Journal of Experimental Psychology: Learning, Memory, and Cognition

Compound Reading in German: Effects of Constituent Frequency and Whole-Word Frequency in Children and Adults

Jana Hasenäcker and Sascha Schroeder

Online First Publication, July 19, 2018. http://dx.doi.org/10.1037/xlm0000623

American Psychological Association

Journal of Experimental Psychology: Learning, Memory, and Cognition

0278-7393/18/$12.00

http://dx.doi.org/10.1037/xlm0000623

Compound Reading in German: Effects of Constituent Frequency and Whole-Word Frequency in Children and Adults

Jana Hasenäcker and Sascha Schroeder
Max Planck Institute for Human Development, Berlin, Germany

Current models of morphological processing differ in their assumptions about the recognition of compound words. The relative contribution of the first and second constituent and the whole-word remains unsolved. Particularly for beginning readers, the first constituent might have a privileged role attributable to more sequential decoding strategies. In a series of lexical decision experiments, the influence of constituent and whole-word frequencies on compound recognition was examined in German developing readers as well as adults. Results showed that whole-word and first constituent frequency interactively influenced response times in children. For adults, an effect of whole-word frequency only was obtained for the children's stimuli set, and noninteracting effects of whole-word frequency and first constituent frequency were found when using adult frequency measures. Together, the results suggest that developing readers already decompose compounds and that hybrid interactive models of morphological processing are most suitable to explain compound recognition across development. The applicability of amorphous models is also discussed.

Keywords: compounds, morphological processing, reading development, visual word recognition

In learning to read, each language comes with its own characteristics that pose particular challenges. One of these special challenges in German reading acquisition is the prevalence of long morphologically complex words. Many words that are encountered by children for the first time during the elementary school years in this language are, in fact, compounds (Segbers & Schroeder, 2016). Thus, the youngest readers are already regularly faced with the task of decoding those long and complex words on a regular basis. Despite this omnipresence of compounds, investigations of morphological effects in reading development have focused on derivations, while research on children's compound reading is limited and a direct comparison of the mechanisms in children and adults is absent. However, such a comparison of beginning and skilled readers can ultimately provide new insights to advance theories of compound processing.

In contrast to the sparse compound word processing research with children, research with skilled adult readers has sparked a lively debate in the past decades, centering on the question whether compound words are processed as wholes or decomposed into their constituent morphemes. Different models of complex word processing have emerged that vary in their assumptions concerning decomposition (for a more comprehensive overview of models see also Milin, Smolka, & Feldman, 2017). Full-listing accounts claim that all known complex words are stored as full forms in the mental lexicon and thus retrieved as such (e.g., Butterworth, 1983). In contrast, full-parsing hypotheses assume obligatory decomposition prior to lexical access (e.g., Taft & Forster, 1976), with access to the first constituent initiating activation of compound candidates depending on their whole-word frequency (Taft, 1994). In addition, there are several hybrid accounts that combine the two former hypotheses, assuming that access is possible both via the whole-word and the constituents. Hybrid accounts vary in their assumptions about whether one route is chosen, depending on factors such as length, lexicality, or familiarity (e.g., Caramazza, Laudanna, & Romani, 1988), or whether the routes operate in parallel, either with the faster route "winning" in a race-like fashion (e.g., Schreuder & Baayen, 1995) or with the activation of the constituents and whole-word adding to each other (Andrews, Miller, & Rayner, 2004; Kuperman, Schreuder, Bertram, & Baayen, 2009). Another important class of models are amorphous approaches that deny the use of constituents and whole-words as discrete representational units. Distributed-connectionist theories assume distributed patterns of activation across processing units instead of discrete lexical representations. In these accounts, morphological effects arise as a consequence of activation overlap over hidden units when form and meaning converge (e.g., Plaut & Gonnerman, 2000; Rueckl & Raveh, 1999; Seidenberg & Gonnerman, 2000). The Naïve Discriminative Learning (NDL) framework hinges on dynamically learned associations between co-occurrence of orthographic cues (i.e., letter bigrams or trigrams) and semantic outcomes (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011). No explicit decomposition processes and no discrete representations of constituents are assumed in this framework, nevertheless, "morphological" effects in this amorphous model arise on the basis of the probability of the outcome given the co-occurrence probabilities of the input bi/trigrams.

One typical paradigm to investigate morphological processing in skilled readers involves the systematic manipulation of constituent frequencies and whole-word frequencies of compounds. Re-

Jana Hasenäcker and Sascha Schroeder, MPRG Reading Education and Development (REaD), Max Planck Institute for Human Development, Berlin, Germany.

Correspondence concerning this article should be addressed to Jana Hasenäcker, who is now at Cognitive Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, 34136 Trieste, Italy. E-mail: janahasenaecker@gmail.com

sults from such studies are mixed with regard to the contribution of constituent and whole-word frequency. Some studies point solely to a role for whole-word frequency, at least for lexicalized and/or short compounds (e.g., Bertram & Hyönä, 2003; van Jaarsveld & Rattink, 1988). Evidence from other studies accumulates in favor of parallel influences of whole-word and constituent frequencies (e.g., Bertram & Hyönä, 2003; Kuperman et al., 2009), which operate either in a horse race fashion (e.g., Schreuder & Baayen, 1995) or interactively (e.g., Libben, 2006; Taft, 1994). Evidence from both lexical decision (van Jaarsveld & Rattink, 1988) and eye-tracking (Hyönä & Pollatsek, 1998) suggests a privileged role for the first constituent over the second constituent. In contrast, several other studies found evidence for the second constituent as the primary processing unit (e.g., Duñabeitia, Perea, & Carreiras, 2007; Juhasz, Starr, Inhoff, & Placke, 2003). Finally, Kuperman et al. (2009) even provided evidence for relevance of both constituents. Without assuming decomposition processes and constituent representations, a computational implementation of the NDL also predicted facilitation effects based on lexical characteristics of the constituents and this prediction was supported by behavioral data (Baayen et al., 2011). The NDL also correctly predicted inhibitory effects of constituent frequency in a non-European language, namely Vietnamese, which were especially pronounced when constituents had higher frequencies in relation to the whole-word (Pham & Baayen, 2015). This is interpreted as a conflict between the constituents as words of their own and the whole-word. The relative role of whole-word frequency and first and second constituent frequency has thus not been ultimately resolved and it remains to be decided which model of compound word recognition is most suitable.

Importantly, models of compound recognition have aimed to explain the processes in skilled adult readers and have largely neglected how these processes might develop. An exception to this is the NDL perspective, because learning is at the core of this model. The connections between cues and outcomes are constantly adjusted, driven by both positive and negative evidence (cf. Milin et al., 2017). For example, encountering the letter string "honeybee" would strengthen the connection between the letters contained in "honey" and the semantic outcome of something related to the edible substance produced by bees, whereas encountering the letter string "honeymoon" would weaken that connection. Thus, it is the repeated encounter with letter combination in certain contexts that impact the processing of complex words. In light of this, the neglect of a developmental perspective on compound reading is surprising. But also in the framework of the more classical decompositional theories, the lack of data on children's compound reading is astonishing, because especially for beginning readers decomposing long compounds into their smaller constituents seems like a helpful operation. Not only are the smaller constituents less demanding with regard to visual processes, but they also can be used to determine a compound's meaning. In contrast to the much more intensively studied derivations that are often assumed to be segmented via affix-stripping (Taft & Forster, 1976), compound decomposition needs to rely on a different mechanism. Stem recognition, which has recently also been suggested for derivations (Grainger & Beyersmann, 2017), is a sensible operation, because it directly activates meaning-carrying parts. Word explanation tasks have shown that already preschool children use the constituents to derive the meaning of a compound (e.g., Krott & Nicoladis, 2005). In written texts, children encounter many compounds for the first time, whereas they might have previous experience with the single constituents from different contexts. Word knowledge seems to generally help the processing of morphologically complex words (Carlisle & Fleming, 2003; Carlisle & Katz, 2006; Goodwin, Gilbert, Cho, & Kearns, 2014). Hasenäcker, Schröter, and Schroeder (2017) found that German children benefit from compound structure in reading as early as in second grade and that the development and size of this benefit is related to vocabulary knowledge. In a cross-sectional analysis from Grade 2 to 6, the authors compared the recognition of mono- and multimorphemic word recognition data. The results suggest that there might be a privileged role for the first constituent, as reading still proceeds rather sequentially from left to right in beginning readers. Häikiö, Bertram, and Hyönä (2011) used eye-tracking of sentences to explore the role of hyphenation in compound processing. Although it was not the aim of the study to examine decomposition of concatenated compounds, the results nevertheless imply that both decomposition and whole-word processing are at play in compound reading in Finnish children: slow beginning readers profit from a decomposition strategy and more advanced child readers prefer to use a whole-word strategy, suggesting development toward more holistic processing. However, neither the study design by Häikiö et al. (2011) nor by Hasenäcker et al. (2017) allows to make conclusions about the relative contribution of whole-word, first and second constituent frequencies. Despite the vast, although inconsistent evidence on frequency effects in compound processing in skilled adult readers, there are few corresponding studies using a similar design with children. De Zeeuw, Schreuder, and Verhoeven (2015) used a lexical-decision task to investigate children's use of whole-word and constituent frequencies in compound reading focusing on differences between Dutch monolinguals and Turkish-Dutch bilinguals. They used a set of 80 compounds and included whole-word and first and second constituent frequencies as continuous predictors in a regression analysis. Albeit the emphasis of this study were processing differences between L1- and L2-learners, the study provides some hints for the effect pattern that we might expect in our study. The results of de Zeeuw et al. (2015) overall indicate a clear role of whole-word frequency for second- to sixth-graders. The effects of the constituent frequencies were less decisive: the influence of the first constituent and its interaction with whole-word frequency differed by grade and the influence of the second constituent and its interaction with whole-word frequency differed by home language. As a consequence, it is still unclear which constituent—the first or the second—plays a stronger role in children's compound processing (in L1). Recall that the authors used a regression design with frequencies as continuous predictors. Although this approach can have several advantages (cf. Baayen, 2004), it is less controlled with regard to other possible factors, which then need to be controlled for statistically, requiring more data points to reach the same level of power. Moreover, the between-item comparison design might have introduced high across-item variance. Also, the compounds and their frequency measures were taken from an adult corpus (Celex: Baayen, Piepenbrock, & Gulikers, 1995), which might not ideally represent children's experience with written language. All this might have resulted in the inconsistent effects of constituent frequencies. Overall, we can learn form the de Zeeuw et al. (2015) study that (a) whole-word frequency is the most important factor and needs to be taken into account when investigating compound processing, (b) interindividual differences in language skills might additionally influence compound processing, and (c) a more controlled manipulation of constituent frequencies might be useful to gain a clearer picture of the influence of first and second constituent.

In studies with adults, the most convincing experimental design to tackle the question of the role of constituent frequencies is their orthogonal manipulation in a set of compounds that are matched on other lexical characteristics, such as length (e.g., Andrews et al., 2004; Duñabeitia et al., 2007; Juhasz et al., 2003) and, ideally, uses the same constituents in different constituent-frequency combinations (i.e., *Papierhut* vs. *Zauberhut* as in Bronk, Zwitserlood, & Bölte, 2013) to make the comparison stronger. Employing such an experimental design with children not only presents a more straightforward test of constituent frequency, but also allows relating the results for children more directly to the findings for adults.

Therefore, the present study aims at disentangling the relative contribution of first and second constituent frequencies and their possible interaction with whole-word frequency in children's and adults processing of compound words. To this end, we manipulated the constituent frequencies of compounds in an orthogonal design (frequency/constituent). For the experiments with children, we used a stimuli set chosen from a written child language corpus (childLex: Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015) as those frequency measures can be expected to more precisely approximate the experience with written words that children have. With adults, we used the child stimuli set in one experiment to evaluate how adults process the *same* words that children were tested on. In a second experiment with adults, we used a stimuli set chosen from an adult written language corpus (DWDS: Geyken, 2007) that might present adults' experience more precisely. To minimize across-item variance within each stimuli set, we used pairs of compounds that shared one constituent, while the other constituent differed in frequency (see also Bronk et al., 2013). Given the evidence for the impact of whole-word frequency as a continuous predictor for both adults (Kuperman et al., 2009) and children (de Zeeuw et al., 2015), we also included this measure.

If responses are influenced by whole-word frequency only, this would indicate whole-word processing. If responses are influenced by constituent frequencies, this would support decomposition accounts; a first constituent frequency effect would point to recognition via the first constituent, a second constituent frequency effect would suggest a privileged role for the second constituent. If first and second constituent frequencies interact, this would be evidence for parallel processing of the constituents. Finally, interaction effects with whole-word frequency would support the combined use of any information that is available to maximize opportunity for accomplishing the demanding task of reading a complex word (Libben, 2006). If lexicalized compounds are recognized as a whole and decomposition is mainly important for compounds that are not (yet) lexicalized (Caramazza et al., 1988; van Jaarsveld & Rattink, 1988), the processing of the same words should change with time. Under decompositional accounts, a stronger effect of the first constituent for children would indicate more sequential processing in beginning readers. Under the assumption that whole-word processing takes place for all lexicalized compounds, we would expect to see a development from decompositional toward holistic processing from childhood to adulthood for the child material as the compounds become more and more lexicalized. Deriving hypotheses from the amorphous accounts is less straightforward without a computational simulation of compound processing in German. Based on the available evidence, one would expect a facilitatory effect of whole-word frequency and effects of constituent frequencies that are dependent on the degree of conflict between constituents and whole-words: this could lead to facilitation, as found in English (Baayen et al., 2011), or inhibition from the constituents, as found for Vietnamese (Pham & Baayen, 2015).

The structure of the experiments is as follows: Experiment 1 provides a first investigation of frequency effects in compound reading in a group of elementary schoolchildren from different grades (2nd to 4th) to test the validity of such a frequency manipulation in children. However, reading develops considerably throughout the elementary school years and morphological processing undergoes important changes during this time (de Zeeuw et al., 2015; Häikiö et al., 2011; Hasenäcker et al., 2017). Experiment 2 therefore presents a replication of Experiment 1 with a much larger group of children, all attending Grade 4. This allows also taking into account interindividual differences in vocabulary, which have been shown to modulate the benefit of morphological structure in word recognition (Hasenäcker et al., 2017). Experiment 3 then examines adult's compound reading with the child stimuli set to directly compare how adults process the *same* words that children were tested on. Experiment 4 presents adult data from a similar experiment to test whether those words are processed differently than the highly familiar "child words." This structure of experiments provides key evidence to understand how the processing of compound words develops with increasing experience with compound words and allows us to gain more insight into the underlying representational mechanisms.

Experiment 1

Method

Participants. Twenty-two elementary schoolchildren (13 girls, M_{age} = 7.8 years, SD_{age} = 0.9, age range: 7–9 years) attending Grades 2–4 in the Berlin area were recruited to participate in the study. The study was approved by the ethics committee of the Max Planck Institute for Human Development and testing took place at the institute's test center. Written consent was obtained from the parents and oral consent was asked from the children prior to participation. One child had to be excluded from the analysis as this child was not capable of carrying out the full experimental session.

Materials. Thirty-two pairs of compounds were selected from the childLex corpus (Schroeder et al., 2015). All compounds consisted of exactly two concatenated stems, written without interword space according to German orthographic rules, and were 7–11 letters in length. A compound pair always shared one constituent (see Appendix A for a full list of all pairs); for half of the pairs the first constituent was shared (e.g., *Handschuh* and *Handtuch*), for the other half the second constituent was shared (*Autobahn* and *Eisenbahn*). Constituent frequency was manipulated in a 2 × 2-design (first/second constituent, higher/lower). Thus, four combinations emerged with 16 compounds in each group: higher-higher (*h-h*), higher-lower (*h-l*), lower-higher (*l-h*), lower-lower (*l-l*). Higher frequency constituents had a normalized lemma frequency (per million tokens) above 100 and lower frequency constituents below 100 (higher: $M = 287.82$, $SD = 228.26$, min = 105.37, max = 1069.97; lower: $M = 43.73$, $SD = 29.91$, min = 2.54, max = 99.48). Note that the cut-offs deviate from the typical cut-offs used for "low" frequency in lexical decision experiments with monomorphemic words. This is because the constituent frequency categories in this experiment represent the higher and lower parts of the distribution of constituent frequencies of typical German compounds. They can thus be regarded as "higher" and "lower" relative to the distribution found for constituents in German compounds. A lower cut-off as typically used for monomorphemic lexical decision would have resulted in the necessity to choose extremely uncommon compounds. Instead, we chose to ensure (a) a good representation of typical compounds in German and (b) children's knowledge of the words. Hence, the chosen whole-word frequency range adequately represents the distribution of the frequencies for the majority of noun-noun compounds in German, being lower than the constituent frequencies, as is usually the case for German compounds, where very common stems are used to create specialized compound meanings. The normalized lemma frequency of the whole compounds used in our experiment thus ranged between 0.71 and 38.68 ($M = 5.66$, $SD = 6.18$). Across the four groups (h-h, h-l, l-h, l-l), compounds were matched on whole-word frequency, bigram frequency, neighbors (OLD20), number of letters, and number of syllables, all $t < 1$, $p > .05$. Item characteristics are summarized in Table 1.

A list design was used, such that each participant saw a given constituent only in one combination and saw 32 compounds in total.

In addition to the compound words, 32 pseudowords were created by selecting 64 stems, changing one letter in each stem and then combining two resulting pseudostems into a pseudocompound (e.g., *Stock* "stick" and *Wolf* "wolf" were made into *Stackwalf*). Pseudowords and words were also matched on bigram frequency, number of letters, and number of syllables, all $t < 1$, $p > .05$.

Procedure. Testing took place individually in a quiet room on a laptop with a 15" monitor and a refresh rate of 60 Hz. The stimuli were presented in white 20-point Courier New font on black background. Each trial started with a 500-ms fixation cross in the center of the screen, followed directly by a stimulus, which remained on screen until a response was made by the participant. Participants were instructed to decide as quickly and as accurately as possible whether the presented stimulus was an existing German word or not and indicate their decision by pressing the *D* or the *K* key on a standard keyboard, marked red and green. Prior to the experimental trials, four practice trials with feedback (right or wrong answer) were given. For the children, a short break timed by the experimenter was included after half of the experimental trials. Accuracy and RTs were recorded.

Results

Main data analyses for words were performed using (generalized) linear mixed-effects models as implemented in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in the statistical software R. For the response time analysis, incorrect responses (12.65%) and response times below 200 ms or above 8000 ms (1.70%) were removed first and the remaining response times were logarithmically transformed. Next, model criticism based on a simple model including random effects for subject and item was used for further outlier trimming, excluding all data points with residuals exceeding 2.5 standard deviations for the main analyses (2.25%). Then, a model was fitted to the data including first constituent frequency and second constituent frequency as categorical predictors (higher vs. lower) and whole-word frequency as a continuous centered predictor (logarithmically transformed to the base 10). Their interactions were also entered as fixed effects. Random intercepts were included for participants and items. A parallel model was fitted to the error data. Post hoc comparisons were carried out using cell means coding and single *df* contrasts with the glht function of the multcomp package (Hothorn, Bretz, & Westfall, 2008) and were evaluated using a normal distribution. Mean response times are shown in Table 2. Results for the overall effects tests using contrast coding and Type III sum of squares (using the Anova function in the car package) are summarized in Table 3.

The response time analysis revealed a main effect of whole-word frequency, indicating that compounds with a higher whole-word frequency were responded to faster. There were no main effects of first and second constituent frequency. However, first constituent frequency and whole-word frequency interacted. Post hoc contrasts showed that whole-word frequency affected response times when the first constituent frequency was high, $b = 0.18$, $t = 4.26$, $p < .001$, but not when it was low, $b = 0.009$, $t = 0.14$, $p = .89$. The interaction is depicted in Figure 1.

Parallel analyses were conducted on the accuracy data. A significant interaction of first constituent frequency and whole-word frequency was observed. Post hoc contrasts showed that whole-word frequency affected error rates, leading to fewer errors, when the first constituent frequency was high, $b = 1.10$, $t = 3.04$, $p = .002$, but not when it was low, $b = 0.05$, $t = 0.12$, $p = .90$.

Table 1
Overview of Lexical Characteristics in the Four Frequency Groups and in the Entire Set of Words

Characteristic	h-h	h-l	l-h	l-l	All
Whole-word frequency	8.41 (9.80)	6.03 (5.71)	4.55 (3.68)	3.66 (2.10)	5.66 (6.18)
First constituent frequency	279.4 (226.73)	279.4 (226.73)	44.90 (31.13)	44.90 (31.13)	162.20 (199.36)
Second constituent frequency	296.20 (233.10)	42.56 (29.09)	296.20 (233.10)	42.56 (29.09)	169.40 (208.55)
Length in letters	9.25 (1.00)	9.25 (1.13)	9.13 (1.20)	9.13 (.96)	9.19 (1.05)
Neighbors (OLD20)	2.91 (.36)	2.98 (.38)	2.91 (.45)	3.08 (.39)	2.97 (.39)
Summed Bigram frequency	82999 (46738)	85785 (36269)	98705 (42479)	90426 (43143)	89478 (41738)

Note. Means are shown with standard derivations in parentheses.

Table 2
Overview of Mean Response Times for All Conditions in Experiment 1 (Standard Deviations in Parentheses)

Measure	h-h	h-l	l-h	l-l
Mean RT in ms	2100 (288)	2142 (293)	2087 (285)	2210 (303)

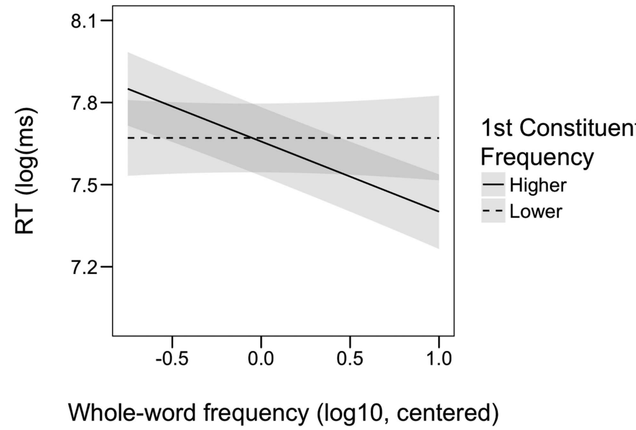

Figure 1. Mean response times (and standard errors) as a function of whole-word and first constituent frequency in Experiment 1.

Discussion

Experiment 1 investigated whole-word frequency and first and second constituent frequencies in the processing of compound words in a group of second- to fourth-graders. The results indicate that the whole-word frequency of the compound plays a major role in word recognition. The effect of whole-word frequency additionally interacted with first constituent frequency: whole-word frequency affected processing when the first constituent was of high frequency, but not when it was of low frequency. One way to think of this process is that upon presentation with a compound (e.g., *toothbrush*), the initial constituent (*tooth*) is activated and so are morphologically related words (*toothless, toothache, toothpaste*), among them the target word. In the case that both the first constituent and the whole-word are of high frequency, then activation of the presented compound is fast and strong. If the first constituent is of high frequency, but whole-word frequency is low, then there might arise inhibition from the constituent and/or higher frequent morphological relatives. If the first constituent is of low frequency, its activation is weak and slow and feeding forward of activation to morphologically related words is limited. Indeed, this fits perfectly with Kuperman et al.'s (2009) suggestion that "The higher the frequency of a complex word in language, the stronger the association between that word and its morphemes, and the more experience the reader has with integrating a given morpheme into that embedding word. If so, a high-frequency compound may benefit more from identification of one of its constituents than a low-frequency compound" (p. 885).

Overall, the results from Experiment 1 converge with previous evidence on the special role of the first constituent (Hyönä & Pollatsek, 1998; Taft & Forster, 1976; van Jaarsveld & Rattink, 1988) and are in favor of the view that compound processing is decompositional from left to right at the beginning of reading development (Häikio et al., 2011). Furthermore, de Zeeuw et al. (2015) in their study with L1 and L2 Dutch children also found an interaction of whole-word and first constituent frequency. However, in their study, the direction of the interaction was opposed to that in our experiment: the first constituent exerted more influence on compound recognition when the whole-word frequency was low. Moreover, this interaction only emerged in second, but not in fourth and sixth grade. The differences in results between de Zeeuw et al.'s (2015) and our study are surprising. They might be attributable to the different design and focus. Major differences in the experimental set up are that de Zeeuw et al. (2015) used an unmatched item set with a wide frequency range and statistically

Table 3
Results From Mixed-Effect Models for Experiment 1 With First Constituent Frequency, Second Constituent Frequency, and Whole-Word Frequency as Fixed Effects, and Participant and Item as Random Intercepts

Measure	RTs χ^2	RTs p	Errors χ^2	Errors p
Fixed effects (all df = 1)				
Intercept	3792.25	<.001*	95.86	<.001*
First constituent freq	<1	.733	3.10	.078
Second constituent freq	1.13	.287	2.56	.109
Whole-word freq	5.13	.023*	3.54	.060
First Constituent Freq × Second Constituent Freq	<1	.606	2.09	.148
First Constituent Freq × Whole-word Freq	6.25	.012*	4.27	.039*
Second Constituent Freq × Whole-word Freq	<1	.606	<1	.641
First Constituent Freq × Second Constituent Freq × Whole-word Freq	<1	.699	2.90	.089
Random effects				
Participants	737.9	<.001*	17.83	<.001*
Items	16.2	<.001*	0	1

Note. Tests are based on Type III sum of squares and χ^2 values with Kenward-Roger *df*.
* $p < .05$.

accounted for item differences, whereas we used a tightly matched item set. Furthermore, de Zeeuw et al. (2015) focused on L2 readers and compared them between grades (Grade 2 vs. 4. vs. 6), whereas we investigated L1 readers from a grade range spanning from Grade 2 to 4.

Children's reading skills and strategies develop considerably throughout the elementary school years. Especially the use of morphemes in complex word reading has been shown to undergo important changes in those years. Furthermore, vocabulary size has been shown to influence the extent to which morphological structure is utilized (Hasenäcker et al., 2017). A larger number of participants and at the same time restricted to a smaller range of years of reading instruction would allow to take into account interindividual differences in vocabulary size and present a very powerful test of the stability and reliability of the results of Experiment 1. In Experiment 2, we therefore examine a four-times-larger group of children who all attend fourth grade, and thus all have experienced the same number of years of reading instruction but differ in their vocabulary size.

Experiment 2

Method

Participants. One hundred five children (55 girls, M_{age} = 8.83 years, SD_{age} = 0.49, age range: 8–10 years) from the Berlin area participated in Experiment 2. The lexical-decision task was administered to each child in a single session in the middle (November/December) of fourth grade. The single sessions took place in a separate room in the children's schools as part of a larger longitudinal project. The longitudinal project was approved by the ethics committee of the Max Planck Institute for Human Development. Written consent for the children's participation was obtained from the parents at the beginning of the project, and oral consent was asked from the children at the start of the experimental session. In addition, vocabulary size was assessed in a group session using the standardized pen-and-paper vocabulary subtest of the CFT-20R (Weiß, 2006). Note that, because of the design of the longitudinal project, different tasks were administered at different testing points, so that the vocabulary measure was obtained six months prior to the lexical decision data. The vocabulary test consisted of 30 words for each of which a synonym had to be selected from five alternatives with increasing item difficulty. Correct answers were summed up and the resulting scores in our sample were normally distributed (M = 14.68, SD = 5.22, min = 1, max = 27) and were comparable with the age-appropriate percentile norm values (M = 49.49, SD = 10.41).

Materials and procedure. Materials and procedure of the lexical-decision task were the same as in Experiment 1.

Results

Analyses of the response time and accuracy data paralleled those of Experiment 1, and z-transformed vocabulary size was additionally included as a fixed main effect and in interaction with all frequency measures and interactions. Prior to the analysis, 9.26% of response times were removed because they were incorrect and response times below 200ms or above 5000ms (1.18%) were also removed, as were another 1.96% that were identified as outliers by the model criticism procedure (see Experiment 1). Mean response times are shown in Table 4. Model results are summarized in Table 5.

The response time analysis revealed a main effect of whole-word frequency, indicating that compounds with a higher whole-word frequency were responded to faster. There were no main effects of first and second constituent frequency, but an interaction of first constituent frequency and whole-word frequency. Post hoc contrasts showed that whole-word frequency affected response times when the first constituent frequency was high, $b = 0.19$, $t = 5.71$, $p < .001$, but not when it was low, $b = 0.02$, $t = 0.50$, $p = .62$. This interaction is shown in Figure 2. There was also a main effect of vocabulary size, indicating that larger vocabulary size resulted in faster response times. There were no interactions of vocabulary size with any of the frequency measures.

Analysis of the accuracy data showed a significant main effect of whole-word frequency with fewer errors when whole-word frequency was high. There was also a main effect of vocabulary size with larger vocabulary leading to fewer errors. There were no significant interactions.

Discussion

Experiment 2 replicated the findings of Experiment 1 with a much larger group of readers, all attending Grade 4. The results confirm a major role for whole-word frequency in children's compound reading and an interaction with first constituent frequency in the way that whole-word frequency affected processing times only when the first constituent was of high frequency.

In Experiment 2, interindividual differences in vocabulary size were additionally examined, because they have been previously found to help the processing of morphologically complex words. In Experiment 2, larger vocabulary size indeed had a positive effect on overall speed and accuracy of compound word recognition. However, it did not interact with any of the frequency measures. This suggests that the relative importance of whole-word frequency and first constituent frequency is independent of vocabulary size: Regardless of vocabulary size all 4th graders use information of the whole-word and the first constituent; children with larger vocabularies are just faster and more accurate in doing so.

Taken together, the simultaneous importance of whole-word frequency and first constituent frequency that was found in both Experiment 1 and 2 presents compelling evidence against full-parsing theories and for some form of decomposition. The results fit best with an interactive activation framework, either a hybrid or a purely decompositional one. In a hybrid interactive activation account, representations of the constituents and the whole-word are activated and used cooperatively (e.g., Baayen & Schreuder, 2000; Kuperman et al., 2009), and the first constituent has a greater role in this than the second constituent because of the pronounced

Table 4
Overview of Mean Response Times for All Conditions in Experiment 2 (Standard Deviations in Parentheses)

Measure	h-h	h-l	l-h	l-l
Mean RT in ms	1334 (54)	1249 (50)	1296 (52)	1342 (55)

Table 5
Results From Mixed-Effect Models for Experiment 2 With First Constituent Frequency, Second Constituent Frequency, Whole-Word Frequency, and Vocabulary as Fixed Effects, and Participant and Item as Random Intercepts

Measure	RTs χ^2	RTs p	Errors χ^2	Errors p
Fixed effects (all $df = 1$)				
Intercept	50615.42*	<.001*	386.44	<.001*
Vocabulary	65.6*	<.001*	7.32	.007*
First constituent freq	<1	.417	3.25	.072
First Constituent Freq × Vocabulary	<1	.321	<1	.502
Second constituent freq	<1	.578	1.11	.292
Second Constituent Freq × Vocabulary	<1	.918	<1	.342
Whole-word freq	13.17*	<.001*	9.10	.003*
Whole-word Freq × Vocabulary	2.33	.127	<1	.964
First Constituent Freq × Second Constituent Freq	3.52	.061	<1	.895
First Constituent Freq × Second Constituent Freq × Vocabulary	<1	.398	1.80	.180
First Constituent Freq × Whole-word Freq	7.77*	.005*	3.40	.065
First Constituent Freq × Whole-word Freq × Vocabulary	<1	.625	<1	.752
Second Constituent Freq × Whole-word Freq	1.93	.165	1.38	.240
Second Constituent Freq × Whole-word Freq × Vocabulary	<1	.965	1.80	.180
First Constituent Freq × Second Constituent Freq × Whole-word Freq	<1	.495	1.43	.232
First Constituent Freq × Second Constituent Freq × Whole-word Freq × Vocabulary	<1	.639	1.14	.285
Random effects				
Participants	1771*	<.001*	44*	<.001*
Items	190*	<.001*	29*	<.001*

Note. Tests are based on Type III sum of squares and χ^2 values with Kenward-Roger *df*.
* $p < .05$.

left-to-right bias in beginning readers. In the decompositional interactivation account of Taft (1994), the first constituent feeds forward activation to possible compound candidates starting with this constituent. The activation strength and speed of the compounds then depends on whole-word frequency such that the compound with the highest whole-word frequency receives most activation. The pattern of results can also be explained in the framework of the NDL (Baayen et al., 2011): as reading proceeds from left to right, the letters of the first constituent are read first, leading to activation of the output of the first constituent as its own word and this is faster and stronger based on the constituents frequency. At the same time, parafoveal information allows the reader to recognize that the first constituent is not followed by a space, but by a letter, thus whole-word frequency comes into play (cf. Baayen et al., 2011, p. 62). Depending on how strong the connections between the cues (letter input) and the different outcomes (meaning of the constituent as words on its own vs. meaning of the compound) are relative to each other, they either help or hinder word recognition of the whole-word.

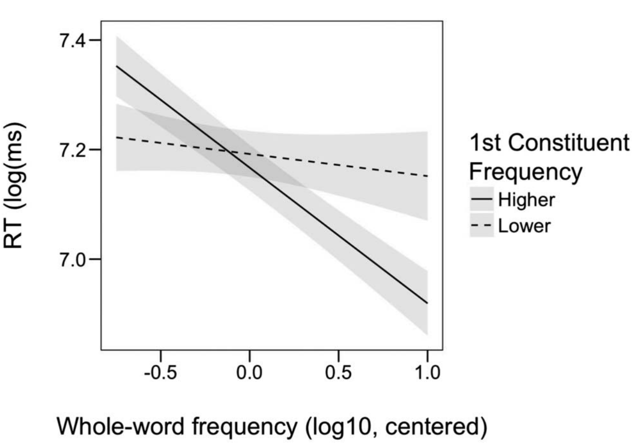

Figure 2. Mean response times (and standard errors) as a function of whole-word and first constituent frequency in Experiment 2.

Although the influence of whole-word and first constituent frequency in children is in line with some previous adult studies (Bertram & Hyönä, 2003; Kuperman et al., 2009), some other adult studies have found evidence in favor of an influence of the second constituent frequency (Duñabeitia et al., 2007; Juhasz et al., 2003). It is thus not clear whether the pattern we found for children would be the same for adult readers in German. Especially the privileged role of the first constituent might be restricted to children's reading, which is still more sequential from left to right. Adults recognize words more in parallel. Thus, they might not show the same privileged role for the first constituent. Instead, they might either rely on first and second constituent frequencies equally or more on whole-word frequency alone. Moreover, the words used in the experiments with children might be lexicalized in adults, which might also lead to stronger reliance on whole-word processing. To directly compare compound reading in children and adults, we conducted the same experiment that we did with children also with a group of adult readers in Experiment 3.

Experiment 3

Method

Participants. Twenty-two young adults (12 women, M_{age} = 26.0 years, SD_{age} = 2.6, age range: 21–32 years) from the Berlin area were recruited to participate in the study. The study was approved by the ethics committee of the Max Planck Institute for Human Development and testing took place at the test center of the institute. All participants gave informed written consent prior to participation.

Materials and procedure. Materials and procedure were the same as in Experiment 1 and 2.

Results

Analyses of the response time and accuracy data for the adults paralleled the one for the children in Experiment 1 and 2. However, because overall accuracy was extremely high for adults (98.86%), we only analyzed response times. Incorrect response times to words were removed (1.42%), as were response times below 200 ms or above 1200 ms, which amounted to 2.16%, and further 0.74% were identified as outliers by model criticism (see Experiment 1 and 2). Means response times are shown in Table 6. The model results are summarized in Table 7 (Exp 3).

The response time analysis revealed only a main effect of whole-word frequency, attributable to faster responses to compounds with a higher whole-word frequency. There were no other significant effects. The response times are depicted in Figure 3.

Discussion

The results for adults in Experiment 3 indicate only a role for whole word-frequency. This suggests a more holistic and less left-to-right processing in skilled readers and is generally in line with developmental findings by Häikiö et al. (2011). The effect of whole-word frequency independent of first constituent frequency in adults is compatible with full-parsing and hybrid accounts. The lack of any constituent frequency effects might seem surprising at first and seemingly stands in contrast to findings by Bronk et al. (2013) with adult readers in German (see also Andrews, 1986; Andrews et al., 2004; Taft & Forster, 1976 for evidence from English). However, it needs to be kept in mind that the stimulus set was designed with frequency measured from a child corpus. For the experienced adult readers, the words were likely highly lexicalized. Caramazza et al. (1988) and van Jaarsveld and Rattink (1988) suggest that highly lexicalized compounds do not require decomposition and Juhasz (2018) found that a reader's familiarity with compound words significantly influences their recognition. Adults seem to process highly lexicalized words in a more holistic fashion. Hence, it is likely that the *same* words that children process via a combination of first constituent and whole-word frequency can be processed by adults *without* resorting to first constituent frequency.

Another possible explanation for the pattern of effects might be that the frequency categories were just not adequate for the adult readers. We hence extracted adult frequencies for the stimulus set from the DWDS corpus (Geyken, 2007). The constituent frequencies were rather highly correlated across corpora (1st Constituents: $r = .74$; 2nd Constituents: $r = .81$), whereas whole-word frequencies were only moderately correlated ($r = .37$). For example, the compound *Eisdiele* (Engl. *ice cream parlor*) has a normalized lemma frequency of around 8 in childLex, but smaller than 0.1 in DWDS. Similarly, the constituent *Hexe* (Engl. *witch*) is highly frequent in childLex (norm lemma freq: 128), but very rare in DWDS (norm lemma freq: 8). Consequently, the matched categories of the original 2 × 2 manipulation did not hold up anymore when considering adult measures. Therefore, we reran the analysis with the adult frequencies as continuous predictors. Results are presented in Table 7 (Exp 3b). None of the predictors reached significance, but whole-word frequency clearly had the highest chi-square value, in line with the results from the analysis using the child frequency measures. This speaks against decomposition of the child material by adults. It should, however, not be concluded from this that adults do not use decomposition at all. Previous studies with compounds chosen from adult corpora based on adult-appropriate frequency measures do point to some form of decomposition in skilled readers. Possibly, the material and its frequency ranges in this experiment did not allow us to observe any effects. Experiment 4 therefore tests the possibility that constituent frequency effects arise in adults when words and their frequencies are taken from an adult corpus more precisely reflecting adult's written word experience and are matched to the child frequencies from Experiments 1 and 2.

Table 6
Overview of Mean Response Times for All Conditions in Experiment 3 (Standard Deviations in Parentheses)

Measure	h-h	h-l	l-h	l-l
Mean RT in ms	629 (20)	604 (19)	631 (20)	637 (21)

Experiment 4

Method

Participants. Thirty young adults (16 women, $M_{\text{age}} = 23.93$ years, $SD_{\text{age}} = 4.36$, age range: 18–34 years) from the Berlin area were recruited to participate in the study. The study was approved by the ethics committee of the Max Planck Institute for Human Development, and testing took place at the test center of the institute. All participants gave informed written consent prior to participation.

Materials. Thirty-two pairs of compounds were selected from the DWDS corpus (Geyken, 2007). The words were chosen such that they resembled the word material from Experiment 1–3 as closely as possible with regard to normalized lemma frequency, bigram frequency, neighbors (OLD20), number of letters, and number of syllables. All compounds consisted of two concatenated stems without interword spacing. Each compound pair shared one constituent (see Appendix B for a full list of all pairs); for half of the pairs this was the first constituent (e.g., *Notlage* and *Notwehr*), for the other half it was the second constituent (*Straßenbau* and *Ackerbau*). Constituent frequency was manipulated in a 2 × 2-design, yielding the four combinations higher-higher (*h-h*), higher-lower (*h-l*), lower-higher (*l-h*), lower-lower (*l-l*). Higher constituents had a normalized lemma frequency above 100 and lower constituents below 100 (higher: M = 254.68, SD = 147.76, min = 114.43, max = 830.93; lower: M = 49.30, SD = 32.52, min = 3.07, max = 95.75). The normalized lemma frequency of the whole compounds ranged between 0.99 and 9.45 (M = 3.45, SD = 1.71). Compounds were matched across conditions on whole-word frequency, bigram frequency, neighbors (OLD20), number of letters, and number of syllables, all $t < 1$, $p > .05$. As a result of using a list design, each participant saw a given constituent only in one combination and saw 32 compounds in total.

Table 7
Results From Mixed-Effect Models for Experiment 3 and the Reanalysis 3b With First Constituent Frequency, Second Constituent Frequency, and Whole-Word Frequency as Fixed Effects, and Participant and Item as Random Intercepts

	RTs			
	Exp 3		Exp 3b	
Measure	χ^2	*p*	χ^2	*p*
Fixed effects (all *df* = 1)				
Intercept	53475.32	<.001*	57398.74	<.001*
First constituent freq	1.83	.096	<1	.433
Second constituent freq	<1	.376	<1	.646
Whole-word freq	7.64	.002*	3.13	.077
First Constituent Freq × Second Constituent Freq	2.62	.152	<1	.989
First Constituent Freq × Whole-word Freq	1.42	.227	<1	.646
Second Constituent Freq × Whole-word Freq	<1	.537	<1	.431
First Constituent Freq × Second Constituent Freq × Whole-word Freq	<1	.732	<1	.949
Random effects				
Participants	221	<.001*	208	<.001*
Items	18	<.001*	32	<.001*

Note. Tests are based on Type III sum of squares and χ^2 values with Kenward-Roger *df*.
* $p < .05$.

In addition to the compound words, the 32 pseudowords from Experiment 1–3 were used.

Procedure. The procedure was the same as in Experiments 1–3.

Results

Analyses of the response time and accuracy data paralleled the one in Experiment 3. Again, because accuracy was extremely high, no analysis on the accuracy data was conducted. For the response time analysis, incorrect response times that were removed (1.98%), as were response times below 200 ms or above 1000 ms (2.34%), model criticism was used to exclude outliers (2.07%). Mean response times are shown in Table 8. The model results are summarized in Table 9.

The response time analysis revealed a main effect of whole-word frequency, attributable to faster responses to compounds with a higher whole-word frequency. There was also a main effect of first constituent frequency and a main effect of second constituent frequency, both indicating that responses were faster when the respective constituent was of high frequency. No interactions reached significance. The response times are depicted in Figure 4.

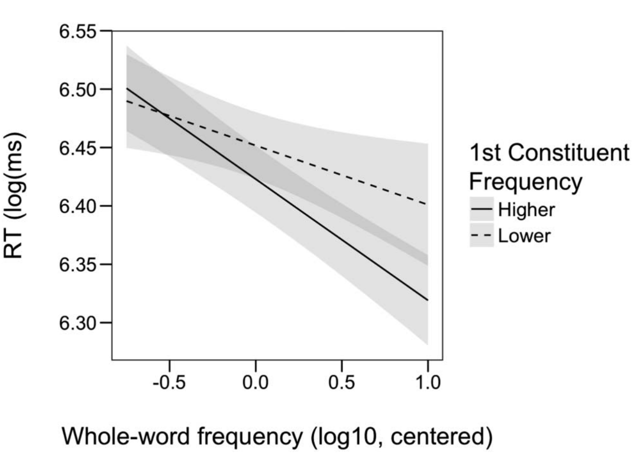

Figure 3. Mean response times (and standard errors) as a function of whole-word and first constituent frequency in Experiment 3.

Discussion

The results of Experiment 4 suggest that experienced readers of German in fact do use decomposition when processing compound words. Using a stimuli set matched on adult frequency measures yielded effects of whole-word and both first and second constituent frequencies. Thus, when compounds are not highly lexicalized adults, too, attend to first constituent frequency in addition to whole-word frequency. In contrast to the children, they also take into account second constituent frequency. Thus, adults do not show a privileged role of the first over the second frequency, presumably because their reading is more holistic than the children's reading, enabling them to activate both constituents quickly upon presentation. The pattern of results observed in Experiment 4 is most compatible with hybrid models (e.g., Andrews et al., 2004; Kuperman et al., 2009; Schreuder & Baayen, 1995). In contrast to the pattern for children, whole-word and first constituent frequency did not interact, which does not fit with access via the first constituent as proposed by Taft (1994) and is also difficult to reconcile in the NDL framework (Baayen et al., 2011). The consequences that the findings of our four experiments have when considered together will be discussed in detail below.

Table 8
Overview of Mean Response Times for All Conditions in Experiment 4 (Standard Deviations in Parentheses)

Measure	h-h	h-l	l-h	l-l
Mean RT in ms	612 (20)	593 (19)	639 (21)	617 (20)

Table 9
Results From Mixed-Effect Models for Experiment 4 With First Constituent Frequency, Second Constituent Frequency, and Whole-Word Frequency as Fixed Effects, and Participant and Item as Random Intercepts

Measure	RTs χ^2	RTs p
Fixed effects (all $df = 1$)		
Intercept	49809.29	<.001*
First constituent freq	7.82	.005*
Second constituent freq	5.01	.025*
Whole-word freq	4.66	.031*
First Constituent Freq × Second Constituent Freq	<1	.928
First Constituent Freq × Whole-word Freq	<1	.861
Second Constituent Freq × Whole-word Freq	<1	.738
First Constituent Freq × Second Constituent Freq × Whole-word Freq	<1	.720
Random effects		
Participants	557	<.001*
Items	28	<.001*

Note. Tests are based on Type III sum of squares and χ^2 values with Kenward-Roger *df*.
* $p < .05$.

General Discussion

The present research investigated the role of whole-word frequency and first and second constituent frequencies in the processing of compound words in beginning and experienced readers in German. Despite the extremely rich and productive compounding system and the resulting omnipresence of compounds in reading German, previous research on morphology in reading development has targeted derivations and largely neglected compounds. Models on compound processing, reversely, have focused on skilled reading and disregarded development, albeit this can be utterly informative in deciding between different models.

To address this research gap, we conducted a series of experiments, in which readers were presented with a lexical-decision task on compound words with varying constituent and whole-word frequencies. Overall, the findings point to an influence of whole-word frequency and constituent frequency in compound recognition that undergoes some subtle changes with increasing reading experience. The experiments together allow new insights into compound processing in inexperienced and experienced readers. The results and their consequences for models of morphological processing are discussed below.

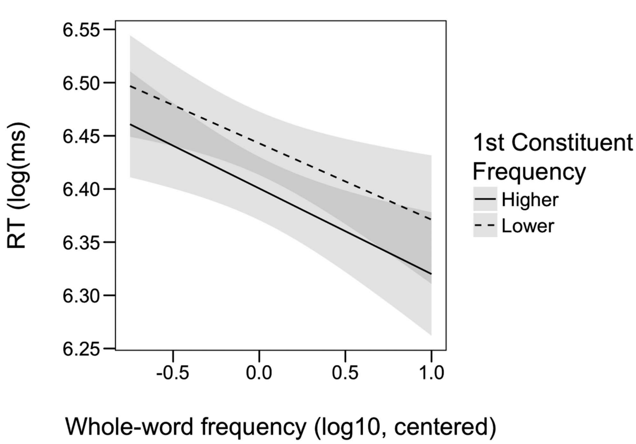

Figure 4. Mean response times (and standard errors) as a function of whole-word and first constituent frequency in Experiment 4.

First, a smaller group of children from Grade 2–4 performed a lexical-decision task on compounds taken from a child written language corpus (childLex: Schroeder et al., 2015). The experiment was then replicated with a much larger group of children all from Grade 4, additionally including interindividual differences in vocabulary size. In both experiments, we found an interaction effect of first constituent and whole-word frequency. Vocabulary size did influence overall speed and accuracy of compound word recognition, but did not influence the relative contribution of whole-word and constituent frequencies in Grade 4. The observed interaction of first constituent and whole-word frequency can be explained in an interactive hybrid framework with a left to right preference that is attributable to children's more sequential reading. The first constituent and the whole-word would both be activated and strengthen activation of each other interactively, making compound recognition especially fast when they are both of high frequency. Alternatively, compound word processing via the first constituent as suggested by Taft's (1994) interactive decomposition model could also explain the pattern of effects. Upon presentation with a compound, children activate all compounds that they know starting with this same constituent. The strength of the resulting compound activation then depends on whole-word frequency. If the first constituent is of high frequency, activation is quickly fed forward and compounds with high whole-word frequency are activated easily, whereas compounds with low whole-word frequency take longer to be activated. If the first constituent is of low frequency, it takes already longer to activate the constituent itself and the activation to be fed forward is slower, such that no differences arise at the whole-word level. It is also possible that low constituent frequency is correlated with low productivity, so that there are less compound candidates to receive activation—maybe only one candidate in many cases—and hence whole-word frequency is not pivotal.

Indeed, morphological family size, measuring constituent productivity, has been reported to influence response times to compounds (e.g., Juhasz & Berkowitz, 2011; Kuperman et al., 2009). Family size tends to correlate with constituent frequency, especially in German with its highly productive compounding system. For the stimuli used in the present study, the correlation was very high: for the child stimuli set (Experiments 1–3) it was 0.56 for first constituents and 0.78 for second constituents, and for the adult stimuli set (Experiment 4) it was 0.69 for first and 0.77 for second constituents (means and standard deviations of morphological family sizes of the stimuli are summarized in Table C1). This makes it difficult to disentangle the influence from frequency and family size post hoc, as it is problematic to examine these highly correlated variables within the same analysis. To nevertheless examine this issue, we reanalyzed the data from all four experiments with family size (as a continuous variable) instead of constituent frequency. Results of the models are summarized in Table C2. For the data of Experiment 1, a main effect of whole-word frequency and an interaction of first constituent family size emerged, mirroring the pattern obtained with constituent frequency. This interaction was modulated further by second constituent family size, as indicated by a three-way interaction. Responses were fastest when first constituents family size, second constituent family size, and whole-word frequency were all higher, whereas they were slowest when the constituent family sizes were higher but the whole-word frequency was lower. For the data of Experiment 2, again a main effect of whole-word frequency was present, the interaction of whole-word frequency and first constituent family size failed to reach significance, but the three-way interaction of whole-word frequency, first constituent family size, and second constituent family size was replicated.

Although it needs to be kept in mind that these analyses were post hoc and were not considered in our experimental manipulation, the main result, namely the interactive use of whole-word and first constituent, corresponds to the findings with frequency in Experiment 1 and 2. They support the assumption that children activate the initial constituent (e.g., *tooth*) and morphologically related words (e.g., *toothbrush, toothless, toothache* . . .). The speed and ease of compound recognition then depends on how often the constituent appears on its own or as a constituent in complex words and whether the target compound is a frequent candidate among them. If the constituents and the whole-word are of high frequency, then activation of the presented compound is especially fast and strong. However, if the constituents are of high frequency, but the whole-word is not, a conflict can arise such that there is a "tug of war" between the whole-word and its constituents (see also Smolka & Libben, 2017). Indeed, this can also be explained in the framework of Naïve Discriminative Learning (NDL: Baayen et al., 2011): the letters in the first part of *toothbrush* might be a good cue for the meaning of *tooth* as a single word, because of its high frequency as such, but maybe a less good cue for the meaning of *toothbrush*, which is more low frequent (see also Pham & Baayen, 2015). The interactive pattern of whole-word and first constituent frequency that we observed in children is, however, not compatible with full-listing accounts (Butterworth, 1983) nor with horse-race accounts (Schreuder & Baayen, 1995).

To compare the developing readers' performance with that of skilled readers, adults were tested on the same words as the children. With the child material, we found only an effect of whole-word frequency. Another group of adults conducted a similar lexical-decision task, but with words and their frequencies taken from an adult corpus (DWDS: Geyken, 2007) and matched to the child stimulus set. With the adult material, we found effects of first constituent frequency, second constituent frequency, and whole-word frequency, but no interaction effects. As for the children, we reanalyzed the data from the adult experiments with family size instead of constituent frequency. Results of the models are summarized in Table C2. For the data of Experiment 3, there was only an effect of whole-word frequency, mirroring the analysis with constituent frequencies. In addition, we also calculated the constituent family sizes for the words in Experiment 3 (child stimulus set) using the adult corpus (DWDS: Geyken, 2007) and reanalyzed the data using the adult measures. Again, this yielded only a significant effect of whole-word frequency. For Experiment 4, the adult stimulus set, there was only a main effect for first constituent family size. This deviates slightly from the pattern of results obtained with constituent frequencies, but recall that the family size analysis was post hoc and just intended as an additional check. Despite the high correlation between frequency and family size, they might be used slightly differently in processing. Nevertheless, our results suggests that adults access compounds as whole-words if they are highly lexicalized and/or familiar, whereas they also use information from the constituents if the compounds are not lexicalized, in line with Caramazza et al. (1988; see also van Jaarsveld & Rattink, 1988). One problem with this idea is that it is unclear how one route—decomposition or whole-word—can be chosen based on characteristics such as lexicality, familiarity, or frequency *before* either the entire word or its parts are recognized. Bertram and Hyönä (2003) propose that visual features such as length determine the route. However, compound length was matched across the child material and adult material in our study.

Taken together, over all four experiments, hybrid interactive models (Kuperman et al., 2009; Libben, 2006) are best able to explain the pattern of results. Whole-word and constituent information is both taken into account in compound recognition. Interestingly, the reliance on first constituent frequency and the resulting interaction of this measure with whole-word frequency is what changes the most with increasing reading experience. For children, lower first constituent frequencies seem to not boost compound recognition, whereas they do in adults. Moreover, the compound recognition process seems to be more sequential from left-to-right with a privileged use of the first over the second constituent in children and more parallel in adults with use of both first and second constituents. Amorphous models might be a suitable alternative to explain our pattern of results. Because of their learning mechanism and the continuous update of connections between letter cues and outcomes, they might also explain the absence of interaction effects in the adults. To explore the changing dynamics of processing across development, simulations for the NDL with German could prove very insightful. Although this was beyond the scope of the present study, it would be a promising enterprise for the future.

One possible criticism of our study concerns the use of pseudowords with no real word constituent. This would have made it sufficient for participants to decode only the first part, that is the first pseudoconstituent, to make a lexical decision, thus artificially amplifying the role of the first constituent. Some of the previous studies with similar designs to ours used pseudocompounds made of a new combination of two real stems, such as *Pianotasse* (Engl. *Pianocup*; Bronk et al., 2013). Owing to the high productivity of compounding in German, which makes any new combination more or less possible, the rejection of such pseudowords is really hard for children and would have resulted in extremely prolonged RTs and high false alarm rates. Therefore, these kinds of pseudowords were not used here. However, the omnipresent effect of whole-word frequency in all four experiments is a strong indicator that participants did not make decisions based on the lexicality of the first constituent only.

Overall, our series of experiments contributes to disentangling the relative contribution of first constituent, second constituent, and whole-word frequency in children's and adults' processing of compound words in German. The reading of long morphologically complex words is a challenge particularly central to German reading acquisition because of the extremely high compound productivity. Despite this, investigations of compound processing in developing readers have been sparse. The experiments presented here therefore are an important step forward in understanding the use of word morphology in written language acquisition. The results show that developing readers already use information from the first constituent for compound recognition. The changes we observed with increasing reading expertise offer new impulses for investigations to improve models of morphological processing and its development. In particular, the present study demonstrates that single- or dual-route models are too simplistic to capture the processes of compound recognition and that accounts based on the continuous learning of probabilities and maximization of opportunity are more suitable. Whether models denying morphological representations are superior to those explicitly assuming them is an important issue for future investigations.

References

Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects? *Journal of Memory and Language, 25,* 726–740. http://dx.doi.org/10.1016/0749-596x(86)90046-x

Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology, 16,* 285–311. http://dx.doi.org/10.1080/09541440340000123

Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers, 1,* 1–45.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118,* 438–481. http://dx.doi.org/10.1037/a0023851

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baayen, R. H., & Schreuder, R. (2000). Towards a psycholinguistic computational model for morphological parsing. *Philosophical Transactions of the Royal Society, Series A: Mathematical, Physical and Engineering Sciences, 358,* 1–13.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Usinglme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language, 48,* 615–634. http://dx.doi.org/10.1016/S0749-596X(02)00539-9

Bronk, M., Zwitserlood, P., & Bölte, J. (2013). Manipulations of word frequency reveal differences in the processing of morphologically complex and simple words in German. *Frontiers in Psychology, 4,* 546. http://dx.doi.org/10.3389/fpsyg.2013.00546

Butterworth, B. (Ed.). (1983). Lexical representation. In *Development, writing and other language processes* (Vol. 2, pp. 257–294). London, UK: Academic Press.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28,* 297–332. http://dx.doi.org/10.1016/0010-0277(88)90017-0

Carlisle, J. F., & Fleming, J. (2003). Lexical processing of morphologically complex words in the elementary years. *Scientific Studies of Reading, 7,* 239–253. http://dx.doi.org/10.1207/S1532799XSSR0703_3

Carlisle, J. F., & Katz, L. A. (2006). Effects of word and morpheme familiarity on reading of derived words. *Reading and Writing, 19,* 669–693. http://dx.doi.org/10.1007/s11145-005-5766-2

de Zeeuw, M., Schreuder, R., & Verhoeven, L. (2015). Lexical processing of nominal compounds in first- and second-language learners across primary grades. *Writing Systems Research, 7,* 133–156. http://dx.doi.org/10.1080/17586801.2014.926806

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2007). The role of the frequency of constituents in compound words: Evidence from Basque and Spanish. *Psychonomic Bulletin & Review, 14,* 1171–1176. http://dx.doi.org/10.3758/BF03193108

Geyken, A. (2007). The DWDS corpus: A reference corpus for the German language of the 20th century. In C. Fellbaum (Ed.), *Collocations and idioms: Linguistic, lexicographic, and computational aspects* (pp. 23–41). London, UK: Bloomsbury.

Goodwin, A. P., Gilbert, J. K., Cho, S.-J., & Kearns, D. M. (2014). Probing lexical representations: Simultaneous modeling of word and reader contributions to multidimensional lexical representations. *Journal of Educational Psychology, 106,* 448–468. http://dx.doi.org/10.1037/a0034754

Grainger, J., & Beyersmann, E. (2017). Edge-Aligned Embedded Word Activation Initiates Morpho-orthographic Segmentation. *Psychology of Learning and Motivation,* 285–317. http://dx.doi.org/10.1016/bs.plm.2017.03.009

Häikiö, T., Bertram, R., & Hyönä, J. (2011). The development of whole-word representations in compound word processing: Evidence from eye fixation patterns of elementary school children. *Applied Psycholinguistics, 32,* 533–551. http://dx.doi.org/10.1017/S0142716411000208

Hasenäcker, J., Schröter, P., & Schroeder, S. (2017). Investigating developmental trajectories of morphemes as reading units in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1093–1108. http://dx.doi.org/10.1037/xlm0000353

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal Biometrische Zeitschrift, 50,* 346–363. http://dx.doi.org/10.1002/bimj.200810425

Hyönä, J., & Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1612–1627. http://dx.doi.org/10.1037/0096-1523.24.6.1612

Juhasz, B. J. (2018). Experience with compound words influences their processing: An eye movement investigation with English compound words. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 71,* 103–112. http://dx.doi.org/10.1080/17470218.2016.1253756

Juhasz, B. J., & Berkowitz, R. N. (2011). Effects of morphological families on English compound word recognition: A multitask investigation. *Language and Cognitive Processes, 26*(4–6), 653–682. http://dx.doi.org/10.1080/01690965.2010.498668

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology, 94,* 223–244. http://dx.doi.org/10.1348/000712603321661903

Krott, A., & Nicoladis, E. (2005). Large constituent families help children parse compounds. *Journal of Child Language, 32,* 139–158. http://dx.doi.org/10.1017/S0305000904006622

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 876–895. http://dx.doi.org/10.1037/a0013484

Libben, G. (2006). Why study compound processing? In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 1–23). Oxford, UK: Oxford University Press.

Milin, P., Smolka, E., & Feldman, L. B. (2017). Models of lexical access and morphological processing. *The handbook of psycholinguistics* (pp. 240–268). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9781118829516.ch11

Pham, H., & Baayen, R. H. (2015). Vietnamese compounds show an anti-frequency effect in visual lexical decision. *Language, Cognition and Neuroscience, 30,* 1077–1095. http://dx.doi.org/10.1080/23273798.2015.1054844

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes, 15*(4–5), 445–485. http://dx.doi.org/10.1080/01690960050119661

Rueckl, J. G., & Raveh, M. (1999). The influence of morphological regularities on the dynamics of a connectionist network. *Brain and Language, 68,* 110–117. http://dx.doi.org/10.1006/brln.1999.2106

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ: Erlbaum.

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). childLex: A lexical database of German read by children. *Behavior Research Methods, 47,* 1085–1094. http://dx.doi.org/10.3758/s13428-014-0528-1

Segbers, J., & Schroeder, S. (2016). How many words do children know? A corpus-based estimation of children's total vocabulary size. *Language Testing, 34,* 297–320. http://dx.doi.org/10.1177/0265532216641152

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences, 4,* 353–361. http://dx.doi.org/10.1016/S1364-6613(00)01515-1

Smolka, E., & Libben, G. (2017). "Can you wash off the hogwash?" – semantic transparency of first and second constituents in the processing of German compounds. *Language, Cognition and Neuroscience, 32,* 514–531. http://dx.doi.org/10.1080/23273798.2016.1256492

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes, 9,* 271–294. http://dx.doi.org/10.1080/01690969408402120

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning & Verbal Behavior, 15,* 607–620. http://dx.doi.org/10.1016/0022-5371(76)90054-2

van Jaarsveld, H. J., & Rattink, G. E. (1988). Frequency effects in the processing of lexicalized and novel nominal compounds. *Journal of Psycholinguistic Research, 17,* 447–473. http://dx.doi.org/10.1007/BF01067911

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 (CFT 20-R) mit Wortschatztest (WS) und Zahlenfolgentest (ZF)* [Basic test of intelligence scale 2 (CFT 20-R) with vocabulary test (WS) and digit sequence test (ZF)]. Göttingen, Germany: Hogrefe.

Appendix A

Words Used in Experiments 1–3

Higher-higher—higher-lower pairs: *Baumhaus-Baumstamm, Handschuh-Handtuch, Hundehaar-Hundefutter, Landstraße-Landkarte, Steinboden-Steinbruch, Traummann-Traumfee, Wasserfall-Wasserhahn, Weltmeer-Weltmeister*. Higher-higher—lower-higher pairs: *Brieftasche-Aktentasche, Grundstück-Goldstück, Herzblatt-Kleeblatt, Hexenbuch-Zauberbuch, Königshof-Pausenhof, Seitenwand-Kellerwand, Spielplatz-Marktplatz, Walsweg-Kiesweg*. Lower-higher—lower-lower pairs: *Eisberg-Eisdiele, Glasauge-Glaskugel, Holzbein-Holzkiste, Lagerraum-Lagerhalle, Milchzahn-Milchreis, Mondstein-Mondschein, Schneemann-Schneesturm, Tanzschritt-Tanzsaal*. Higher-lower—Lower-lower pairs: *Angsthase-Osterhase, Autobahn-Eisenbahn, Fingerhut-Federhut, Glückspilz-Giftpilz, Halsband-Gummiband, Pferdestall-Hühnerstall, Seereise-Heimreise, Schulheft-Matheheft*

Appendix B

Words Used in Experiment 4

Higher-higher—higher-lower pairs: *Abendkleid-Abendbrot, Erdboden-Erdball, Grundrecht-Grundstein, Haustür-Hauswand, Kraftwerk-Kraftwagen, Landkreis-Landkarte, Lebensgefahr-Lebenslauf, Weltraum-Weltmeer*. Higher-higher—lower-higher pairs: *Arbeitszimmer-Hotelzimmer, Luftzug-Kreuzzug, Musikstück-Beweisstück, Sachgebiet-Fachgebiet, Sinnbild-Sternbild, Stadtteil-Bruchteil, Tierart-Tonart, Wasserfall-Schneefall*. Lower-higher—lower-lower pairs: *Blickpunkt-Blickfeld, Dienstzeit-Dienstbote, Ehrenwort-Ehrengast, Festschrift-Festsaal, Heimweg-Heimweh, Notlage-Notwehr, Sonnenlicht-Sonnenschein, Titelseite-Titelblatt*. Higher-lower—lower-lower pairs: *Bildschirm-Regenschirm, Brieftasche-Aktentasche, Fußnote-Banknote, Herzschlag-Ratschlag, Kopfhaut-Netzhaut, Liebespaar-Brautpaar, Straßenbau-Ackerbau, Tischplatte-Glasplatte*

(*Appendices continue*)

Appendix C

Table C1

Means (Standard Deviations in Parentheses) of Family Size for First and Second Constituents

Experiment	h-h	h-l	l-h	l-l	all
Experiment 1–3: First constituent family size	42.25 (23.16)	44.52 (28.89)	32.18 (26.17)	29.07 (21.32)	37.01 (25.87)
Experiment 1–3: Second constituent family size	47.89 (24.52)	20.29 (9.46)	45.24 (18.49)	17.93 (8.91)	32.84 (21.63)
Experiment 4: First constituent family size	546.85 (270.18)	478.21 (258.16)	172.63 (121.10)	138.41 (191.75)	347.36 (266.79)
Experiment 4: Second constituent family size	383.18 (196.98)	226.98 (163.04)	381.96 (179.68)	196.18 (166.64)	297.08 (196.71)

Note. For Experiment 1–3 (child stimulus set) the counts are based on ChildLex (Schroeder et al., 2015). For Experiment 4 (adult stimulus set) the counts are based on DWDS (Geyken, 2007).

Table C2

Results From Mixed-Effect Models for Experiment 1 With First Constituent Family Size, Second Constituent Family Size, and Whole-Word Frequency as Fixed Effects, and Participant and Item as Random Intercepts

	χ^2				
Measure	Exp 1	Exp 2	Exp 3	Exp 3b	Exp 4
Fixed effects (all *df* = 1)					
Intercept	3806*	33311*	55458*	56649*	49727*
First constituent family size	1.51	2.80	1.25	1.16	10.42*
Second constituent family size	<1	3.09	1.44	<1	2.86
Whole-word freq	6.71*	17.53*	11.85*	4.45*	2.27
First Const Fam Size × Second Const Fam Size	<1	<1	<1	1.44	<1
First Const Fam Size × Whole-word Freq	5.13*	3.14	2.62	<1	<1
Second Const Fam Size × Whole-word Freq	1.36	3.56	1.65	<1	<1
First Const Fam Size × Second Const Fam Size × Whole-word Freq	4.95*	5.99*	<1	1.03	<1
Random effects					
Participants	740*	2537*	261*	207*	557*
Items	13*	166*	19*	28*	23*

Note. Experiment 3b refers to the data of Experiment 3 (child stimulus set) with frequency and family size measures based on an adult corpus (DWDS: Geyken, 2007). Tests are based on Type III sum of squares and χ^2 values with Kenward-Roger *df*.

* $p < .05$.

Received September 20, 2017
Revision received March 30, 2018
Accepted April 12, 2018 ■