four corpora were taken into consideration. The FLOB Corpus of British English includes a wide variety of written genres and consists of 500 text samples of an approximate length of 2000 words each. The Hamburg Corpus of German was modelled on the FLOB Corpus and thus provides an excellent basis for cross-language investigation. Emphasis was therefore laid on the comparison of these two data sets. Two further corpora were consulted – LOB for English and DEREKO (“Deutsches Referenzkorpus”) for German. The German corpus was compiled at the “Institut für Deutsche Sprache” (IDS) in Mannheim, Germany. With more than eight billion words, it is the largest German corpus currently in existence. It consists exclusively of written language and covers a wide array of genres ranging from prose to journalese. Finally, it was decided to include the corpus on which FLOB was modelled, viz. LOB. The addition of the LOB Corpus, which was culled thirty years before FLOB, provides an opportunity of testing for diachronic shifts in word-class frequency.

The results of the token-frequency analysis are presented in Table 4, which is organized along the same lines as Table 3. As before, adverbs were assigned to the closed-class category. All data were computed anew even though LOB and FLOB

had been previously examined for word-class frequencies by Mair et al. (2002). This was done in an attempt to reconstruct the many substantive and technical decisions that had gone into the computer search. Neither between our data and Mair et al.'s nor between Mair et al.'s study and Johansson & Hofland's (1989) analysis of LOB and BROWN was there perfect agreement. Note that the Hamburg Corpus was originally untagged. Automatic part-of-speech tagging was done on the basis of the Maximum Entropy Model within the programming environment R (using R package: Open NLP).

Table 4. Token frequency of various word classes in German and English

Language	Word class				
	nouns	verbs	adjectives	closed-class items	total
English:	221,801	178,262	80,040	521,794	1,001,897
FLOB	22.1%	17.8%	8.0%	52.1%	100%
German:	223,360	151,747	96,903	549,730	1,021,740
Hamburg	21.9%	14.9%	9.5%	53.8%	100.1%
German:	2,842,318,631	1,130,004,141	761,142,694	3,827,589,647	8,561,055,113
DEREKO	33.2%	13.2%	8.9%	44.7%	100%
English: LOB	206,115	179,975	73,574	623,404	1,083,068
	19.0%	16.6%	6.8%	57.6%	100%

From the perspective of the research hypothesis, Table 4 seems to display a disappointing result. The critical comparison between the FLOB and the Hamburg Corpus produces virtually no difference in the token frequency of English and German nouns. By contrast, verbs are considerably more common in English than in German while adjectives are somewhat more common in German than in English. If anything, this result suggests a certain independence between the token frequency of nouns and verbs. It is not the case that a higher rate of nouns is accompanied by a lower rate of verbs and vice versa. Quantitatively speaking, there is no compensatory relationship between nouns and verbs. There is no question of linking the difference in verb-token frequency to the propensity for nominal compounding. Thus, the interlingual difference in nominal-compound propensity remains unaccounted for.

When the other, much larger corpus of German is taken into consideration, we are taken aback by an almost dramatic difference between the two German corpora. While the rate of verbs and adjectives is rather similar, DEREKO includes more than 11% more nouns than the Hamburg Corpus. We will make no attempt to speculate on the reasons for this discrepancy. The least that this

disparity suggests is that it is unwise to rely on a single corpus and take the data derived from it at face value, i.e. as representative of the language at large. Note in passing that the comparison between the DEREKO and the Hamburg Corpus confirms the above-mentioned relative independence between nouns and verbs.

Obviously, when FLOB is compared to DEREKO, a major difference emerges between German and English nouns. In keeping with the research hypothesis, German manifests itself as the more nominal language compared to English. However, in view of the fact that DEREKO is less directly comparable to FLOB than the Hamburg Corpus is, it seems imprudent to make much of this difference.

As noted before, the inclusion of LOB allows us to assess the diachronic stability of word-class frequencies. The comparison between LOB and FLOB reveals an appreciable increase in nouns from 1960 to 1990 but also a minor increase in verbs and adjectives. This by and large replicates the trends reported for nouns by Mair et al. (2002) in their comparison of LOB and FLOB, even though the absolute numbers are far from identical.

When the Hamburg Corpus is compared to LOB rather than FLOB, German comes out as the more nominal language. Again, it appears unjustified to base strong theoretical claims on this comparison because the two corpora were collected at different times and therefore ignore possible diachronic changes.

In the following, a tentative proposal will be offered which is compatible with all of the frequency data contained in Table 4 and which takes the diachronic dimension into account. We submit that there used to be a frequency difference between German and English, with nouns being used more often in German than in English. By hypothesis, this difference contributes to the higher compound propensity in the former than the latter language. English has recently caught up with German by increasing its use of nouns to levels similar to German. Provided that there is a connection between single-noun use and nominal-compound propensity, this account predicts a diachronic increase in the use of nominal compounds in English. This is precisely what we find. Biber & Clark (2002) and Pastor-Gómez (2011) document a major increase in nominal-compound use in the second half of the 20th century (at least in some genres). It stands to reason that there is a certain time lag between the increase in single-noun use and the increase in nominal-compound use. This proposal generates the prediction that the difference in the use of nominal compounds between the two languages will diminish over time.

Essentially the same argument can be made for adjectives although the ratios and implicationally, the differences between the ratios are smaller than in the case of nouns. Table 4 shows a slight increase in the textual frequency of English adjectives from 1960 to 1990 (see also Mair et al. 2002). It is noteworthy in this connection that the frequency of at least one type of adjectival compound (e.g.

EU-funded) has substantially increased in the second half of the 20th century (Hilpert 2015). Provided there is an effect, the advantage of German as compared to English adjective use may be just strong enough to promote the use of adjectival compounds in the former language.

It is also conceivable that there is a “spillover effect” from nouns to adjectives. Owing to a considerable overlap between the two word classes (Berg 2000), adjectives may profit from the heightened probability of nominal compounding and thereby increase their availability for compounding. Since both German and English verbs resist compounding quite strongly (see Table 2), this equal resistance probably occurs for reasons which are unrelated to the usage frequency of verbs. In other words, it is unlikely that the relatively low frequency of verbs underlies their resistance to compounding.

While the preceding account is no more than suggestive, it may not go amiss to draw attention to its plausible psycholinguistic basis. The processing perspective gives us an idea of the mechanism which may create a link between compound propensity and token frequency. Compounds require the co-activation of (at least) two independent words.⁸ Co-activation is brought about by spreading activation from one mental node to another. One principle of activation spreading which is of particular relevance in the present context is the similarity constraint. It states that activation is predominantly relayed to nodes sharing a property with the target node. For example, lexical nodes activate other lexical nodes of the same word class rather than of different word classes (e.g. Fay & Cutler 1977, Zwicky 1978/79, Berg 1992). The second relevant principle, which goes back to Morton (1969), is that the repeated use of a node lowers its production threshold. Therefore, the more often a given node is accessed, the more activation it spreads to connecting nodes.

Together, these two principles generate the assumed link between compound propensity and token frequency. The more often simple nouns are used, the higher the availability of nouns as a class. So when a complex idea has to be cast in a linguistic mould, the activation of a given noun facilitates the near-concurrent activation of another noun. The higher the co-activation of this other noun, the higher its likelihood of being fitted into a common schema with the “first” noun, i.e. the higher the likelihood of creating a compound noun. This account describes

8. This analytic perspective on compound processing is supported by both production and perception studies (e.g. Roelofs 1997, Juhasz, Starr, Inhoff & Placke 2003, Andrews, Miller & Rayner 2004, Bien, Levelt & Baayen 2005). This view in no way contradicts the holistic effects also observed in compound research. Analytic and holistic effects arise at different levels of processing and should be regarded as complementary in nature (e.g. Kuperman, Schreuder, Bertram & Baayen 2009).

the situation prevailing in German. In English, by contrast, the probability of fitting two nouns into a compound scheme is lower and therefore, the degree of co-activation is also lower (Berg 2012). Note that this account explains not only the difference between German and English but also the fact, observed in both languages, that noun-noun compounds occur far more often than adjective-noun compounds.

It should be emphasized that, provided token frequency plays a role in the compound game, it is only one among a number of factors involved in compound propensity. It is conceivable that token frequency is a necessary, though not a sufficient criterion for compounding. When a particular word class is infrequently used, the combined use of this word class with itself or any other word class may be expected to be fairly infrequent. At the same time, the frequent use of a particular word class is no guarantee that it will frequently serve as an input to compounding. Take the case of the Romance languages, which are reluctant to use nominal compounding. This property does not, of course, imply that they use nouns only sparingly or that there cannot be differences in noun use among the individual Romance languages. It makes more sense to argue that an elevated use of simple words facilitates their use as parts of compounds, i.e. the creation of compound words.

Let us finally return to Table 4. While the frequency analysis of nouns was not without its problems, German verbs are consistently less common than their English counterparts. Therefore, there is a sounder quantitative basis for arguing that English is a (more) verbal language than for arguing that German is a (more) nominal language. It should be noted in this connection that these claims about the nominal or verbal nature of a language are based on a between-language rather than a within-language comparison. Even though English is assumed to be a verbal language, its rate of verb tokens is still lower than its rate of noun tokens. The same is true for German.

6. Outlook

In an effort to come to grips with cross-linguistic differences in compound propensity, the present paper has mainly addressed quantitative issues of language usage. An attempt has been made to ascribe differences in compound propensity to differences in textual frequency. This assumed link is probably not particularly strong. It therefore has to be seen in conjunction with other principles. In addition to the usage factor, at least two aspects seem worth exploring in future work – the semantic and the cognitive. It is possible that there are semantic differences between German and English compounds and that these differences give rise to variable

compound propensity. To be more specific, one language may make heavier use of certain semantic relations between the constituents of compounds than another and as a consequence, be more inclined towards compounding. The cognitive factor is the least tractable but potentially the most rewarding one. From a cognitive viewpoint, compounds are categories, i.e. they represent unitary concepts. Could it be that languages (speakers, that is) differ in their strategies of (non-linguistic) categorization? Is one language more likely to create unitary concepts than another? And if so, how could this difference come about? Answers to these and related questions may significantly advance our understanding of cross-linguistic differences and ultimately, the relation between language and thought.

Acknowledgements

I wish to express my gratitude to quite a large number of helping hands: to Janine Hutter for her unflinching support in data collection, to Arne Zeschel for putting me through to Marc Kupietz, to Marc Kupietz for extracting token-frequency data from the “Deutsches Referenzkorpus”, to Martin Hilpert for kindly placing the Hamburg Corpus at my disposal, to Martin Schweinberger for tagging the Hamburg Corpus, to Marion Neubauer for her diligent corpus work and her close reading of an earlier version, to Katerina Stathi for sharing with me her insights into the possible reasons for variable compound propensity and last but not least, to the reviewers whose comments have helped me to strengthen my analysis and at the same time convinced me to tone down some of my claims.

References

- Ahlemeyer, Birgit and Inga Kohlhof 1999. Bridging the cleft: An analysis of the translation of English “it”-clefts into German. *Languages in Contrast* 2: 1–25. doi:10.1075/lic.2.1.03ahl
- Andrews, Sally, Brett Miller and Keith Rayner 2004. Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology* 16: 285–311. doi:10.1080/09541440340000123
- Baayen, R. Harald, Richard Piepenbrock and Leon Gulikers 1995. *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium.
- Bauer, Laurie 1978. *The grammar of nominal compounding*. Odense: Odense University Press.
- Bauer, Laurie 1998. When is a sequence of two nouns a compound in English? *English Language and Linguistics* 2: 65–86. doi:10.1017/S1360674300000691
- Becker, Thomas 1992. Compounding in German. *Rivista di Linguistica* 4: 5–36.
- Berg, Thomas 1992. Prelexical and postlexical features in language production. *Applied Psycholinguistics* 13: 199–235. doi:10.1017/S0142716400005567
- Berg, Thomas 2000. The position of adjectives on the noun-verb continuum. *English Language and Linguistics* 4: 269–293. doi:10.1017/S1360674300000253

- Berg, Thomas 2012. The cohesiveness of English and German compounds. *The Mental Lexicon* 7: 1–33. doi:10.1075/ml.7.1.01ber
- Berg, Thomas 2016. The multiplanar nature of frequency. *Glottology* 7: 1–19.
- Berg, Thomas, Sabine Helmer, Marion Neubauer and Arne Lohmann 2012. Determinants of the extent of compound use: A contrastive study. *Linguistics* 50: 269–303. doi:10.1515/ling-2012-0010
- Biber, Douglas and Victoria Clark 2002. Historical shifts in modification patterns with complex noun phrase structures. In *English historical syntax and morphology*, Teresa Fanego, María José López-Cousa and Javier Pérez-Guerra (eds), 43–66. Amsterdam: Benjamins. doi:10.1075/cilt.223.06bib
- Bien, Heidrun, Willem J. M. Levelt and R. Harald Baayen 2005. Frequency effects in compound production. *Proceedings of the National Academy of Sciences* 102: 17876–17881. doi:10.1073/pnas.0508431102
- Deng, Yaochen 2009. A synergetic linguistic study on compounding propensity in technical English. *Glottology* 2: 54–61.
- Djirackor, Yao 1982. Untersuchungen zur Wiedergabe deutscher Substantivkomposita im Englischen. Eine kontrastive Studie. Unpublished Ph.D. thesis, University of Hamburg, Germany.
- Doherty, Monika 1993. Parametrisierte Perspektive. *Zeitschrift für Sprachwissenschaft* 12: 3–38. doi:10.1515/zfs.1993.12.1.3
- Donalies, Elke 2004. *Grammatik des Deutschen im europäischen Vergleich: Kombinatorische Begriffsbildung. Teil 1: Substantivkomposition*. Mannheim: Institut für deutsche Sprache.
- Donalies, Elke 2008. Komposita, Derivate und Phraseme des Deutschen im europäischen Vergleich. *Deutsche Sprache* 36: 305–323.
- Dressler, Wolfgang U., Gary Libben, Jacqueline Stark, Christiane Pons and Gonia Jarema 2001. The processing of interfixed German compounds. In *Yearbook of Morphology 1999*, Geert Booij and Jaap van Marle (eds), 185–220. Dordrecht: Kluwer. doi:10.1007/978-94-017-3722-7_8
- Erben, Johannes 2003. Hauptaspekte der Entwicklung der Wortbildung in der Geschichte der deutschen Sprache. In *Sprachgeschichte*, Vol. 3, Werner Besch, Anne Betten, Oskar Reichmann and Stefan Sonderegger (eds), 2525–2539. Berlin: Walter de Gruyter.
- Ermlich, Karsten 2004. Zur Wiedergabe deutscher Substantivkomposita im Englischen. *Deutsch als Fremdsprache* 41: 206–212.
- Fabb, Nigel 1998. Compounding. In *The handbook of morphology*, Andrew Spencer and Arnold Zwicky (eds), 66–83. Oxford: Blackwell.
- Fan, Fengxiang and Gabriel Altmann 2007. Some properties of English compounds. In *Problems of typological and quantitative lexicology*, Volodymyr Kaliuscenko, Reinhard Köhler and Viktor Levickij (eds), 177–189. Chernivtsi: Puta.
- Fay, David and Anne Cutler 1977. Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry* 8: 505–520.
- Fischer, Klaus 1999. Englische und deutsche Satzstrukturen: ein valenztheoretischer Vergleich mit statistischen Anmerkungen. *Sprachwissenschaft* 24: 221–255.
- Fischer, Klaus 2007. Komplexität und semantische Transparenz im Deutschen und Englischen. *Sprachwissenschaft* 32: 355–405.
- Gaeta, Livio and Davide Ricca 2009. “Composita solvantur”: Compounds as lexical units or morphological objects? *Rivista di Linguistica* 21: 35–70.

- Greenberg, Joseph H. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26: 178–194. doi:10.1086/464575
- Gunkel, Lutz and Gisela Zifonun 2008. Constraints on relational-adjective noun constructions: A comparative view on English, German and French. *Zeitschrift für Anglistik und Amerikanistik* 56: 283–302. doi:10.1515/zaa.2008.56.3.283
- Hilpert, Martin 2015. From *hand-carved* to *computer-based*: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics* 26: 113–147. doi:10.1515/cog-2014-0001
- Hüning, Matthias and Barbara Schlücker 2010. Konvergenz und Divergenz in der Wortbildung – Komposita im Niederländischen und Deutschen. In *Kontrastive germanistische Linguistik. Band 2*, Antje Dammel, Sebastian Kürschner and Damaris Nübling (eds), 783–825. Hildesheim: Olms.
- James, Carl 1980. *Contrastive analysis*. Harlow: Longman.
- Johansson, Stig and Knut Hofland 1989. *Frequency analysis of English vocabulary and grammar*. Oxford: Clarendon.
- Juhász, Barbara J., Matthew S. Starr, Albrecht W. Inhoff and Lars Placke 2003. The effects of morphology on the processing of compound words: Evidence from naming, lexical decision and eye fixations. *British Journal of Psychology* 94: 223–244. doi:10.1348/000712603321661903
- Klinge, Alex 2005. The structure of English nominals. Unpublished postdoctoral thesis. Copenhagen.
- Koester, Dirk, Thomas C. Gunter, Susanne Wagner and Angela D. Friederici 2004. Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience* 16: 1647–1668. doi:10.1162/0898929042568541
- Kortmann, Bernd and Paul Georg Meyer 1992. Is English grammar more explicit than German grammar, after all? In *New departures in contrastive linguistics*. Vol. 1, Christian Mair and Manfred Markus (eds), 155–166. Innsbruck: Innsbrucker Beiträge zur Kulturwissenschaft.
- Krott, Andrea, Robert Schreuder and R. Harald Baayen 1999. Complex words in complex words. *Linguistics* 37: 905–926. doi:10.1515/ling.37.5.905
- Krzeszowski, Tomasz P. 1990. *Contrasting languages. The scope of contrastive linguistics*. Berlin: Mouton de Gruyter. doi:10.1515/9783110860146
- Kuperman, Victor, Robert Schreuder, Raymond Bertram and R. Harald Baayen 2009. Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 35: 876–895.
- Lieber, Rochelle 1992. Compounding in English. *Rivista di Linguistica* 4: 79–96.
- Mair, Christian, Marianne Hundt, Geoffrey Leech and Nicholas Smith 2002. Short-term diachronic shifts in part-of-speech frequencies. *International Journal of Corpus Linguistics* 7: 245–264. doi:10.1075/ijcl.7.2.05mai
- Marchand, Hans 1969. *The categories and types of present-day English word-formation*. Munich: Beck.
- Matthews, Stephen and Virginia Yip 2004. *Cantonese: A comprehensive grammar*. London: Routledge.
- Morton, John 1969. Interaction of information in word recognition. *Psychological Review* 76: 165–178. doi:10.1037/h0027366
- Mühlhäusler, Peter 1979. *Growth and structure of the lexicon of New Guinea Pidgin*. Canberra: Department of Linguistics, Australian National University.

- Pastor-Gómez, Iria 2011. *The status and development of N + N sequences in contemporary English noun phrases*. Bern: Lang. doi:10.3726/978-3-0351-0193-5
- Potter, Simeon 1957. *Modern linguistics*. London: Andre Deutsch.
- Roelofs, Ardi 1997. Morpheme frequency in speech production: Testing WEAVER. In: *Yearbook of Morphology 1996*, Geert Booij and Jaap van Marle (eds), 135–154. Dordrecht: Kluwer. doi:10.1007/978-94-017-3718-0_9
- Rohdenburg, Günter 1998. Subordinate clauses introduced by interpretative verbs in English and their less explicit counterparts in German. In *Kontrast und Äquivalenz*, Wolfgang Börner and Klaus Vogel (eds), 233–249. Tübingen: Narr.
- Sadeniemi, Markus, Kimmo Kettunen, Tiina Lindh-Knuutila and Timo Honkela 2008. Complexity of European Union languages: A comparative approach. *Journal of Quantitative Linguistics* 15: 185–211. doi:10.1080/09296170801961843
- Sanchez, Christina 2008. *Consociation and dissociation*. Tübingen: Narr.
- Schlücker, Barbara 2012. Die deutsche Kompositionsfreudigkeit. Übersicht und Einführung. In *Das Deutsche als kompositionsfreudige Sprache*, Livio Gaeta and Barbara Schlücker (eds), 1–15. Berlin: Walter de Gruyter. doi:10.1515/9783110278439.1
- Steiner, Erich 2012. A characterization of the resource based on shallow statistics. In *Cross-linguistic corpora for the study of translations*, Silvia Hansen-Schirra, Stella Neumann and Erich Steiner (eds), 71–89. Berlin: Walter de Gruyter. doi:10.1515/9783110260328.71
- Stöhr, Ingo R. 1987. Pragmatic factors and nominal compounding: the influence of „deictic” reference in English and German. In *Neuere Forschungen zur Wortbildung und Historiographie der Linguistik*, Brigitte Asbach-Schnitker and Johannes Roggenhofer (eds), 31–39. Tübingen: Narr.
- Twain, Mark 1880/1997. *A tramp abroad*. London: Penguin.
- Wandruszka, Mario 1968. Englische und deutsche Nominalkomposition. In *Wortbildung, Syntax und Morphologie*, Herbert E. Brekle and Leonhard Lipka (eds), 242–250. The Hague: Mouton.
- Wandruszka, Mario 1969. *Sprachen. Vergleichbar und unvergleichlich*. Munich: Piper.
- Zimmer, Karl E. 1981. Some general observations about nominal compounds. In *Wortbildung*, Leonhard Lipka and Hartmut Günther (eds), 233–257. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Zwicky, Arnold M. 1978/79. Classical malapropisms. *Language Sciences* 1, 339–348. doi:10.1016/S0388-0001(79)80022-4

Primary Works

German → English

- Schirach, Ferdinand von: *Der Fall Collini*. Munich: Piper, 2011 (up to middle of p. 98). – English translation: *The Collini Case*. Harmondsworth: Penguin, 2012. (up to and including p. 105)

English → German

- Rankin, Ian: *Exit music*. London: Orion, 2007 (up to and including p. 211). – German translation: *Ein Rest von Schuld*. Munich: Goldmann, 2008 (up to and including p. 253).

English

- Christie, Agatha: Hercule Poirot, Miss Marple and Three Whodunits. – Hercule Poirot, Miss Marple und Drei Fälle. Munich: Deutscher Taschenbuch Verlag, 1984/1998.
- Baron, Alexander *et al.*: English Short Stories – Englische Kurzgeschichten. Munich: Deutscher Taschenbuch Verlag, 1984/1994.

Author's address

Thomas Berg
Department of English
University of Hamburg
Von-Melle-Park 6
20146 Hamburg
Germany
thomas_berg@uni-hamburg.de