Triangulating the effects of interfixation in the processing of German compounds

Gary Libben,
University of Alberta

Gonia Jarema,
University of Montreal and Centre Universitaire de Geriatrie de Montréal

Wolfgang Dressler
University of Vienna and the Austrian Academy of Sciences

Jacqueline Stark
Austrian Academy of Sciences

Christiane Pons
Austrian Academy of Sciences

Address correspondence to:
Gary Libben (gary.Libben@ualberta.ca)
Department of Linguistics
4-32 Assiniboia Hall
University of Alberta
Edmonton, Alberta T6G 2E7
Canada

Abstract

The German language shows a high degree of compounding. The process is very productive in the language and compounds may be formed from the combination of four or more roots. In addition, German compounding also shows the presence of interfixes between constituents in about 35% of all compounds. We report on a compound composition experiment that investigates the role of these interfixes in on-line compound processing. The study reveals that interfixation carries with it a processing cost in German and that this processing cost is elevated in cases in which the interfix attaches to a truncated form of the initial compound constituent. Moreover, we find that response times are increased when initial compound constituents show inconsistent interfixation patterns across the language. These results support the view that German compounds are represented in terms of their constituents and that interfixation choices are made on-line.

Triangulating the effects of interfixation in the processing of German compounds

The German language offers a rich source of evidence concerning how compounds may be represented and accessed in the mind. Compounding is a very productive word formation process both diachronically and synchronically, so that native speakers of German are likely to possess a large store of compound forms and are also likely to encounter new compounds on a regular basis.

In addition to being frequent and productive in the language, German compounding shows a high level of morphological complexity. Although most compounds consist of simple root+root combinations such as *Handschuh* (literally 'hand + shoe = glove'), more complex forms can be made by embedding simple compounds into larger lexicalized structures such as *Handschuhleder* (=glove leather) or *Lederhandschuh* (=leather glove). These can also be recombined productively to form novel, but fully comprehensible strings such as the five-member compound *Lederhandschuhreinigungsmittel* = leather glove cleanser).

Another potentially complicating factor in both the representation and processing of German compounds is that they may contain interfixes (or linking morphemes) between the roots of the compound. Thus, whereas a compound such as *Handschuh* simply involves two constituent morphemes, other compounds such as *Liebesbrief* (= love letter) are composed of the units *Liebe* (love) + *s* (interfix) + *Brief* (=letter).

Interfixes such as the –*s*– in *Liebesbrief*, have been characterized using a variety of labels in the literature. Krott, Schreuder & Baayen (this volume) refer to them as 'linking elements' for both Dutch and German. Kehayia, Jarema, Tsapkini, Perlak, Ralli & Kadzielawa (1999) employed the term 'linking morpheme' for Greek and Polish. In the German literature, the terms 'Fugenelement' (Fuhrhop, 1996) or 'Fugenmorphem' (Fleischer 1976: 121-131; Ortner &

Müller-Bollhagen 1991: 73-111, Žepić 1970) are often used. Both these latter terms employ the German morpheme *Fuge*, meaning 'seam or joint'. Finally, the term 'interfix', which goes back to Malkiel (1958) and has been used most recently by Dressler Libben, Stark, & Pons (1999) and Jarema, Libben, Kehayia, & Dressler (in press), is the one that we employ throughout this paper.

According to Krott et al. (in press) approximately 35% of German compounds contain interfixation. This calculation accords well with that of Wellmann (1991), (cited in Fokuhl 1999) who places the percentage of interfixed compounds at 27% for noun-noun compounds and at 30% for adjective-noun compounds. Thus, it seems safe to conclude that although interfixation in German has a relatively high profile in morphological characterizations, it does not dominate compound formation in the language. This places German in contrast with languages such as Greek and Polish, for which compound interfixation is almost ubiquitous.

German, like its linguistic relative Dutch, finds itself between languages that have no native interfixation (e.g., English) and languages for which interfixation is integral to the process of compounding (e.g., Greek). German interfixation is also distinguished in terms of the variety of forms possible. In this way, it differs markedly from Greek and Polish, which have essentially only one interfix (the vowel-*o*-). It also differs significantly from Dutch, in which interfixation is restricted to the forms *-s-* and *-en-* (Krott et al, this issue). In Figure 1, a classification of the main morphological subtypes of noun-noun German compounds is provided. This classification, based on the analysis of Dressler et al. (1999), contrasts simple constituent + constituent compounding with nine other subtypes. The first may be described as root + constituent compounding. German compounds such as *Sprachlabor* (language laboratory) are formed from a truncation of the first constituent *Sprache* to *Sprach*. Thus, although this type of compounding does not involve interfixation, it does show an alteration of the initial element at the inter-constituent seam.

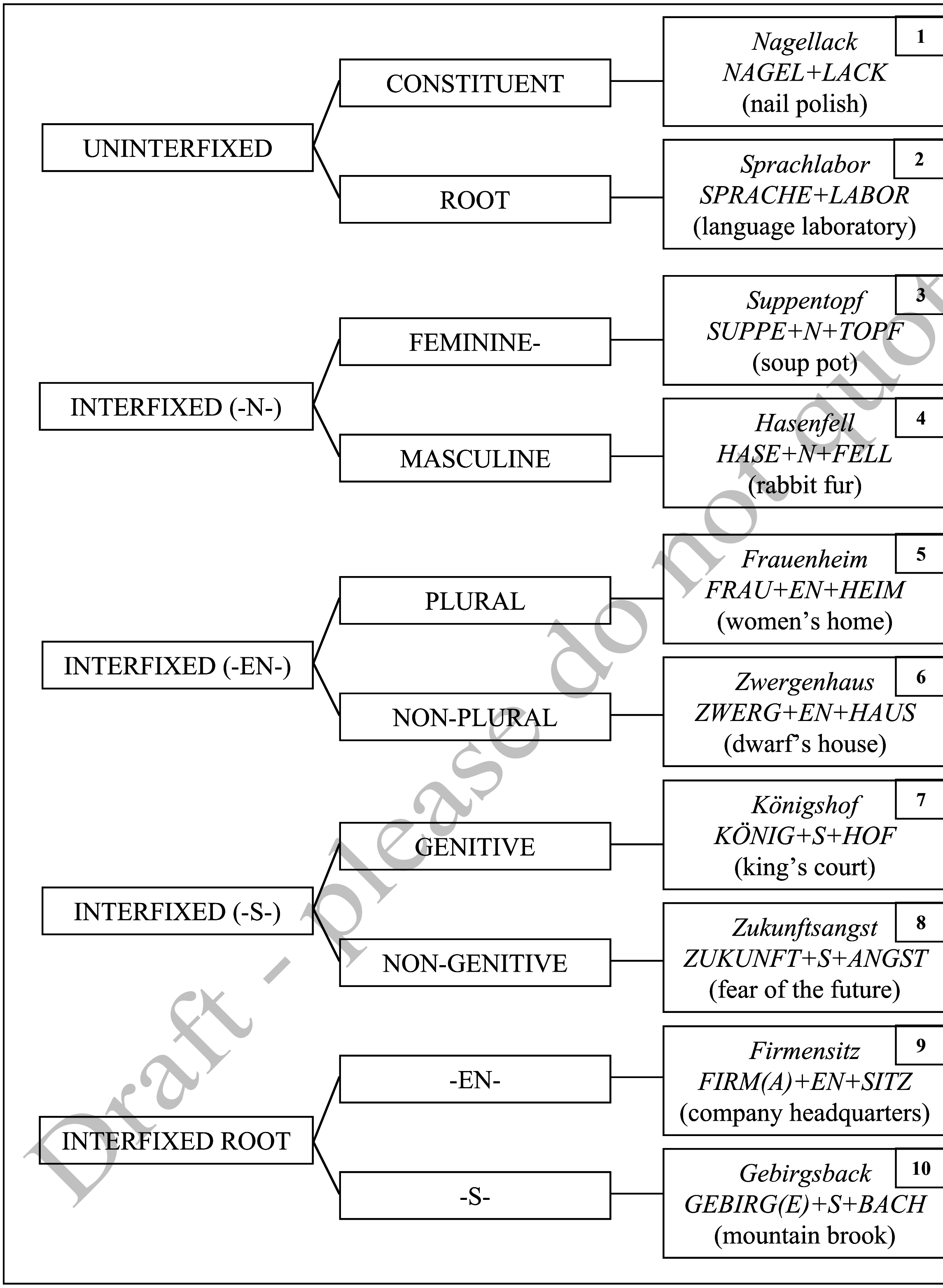

Figure 1. A classification of interfixation in German compounds.

Constituent alteration is also evident in compounds such as *Firmensitz* (company headquarters) and *Gebirgsbach* (mountain brook). Here too, the final vowel of the first constituent has been dropped. In addition, however, an interfix (*-en-* or *-s-*) has been added to the truncated forms. Both Dressler et al. (1999) and Jarema et al. (in press) found that these types of compounds resulted in longer processing times, possibly because root truncation plus interfixation makes recovery of the full initial constituent form more difficult. In the Jarema et al. (in press) study, this interpretation was based on patterns of morphological priming within compound types. Their priming paradigm allowed a compound to be preceded by one of four primes for each participant -- an unrelated prime, a root prime (e.g., FIRM) a constituent prime (e.g., FIRMA) and an interfixed prime (e.g., FIRMEN). For compounds such as *Firmensitz* and *Gebirgsbach*, they found that the full constituent was a better prime of the compound than the truncated root, even though the latter is not present in the actual compound form.

Of additional relevance, is the fact that priming effects for constituents that were not fully present in the compound (e.g. *Firma* → *Firmensitz*; *Gebirge* → *Gebirgsbach)* primed their compounds just as effectively as did constituents that were overtly present in the compound form (e.g. *Suppe* → *Suppentopf, Hase--> Hasenfell, Zwerge-->Zwergenhaus, König-->Königshof, Zukunft-->Zukunftsangst).*

The experimental paradigm employed by Dressler et al. measured constituent recoverability from compound forms in a very direct manner. Participants were shown compounds on a computer screen. After 250 milliseconds of presentation, an arrow appeared below the compound. If the arrow pointed to the right, participants were to name the right compound constituent as quickly as possible (e.g., *Sitz* for *Firmensitz*). If the arrow pointed left (the critical condition) they were to name the left constituent (e.g., *Firma* for *Firmensitz*). Constituent naming latency served as the measured variable and showed a significant relative delay when the interfixed compound was root-based. The other categories of interfixation, however, showed less clear-cut results. In general, shorter interfixes (-*n*- and -*s*-) showed faster decompositions times, suggesting that characteristics of the interfix affect the ease with which constituents may be extracted from their full compound forms. However, the primary basis of classification, the relation of interfixes to inflectional suffixes, did not significantly affect response times in the task. In particular, the expected difference between compounds such as *Königshof* and *Zukunftsangst* did not obtain. A response time difference was expected for these two categories because, in the case of *Königshof*, the -*s*- interfix is homophonous with the genitive singular form appropriate to the initial constituent. In the case of *Zukunftsangst*, however, the -*s*- interfix is morphologically illegal as a genitive marker. The reason for this is that genitive -*s*- can only attach to masculine or neuter nouns and not to feminine nouns such as *Zukunft*. In fact, however, this category showed the fastest response times along with regular -*s*- interfixation, simple compounding, and -*n*- interfixation for feminine nouns ending in shwa (e.g., *Suppentopf*).

The overall pattern of results obtained by Dressler et al. (1999) is reproduced in Figure 2 below.

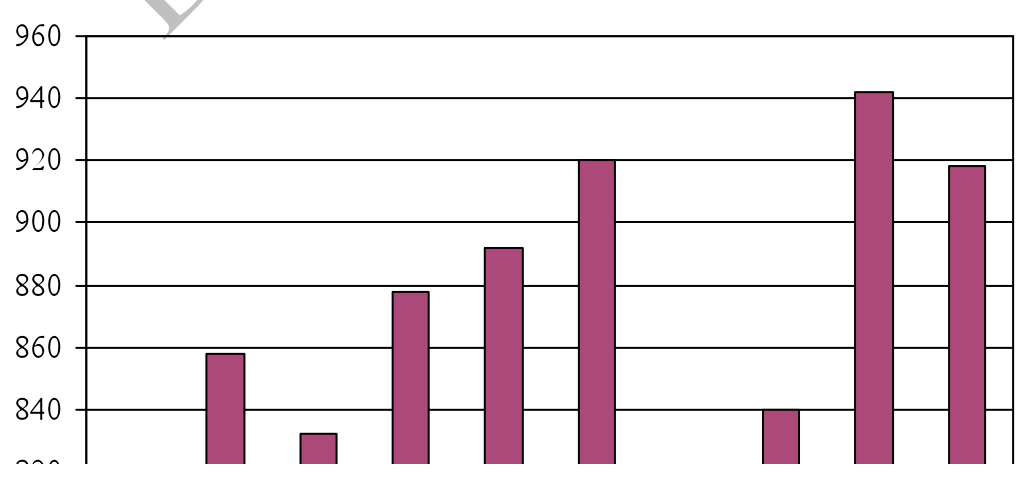

Figure 2. Mean response times in milliseconds for the root extraction task employed by Dressler et al. (1999).

As can be seen in Figure 2, compounds that involve alterations to the constituents show elevated response times. This finding, together with the priming results of Jarema et al. (in press), suggest that German compounds may be represented in terms of their constituents (e.g., *Firma* + *Sitz*) rather than in terms of their surface components (e.g., *Firmen* + *Sitz*). This would account for the fact that *Firma* serves as a better prime of the compound form than does the substring *Firmen*.

The question of whether compounds are indeed represented in such a decomposed form has been the focal point of much research on compound processing (see Libben, Derwing, & de Almeida 1999; McQueen & Cutler 1998; Sandra 1990). At issue is whether compound words are assembled and dissembled into their constituent morphemes during lexical processing and how such morphological processing might interact with whole word processing. Because interfixation in German adds a level of morphological complexity to the process of compound formation and because it shows variation across subcategories of compounds, it may offer a key opportunity to learn how morphological variation interacts with processing complexity.

In the experiment reported below, we investigated the role that interfixation plays in German compound processing by employing what may be termed a 'compound composition' task. Like the Dressler et al. decomposition task, the composition task was designed to target the role of interfixation in compound formation. In this case, however, the experiment focuses on the other side of the processing coin, by measuring the ease with which compounds can be created from their constituents. Thus, instead of measuring the amount of time required to recover *Firma* from *Firmensitz*, we measured the amount of time required to construct *Firmensitz* from its constituents *Firma* + *Sitz*). We reasoned that such comparison data would serve to distinguish between representational, task independent effects and those that are specifically related to either composition or decomposition tasks. The experiment was conducted with the same participants as those that took part in the Dressler et al (1999) study and employed the identical stimulus set, thus serving to target task differences against a common participant and stimulus background.

Method

Participants

Twenty-four native speakers of German participated in the study. All were students of linguistics at the University of Vienna and were remunerated for their participation. All participants had completed the decomposition experiment reported in Dressler et al. (1999) prior to the composition task reported here. The two experiments were separated by an experiment on bilingual processing that employed stimuli unrelated to the processing of interfixed compounds.

Stimuli

The critical stimuli in the experiment consisted of 80 compounds representing the 10 categories of interfixation described above, balanced across categories for both whole word and initial constituent frequency. All compounds were trisyllabic with word-initial main stress, wherever possible. Identical overall morphological structure was maintained by ensuring that all compounds were right-headed, endocentric, noun-noun constructions in which the first constituent was bisyllabic and the second was monosyllabic. The only exception to this uniformity was in the category of non-interfixed root-based compounds (e.g., *Sprachlabor*) for which truncation did not allow for bisyllabic initial constituents. Because a voice key was used to collect response time latencies, it was necessary to ensure that all compounds had consonantal onsets. The entire stimulus set is provided in Appendix 1.

Apparatus

All testing was conducted in a quiet room using the Macintosh G3 computer running PsyScope 1.1 (Cohen, MacWhinney, Flatt, & Provost, 1993). A button box was used to obtain microphone input and to calculate millisecond naming latency.

Procedure

Participants were seated in front of a computer screen on which two words appeared in the following manner: "WORD 1 + WORD 2". The participant's task was to say, as quickly as possible, the compound that resulted from the combination of these words. So, for example, FIRMA + SITZ should have elicited the compound *Firmensitz*. Naming latency served as the dependent variable in the experiment and was calculated as the time elapsed from visual stimulus onset to oral response onset. Each trial consisted of the following:

(a) a blank screen for 250 ms;

(b) a fixation point presented in the middle of the screen for 500 ms;

(c) the puzzle, (e.g. FIRMA + SITZ), which remained on the screen until response onset.

One thousand milliseconds subsequent to oral response, the correct answer (e.g., *Firmensitz*) appeared in the middle of the screen for 1000 ms.

Results and Discussion

Response latency was calculated as the time elapsed from the appearance of the compound constituents on the computer screen to the onset of naming the full compound. Prior to analysis, voice key errors (3%) and erroneous responses (2%) were eliminated from the dataset. One item, *Mondenschein*, was eliminated from all analyses because the majority response was the uninterfixed form *Mondschein*, rather than the intended target. Response latencies greater than 1500 ms (4%) were recoded to 1500 ms, to preserve the data points and their position on the latency continuum.

The latency data were analyzed in a one-way ANOVA in which the ten categories of interfixation constituted a within-subjects factor in the participants analysis and a between-items factor in the items-analysis. In both the participants and items analysis, there was a significant effect of the independent variable compound type ($F_{participants}$ (9, 207)=14.3, p<.0001; F_{items}(9, 69)=6.7, p<.0001) The overall pattern of results is provided in Figure 3.

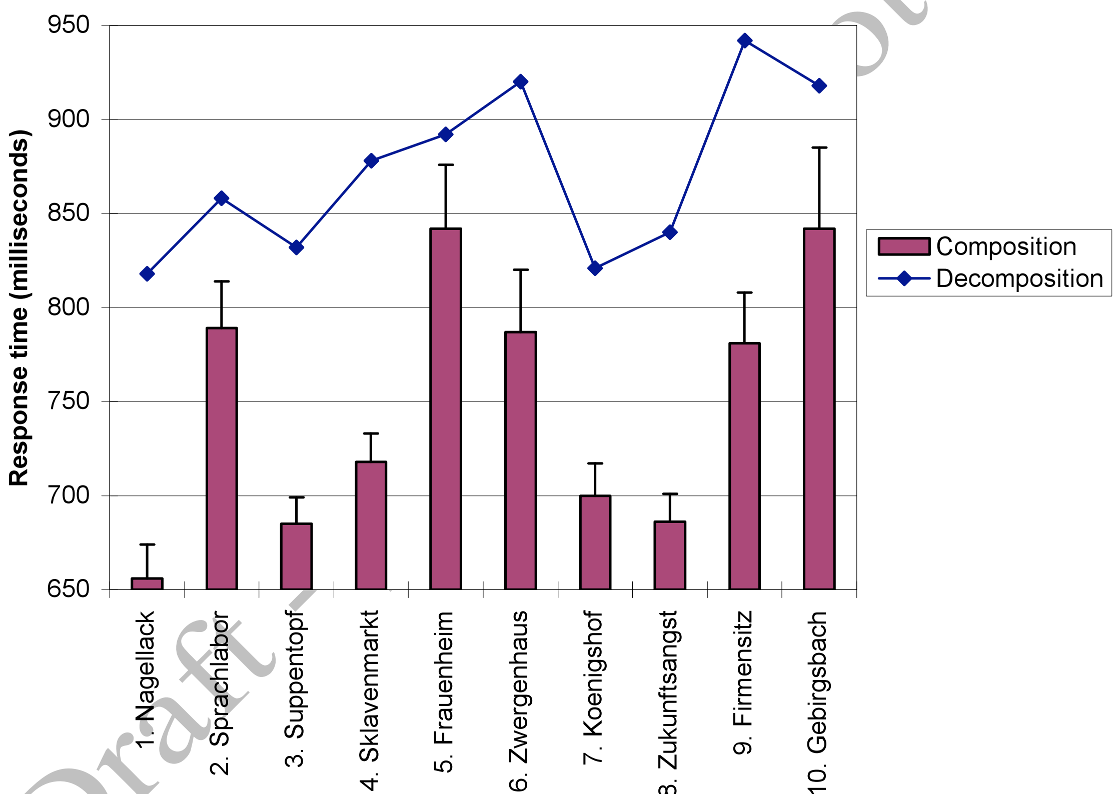

Figure 3. Compound composition response times and standard error bars in milliseconds. The line above the bar graph shows the comparison decomposition response times obtained by Dressler et al . (1999).

As can be seen in Figure 3, there is considerable similarity in the pattern of response obtained in this compound composition task and the pattern obtained in Dressler's et al. (1999) decomposition task, using the identical stimuli and participants. This similarity was evident despite the fact that the decomposition task clearly was more difficult for participants, showing considerably longer response times. (mean RT difference = 105 ms; $t(77)=12.3$, $p<.0001$).

In both tasks, the simple uninterfixed forms (e.g., *Nagellack*) showed the fastest response times. Again, there was no difference between the legal genitive forms (e.g., *Königshof*) and the illegal genitive forms (e.g., *Zukunftsangst*). Root-based forms, were associated with longer response times overall. Finally, the *-n-* and *-s-* interfixes showed shorter response times than did the *-en-* interfixes.

Overall, then, we seem to have evidence that the pattern of responses obtained by Dressler et al. do indeed reflect characteristics of compound forms that are stable across decomposition and composition tasks. Thus it may be concluded, for example, that interfixation slows processing in general, and that root-based rather than constituent-based compounding results in greater processing cost. These statements, while characterizing the pattern of results, do not explain them. Ultimately what we wish to know is what factors generate the pattern we observe across these two tasks. In order to move toward this goal, it is worthwhile to explore the cases in which the response time patterns diverge across the tasks and the cases in which there is marked variation within apparently similar categories. In the sections below, we therefore analyse the composition data patterns taking each pair of categories individually, following the classification presented in Figure 1.

Uninterfixed compounds

As the response times for uninterfixed forms such as *Nagellack* indicate, native speakers of German are fastest at simply concatenating two constituents to form noun-noun compounds. There are two ways in which this finding might be accounted for: The first is to appeal to processing complexity. Simple concatenation is the computationally least complicated of all compound formation operations. It is possible, then, that fewer steps may be required to assemble these compounds from their constituents, thus requiring less processing time. However, there is another approach to an account of the processing ease for this class of stimuli. As we have noted above, simple concatenation is the most frequent means by which German compounds are formed. Thus the speed with which these compounds are processed may reflect the fact that they represent the unmarked case for compound formation.

The account in terms of processing steps seems also to be unable to provide a convincing account for why composition times for the root-based uninterfixed compounds (e.g., *Sprachlabor*) would be more than one hundred milliseconds greater than those for simple compounding. It seems unlikely that simple truncation would induce a processing cost of this magnitude. Moreover, if computational complexity were indeed the reason for which the differences obtained between these two categories, we would expect that the magnitude of the difference would be greater for the decomposition task than for the composition task. The reason for this is that decomposition would involve a reconstruction of the entire constituent from its truncated form (e.g., *Sprach* → *Sprache*). In fact, however, the difference between these two categories failed to reach significance in the decomposition task (p=.09). By contrast, the composition task showed a highly significant difference in both the items and participants analyses (p<0001).

Thus, we are left with the following question: Why is the composition of compounds such as *Sprachlabor* so much more difficult than the composition of *Nagellack*? The answer that we propose to this question is the following:

Performing the operation of truncation or interfixation is not the problem.

Rather, the problem is deciding which operation to perform.

This line of reasoning guides the interpretation of our results for all categories of compounding. An important source of evidence in favour of this approach comes from our reanalysis of the results obtained in an off-line task reported in Dressler et al. (1999). Although this off-line task functioned essentially as a pretest in their study, it has important consequences for our present analysis. We therefore summarize the paradigm they employed: Participants were presented with three initial constituents taken from each of the ten categories of interfixation. Each initial constituent was paired with a nonsense word to create a puzzle that had the form (word+nonsense=?). Dressler et al. reasoned that if the choice of interfix is determined by the initial constituent in German (as also corroborated in the lexical statistics study of Krott et al. in press), then the obtained answers to these puzzles would correspond to the forms of the initial constituents in existing compounds. Thus, the correct answer for *Nagel+Pratsch=?* would be *Nagelpratsch*, whereas the correct answer for *Sprache+Litz=?* would be *Sprachlitz*, and the correct answer for *Firma+Grutz=?* would be *Firmengrutz* etc.

In general, Dressler et al. found a very high rate of correspondence between the interfixation choices that participants made on the basis of initial constituents paired with nonsense words and the forms of constituents in the existing compounds from which they were derived. Overall, the accuracy rate was 82%. However, the least accurate of the categories was the uninterfixed roots, which showed a correspondence of only 46%. Dressler et al. suggested that this reflects the fact that this category of compounding shows considerable variation. That is, in addition to truncated forms such as *Sprachlabor*, there exist interfixed forms in the language with the same initial constituent (e.g., *Sprachenwechsel*). By contrast, the simple interfixed forms in the stimulus set show no variation in the language (i.e., they are always uninterfixed) and showed 100% accuracy in the nonsense-word off-line task.

We suggest that it is this variability of form within the category of uninterfixed root-based compounds that accounts for their elevated processing times in general and, in particular, for their difficulty in the composition task. It is in this task that the uncertainty concerning which form to create would have the greatest effect. In the following sections, we will discuss how this factor of certainty may account for most of the variation among the compound categories.

Constituents interfixed by -n-

The next two categories of compounds (*Suppentopf* and *Hasenfell*) showed relatively low response times in Figure 3. Although the pattern of response times within this category is similar across both the composition and decomposition tasks, they are much faster in the composition task relative to *-en-* interfixed compounds than they are in the decomposition task. Again, it seems that a plausible reason for this is that the choice of interfixed form for these compounds is consistent in the language. It is almost always the case that a noun ending in shwa will take the *-n-* interfix. In the case of feminine nouns, this rule is almost invariant. In the nonsense word puzzles employed by Dressler et al., all three nouns in this category showed 100% *-n-* interfixation (that is, all participants created the string *Suppenfend* from the puzzle *Suppe+Fend=?*)

Constituents interfixed by -en-

This unanimity of response was not seen in among the *-en-* interfixed compounds.

Although *-(e)n-* compounding has been claimed by Dressler et al. to be the dominant interfix form in German, their participants in the off-line task showed only a 52% accuracy rate for the three constituents from the *Zwergenhaus* category, preferring no interfixation (e.g., *Zwerg+ Krutsch=Zwergkrutsch*) 48% of the time. The accuracy score for the *Frauenheim* category was 88%. We must note, however, that if the factor of certainty indeed accounts for the elevated response times for these two categories, it remains unclear whether degrees of uncertainty play a role in influencing response times in the composition task. It may be that latencies are influenced by how certain one is; it may also be the case that all that matters is whether one is completely certain of the appropriate form for the initial compound constituent.

Constituents interfixed by -s-

As we have noted above, the two categories of -s- interfixation did not differ significantly in response times. Of note, however, is that, as in the decomposition task, -s- interfixed compounds showed very fast response times. Again, we might look to the results obtained for the off-line nonsense word task for evidence concerning whether or not the items for each of these categories showed less variation in interfix choice. Here we do not find supporting evidence. In total, Dressler et al. employed three stimuli from these two categories of -s- interfixation that could be used for comparison purposes (*Himmel, Zweifel, Fabrik*). Together they showed a success rate of 81% (roughly equal to the average of 82% across all categories).

Interfixed roots

In contrast to the *-s-* interfixed compounds, the interfixed root constructions showed slow response times. These are unquestionably the most difficult constructions from a computational processing perspective. In the decomposition task, the extraction of the constituent *Firma* from the interfixed root *Firmen* requires both removal of the interfix as well as a reconstruction of the constituent from its truncated form. The composition task also would involve two operations—truncation plus interfixation. Thus, there appears to be a sound account of the difficulty of these compounds in terms of processing steps. However, when we consider the pattern of data across both tasks and both subcategories of root-based interfixation, we see a puzzling pattern. According to the processing steps account, the –s- and –en- interfixed roots should behave rather the same way. Moreover, that pattern should be consistent across the composition and decomposition tasks. However, as can be seen in Figure 4, the data for these categories show a interaction which was significant in the analysis by participants ($F_{1,22}=5.9$, $p=.02$) but not in the analysis by items ($p=.12$).

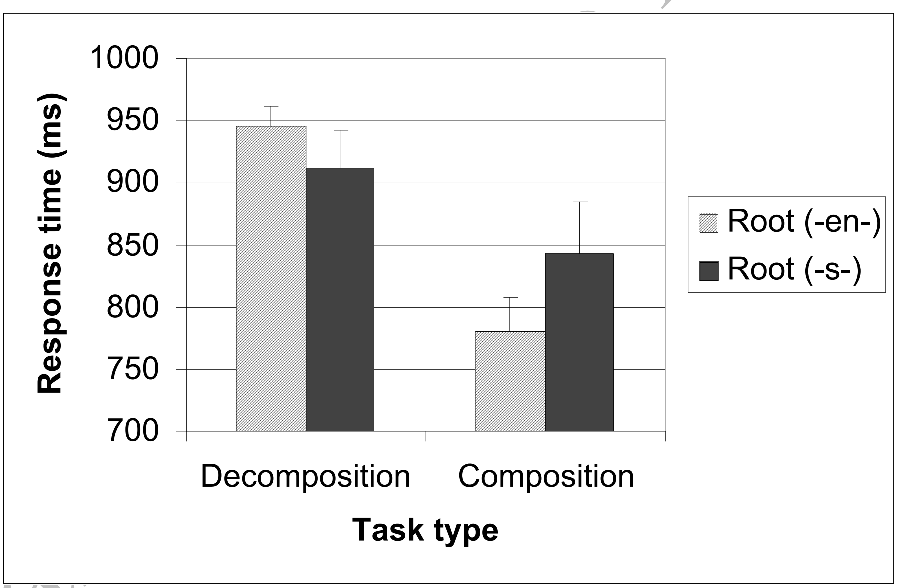

Figure 4. Response times (by items) for the two categories of interfixed roots across the decomposition and composition tasks. Error bars represent standard errors.

Taken together, the pattern of results that we have reviewed show that there are substantial differences in the manner in which members of the ten compound categories behave in this composition task. These differences do not seem to be easily accounted for by an approach that appeals to the complexity of composition procedures themselves. On the other hand, it seems that an approach that appeals to the extent to which the form of the initial constituent is invariant in compounding may hold some promise. It provides a potential reason why compounds in the *Frauenheim* and *Zwergenhaus* category show elevated response times despite the fact that their formation, seen in terms of processing complexity, should be very easy.

The first step in our exploration of the certainty account involved a closer look at the exact relation between participant's performance in Dressler et al.'s off-line nonsense compound task and our obtained composition latencies for the same constituents (a subset of our stimulus set). This relation is shown in Figure 5. We classified the off-line responses into two groups. The consistent responses (shown in Figure 5 as filled circles) represented stimuli for which all participants chose the correct form. We coded these as "consistent" stimuli. Stimuli that showed variation among participants were coded as "variable" and are represented in Figure 5 as empty circles. The stimuli are ordered from left to right on the x-axis to correspond to the stimulus classification presented in Figure 1. The y-axis represents the obtained latencies for each stimulus in our composition task.

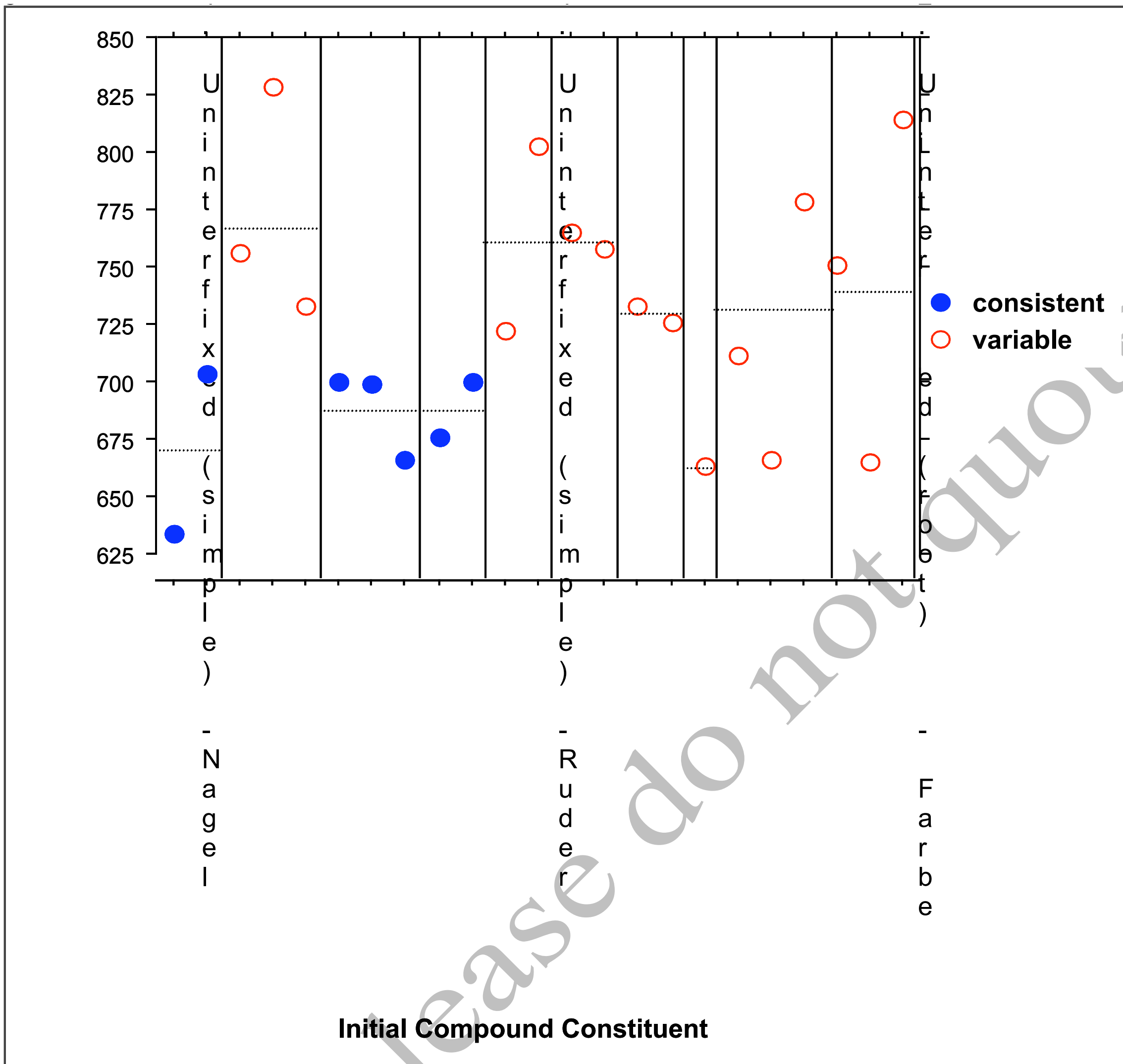

Figure 5. The relation between form variability in the nonsense word task employed by Dressler et al. (1999) and RTs for compounds possessing the same initial constituents in our composition task. The compound categories correspond to those in Figure 1 and are separated by vertical lines. The horizontal dotted lines display the mean response times for the selected stimuli in each category.

The data pattern shown in Figure 5 show three important characteristics. The first is that the consistent responses are associated with lower response times. The second is that in addition to being lower on average, the consistent responses are associated with much less RT variability, clustering between 630 and 705 milliseconds. The constituents that showed variability of interfixation choice, on the other hand, are associated with response times that vary between 670 and 830 milliseconds. Finally, it should be noted that although these stimuli represent only about 25% of the full stimulus set in our composition experiment (23 of 80 items), the average RTs for these items (shown as horizontal dotted lines) are very similar to those presented in Figure 3 for the entire stimulus set.

Thus, based on this subset data, we see some evidence that there is a relation between whether or not participants show variability in the choice of interfixation form for initial constituents of compounds and their composition RTs for existing compounds that contain these constituents. In order to explore this possibility more fully, we consulted two print-based frequency counts for German (Ruoff, 1981, Wahrig 1986) and calculated two measures for the initial constituent of each compound. The first was the degree of variation that the constituent showed in terms of interfixation form. The second was the number of compounds listed for which the lexical item served as an initial constituent. This measure approximates the positional family size for the compound constituent (see de Jong, Schreuder, & Baayen, 1999).

Our calculations were conducted in the following manner. Any constituent that was listed in either frequency count more than twice was deemed to have a positional family, all others were classified as having “no family”. We reasoned that, for the “no family” constituents, participants would have very few stored representations for the lexical element as an initial constituent of a compound. We presumed that when more than 2 items were listed in the frequency counts, participants were likely to have stored representations that would influence their processing. We classified such compounds in terms of the forms that they took in the frequency counts. If more than 95% showed the same choice of interfix as was present in the stimulus compound, we classified that compound as having a "consistent family". Those that showed between 75% and 95% accordance were classified as having a "majority family". If the accordance was less than 75% we classified the stimulus as "confusing". Table 1 exemplifies these calculations for the category of *-s-* interfixed masculine constituents.

Table 1. The calculation of stimulus categories based on positional family size and interfixation variation.

		Interfix				
Constituent	**Compound**	**none**	**-n-**	**-en-**	**-s-**	**Category**
HANDEL	HANDELSRECHT	0	0	0	76	consistent family
FEUER	FEUERSBRUNST	106	0	0	3	confusing
MONAT	MONATSFRIST	0	0	0	10	consistent family
HIMMEL	HIMMELSBRAUT	7	0	0	35	majority
HUNGER	HUNGERSNOT	16	0	0	1	confusing
KOENIG	KÖNIGSHOF	1	0	0	26	majority
ZWEIFEL	ZWEIFELSFALL	0	0	0	1	no family

Our next analysis addressed the question of whether the classification of stimuli in this manner would predict response times in the composition task. We expected the following: Compounds with no initial constituent family would show long composition times. The reason for this is that with no family upon which to base the choice of interfix, participants would show the effects of uncertainty in their response latencies. We also expected that stimuli in the "confusing" group would show long response times, but for a different reason. Here, we expected that response times would be elevated as a result of interference from forms that are incompatible with the correct response.

It was expected that the "consistent family" stimuli would show the fastest response times. These represent the ideal case: a family of experience with the form of the initial consistent and no variation among the forms that the constituent will take. Finally, we reasoned that whether the "majority" stimuli pattern with the "consistent" group or with the "confusing" group would inform us as to how fine-grained the sensitivity to interfix variation may be. The results of this analysis are presented in Figure 6.

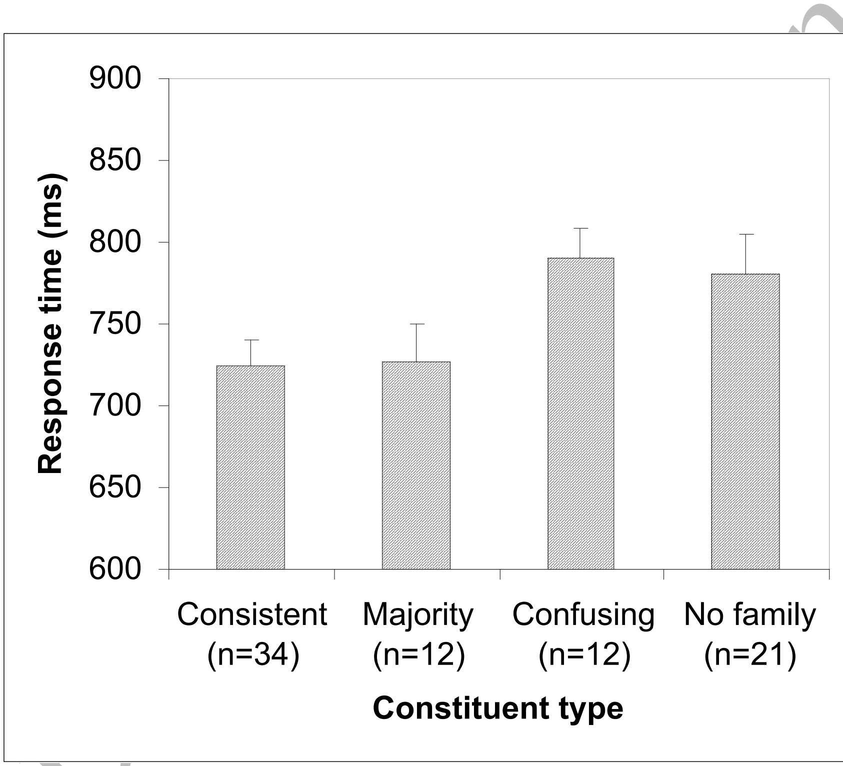

Figure 6. Response times and standard errors (by items) for the composition task. Stimuli are classified in terms of the presence of a positional family for the first constituent and the consistency of that family.

The response times shown in Figure 6 were analyzed in a one-way ANOVA by items that just reached significance ($F(13,75)=2.7$, $p=.05$). As the figure shows, the response times fall clearly into two groups. Compounds with no initial constituent family and those with a family that is not consistent with the correct answer for the stimulus compounds showed higher response times than the consistent and majority groups. It seems, therefore, that the pattern may be reduced to a relatively simple conclusion: Uncertainty creates elevated response times.

Back to the ten compound categories

In the final section of this analysis, we turn our attention back to the ten categories of interfixation with which we began. As we have seen, an important determinant of response time in this composition task is simply whether a participant can be confident in the choice of interfix for a given initial constituent of a compound. In light of the data pattern shown in Figure 6, therefore, we collapsed the four categories of certainty into two categories. Now, the final question we ask is: How does this factor of certainty relate to the pattern of response times for the ten compound categories? In other words, can the factor of certainty clarify the pattern of results with which we began this analysis? To find out, we simply tabulated, for each of the ten categories, the number of uncertain compounds that were contained within each set. This number was set against the average composition RTs for each category shown in Figure 3. These were submitted to a regression analysis with the number of uncertain stimuli serving as the independent variable and with composition RT serving as the dependent variable. The results of this analysis are shown in Figure 7.

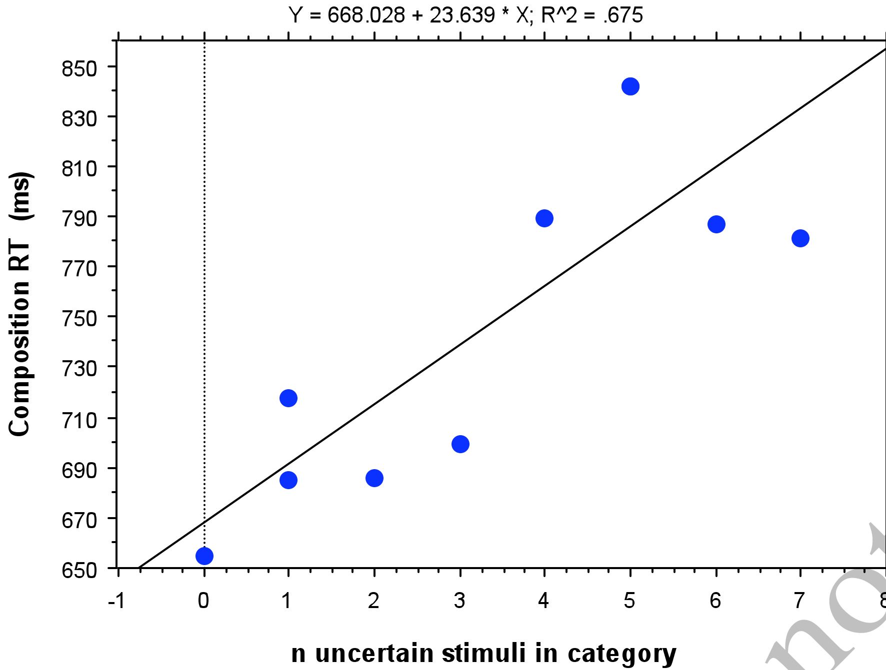

Figure 7. Regression plot for the number of uncertain initial constituents in a stimulus category and the composition RT for that category.

As can be seen in Figure 7, the relationship between the two variables is very strong (r squared = .675; $F(1,8)=16.6$, $p=.003$). Moreover, the data pattern displays the either/or threshold that we have noted throughout this analysis. The response times for the ten categories of interfixation seem to fall into two categories—fast and slow. These two categories seem to be related to two categories of interfixation choice—clear and unclear.

In summary, we have found that the evidence from a number of perspectives point to the view that clarity of choice is responsible for the majority of the variance in this compound composition experiment. In addition to offering an account for the RT data in this experiment, this account has substantial consequences for our views of how these multimorphemic words are produced on-line and the role that rules and analogy might play in this production. These issues are discussed in the following section.

General Discussion

As we noted at the outset of this paper, German compounding is made complex by the fact that it shows a variety of interfixation patterns. The results and analyses we have presented above suggest that it is precisely this variety of interfixation that creates a processing cost for particular types of compounds. In other words, while interfixation may help in compound processing for languages such as Greek and Polish which have a single interfix form (-o-), in German, it often is a source of processing confusion.

This processing confusion that we have explored leads to the following conclusion about compound processing in German—it must involve on-line computation of interfixed or uninterfixed forms of the initial constituents. Indeed, none of the effects that we have observed in this study would be possible if participants were simply extracting the full compound form from memory. Rather, it must be the case in this experiment that although the stimuli were existing words of German, and although they had been seen in their correct interfixed form (as part of the decomposition experiment) by all participants less than one hour before performing this composition task, they showed effects that were attributable to the characteristics of the initial constituents alone.

This conclusion constitutes the claim that German compounds are produced from their individual constituents and that the selection of the appropriate interfix is computed during the production process.

With respect to the question of what might guide this computation, we have noted two apparently contradictory features in our results: On the one hand, the statistical properties of the interfixation choices play an important role in determining composition response times. In other words, during the composition process, there is an effect of the experience that a native speaker has had with the forms that a particular lexical item can take as the initial constituent of a compound. Yet, this effect seems to be binary—the participant is either sure, in which case response is quick, or unsure, in which case response is slow. Another way of stating this is that if there is a rule of interfixation for a particular initial constituent, that rule is employed. Rules (i.e., relatively invariant patterns) such as "feminine nouns ending in shwa take –n- interfixation" result in fast response times and may account for the relative processing homogeneity that we see with these categories of compounds (e.g., -n- interfixation and simple concatenation).

In this report, we have attempted to employ the technique of triangulation by comparing decomposition response times, off-line interfixation choices and patterns of interfixation in the language in order to isolate the main sources of variance in our data. We suggest that further triangulation across participant populations and related languages might allow us to evaluate the claims above more completely. We expect, for example, that the study of interfixation production among Broca's aphasics would constitute an important source of evidence. If indeed, interfixation is assembled online in the production of German compounds, we should see a propensity for Broca's aphasics to delete or substitute them in both production and repetition. The relative vulnerability of interfixes to deletion and the patterns of substitution among interfixed forms would constitute an important source of evidence in the understanding of how interfixes are represented in the mind and the manner in which these representations are employed in on-line processing.

References

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. 1993. “PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers”. Behavior Research Methods, Instruments and Computers*, 25*.2, 257-271.

de Jong, N. H., Feldman, L.B., Schreuder, R., Pastizzo, M., Baayen, H. in press. “The processing and representation of Dutch and English compounds: peripheral morphological, and central orthographic effects”. Brain and Language.

Dressler, W.U. 1998. "What is the core of morphology". In J. Niemi et al. (eds.) *Language Contact, Variation, and Change*. University of Joensuu, 15-32.

Dressler, W. U., Libben, G., Stark, J., Pons, C., & Jarema, G. 2000. “The Processing of Interfixed German Compounds”. In J. Van Marle & G. Booj (Eds.) Yearbook of Morphology, Amsterdam: Foris.

Fleischer, Wolfgang. 1976. Wortbildung der deutschen Gegenwartssprache. Leipzig: Bibliographisches Institut.

Fokuhl (1999). H.-D. (1999) . “Probleme der Fugenmorphologie bei Determinativkomposita. Ein Forschungsbericht“. [http://www.fbls.uni-hannover.de/sdls/schlobi/hal/docs/hal-9.htm] Hal-9.

Fuhrhop, Nanna. 1996. “Fugenelemente”. In E. Lang & G. Zifonun (eds.) Deutsch - typologisch. Berlin: de Gruyter, 525-550.

Jarema, G. Libben, G., Dressler, W.U., & Kehayia, E. in press. "The role of typological variation in the processing of interfixed compounds: Evidence from German, Polish and Greek". Brain and Language.

Kehayia, E., Jarema, G., Tsapkini, K., Perlak, D., Ralli, A. & Kadzielawa, D. 1999. "The role of morphological structure in the processing of compounds: the interface between linguistics and psycholinguistics". Brain and Language 68, 370-377.

Kempcke, Günter. 1984. (Ed). Handwörterbuch der deutschen Gegenwartssprache. Berlin: Akademieverlag.

Krott, A., Krebbers, L., Schreuder, R., & Baayen, R. H. in press. Semantic influence on linkers in Dutch noun-noun compounds. Folia Linguistica.

Krott, A., Schreuder, R., & Baayen, R. H. 2001. Linking elements in Dutch noun-noun compounds: constituent families as analogical predictors for response latencies.

Krott, A., Schreuder, R., and Baayen, H. (this volume).

Krott, A., Schreuder, R., Dressler and Baayen, H. (in press).

Libben, Gary, Derwing, Bruce, & de Almeida, Roberto. 1999. "Ambiguous novel compounds and models of morphological parsing". *Brain and Language*, 68, 378-386.

McQueen, James & Cutler, Anne 1998. "Morphology in word recognition." In A. Spencer and A. Zwicky (eds) *Handbook of Morphology*. Oxford: Blackwell, 406-427

Malkiel, Yakov. 1958. "Los interfijos hispánicos". *Miscelánea homenaje a A. Martinet, II*. Madrid: Gredos, 107-199.

Mel'čuk, Igor A. 1982. *Towards a language of linguistics*. Munich: Fink.

Ortner, Lorelies and Elgin Müller-Bollhagen (eds.). 1991. *Substantivkomposita: Deutsche Wortbildung, IV*. Berlin: de Gruyter.

Ruoff, Arno. (1981). *Häufigkeitswörterbuch gesprochener Sprache.* Tübingen: Niemeyer.

Sandra, Dominiek. 1990. “On the representation and processing of compound words: Automatic access to constituent morphemes does not occur”. *The Quarterly Journal of Experimental Psychology*, 42a, 529-567.

Schreuder, Robert, Anneke Neijt, Femke van der Weide and R. Harald Baayen. 1998. "Regular plurals in Dutch compounds: linking graphemes or morphemes". *Language and Cognitive Processes* 13, 551-573.

Wahrig, Gerhard. 1986. *Deutsches Wörterbuch. mit einem Lexikon der deutschen Sprachlehre*. München: Mosaik Verlag.

Wellmann (1991)

Žepić, Stanko. 1970. *Morphologie und Semantik der deutschen Nominalkomposita*. Zagreb: Filozofski Fakultet Sveuèilišta.

Appendix 1

Stimuli

1. Uninterfixed (simple)

Nagellack	Ruderboot	Lederstuhl	Wasserrand
Ziegelstein	Segelschiff	Schlüsselloch	Fingerhut

2. Uninterfixed (root)

Sprachlabor	Kirschkuchen	Farbkübel	Palmwedel
Pappschachtel	Schulstufe	Wolldecke	Tuschfeder

3. Interfxed (-n-) constituent (feminine)

Suppentopf	Flaschenbier	Börsenkrach	Rosenzucht
Schlangenbiss	Kerzenduft	Fichtenharz	Hürdenlauf

4. Interfxed (-n-) constituent (masculine)

Riesenrad	Botengang	Rabennest	Falkenblick
Löwenmut	Sklavenmarkt	Hirtenhund	Hasenfell

5. Interfxed (-en-) constituent (non-plural)

Zwergenhaus	Gurtenpflicht	Mondenschein	Sternenzelt
Starenschwarm	Krebsenfang	Schelmenstreich	Straussenei

6. Interfxed (-en-) constituent (plural)

Frauenheim	Tatendrang	Formensinn	Schriftentausch
Saatenwahl	Trachtenfest	Farmenkauf	Schlachtenfront

7. Interfxed (-s-) constituent (genitive)

Königshof	Handelsrecht	Zweifelsfall	Hungersnot
Bürgersmann	Monatsfrist	Himmelsbraut	Feuersbrunst

8. Interfxed (-s-) constituent (non-genitive)

Zukunftsangst	Liebesbrief	Fabrikstor	Arbeitsdrang
Geburtsort	Heiratsgut	Armutsmass	Anstaltskleid

9. Interfixed (en) roots

Firmensitz	Themenkreis	Skalenstrich	Villenbau,
Dogmenkrieg	Prismenlicht,	Mensenplatz	Dramenpreis

10. Interfixed (-s-) roots

Gebirgsbach	Geschichtsband	Gefolgsherr	Hilfsmittel
Gehegswild	Getriebslärrm	Gewerbszweig	Gewebstod

Acknowledgements

This study was supported by special funding by the Austrian Academy of Sciences as well as a Major Collaborative Research Initiative (MCRI) grant from the Social Sciences and Humanities Research Council of Canada (Grant # 412-95-0006) awarded to Gonia Jarema (director), Eva Kehayia, (co-principal investigator), and Gary Libben (co-principal investigator). We also gratefully acknowledge the contribution of Caroline Brew in the testing phase of this research and the contribution of Matthias Schirmeier in the stimulus analysis.