Bleaching, productivity and debonding of prefixoids. A corpus-based analysis of 'giant' in German and Swedish

Muriel Norde & Kristel Van Goethem Humboldt Universität zu Berlin; F.R.S.-FNRS & Université catholique de Louvain

Introduction

The past three decades have seen a keen interest in the emergence of grammatical words and structures, especially within the framework of grammaticalization studies (Norde et al., 2013, p. 1). In advanced stages of grammaticalization, free morphemes may become bound, a process that has long been considered irreversible (Haspelmath, 1999). More recent research, however, has shown that bonding is not irreversible – under specific circumstances, bound morphemes, both clitics and affixes, may 'debond' into free morphemes again (Norde, 2009, p. 186ff.). This article discusses some issues related to debonding of one particular type of bound morpheme, viz. prefixoids, which are compound members that also occur as free morphemes, but have a different meaning when bound. Drawing on earlier research on the emergence of free uses of affixes and affixoids (Norde & Van Goethem, 2014; Van Goethem, 2014; Van Goethem & Hiligsmann, 2014; Van Goethem & De Smet, 2014), this study offers a quantitative analysis of the distribution, semantics and productivity of bound and free uses of two morphemes both originally meaning 'giant': German *Riesen(-)* and Swedish *jätte(-)*.

As prefixoids, these morphemes collocate with different parts of speech, e.g. nouns or adjectives. Basically, prefixoids have two different functions, as illustrated in (1)-(2). In (1), the prefixoids have simile meaning; a *Riesenpizza*, for example, is not a pizza cooked by a giant, or a pizza topped with pieces of giant, but a pizza of exceptional size, just like a giant is a creature of exceptional size. Similarly, adjectives such as German *riesengroβ*, or Swedish *jättestark*, may be paraphrased as 'as big as a giant', or 'as strong as a giant' respectively. In (2), on the other hand, the 'giant' morphemes serve to intensify the meaning of the noun or adjective they collocate with; a *Riesenproblem*, for instance, is not a large problem in terms of size, but in terms of degree. In both constructions, similes and intensifiers, the prefixoid may be written separately, as in the Swedish examples in (3)-(4).

- (1) Gm. *Riesenpizza* (giant pizza), Sw. *jättebanan* (giant banana)
- (2) Gm. *Riesenproblem* (lit. giant-problem; huge problem), Sw. *jättetalang* (lit. giant-talent; huge talent); Gm. *riesenpositiv* (very positive), Sw. *jättekul* (very cool)
- (3) Sw. han har blivit jätte stooor och jätte tung (he has become very big and very heavy) (SECOW2011X_2827085)
- (4) Sw. tror verkligen du behöver en mentor och mycket kärlek, för du verkar verkligen vara jätte arg.
 (I really do think you need a mentor and lots of love, because you do seem very angry.)
 (SECOW2011X_929023)

The central objective of this paper is to contrast the bound and free uses of Riesen(-) and $j\ddot{a}tte(-)$, both in terms of the parts of speech they collocate with, their semantics, and their productivity. More specifically, we aim to explore whether synchronic corpus data can be used to determine whether the free uses in (3)-(4) are simply a matter of (erroneous) spelling, or whether they reflect a process of "debonding" (Norde 2009).

Before proceeding to our case studies in Section 3, we will briefly discuss some concepts that are central to our analysis: 'similes' (1.1), 'affixoids' (1.2), and 'debonding' (1.3). In section 2, we will present our sources and methods, and in section 3 we present the empirical results of our corpus study. The implications of our findings, finally, will be discussed in section 4.

1. Theoretical preliminaries

1.1. Similes and non-similes

Simile compounds are compounds in which the first compound member serves as a point of comparison. Hoeksema (2012) includes similes in a broader type of so-called 'elative compounds', which "indicate a high degree of a property that is expressed by their right-hand member, the head of the compound, usually by making use of some kind of conventionalized comparison" (Hoeksema, 2012, p. 97). *Riesen-* and *jätte-* also occur as first parts of simile compounds, both in adjectival ((5)-(6)) and nominal ones ((7)-(8)). As such, the morphemes preserve their original 'giant' meaning and the compounds can be paraphrased as 'as Adj as a giant' or 'as N as (of) / like a giant' respectively.

(5)	Die Insel ist natürlich riesengroß .
	(The island is huge, of course.)
	(DECOW2012-03X_20349749)
(6)	En jättestor påse som räcker hela helgen.
	(A huge bag that will last the entire weekend)
	(SECOW2011X_610906)
(7)	wie wär 'S mit nem riesenpott chili con carne?
	(What about a huge pot of chili con carne?)
	(DECOW2012-03X_15585210)
(8)	iao är inte en skitsnygg hlondin med jättehrö st

(8) *jag är inte en skitsnygg blondin med jättebröst* (I am not a gorgeous blonde with huge breasts) (SECOW2011X_1209786)

However, as we already noted in our introductory section, the two morphemes may lose this lexical meaning and develop into pure intensifiers, in which case a paraphrase with 'giant' is no longer available. In such 'non-similes', *Riesen-* and *jätte-* can again modify adjectives ((9)) and nouns ((10)-(11)), but the former is marginal in German, with only one non-simile construction occurring in the sample of free usages of *riesen* (see further section 3.2).

(9) Jag gick i en liten skola på 25 elever (4 i min klass) och det var jättemysigt!
(I went to a small school with 25 pupils [4 in my year] and that was very cosy!)
(SECOW2011X_679723)

- (10) Haben Sie auch Panik und denken Halloween-Kübis schnitzen, das kann nur eine Riesen-Sauerei werden?
 (Are you also panicking, and thinking: 'cutting halloween pumpkin, that can only turn into a huge mess?')
 (DECOW2012-03X_4085864)
 (11) Idag vaknade jag med en jätteförkylning.
- (Woke up this morning with a terrible cold) (SECOW2011X_528823)

1.2. Affixoids

In the examples (9)-(11), riesen- and jätte- occur as left-hand members of nonsimile compounds, with a bleached meaning as intensifiers. Such morphemes "which look like parts of compounds, and do occur as lexemes, but have a specific and more restricted meaning when used as part of a compound" are called 'affixoids' (Booij, 2009, p. 208; see also Booij, 2010; Booij & Hüning, 2014; Klara, 2012; Leuschner, 2010; Oebel, 2012; Sachs 2012). It has been argued that the emergence of prefixoids can be seen as an instance of grammaticalization, since they "have lost their original lexical meaning more or less, and have acquired a more general and abstract meaning of intensification" (Booij, 2010, p. 58; see also Van Goethem, 2008, 2011; Van Goethem & Hiligsmann, 2014). As a result, a cline can be drawn from determinative compounds with literal reference to a giant (e.g. Gm. Riesenhand (hand of a giant)), simile compounds with reference to the prototypical physical properties of giants (e.g. Riesenbühne (giant stage)) and intensifying compounds (e.g. Riesenfan (huge fan)). A similar cline can be assumed for adjectival compounds, likewise ranging from determinative compounds (e.g. Sw. jättelik (resembling a giant)) to simile compounds (e.g. jättestor (as big as a giant)) to intensifying compounds (e.g. jättegullig (very cute)). Since the latter two types of compound involve loss of lexical content, we argue that the 'giant' morpheme functions as a prefixoid in intensifying constructions as well as in similes.

1.3. Debonding

Debonding is one of the subtypes of degrammaticalization distinguished by Norde (2009). The process is defined as "a composite change whereby a bound morpheme in a specific linguistic context becomes a free morpheme" (Norde, 2009, p. 186). It is typically marked by the following parameters: *severance*, *recategorialization*, *scope expansion*, and *flexibilization* (Norde, 2009, p. 186-

227). Earlier studies (Norde & Van Goethem, 2014; Van Goethem, 2014; Van Goethem & Hiligsmann, 2014; Van Goethem & De Smet, 2014) have shown that free uses of prefixoids may be due to a process of debonding. For instance, for the Dutch prefixoid *reuze* (cognate with German *Riesen-*) Van Goethem & Hiligsmann (2014) have shown that its autonomous uses may display all of the typical parameters of the process, including scope expansion and flexibilization, i.e. extension to new syntactic contexts.

The studies mentioned above suggest that the free forms are not merely spelling variants of the bound prefixoids. Nevertheless, the process may have started out as a spelling phenomenon, subject to formal factors such as vowel clashes, but a subsequent reanalysis has led to the creation of a free morpheme with two main functions, intensifying adverb and attributive adjective (cf. Van Goethem & Hiligsmann, 2014). Crucially, however, this does not mean that separate spelling is necessarily a sign of debonding – the results of Dutch *reuze*-cannot be extrapolated to similar prefixoids in other languages; each prefixoid needs to be examined in its own right.

2. Sources and methods

Data for this study are drawn from the COW corpusⁱ, a gigatoken database of tagged and lemmatized texts from the web, compiled at the FU Berlin (cf. Schäfer & Bildhauer, 2012 and 2013). This corpus is eminently suited to the study of language change in progress: it provides similar data sets from different languages, among them German and Swedish (see Table 1), and a substantial portion of these data are from informal sources, which is the typical locus of recent and innovative constructions. The subcorpora used for this study are given in Table 1.

COW Subcorpora	Number of tokens	Number of documents
DECOW2012-03X	9,108,097,177 tokens	7,632,384 documents
(German subcorpus)		
SECOW2011X	1,749,846,457 tokens	1,912,757 documents
(Swedish subcorpus)		

Table 1: Subcopora used in this study

To collect our data, we proceeded as follows: using the COLibrI query interfaceⁱⁱ, we selected 1000 random results for both the bound and free forms of German $r/Riesen^{iii}$ and Swedish *jätte*. All results were checked manually, because

irrelevant hits would have to be discarded. In the German sample of free forms, a substantial number of results featured the noun *Riese* 'giant', which is inflected *Riesen* in all forms except the nominative singular. Other German results that have been excluded from the analysis are derived adjectives such as *riesenartig* and *riesenhaft* (both meaning 'gigantic'), or determinative compounds such as *Riesenheim* 'home of giants'. The difference between determinative compound and prefixoid construction is illustrated in (12)-(13): (12) is a determinative compound in which *Riesen* has its original meaning ('a joke about a giant'), whereas in (13), *Riesen* intensifies the meaning of the second compounding element ('a big joke').^{iv} Thus, examples such as (13) have been retained in our data set, but constructions of the type illustrated in (14) have been removed.

- (12) Ach ja, und über Zwergenwitze lacht sie zwar findet sie aber genauso platt wie unsereiner manchen Riesenwitz.
 (Well, yes, she does laugh at jokes about dwarfs, but she finds them every bit as vulgar as many a joke about giants.)
 (DECOW2012-03X_25024656)
- (13) Anfangs hält Pietro das Ganze natürlich für einen Riesenwitz.
 (Initially, Pietro thought the whole thing was a big joke, of course) (DECOW2012-03X_18003276)

The Swedish samples, by contrast, contained only very few examples that were not relevant to our investigation; the noun *jätte* (giant) did not occur in the sample. Those few instances that did have to be removed were repeated free forms (e.g. *jätte jätte mysigt* (very, very cosy)); which occurred twice in the sample but was counted only once) or the adjective *jättelik* (gigantic).^v

The remaining examples were entered into a database and each occurrence was tagged for R1 (i.e. the second compounding element in bound constructions, or the first word to the right in free constructions), part of speech of the R1, and semantic type (e.g. simile). This database forms the basis for the quantitative analyses in the next section. Table 2 summarizes the number of analysed bound and free forms for both languages.

	Bound forms	Free forms	Totals
DECOW2012-03X	864	661	1525
SECOW2011X	990	977	1967

 Table 2: Number of analysed tokens per language

3. Contrastive analysis

In this section, we present the results of our German and Swedish case studies. As we stated in the introduction, it is the purpose of these case studies to examine whether synchronic corpus data can reveal whether a prefixoid has debonded, or whether it is just a spelling variant of the bound form. In our view, the bound and free constructions are to be considered as two separate constructions if distributional and semantic differences can be shown to be statistically significant. To test this, we have analysed three sets of properties: (i) formal properties, i.e. part of speech of the R1 (i.e. the second compounding element in bound constructions and the first word to the right in free constructions), and for German, use of upper and lower case R/r (section 3.1); (ii) semantic properties, i.e. meaning of the R1, and the degree of bleaching of the prefixoid (section 3.2); (iii) productivity (section 3.3). Our analysis will show that the German bound and free construction are in fact two separate constructions, whereas in Swedish the free forms are merely reflections of a general tendency in Swedish to write compounds as two words (cf. SRB, 2005, p. 43).

3.1. Formal properties

In this section, we will first examine whether the bound and free morphemes corresponding to 'giant' in German and Swedish occur in the same construction types. To this end, the part of speech (POS) of the R1 of the bound and free morphemes will be compared. Secondly, we will look at the usage of upper case and lower case spellings of [R/r]iesen(-). Both properties will be analysed statistically, in order to establish whether the free uses are only a matter of spelling, or whether the bound and free uses reflect two separate constructions.

Figure 1 summarizes the results with respect to the R1 of bound and free morphemes in both languages. German clearly has a strong preference for nouns as R1 of [R/r]*iesen*, both in bound (88%) and free (97%) constructions. In nominal constructions, R1 is mostly a single noun, but there are also a few cases where it is clear that *riesen* takes scope over an NP:

(14) das spiel hat eine riesen deutsche community wo man genug hilfe findet.
(the game has a huge German community where one can find plenty of support)
(DECOW2012-03X 77153929)

In example (14), *riesen* modifies the NP *deutsche community*, not just the adjective *deutsche*, since it is clear from the context that a big community is being referred to, not a community that is 'very' (i.e. typically) German.



Figure 1: Comparison R1s of [*R*/*r*]*iesen*(-) *and jätte*(-)

A striking property of the Swedish prefixoid is that, at a first glance, bound and free constructions are highly similar in every respect. As can be seen in Figure 1, the parts of speech with which $j\ddot{a}tte(-)$ collocates differ only slightly from one another. An overwhelming majority of collocations is adjectival (76% in bound constructions, 77% in free constructions); followed by adverbial constructions (11% in bound constructions, 13% in free constructions).^{vi} The third largest group consists of collocations with the quantifiers *mycket* (much), or *många* (many); 8% in both bound and free constructions). Predicative usage of *jätte* is very rare – only one example occurs (see example (15)), but it is not clear what *jätte* could mean, or which adjective may have been elided (the example could not be retrieved by Google either).^{vii}

(15) Ja, den är kanske inte så jätte, men jag tycker det.
(Well, it may not be so great / very ??, but I think so)
(SECOW2011X_1826573)

For both languages, we tested whether the differences for free forms and bound forms were significant, and if so, whether the correlation was weak, moderate, or strong.^{viii} The results are summarized in Table 3: for German, the difference

between bound forms and free forms is significant, even if the effect size is only small. In Swedish, there is no significant difference between bound and free constructions.

	German	Swedish
Pearson's χ ²	40.615	6.079
p-value	p < 0.001	p = 0.108
Cramér's V	0.166	-

Table 3: χ^2 analysis of POS of R1

For German, a second property of the R1 can be examined, i.e. the use of upper case *R* or lower case *r*. Nouns in German are written with upper case initials, which means that, following German spelling conventions, nominal compounds should be spelled with upper case R.^{ix} As is evident from Figure 2, this is indeed the case in the overwhelming majority of bound constructions (96%). Interestingly however, a completely different picture emerges from the free constructions, where upper case *R* is only used in 10% of all attestations. The difference is statistically significant ($\chi^2 = 1053.797$; p < 0.001); the effect size is large (Cramér's V = 0.867).



Figure 2: Use of upper case *R* and lower case *r* in German [R/r]iesen(-) constructions with R1 = Noun

To conclude this section, the statistical analysis shows that German bound and free constructions are two separate constructions, whereas the Swedish constructions are not.

3.2. Semantic properties

After having compared the formal properties of bound and free [R/r]iesen(-) and $j\ddot{a}tte(-)$ constructions, we will now contrast the meaning of the R1. First, we will provide a survey of the most frequent R1s, both nouns and adjectives. Secondly, we will examine the degree of bleaching of the prefixoid, which will be determined on the basis the relative proportions of three different semantic types: the intensifier uses (e.g. Gm. *Riesenspaß* (big fun)), similes (e.g. Gm. *Riesenhaus* (a very large house),^x and classifying uses (names of huge species of plants and animals (e.g. Gm. *Riesenschildkröte* (giant tortoise)). While in similes and classifying constructions reference to the prototypical (geometric) properties of giants persists (1.1 and 1.2), this is no longer the case in the intensifying uses which can therefore be considered to have undergone a more advanced degree of semantic bleaching.

R1:	Bound [R/r]iesen-		Free [R/r]iesen		
frequency					
N	Rad 'wheel'	45	<i>Spaβ</i> 'fun'	48	
	Erfolg 'success'	29	Unterschied 'difference'	19	
	<i>Gebirge</i> (place name)	28	Problem 'problem'	13	
	<i>Spaβ</i> 'fun'	27	Auswahl 'choice'	13	
	Schritt 'step'	20	Erfolg 'success'	13	
	<i>Chance</i> 'opportunity'	18	Freude 'joy'	10	
	Problem 'problem'	9	Aufwand 'effort'	8	
	Berg 'mountain'	8	Fan 'fan'	8	
	n < 8	579	n< 8	511	
Adj	<i>groβ</i> 'big'	101	<i>groβ</i> 'big'	17	
9			<i>positiv</i> 'positive'	1	
Totals		864		661	

Table 4: Most frequent R1s of bound and free [*R*/*r*]iesen(-)

Table 4 shows the most frequent R1s of bound and free [R/r] iesen. The most frequent collocation in the bound forms is the word for 'Ferris wheel', *Riesenrad*, which only occurs once in the sample of free forms. The third most frequent form is a geographical name (*Riesengebirge*), which does not occur in the free

constructions at all. A further difference between bound and free forms is that in the latter, 11% of the R1s consists of English loan words, e.g. *Party* or *Cliffhanger*, whereas only 4% of bound constructions involve English loans. On the other hand, there are R1s that frequently occur in both bound and free forms, such as *Problem* (problem) and *Spa* β (fun). As far as adjectives are concerned, there is either very little variation or none at all, so these will not be discussed any further at this point.

As we can see in Table 5, bound and free $j\ddot{a}tte(-)$ are very similar with respect to R1 frequencies as well. For nouns, the numbers are relatively small, so the similarities may be a matter of coincidence, but the parallels in the adjectival constructions are striking: the top 4 is ranked identically (the frequencies of number 1, *bra* (good) are almost the same even), and the other R1s may be ordered somewhat differently, but they are mostly the same types.

R1: frequency	Bound jätte-		Free <i>jätte</i>	
Ν	ont 'pain'	7	ont	10
	mys 'cosiness'	4	mys	3
	skoj 'fun' 4		n < 3	12
	bråttom 'hurry'	3		
	grattis 'congrats'	2		
	bröst 'breast'	2		
	n < 2	22		
Adj / Adv / Quant	bra 'good'	154	bra 'good'	150
0 0	fin 'fine'	124	fin 'fine'	79
	<i>mycket</i> 'much'	75	mycket 'much'	71
	kul 'cool'	50	kul 'cool'	53
	glad 'happy'	35	rolig 'nice'	42
	god 'good, tasty'	30	mysig 'cosy'	32
	rolig 'nice'	28	trött 'tired'	29
	stor 'big'	25	trevlig 'nice'	29
	<i>söt</i> 'sweet, cute'	21	<i>länge</i> 'long', adv	25
	mysig 'cosy'	20	söt 'sweet, cute'	22
	n < 20	384	n < 22	419
No R1		0		1
(predicative use /				
interjection)				
Totals		990		977

 Table 5: Most frequent R1s of bound and free jätte(-)

The results pertaining to the degree of bleaching of the R1s are summarized in Figure 3.



Figure 3: Bleaching^{xi}

As is evident from Figure 3, German bound and free forms are strikingly different as far as semantic types are concerned.^{xii} Classifying constructions make up more than 14% of the bound forms; these include botanical and zoological names such Riesenbärenklau (giant hogweed), Riesenhai (basking shark), as well as geographical names such as *Riesengebirge*. In the sample of free forms, only three such names occur. The most striking difference, however, is that intensifying constructions are far more frequent in the free forms - no less than 70% consists of intensified nouns such as riesen Aufwand 'giant effort', riesen Arschloch 'big asshole', riesen Dankeschön 'big thank you', or riesen Freude 'huge delight'. In the bound forms, on the other hand, almost 50% of the tokens are similes, e.g. Riesenschnitzel 'huge schnitzel', Riesenosterhase 'giant Easter Bunny', Riesenwelle 'giant wave', or Riesentrampoline 'giant trampoline'. Considering that the botanical names may be considered similes as well (a giant hogweed is a very large kind of hogweed), it is evident that the German bound forms are strongly associated with simile meaning, whereas the free forms are strongly associated with intensifying meaning, and hence are more bleached.

In Swedish, by contrast, the number of simile constructions is very low, especially in adjectives, suggesting that Swedish *jätte* is far more bleached than its German counterpart.^{xiii} Figure 3 furthermore reveals that bound and free *jätte(-)* are very similar with respect to proportions of similes and intensifying constructions. Simile nominal compounds include *jätteskylt* 'huge sign' and *jätte*

382

mage 'huge belly'; simile adjectival compounds are e.g. *jättehög* 'very high', or *jätte stor* 'very big'. Non-similes which, as said, are far more common, are collocations such as *jättebråttom* 'big hurry', *jätteöverraskning* 'big surprise', *jätterosa* 'very pink', *jätte ledsen* 'very sad'. In fact, *jätte* has been bleached to such an extent that it may even combine with adjectives that we might call "anti-similes", such as *jätteliten* 'very small', or *jättekort* 'very short'. Other remarkable collocations include a comparative form (*jättesnabbare* 'much faster'), prefixed adjectives (*jätteobekvämt* 'very uncomfortable', or *jätte onödigt* 'very unnecessary'), or combinations with other intensifiers (*jättemegatrött* 'very tired'). All this suggests that *jätte*, both free and bound, functions as an intensifier in ways similar to traditional intensifying adverbs such as *mycket* 'very'.

Figure 3 shows that, in general, the bleached intensifier uses are more typically associated with the free forms, and this holds for both languages. Probably language users more easily reanalyse the 'giant' morphemes as free adjectives or adverbs when the original 'giant' meaning is no longer available, and vice versa, prefer to use compounds for similes and classifying uses.

Table 6 summarizes the results of the chi-squared tests and their association strengths. The sample size of German adjectival constructions was too small, so these have been excluded, but the noun constructions show a significant difference, with moderate effect size. For Swedish, finally, there was no significant difference in the nominal constructions, and although p < 0.001 in the adjectival constructions, Cramér's V is < 0.01, which means that the association is very weak. These results correlate with our findings in the preceding section, i.e. that bound and free constructions are significantly different in German, but not, or only very weakly, in Swedish.

	German		Swedish	
	R1=N	R1=AAQ	R1=N	R1=AAQ
Pearson's χ^2	194.892	-	3.176	15.966
p-value	p<0.001	-	p=0.075	p < 0.001
Cramér's V	0.372	-	-	0.094

Table 6: χ^2 analysis of semantic types, with R1=Noun and R1=Adjective / Adverb / Quant

3.3. Productivity

Following Taylor (2002: 290ff.), we will assume that the more "salient" a word formation schema, the more likely it is to be able to sanction new instances of that

schema, i.e. the more productive it will be, which is reflected not only by a relatively high number of types, but also by a relatively high number of hapax legomena (i.e. collocation types that occur only once in a corpus or sample). We therefore used two measures for productivity: type / token ratio (TTR) and Potential Productivity (PP). The latter is discussed in Baayen (2009), and is calculated by dividing the number of hapax legomena of a particular word formation pattern in the corpus by the total number of tokens of that pattern (Baayen, 2009, p. 902). This ratio will allow us to compare the potential growth rate of the bound and free morphemes in both languages.



Figure 4: Type token ratio and potential productivity

For German, type / token ratio and potential productivity have been calculated for both nouns and adjectives. As far as nominal constructions are concerned, the differences between bound and free forms are comparatively small. As is shown in Figure 4, the free forms are slightly more productive than the bound forms This is due to the fact that some collocations have fairly high token frequency (see 3.2), which implies that the number of types is smaller than when all types have low

token frequency. Productivity of adjectival constructions has been added for the sake of completeness, but even without the productivity figures it is evident that a construction with a type frequency of 1 and a token frequency of 111 is not productive. For the free forms, there is one other type, but the number of tokens in general is so small that no conclusions can be drawn from this.

An online survey of Swedish speaker's prefixoid preferences^{xiv} has shown that *jätte*- is by far the most popular prefixoid, both for men and women in all age groups. We may thus expect that token frequency is very high. In nominal constructions, the number of types is relatively large, resulting in high type / token frequency as well as potential productivity. However, there are not many nominal compounds in the sample, as Swedish *jätte* far more often collocates with adjectives. As we see in Figure 4, however, productivity of adjectival constructions is low. This, we think, is less due to schema salience than it is to the fact that there are far more nouns in Germanic languages than there are adjectives, especially since nominal compounding is an extremely common word formation process which gives a significant boost to the number of possible nouns in a language.

4. Discussion and conclusions

On the basis of the quantitative data presented in our case studies, we can now draw the following conclusions.

German bound and free [R/r] iesen, can be said to represent two separate constructions. Both as regards part of speech of the R1, use of upper / lower case, and semantic type of the R1, the two constructions have been shown to be significantly dissimilar. Furthermore, free [R/r] is clearly more bleached than the bound variant. The free variant is also more productive. In nominal constructions, the free forms can often be reinterpreted as an adjective modifying the R1. German adjectives are inflected for three genders and four cases in the singular, four cases in the plural, and definite and indefinite variants for all of those, resulting in 32 adjectival constructions. However, many adjectival forms are the same, and no less than 19 out of those 32 possibilities have a form in -en. In other words, there are 19 potential bridging contexts (Heine, 2002), in which free riesen may be plausibly reanalysed as an adjective. If such a reanalysis would indeed have occurred, we would expect other inflected forms of this new adjective as well, so we carried out some additional searches in the same COW subcorpus. This yielded results of the forms rieser (masculine, nominative, singular, indefinite), as in example (16), rieses (neuter, nominative or accusative, singular,

indefinite), as in example (17), or *riesem* (masculine, dative, singular, indefinite), as in example (18).

- (16) Ich habe mir das Buch nur gekauft, weil ich ein rieser Supernautralfan bin.
 (I only bought the book because I am a huge Supernautral fan.)
 (DECOW2012-03X_534599141)
 (17) Ein rieses Dankeschön nochmal
- (17) Ein rieses Dankeschon nochmal (A huge thank you, once more) (DECOW2012-03X_724419831).
- (18) Mit einem fetten Kopf, und riesem Muskelkater stand ich auf.
 (I got up with a heavy head and terrible muscle pain)
 (DECOW2012-03X_1198527106)

As far as Swedish is concerned, there is no reason to assume that the bound and free forms are two separate constructions. Distributional differences are non-significant. The free forms may simply reflect the general tendency, not yet accepted but very common in informal written Swedish, to write compounds as two separate words.

On the basis of our quantitative analyses, we argue that debonding has taken place in German, but not in Swedish. It is likely that, once debonded, the German morphemes developed in directions different from their bound counterparts, a process which Hopper (1991) has termed "divergence".

Our case studies have shown that debonding presupposes a high degree of bleaching – if the prefixoid develops more abstract meanings, such as 'huge' and 'very' typically assumed by adjectives and adverbs, it can more easily be reanalysed as an adjective or adverb. This is in line with Booij's (2010, p. 61) observation that "[t]he meaning of intensification that is connected to these nouns is a type of meaning expressed prototypically by adjectives [and adverbs, MN & KVG], and hence the categorical reinterpretation of these nouns as adjectives [and adverbs] in this context is a natural development". Note however, that the reverse is not necessarily true, i.e. bleaching does not necessarily result in debonding. Swedish *jätte-* has been bleached substantially, but it is still a prefixoid; the free forms are simply spelling variants, as we have argued above.

It is also likely that debonding correlates with productivity: a high type / token ratio and potential productivity increase the number of contexts in which reanalysis may take place. German adjectival constructions are a case in point, because the type / token ratio in bound forms approaches zero (the only type is

riesengro β), the free forms are probably spelling variants, so expansion to other R1s is extremely limited.

To conclude, we hope to have shown that the quantitative analyses of synchronic data, as advanced in this paper, may be fruitfully used to uncover diachronic changes. If two constructions are significantly different, we may assume that these differences are due to diverging developments. Specifically, the significant differences between German bound and free 'giant' constructions strongly suggest that debonding of the prefixoids has taken place, after which the bound and free constructions went their separate ways. Conversely, from the absence of significant differences it may be inferred that no change has occurred, as in the case of Swedish *jätte*, where the bound and free forms are most probably spelling variants of one and the same construction, in other words, there is no evidence that the Swedish prefixoid has debonded. This, of course, leaves the question of why debonding takes place in some prefixoid constructions, but not in others. We suspect this is related to formal factors, such as vowel clashes or the presence or absence of bridging contexts. Examining such factors in more detail, both qualitatively and quantitatively, we leave for future work.

References

- Baayen, H. (2009). Corpus linguistics in morphology: Morphological productivity. In A. Luedeling, & M. Kyto (Eds.), *Corpus Linguistics. An international handbook* (pp. 900-919). Berlin: De Gruyter.
- Booij, G. (2009). Compounding and Construction Morphology. In R. Lieber, & P. Štekauer (Eds.), *The Oxford Handbook of Compounding* (pp. 201–216). Oxford: Oxford University Press.
- Booij, G. (2010). Construction Morphology. Oxford: Oxford University Press.
- Booij, G., & Hüning, M. (2014). Affixoids and constructional idioms. In R.
 Boogaart, T. Colleman, & G. Rutten (Eds.), *Extending the scope of construction-based grammar* (pp. 77-105). Berlin: De Gruyter.
- Haspelmath, M. (1999). Why is grammaticalization irreversible? *Linguistics*, 37(6), 1043-1068.
- Heine, B. (2002). On the role of context in grammaticalization. In I. Wischer, & G. Diewald (Eds.), *New Reflections on grammaticalization* (pp. 83-101). Amsterdam / Philadelphia: John Benjamins.
- Hoeksema, J. (2012). Elative Compounds in Dutch: Properties and Developments In G. Oebel (Ed.), *Intensivierungskonzepte bei Adjektiven und Adverben im Sprachvergleich* (pp. 97–142). Hamburg: Verlag Dr. Kovac.

- Hopper, P. J. (1991). On some principles of grammaticization. In E.C. Traugott & B. Heine (Eds.) *Approaches to Grammaticalization, vol. I* (pp. 17-35). Amsterdam / Philadelphia: John Benjamins.
- Klara, L. (2012). *Steinreich* ist *saucool*: Adjektivische Steigerungsbildungen des Gegenwartdeutschen. In G. Oebel (Ed.), *Intensivierungskonzepte bei Adjektiven und Adverben im Sprachvergleich* (pp. 143-170). Hamburg: Verlag Dr. Kovac.
- Leuschner, T. (2010). Ausnahmepianist Fettgeschreckt-inbleich! Deutsche, Nierländische und Schwedische präfixoide im Spannungsfeld von Genealogie, Kreativität und Norm. In A. Dammel, S. Kürschner, & D. Nübling (Eds.), Kontrastive Germanistische Linguistik (Germanistische Linguistik, 206–209) (pp. 863–892). Hildesheim: Georg Olms Verlag.
- Norde, M. (2009). Degrammaticalization. Oxford: Oxford University Press.
- Norde, M., Beijering, K., & Lenz, A. (2013). Current trends in grammaticalization research. *Language Sciences*, *36*, 1-6.
- Norde, M. & Van Goethem, K. (2014.). Emancipatie van affixen en affixoïden: degrammaticalisatie of lexicalisatie? To appear in *Nederlandse Taalkunde*.
- Oebel, Guido (2012). Elation im Schwedischen. In G. Oebel (Ed.), Intensivierungskonzepte bei Adjektiven und Adverben im Sprachvergleich (pp. 183-248). Hamburg: Verlag Dr. Kovac.
- Sachs, Emmy (2912). On "steinalt", "stock-still" and similar formations. In G. Oebel (Ed.), *Intensivierungskonzepte bei Adjektiven und Adverben im Sprachvergleich* (pp. 59-79). Hamburg: Verlag Dr. Kovac.
- Schäfer, R., & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. In N. Calzolari, K. Choukri, T. Declerck et al. (Eds.), *Proceedings of the Eight International Conference on Language Resources* and Evaluation (LREC 2010) (pp. 486-493). Valletta, Malta: European language resources distribution agency.
- Schäfer, R., & Bildhauer, F. (2013). Web Corpus Construction (Synthesis Lectures on Human Language Technologies). San Francisco: Morgan & Claypool.
- Schlücker, B. (2013). Non-classifying compounds in German. *Folia Linguistica* 47(2), 449-480.
- SRB (2005) = Språkriktighetsboken. Utarbetad av Svenska Språknämnden. Stockholm: Norstedts.
- Taylor, J. R. (2002). Cognitive grammar. Oxford: Oxford University Press.
- Van Goethem, K. (2008). *Oud-leerling* versus *ancien élève*. A comparative study of adjectives grammaticalizing into prefixes in Dutch and French. *Morphology*, *18*, 27-49.

- Van Goethem, K. (2011). From adjective to affix in Dutch and French. The influence of word order patterns on grammaticalization. *Studies in Language*, 35(1), 194-216.
- Van Goethem, K. (2014). *Reuze bedankt, het was echt top!* Een vergelijkend onderzoek naar loskoppeling van samenstellingscomponenten. *Internationale Neerlandistiek*, 52(1), 27-46.
- Van Goethem, K., & De Smet, H. (2014). How nouns turn into adjectives. The emergence of new adjectives in French, English and Dutch through debonding processes. *Languages in Contrast*, 14(2).
- Van Goethem, K., & Hiligsmann, Ph. (2014). When two paths converge: debonding and clipping of Dutch *reuze* 'lit. giant; great'. *Journal of Germanic Linguistics*, 26(1), 31-64.

Abstract

In this paper, we present a contrastive survey of a morpheme originally meaning 'giant' in German and Swedish. In both languages, this morpheme has developed into a prefixoid with simile or intensifying meaning. More recently, these prefixoids have been shown to occur as free morphemes as well, and it is the purpose of this paper to explore whether a quantitative analysis of synchronic corpus data can be used to determine whether the free forms are spelling variants, or whether they are truly new constructions that are the result of debonding. Drawing data from the COW corpus of contemporary web text, we compare bound and free forms on the levels of R1 collocations, semantic bleaching, and productivity. Our analysis suggests that the German prefixoid has undergone debonding, whereas the Swedish free forms are mere spelling variants.

Keywords: compounding, prefixoids, semantic bleaching, productivity, debonding, corpus linguistics, German, Swedish

Authors

Norde, Muriel Nordeuropa-Institut Humboldt-Universität zu Berlin Unter den Linden 6 10099 Berlin Germany muriel.norde@hu-berlin.de Van Goethem, Kristel F.R.S.-FNRS & Université catholique de Louvain Place Blaise Pascal, 1, box L3.03.33 1348 Louvain-la-Neuve Belgium Kristel.vangoethem@uclouvain.be

ⁱ The corpus is available, after registration, at <u>http://hpsg.fu-berlin.de/cow/</u>. The German subcorpus only contains texts from Germany, the Swedish subcorpus only contains texts from Sweden. ⁱⁱ http://hpsg.fu-berlin.de/cow/colibri/

ⁱⁱⁱ See section 3.1 for the variation between uppercase R and lowercase r.

^{iv} The difference between these two types of constructions is also reflected by their prosody: in (18), stress is on the first syllable (*'Riesenwitz*), whereas in (19), it is on the third (*Riesen'witz*; Katharina Müller, p.c.). On stress patterns in different kinds of German compounds see also Schlücker (2013).

^v One reviewer suggests that *jättelik* is an adjectival compound in which *jätte* had lost the meaning of 'giant', and hence should not have been excluded from the sample. However, *lik* (cognate with English *like*) is an adjective meaning 'resembling', so *jättelik* means 'resembling a giant > gigantic'. Had *jätte* been an intensifying prefixoid, the interpretation would have been ?'very resembling', but that is not what it means.

^{vi} Most of these adverbs, moreover, are derived from adjectives by means of the suffix -t, e.g. *trevligt* 'nicely', *gulligt* 'cutely', or *snabbt* 'rapidly'.

^{vii} The Swedish SAOB dictionary reports that predicative usage was possible, at least in an example from a 1933 newspaper (*Stämningen är jätte* 'the atmosphere is great'), but in Modern Swedish this seems very unusual (Ida Zelic, p.c.).

^{viii} We used SPSS software to calculate whether two variables (bound forms and free forms) are independent of each other. Because of the relatively large sample size (2,000 initial results for each language), we set the significance threshold at p < 0.001.When the contingency table contained only four cells (e.g. in case of POS of R1 in German, which is only Noun or Adjective), we used the continuity correction figure. When the difference proved significant, we have added Cramér's V, which indicates correlation strength: 0.10 - 0.30 indicates a small effect size; 0.30 to 0.50 a moderate one, and > 0.50 a large one.

^{ix} Adjectives are normally written in lower case, but examples such as *eine (Riesen-)Große Wiese* 'a huge meadow' (DECOW2012-03X_30564577) may suggest that language users are unsure when the adjective contains a prefixoid.

^x Simile constructions include figurative uses of the R1; for instance, we have classified Gm.

Riesenschritt (giant step) as a simile, because it is initially based on a comparison.

xi <AAQ> stands for Adjective, Adverb, Quantifier. These have been taken together here.

^{xii} [R/r]*iesen* with adjectival scope has been left out of Figure 2, because, as we have seen, the German prefixoid only collocates with one single adjective (bound forms), or two (free forms).

^{xiii} The Swedish sample contained no instances of classifying compounds, but these do occur in Swedish (e.g. *jätteloka* (giant hogweed)),

xiv http://lingvistbloggen.ling.su.se/?p=2030