Reading Polymorphemic Dutch Compounds: Towards a Multiple Route Model of Lexical Processing

Victor Kuperman*

Radboud University Nijmegen, The Netherlands

Robert Schreuder

Radboud University Nijmegen, The Netherlands

Raymond Bertram

University of Turku, Finland

R. Harald Baayen

University of Alberta, Canada

May 12, 2008

Running Head: Reading of Dutch Compounds

*Corresponding author: Victor Kuperman, Radboud University Nijmegen, P.O. Box 310, 6500 AH, Nijmegen, The Netherlands. E-mail: victor.kuperman@mpi.nl. Phone: +31-24-3612160. Fax: +31-24-3521213

Abstract

This paper reports an eye-tracking experiment with 2500 polymorpemic Dutch compounds presented in isolation for visual lexical decision, while readers' eye-movements were registered. We found evidence that both full-forms of compounds (*dishwasher*) and their constituent morphemes (e.g., *dish*, *washer*, *er*) and morphological families of constituents (sets of compounds with a shared constituent) played a role in compound processing. We observed simultaneous effects of compound frequency, left constituent frequency and family size early (i.e., before the whole compound has been scanned), and also effects of right constituent frequency and family size that emerged after the compound frequency effect. The temporal order of these and other effects that we observed goes against assumptions of many models of lexical processing. We propose specifications for a new multiple route model of polymorphemic compound processing, which is based on time-locked, parallel and interactive use of all morphological cues, as soon as they become (even partly) available to the visual uptake system.

Keywords: morphological structure; lexical processing; eye movements; compounds

Current models of morphological processing and representation in reading have explored a wide range of logically possible architectures. Sublexical models hold that complex words undergo obligatory parsing and that lexical access proceeds via their morphemes (cf., Taft, 1991; Taft & Forster, 1975, 1976). Supralexical models, by contrast, argue that morphemes are accessed only after the compound as a whole has been recognized (e.g., Diependaele, Sandra & Grainger, 2005; Giraudo & Grainger, 2001). Dual route models hypothesize that full-form based processing goes hand in hand with decompositional processing. The two access routes are usually assumed to be independent (Allen & Badecker, 2002; Baayen & Schreuder, 1999; Frauenfelder & Schreuder, 1992; Laudanna & Burani, 1995; Schreuder & Baayen, 1995), although an interactive dual route model has been proposed as well (Baayen & Schreuder, 2000). In connectionist models such as the triangle model (Seidenberg & McClelland, 1989), morphological effects are interpreted as arising due to the convergence of orthographic, phonological and semantic codes. What all these theories have in common is that they were developed to explain data obtained with chronometric measures for isolated reading of bimorphemic complex words. As a consequence, they tend to remain silent about the time-course of information uptake in the reading of complex words.

Establishing the temporal order of activation of full-forms (e.g., *dishwasher*) of complex words and of their morphological constituents (e.g., *dish* and *washer*) is critical for adjudicating between competing models of morphological processing. The present study addresses the time-course of morphological processing by considering the reading of long, polymorphemic Dutch compounds. Importantly, current models of morphological processing offer different

predictions with regard to the visual recognition of such compounds. On supralexical models, one expects activation of the compound's full-form (diagnosed by the compound frequency effect) as the initial step of lexical access. After the full-form of the compound is activated, one expects to observe simultaneous activation of both the left and the right constituent (diagnosed by frequency-based properties of a constituent). On strict sublexical models, the predicted order of activation is as follows: first, the left constituent of a compound, second, its right constituent, and finally (either coinciding with activation of the right constituent, or following it) the full-form. The sublexical model of Taft and Forster (1976) argues that activation of the compound's left constituent is sufficient to trigger the retrieval of the compound's full-form. This model predicts sequential effects of the left constituent frequency and compound frequency, and no effects of the right constituent. On some dual-route models of parallel processing, one expects roughly simultaneous effects of compound frequency and left constituent frequency, since both routes are argued to be pursued simultaneously and independently (e.g., Baayen & Schreuder, 1999). Bertram and Hyönä (2003) have also proposed a dual-route architecture with a headstart for the decomposition route in case of long compounds, which predicts early effects pertaining to the compound's left constituent followed by the compound frequency effect.

Earlier eye-tracking studies not only confirmed the joint relevance of both constituents and full-form representations for reading posited by dual route models (Andrews, Miller & Rayner, 2004; Hyönä, Bertram & Pollatsek, 2004, Zwitserlood, 1994), they have also made more precise information about the time-course of morphological processing available. For

instance, Hyönä *et al.* (2004) found that for long compounds there is early activation of the left constituent (*dish*) and later activation of the right constituent (*washer*). However, two important questions about the time-course of morphological processing are as yet unresolved. First, the temporal locus of compound frequency effects remains unclear. Several eye-tracking studies of compounds (cf., Andrews *et al.*, 2004; Bertram & Hyönä, 2003; Pollatsek *et al.*, 2000) have observed effects of compound frequency for the very first fixation, but these effects failed to reach significance. ERP studies of reading (Hauk & Pulvermüller, 2004; Penolazzi, Hauk & Pulvermüller, 2007; Sereno, Rayner & Posner, 1998) have repeatedly shown early effects of whole word frequency ($<$ 150-200 ms), but they focused on relatively short (4-6 characters) and morphologically simplex words. An early locus for the compound frequency effect in long compounds would challenge strict sublexical accounts of morphological processing, according to which whole word frequency effects would reflect post-access combinatorial processes instead of tapping into early visual information uptake.

Second, it is unclear whether the activation of the compound's full-form precedes, follows or coincides with the activation of the compound's constituents. The present evidence is controversial. For instance, Juhasz, Starr, Inhoff and Placke (2003) argued – on the basis of eye-tracking, lexical decision and naming experiments – that it is the compound's head, the last constituent to be read (e.g., *washer* in *dishwasher*), that plays the decisive role in the late stages of compound recognition, while the effects of the initial constituent emerge early and are weak (see, however, Juhasz, 2007). A possible reason for the dominance of the right constituent is its typical semantic convergence with the meaning of the whole compound (see

also Duñabeitia, Perea & Carreiras, 2007). These results were argued to support models that argue for either co-activation of the right constituent and the full-form (Pollatsek, Hyönä & Bertram, 2000), or activation of the right constituent following activation of the full-form (Giraudo & Grainger, 2001). Their claim contrasts with chronometric studies by e.g., Taft and Forster (1976) who found evidence for the left constituent guiding lexical access to a compound's meaning. Taft and Forster (1976) saw these results as evidence that a compound's full-form gets activated *after* the left constituent of the compound receives activation.

The first aim of the present study is to address the temporal order of lexical access to the full-form and the morphological constituents of compounds. In other words, we explore how soon and in what order do the properties of the compound's full-form, and the properties of the compound's left and right constituents, emerge in the timeline of compound recognition. Second, we broaden the scope of constituent processing by probing whether morphological families of constituents (i.e., sets of compounds sharing a constituent, e.g., *ice pick*, *ice cube*, *ice box*) contribute to the speed of processing over and above properties of full-forms and those of constituents as isolated words. Lexical decision studies argued that the effects of constituent families are semantic in character, and hence emerge late, at the peripheral post-access stages of the complex word processing (e.g., De Jong, Schreuder & Baayen, 2000). In this study we tackle the temporal locus of the effects of constituent families using eye-tracking as a technique with a better temporal resolution than the one offered by lexical decision latencies. Third, we zoom in on the issue of independence of the full-form and

decompositional processing routes claimed in some dual-route parallel processing models by considering the possibility that the effects elicited by the full-form properties might be modulated by constituent properties.

Instead of investigating bimorphemic compounds, we examined compounds with three to six morphemes. Type-wise, such polymorphemic compounds are more common in Dutch than the bimorphemic compounds that are traditionally studied in the experimental literature. For instance, perusal of CELEX (Baayen, Piepenbrock & Gulikers, 1995) shows that 54% of the nominal compounds has more than two morphemes. An additional dimension of morphological processing that we consider as the fourth goal of our study is the role of (free-standing and bound) morphemes deeply embedded in morphological structure (e.g., *wash-* and *-er* in *dishwasher*). Are morphemes at lower levels of morphological hierarchy recognized as independent units of meaning by the human lexical processor and used in compound identification, or are they invariably treated as parts of larger structural units (e.g., *washer*)? If, as we will argue, readers maximize their use of cues available for efficient compound identification, we may expect that the deeply embedded free and bound morphemes are used in the course of processing as well.

In what follows, we report a large regression experiment with 2500 target compounds that combined eye-tracking of isolated word reading with lexical decision as superimposed task to ensure sufficient depth of processing. We opted for this combination since it provides detailed insight into the time-course of morphological processing and it provides sufficient statistical power. In the General Discussion, we return in detail to the methodological consequences of

our decision to make use of lexical decision rather than sentential reading. Here, we restrict ourselves to noting that a parallel study presenting Finnish compounds in sentential contexts (Kuperman, Bertram & Baayen, 2008) yielded a pattern of results that is highly consistent with the morphological effects reported below. Our present experiment provides evidence that current models of morphological processing are too restrictive in their architectures, and that a more flexible framework in which all opportunities for recognition are maximized (Libben, 2006) is called for.

Method

Participants

Nineteen students of the Radboud University of Nijmegen (12 females and 7 males) were paid 20 euro for participation in the study. All were native speakers of Dutch and reported normal or corrected-to-normal vision and right-handedness.

Apparatus

Eye movements were monitored by the head-mounted video-based EYELINK II eye-tracking device produced by SR Research (Mississauga, Canada). The average gaze position error of EYELINK II is $<0.5^o$, while its resolution is 0.01^o. Recording of the eye movements was performed on the left eye only and in the pupil-only mode. The sampling rate of recording used in this study was 250 Hz. The 17-inch computer monitor used for the display of the stimuli had a 60 Hz refresh rate.

Stimuli

In total, 2500 lexical items (1250 existing words and 1250 nonce compounds) were in-

cluded as stimuli. A list of existing polymorphemic Dutch compounds (triconstituent compounds, or biconstituent compounds with at least one and at most four derivational affixes) was selected from the CELEX lexical database (Baayen, Piepenbrock & Gulikers, 1995), for instance, *werk+gev-er*, "work-giver", i.e., "employer". Additionally, a list of multiply complex nonce compounds was created by blending existing words into novel combinations (i.e., combinations that are not registered in the CELEX database), for instance, *alarmijsbaan*, composed of *alarm* "alarm" and the compound word *ijsbaan* "skating ring". At the level of immediate constituents, the resulting targets and fillers represented a mixture of noun-noun, adjective-noun and verb-noun compounds.

The average number of morphemes per stimulus was 3.2 ($SD = 0.4$). The maximum length of a stimulus was set at 12 characters. The resulting range of 8-12 characters (mean length = 11.62, $SD = 0.74$) allowed for a tight experimental control of word length, and kept collinearity of such measures as word length and frequency, and left constituent length and frequency within reasonable bounds. Stimuli were displayed one at a time in a fixed-width font Courier New size 12. With a viewing distance of about 80 cm, one character space subtended approximately 0.36^o of visual angle.

Procedure

Participants were instructed to read words at their own pace. They were also informed that nonce compounds were built of existing Dutch words and were asked to evaluate the *whole* stimulus as an existing word or a non-word by pressing the right button ("Yes" response) or the left button ("No" response) of a dual button box. Prior to the presentation

of the stimuli, the eye-tracker was calibrated using a nine-point grid that extended over the entire computer screen. Prior to each stimulus, a fixation point was presented in the central position of the screen for 500 ms. After each third stimulus a drift correction was performed using the screen-central fixation point as a mark. After 500 ms or after the calibration was corrected, a stimulus was displayed in black lower-case characters on a white background. When one of the dual box buttons was pressed, the stimulus was removed from the screen and a fixation point appeared. If no response was registered after 5000 ms, a stimulus was removed from the screen and the next trial was initiated. Participants' responses and response times were recorded along with their eye movements.

Stimuli were displayed centralized vertically, and slightly off-center horizontally such that the space between the fourth and the fifth characters of a stimulus was always at the center of the screen where the fixation point was shown. This position is closest to the preferred viewing position (the most frequent position where the eyes initially land) reported in eye movements studies for Finnish, English and French words with the lengths that we used, mostly 12 characters, (e.g., Bertram & Hyönä, 2003; McDonald & Shillcock, 2004; Vergilino-Perez, Collins & Doré-Mazars, 2004).

The presentation order of stimuli was randomized. Stimuli were presented in two separate sessions each consisting of three blocks. The order of presentation of the blocks and the order of the words within each block were the same for each participant (see Appendix 2 for the discussion of randomization procedures). For each participant, sessions were run on two different dates, while blocks within one session were separated by a five to ten minute break.

After each break the eye-tracker was calibrated again. A single session lasted 70 minutes at most, and the total time of the experiment lasted a maximum of 130 minutes.

Dependent variables

For the analysis of the lexical decision data, we considered as dependent variables the (natural) log-transformed response times (*RT*), as well as the accuracy of responses (*Correct*).

In the eye-tracking data analysis, we selected as early measures of lexical processing the first fixation duration, *FirstDur*, and the subgaze duration on the compound's left constituent, *SubgazeLeft* (the summed duration of all fixations on the left constituent before exiting it). As measures that tap into later stages of compound recognition, we considered subgaze for the right immediate constituent, *SubgazeRight* (the summed duration of all fixations on the right constituent before exiting it). Gaze duration, *GazeDur*, served as the global measure of processing difficulty. In this study, gaze duration was defined as the summed duration of all fixations on the target word that were completed before one of two events took place: Either the reader fixated away from the word, or the lexical decision was made[1]. All durational measures were natural log-transformed to reduce the influence

[1]Note that *SubgazeLeft* and *SubgazeRight* are not strictly additive in the measure of gaze duration. In the situation where fixation 1 is on the left constituent, fixation 2 on the right one and fixation 3 on the left one, *SubgazeLeft* is equal to the duration of fixation 1, and *SubgazeRight* to the duration of fixation 2. The measure of gaze duration, however, would be equal to the sum of 1, 2 and 3, and could show an effect that differs in size from the sum of effects found for both subgazes. Also we fitted the statistical models to the subgaze measures with the non-zero duration. There are words, however, in which all fixations fall on one constituent, and there is no subgaze duration for the other constituent. In such cases there is only one subgaze component contributing to the composite measure of gaze duration.

of atypical outliers. We considered several other eye-movement measures as well: These included single, second and third fixation durations; initial fixation position; the amplitude of the first within-word saccade; the probability of a given fixation being the last one on the word; the probability of a given fixation being to the left of the previous fixation; and the total number of fixations on a word. The data patterns for these measures were in line with the ones we reported, but did not offer substantial additional insight into our research questions.

Predictors

Morphological variables. The measures of morphological characteristics of stimuli included: whole word (compound) frequency, *WordFreq*; the word frequency of the left constituent as an isolated word, *LeftFreq*; and the word frequency for the right constituent as an isolated word, *RightFreq*. All these frequencies were lemma frequencies, i.e., summed frequencies of a compound word and of its inflectional variants (e.g., sum of frequencies of the singular form *newspaper*, the plural form *newspapers* and the singular and plural genitive forms *newspaper's* and *newspapers'*).

All frequency-based measures in this study, including the ones reported in the remainder of this section, were obtained from CELEX (counts based on a corpus of 42 million word forms) and log-transformed to reduce the influence of outliers.

We also considered measures of morphological connectivity for the constituents of our compounds. We refer to the set of compounds that share the left (right) constituent with the target as the left (right) morphological family of that constituent (e.g., the left constituent

family of *ice cream* includes *ice pick*, *ice cube* and *ice box*). Words that appear as constituents in many compounds (i.e., have large morphological families) or in frequent compounds (i.e., have high family frequency) have been repeatedly shown across languages to elicit shorter lexical decision latencies, whether presented visually or auditorily (cf., e.g., De Jong, Schreuder & Baayen, 2000; De Jong, Feldman, Schreuder *et al.*, 2002; Dijkstra, Moscoso del Prado Martín, Schulpen *et al.*, 2005; Moscoso del Prado Martín, Bertram, Häikiö *et al.*, 2004). Left constituent family size is also known to modulate gaze duration in interaction with semantic opacity of Finnish compounds, cf., Pollatsek and Hyönä (2005)[2].

Morphological family size for the left constituents in our compounds strongly correlated with the frequencies of these left constituents as isolated words. We orthogonalized these collinear measures by fitting a regression model where left constituent family size was predicted by left constituent frequency. We then considered the residuals of this model, *ResidLeftFamilySize*, as our new left family size measure. It was highly correlated with the original

[2]For both the left and the right constituents, the alternative measure of family frequency (the summed token frequency of the members in the morphological family) consistently elicited weaker effects than family size of the respective constituents in all statistical models, in contrast to findings of De Jong *et al.* (2002) for Dutch compounds. The difference in effect sizes was revealed in smaller regression (beta) coefficients for family frequencies, when constituent family frequencies and family sizes were included, separately, as predictors in our statistical models. For instance, in the model for gaze duration, the regression coefficient was -0.026 for left constituent family frequency and -0.036 for left constituent family size. As the distinction between family size and family frequency effects is not crucial for our research questions, we do not discuss this measure further. We rather note that the entropy measure proposed by Moscoso del Prado Martín *et al.* (2004) may be a possible resolution for the relative impacts of the family-based alternatives.

measure ($r = 0.95, p < 0.0001$), but the effects of constituent frequency were now partialled out. Using the same procedure for the right constituent family size and frequency we obtained *ResidRightFamilySize*, which again closely approximated right constituent family size ($r = 0.93, p < 0.0001$), and was orthogonal to *RightFreq*. We decorrelated family size and frequency for analytical clarity, in order to be better able to assess the independent contributions of predictors (beta coefficients) to the model.

The presence of each subconstituent morpheme and its position in the morphological structure were coded by the multi-level factor *Affix* with the following levels: "Initial" (for compounds with prefixed left constituents), "Medial" (for compounds with a suffixed left constituent, an interfix, a prefixed right constituent, or with any combination of these affixes), "Final" (for compounds with suffixed right constituents), "Multiple" (for compounds with multiple affixes[3]) and "Tri" (for 'pure triconstituent' compounds with three word stems and no affixes; for the sake of analytical clarity, we excluded from our analyses 112 compounds with three word stems and further affixes). The resulting counts of stimuli representing each type of morphological complexity are summarized in Table 1.

INSERT TABLE I ABOUT HERE

[3]We classified compounds with more than one affix at the immediate constituent boundary, such as *rover-s-hol*, "robbers' den", as Medial rather than as Multiple. In other words, the category Medial comprises compounds with at least one medial affix, while the category Multiple comprises compounds with affixes at more than one position in the compound. We opted not to differentiate between compounds with different numbers of medial affixes, since the effects of these affixes considered separately were very similar across our analyses.

Table 1: *Counts of compounds partitioned by type of morphological complexity.*

	Type of Complexity	Number of stimuli
1	Triconstituent	580
2	Initial	158
3	Medial	541
4	Multiple	407
5	Final	702

We also considered affix productivity, *AffixProd* (the type count of derived words in which the affix occurs). The total number of morphemes in the compounds was included as an index of the compound's morphological *Complexity*.

Other variables. We also considered word length (*WordLength*) (in the range of 8-12 characters), as well as left constituent length (*LeftLength*). The longitudinal effect of the experimental task on the participants' behavior (e.g., fatigue or habituation as the participant works through the experiment) was estimated by means of the position of the stimilus in the experimental list, *TrialNum*. We also took into account the influence that carried over from trial $N-1$ to trial N (see Baayen, Davidson & Bates, 2008; De Vaan *et al.*, 2007) by considering the log-transformed response time from the trials immediately preceding the current one (*RT1*). Other control predictors that reached significance in codetermining either the lexical decision latencies or reading times as revealed in eye-movements are presented in Appendix 1.

Table 3 in Appendix 1 lists the distributions of the continuous variables used in this study, including their ranges, and mean and median values.

Statistical Considerations

In this study we made use of mixed-effects multiple regression models with random intercepts for *Subject* and *Word* (and occasionally by-participant random slopes and contrasts for item-bound predictors), and the predictors introduced above as fixed effect factors and covariates (cf., Baayen, 2008; Bates & Sarkar, 2005; Pinheiro & Bates, 2000).

Unless noted otherwise, only those fixed effects are presented below that reached significance at the 5%-level in a backwards stepwise model selection procedure. All random effects included in our models significantly improved the explanatory value of those models, as indicated by significantly higher values of the maximum likelihood estimate of the model with a given random effect as compared to the model without that random effect (all ps < 0.0001 using likelihood ratio tests), for detailed treatment of random effects in mixed-effects models see Pinheiro and Bates (2000). Below we report which predictors required random slopes in addition to the random intercepts for *Subject* and *Word*, see Table 9 in Appendix 1.

All models were fitted and atypical outliers were identified, i.e., points that fell outside the range of -2.5 to to 2.5 units of SD of the residual error. Such outliers were removed from the respective datasets (and were not used in the composite eye-movement measures) and the models were refitted in order to avoid distortion of the model estimates due to atypical extreme observations. Below we report statistics of those refitted models.

Due to the large number of models fitted in this study, we only report in Appendix 1 the full specifications of the model for lexical decision latencies for existing words, and of the four models for the eye movements measures (first fixation duration, subgazes for the left and the right constituent, and gaze duration).

Results and Discussion

Lexical Decision

The initial lexical decision data pool consisted of 2500 words x 19 participants = 47500 trials. From this dataset we excluded one word that was misspelled, as well as the trials in which the (log) RT value fell beyond 3 units of standard deviation from the mean. Since no participant exceeded the threshold of a 30% error rate in either nonce compounds or the existing words, none were excluded. The resulting dataset consisted of 47206 trials, of which 41245 were correct replies. The error rate reached 23% for existing words and 3% for nonce compounds. Thus, in the lexical decision task participants exhibited a clear bias towards "no"-responses, which does not come as a surprise given that many of the existing compounds are fairly low-frequency words and also semantically opaque words, the meaning of which is conceptually difficult to construct from the individual constituents, just as is the case with many nonce compounds. For correct replies, the average lexical decision latency was 763 ms ($SD = 246$) for existing words and 801 ms ($SD = 261$) for nonce compounds.

Below we only discuss the analysis of the lexical decision latencies for the 18217 trials with existing compounds that were correctly identified in the lexical decision task.

Morphological Variables. Column *RT* in Table 2 summarizes the effects of compound frequency and frequency-based measures of a compound's constituents on the lexical decision latencies (see Table 4 in Appendix 1 for the full specification of the model). The column provides effect sizes for morphological predictors (see Appendix for the explanation as to how these were computed) and p-values for main effects, as well as indicates interactions between

predictors of interest. For clarity of exposition, we leave out from the table the effects of morphemes deeply embedded in the compound structure: These are discussed separately.

INSERT TABLE 2 APPROXIMATELY HERE

Both compound frequency (*WordFreq*) and morpheme-based frequencies (*LeftFreq*, *RightFreq*), and morphological connectivity measures (*ResidLeftFamilySize*, *ResidRightFamilySize*) entered into negative correlations with the RTs, i.e., higher frequencies or larger families facilitated compound processing.

Of these predictors, compound frequency showed the greatest effect (-96 ms). These facilitatory morphological effects are in accord with previous reports of visual lexical decision experiments with Dutch and English compounds (cf., e.g., Andrews, 1986; De Jong *et al.*, 2000; De Jong *et al.*, 2002; Juhasz *et al.*, 2003).

Interestingly, compound frequency interacted with left constituent frequency in such a way that the effect of compound frequency was strongest in compounds with the low-frequency left constituents and was weaker in compounds where left constituents were relatively frequent, see Fig. 1.

INSERT FIGURE 1 APPROXIMATELY HERE

Suppose, following Libben (2006), that both compound frequency and left constituent frequency are among the morphological cues that the lexical processor may use to facilitate recognition of the compound. Then the observed interaction is the evidence that the magnitude of one such cue (e.g., left constituent frequency) appears to modulate the extent

to which the other cue (e.g., compound frequency) contributes to the identification of the complex word.

We also observed an interaction between right constituent frequency and left constituent family size, see Fig. 2. The effect of right constituent frequency was strongest in compounds with large left constituent families (i.e., with a large number of possible morphemic continuations for the left constituent, e.g., *shoelace*, *shoe cream*, *shoe shop*), and decreased with decreasing morphological family size.

INSERT FIGURE 2 APPROXIMATELY HERE

Apparently, ease of access to the lexical representation of the right constituent (diagnosed by its frequency) speeds up compound recognition more when there is more uncertainty about which candidate to choose from a larger number of possible right constituents. In case the competition in the family is relatively weak, due to a low number of choices, the right constituent may be relatively easy to predict and additional morphological information in the form of right constituent frequency is not as useful for the lexical processor. Again, we find that the magnitude of one cue for compound recognition affects the utility and magnitude of other such cues.

The effects of lower-level, subconstituent, morphemes revealed that compounds with two stems (of which at least one was a derivation) were processed significantly faster than triconstituent compounds (by about 20 ms, averaged across levels of *Affix*). Moreover, stimuli that comprised more morphemes, as measured by *Complexity* elicited longer latencies (effect size = 86 ms), as expected.

Other Control Variables. We observed habituation of participants to the task: The further they were into the experiment (as estimated by the trial position in the experimental list), the faster their lexical decisions were (effect size = -34 ms).

Longer RTs to the immediately preceding trial (*RT1*) went hand in hand with longer lexical decision latency at the current trial (effect size = 223 ms). These findings make a clear case that both the longitudinal effects of the experimental task and those related to immediately preceding trials contribute substantially to modulating lexical decision latencies.

Eye movements

We considered only the first-pass reading (i.e., the sequence of fixations made before the fixation is made outside of the word boundaries) and only those fixations that were completed before a response button was pressed. Trials with blinks and misreadings (i.e., trials for which no fixations were recorded by the eye-tracking device, due to the machine error) were removed, as well as the trials with lexical decision latencies exceeding 3 units of SD from the mean. The resulting dataset comprised 85908 fixations. We also removed from the dataset of fixations and from composite eye-movement measures those fixations that exceeded 2.5 units of SD from the mean log-transformed duration, whereas the mean duration and the standard deviation were calculated separately for each participant. In this way we avoided penalizing very slow or very fast readers. In total, 2227 (2.6%) outliers were removed, and the resulting range of fixation durations was 49 ms to 1197 ms. Subsequently, fixations that bordered microsaccades (fixations falling within same character) were removed (122 x 2 = 244 fixations, 0.1%). The resulting pool of data points consisted of 83437 valid

fixations.

Eighteen percent of the stimuli required a single fixation for reading, 36% required exactly two fixations, 26% required exactly three fixations, and it took four or more fixations to read the remaining 20% of the stimuli. The average number of fixations on a stimulus was 2.6 ($SD = 1.2$). Regressive fixations (within-word fixations located to the left of the previous fixation) constituted 12.6% of our data pool. The average fixation duration was 262 ms (SD 117), and the average gaze duration was 620 ms ($SD = 382$). Eighty-one percent of initial fixations was located either on the fourth or the fifth character of the presented stimulus, which is the area where we intented those fixations to be[4]. Seventy-seven percent of initial fixations were located on the left constituent. Since we had compounds with 2-4 character-long left constituents, a relatively large proportion of initial fixations was located at the right constituent (23%). Seventy-eight percent of second progressive fixations landed on the right constituent.

We further report our findings for the trials with existing compounds and only those that elicited correct responses. Our findings are based on four statistical models: for first fixation duration (14232 data points), for subgaze duration on the left constituent (11684

[4]It should be noted that the positions of almost 90% of initial fixations were within the measurement error ($<0.5^o$ of the visual angle) of EYELINK II, that is no more than 1.4 character away from the displayed fixation point. The shape of the distribution of initial fixation positions was close to normal with the mean of 40.7 pixels (that is, between the 4th and 5th letter) and standard deviation of 8.4 pixels. The initial fixations at the tails of the distribution (in the beginning or the end of the word) may be explained by the somewhat long presentation of the fixation point (500 ms), which may have caused people to occasionally saccade away from that fixation point prior to word presentation.

data points), for subgaze duration on the right constituent (8495 data points), and for gaze duration (14616 data points).

Morphological effects: Compound and immediate constituents. Columns 3 to 6 in Table 2 are a summary of the effects that morphological structure elicits in eye-movements across four statistical models (see full specifications for the models in Tables 5-8 in Appendix 1). Considered jointly, the results of the statistical models in Table 2 outline the temporal flow of compound recognition. First, we found evidence that both immediate constituents and the whole compound affect lexical processing of compound words (cf., e.g., Andrews *et al.*, 2004; Bertram & Hyönä, 2003; Hyönä, Bertram & Pollatsek, 2004). In fact, every single morphological predictor that we considered (compound frequency, constituent frequencies and family sizes, as well as properties of deeply embedded morphemes discussed below) had a role to play in the time-course of visual compound recognition. This hints at the possibility that morphological structure offers more cues for the task of compound identification than previously thought.

Second, properties of the left constituents of compounds showed earlier effects than the respective properties of the right constituents: the latter were only present in the late measures, *SubgazeRight* and *GazeDur*. Moreover, the impact of the right constituent on compound recognition was considerably weaker than that of the left constituent: The effects of the right constituent were smaller in size and often qualified by interactions with other predictors. These findings may reflect that fact that the left constituent is available earlier to the lexical processor than the right constituent. The typical sequence of fixations in our

dataset supported this claim: Initial fixations tended to be located at the left constituent (77% of first fixations), while subsequent fixations mostly landed on the right constituent (78% of progressive second fixations)[5]. We note that the size of the left constituent family codetermined the speed of identification of a compound's right constituent. Apparently, the relative ease of processing of the left constituent spills over to the processing of the right constituent, which is consistent with the spillover effect of word N on word N+1 observed in sentential reading (e.g., Rayner & Duffy, 1986; Reichle, Rayner & Pollatsek, 2004).

Third, the compound frequency effect emerged as early as the first fixation and lingered on throughout the entire time-course of compound processing. That the strong and statistically significant effect of compound frequency shows so early resolves the question raised by Bertram and Hyönä (2003: 627) of whether compound frequency might affect the early stages of visual processing in long compounds. The answer is that it does for 8-12 character-long words[6]. The likelihood that our stimuli, which are mostly 12 character long, are appreciated

[5]Given the lengths of our compounds and the initial fixation positions, it is likely that some characters from the right constituent are identified during an initial fixation on the left constituent. However, the absence of early effects associated with the compound's right constituent implies that the available orthographic information on the right constituent is apparently not sufficient for early activation of that morpheme (cf., Hyönä *et al.*, 2004).

[6]The effect of compound frequency was still significant in the statistical model for the first fixation duration from which single-fixation cases were excluded (model not shown, $p < 0.0001$). We did not observe an interaction of word length by compound frequency, but as the range of word lengths in our study is small, with most words having a length of 12 characters, our data do not shed light on the visual acuity hypothesis of Bertram and Hyönä (2003), according to which compound frequency effects would be more prominent for shorter words with less than 9 characters (Bertram & Hyönä, 2003; cf., also Pollatsek *et al.*,

in one fixation is quite low, in fact, only 18% of our stimuli elicited a single fixation. We conclude that we found evidence that full-form access (diagnosed by the compound frequency effect) is initiated before all characters of the compound have been foveally inspected (for the discussion of the early locus of word frequency effect see also Cleland, Gaskell, Quinlan & Tamminen, 2006).

Fourth, the fact that the effect of compound frequency was simultaneous with the left constituent frequency and family size effect and preceded the right constituent frequency and family effect, poses a problem for strictly sequential sublexical models of morphological processing. In such models, one would expect full-form activation to occur in time after activation of the left and the right constituent. In the Taft and Forster (1976) variant of this model, properties of the right constituent should never exert any influence on compound word identification.activation of the right constituent.

Our set of findings is also problematic for supralexical models, as those models argue for initial activation of the full-form and subsequent spreading activation of constituent morphemes. On this view, the properties of the left and the right constituents are expected to receive activation from the full-form and left and right constituent frequency effects should therefore kick in later than the full-form frequency effect. In fact, however, our data show that at least right constituent effects only emerge in later or global processing measures, i.e., subgaze duration for the right constituent and gaze duration.

Fifth, we observed two surprising effects of constituent morphological paradigms. Left constituent family size effect showed up at the first fixation, which is unexpectedly early

2000; Niswander-Klement & Pollatsek, 2006).

given the traditional interpretation of family size effects as a post-access semantic effect reflecting activation spreading through morphological paradigms (cf., e.g., Bertram, Schreuder & Baayen, 2000; De Jong *et al.*, 2000; De Jong, Feldman, Schreuder *et al.*). To explain the finding one has to assume that either the family size effect is formal rather than semantic in nature, or that semantic effects can emerge earlier than usually claimed. As we outline in the General Discussion, we believe that both the formal and the semantic components contribute to the family size effect. On the other hand, we found a late effect of *ResidRightFamSize* on subgaze duration for the right constituent. Recall that the right constituent family is a set of compounds (e.g., *vanilla cream, ice cream, shoe cream*, etc.) beginning in morphemes that can combine with the given right constituent (*cream*). The effect is surprising since by the time when the right constituent is scanned, it is quite plausible that the one left constituent that actually occurs in the compound (e.g., *vanilla*) has already been (partly) identified and then activation of a paradigm of possible left constituents (e.g., *vanilla, ice, shoe*, etc.) appears unwarranted. It is likely that the effect of the right constituent family may be driven by cases in which lexical processing of the left constituent is not complete at the first fixation (for instance, due to difficult lexical processing of the left constituent or suboptimal visual uptake of word-initial information) and continues as a spillover effect even as the eyes move to the right constituent. We return to the role of morphological families in the General Discussion.

Sixth, the interactions between morphological predictors that we saw in lexical decision latencies were replicated in eye-movement measures. As early as the first fixation, left con-

stituent frequency modulated the compound frequency effect, such that compound frequency contributed most to recognition of those compounds in which left constituent frequency was lower, and the compound frequency effect diminished as the left constituent frequency increased (see Fig. 1). Importantly, compound frequency still has a large role to play even when the left constituent frequency is high and the traditional decompositional route is supposed to be the preferred route of compound processing. This interaction indicates that activation of compounds' full-forms and of morphemes is not independent as claimed in several dual-route models of morphological processing, and that the lexical processor is not identifying compounds by strictly selecting between decomposition or full-form processing. Instead, the processing appears to be flexible and co-operative, taking advantage of both (or more, see below) routes, even when it is prompted to rely more upon one of the routes. Thus, identification of the compound through its full-form is optimal when the other route is less beneficial for identification purposes, and vice versa morphological decomposition preferentially takes place when full-form access is less favorable for compound recognition. Moreover, balanced utilization of the two routes is in place from the earliest stages of complex word recognition.

Also, in subgaze duration for the right constituent we observed the interaction of *ResidLeftFamSize* by *RightFreq*, which showed the strongest effect of right constituent frequency in compounds with large left constituent families, and thus with many potential right constituents that might follow the left constituent (see Fig. 2). As we argued above, we take this interaction as evidence that (morphological or other) properties of morphemes and complex

words serve as cues to recognition of morphologically complex structures and that some cues modulate the presence and magnitude of the effect of other cues.

Morphological effects: Deeply embedded morphemes. Thus far we have considered morphological structure at the level of the whole compound and its immediate constituents. We now consider the effects of the internal structure of these immediate constituents.

Similarly to the lexical decision latencies, triconstituent compounds (i.e., those combining three lexemes) consistently elicited longer reading times in the eye-movement record than compounds with two lexemes (one of which additionally included derivational morphemes). The divergence in the processing of the two compound types did not emerge immediately, at the first fixation, rather it presented itself in subgaze and gaze durations. As effects related to meaning are assumed to occur late, we conclude that the divergence reflects a relative difficulty of semantic integration of three, rather than two, free-standing lexemes (on the temporal order of morphological and semantic effects in compounds, see e.g., Cunnings & Clahsen, 2007).

The role of affix position in a complex word varied in accordance with the temporal order of the visual uptake. Obviously, compound-final affixes are viewed with more acuity when the compound's right constituent, rather than the left one, is under foveal inspection. Indeed, compound-final affixes elicited shorter subgaze durations and gaze durations, but their effect was five times stronger in the model for *SubgazeRight* ($\hat{\beta} = -0.10, p = 0.0001$) than it was in the model for *SubgazeLeft* ($\hat{\beta} = -0.02, p = 0.0001$). Furthermore, multiple affixes appeared to facilitate processing even more than other types of affixation, as revealed

in subgaze duration for the left constituent (see Table 6). This finding is consistent with the hypothesis that affixes function as segmentation cues in locating the boundaries of morphological constituents (Kuperman *et al.*, 2008). The observed advantage of compounds with multiple affixes may indicate the relative ease of identifying a higher-level morphological hierarchy in complex words with multiple segmentation cues.

An analysis of the subset of words with exactly one affix (9790 fixations) showed that more productive affixes (i.e., affixes that occur in more word types) came with shorter gaze durations ($\hat{\beta} = -0.009, t(9790) = -6.403, p < 0.001$; effect size = -15 ms, model not shown). This result converges with lexical decision studies in Finnish (cf., Bertram, Laine & Karvinen, 1999) reporting shorter RTs for derived words with more productive affixes than for words with unproductive affixes.

Orthographic and Visuo-Motor Variables. Compound length (*WordLength*) went hand in hand with shorter first fixations (-37 ms) and with longer gaze durations (26 ms). This trade-off between the number and duration of fixations in correlation with word length is well-attested in the eye-movement literature (cf., Vergilino-Perez *et al.*, 2004 and references therein). Compounds with longer left constituents (*LeftLength*) elicited longer first fixations and subgaze durations for left constituents, which is as expected. In subgaze durations for the right constituents and gaze durations, the effect of left constituent length appeared to be reverse: *LeftLength* correlated negatively with durations. However, since we set the maximum for compound length, longer left constituents implied shorter right constituents. So the longer the compound's left constituent, the shorter its right constituent, and the

faster it takes to complete the visual uptake of the right constituent (hence shorter subgaze duration for the right constituent), which is in line with the direction of the corresponding effect for the left constituent length.

At first fixation, the nonlinear effect of fixation position on fixation duration showed the inverse-U shape (see the linear term *FixPos* and the quadratic term *FixPos2* in Table 5). The fixations between the 4th and the 5th character (i.e., the position of the displayed fixation point in our experiment) had a longer duration (on average by about 70 ms) than did fixations at the word's extremes, the first and the twelfth character of the stimuli. This Inverted-Optimal Viewing Position effect is well attested in the literature on eye-movements for single word recognition and sentential reading (for an overview of available theoretical accounts see Vitu, Lancelin & d'Unienville, 2007). Initial fixation position did not interact with any predictors of our interest.

Other Control Variables. We observed longitudinal effects of the course of the experiment on participants' performance. The more the participants progressed into the experiment (as measured by the position of trial in the experimental list), the shorter their first fixations were (effect size = -9 ms), and their gaze durations were also shorter (effect size = -8 ms). In other words, the eye-movement record, just as the lexical decision latencies, shows that participants become familiarized with the task as the experiment proceeds, in line with e.g., Meeuwissen, Roelofs & Levelt (2003) and De Vaan *et al.* (2007).

The longer the lexical decision latency to the immediately preceding trial was (*RT1*), the longer the first fixations were (effect size = 51 ms). Longer *RT1* also came with a

substantial lengthening of gaze duration (effect size = 282 ms). The "spillover" effect on the current trial of the processing difficulty of the preceding trial is noticeable not only in the visual lexical decision latencies, but apparently co-determines the entire time-course of morphological processing starting from the first fixation onwards. There may be two components to the effect of the RT on the preceding trial. First, this effect may reflect the spillover of the lexical processing load, which is clearly increased in the cases with longer *RT1*. In other words, word N-1 may still be processed even when the lexical decision has been made and word N has been presented. Second, and perhaps more likely, the dynamics of going through the experiment may be such that the local processing speed at word N adapts to the speed developed at previous trials (in our case, the immediately preceding trial). Being fast in a recent decision-making and motoric action of the lexical decision may influence the availability of resources and expected speed of processing for the current trial (regardless of the actual lexical characteristics of the currently presented word). We leave disentangling these possibilities to further research. Yet we note that neglecting this predictor in the statistical analysis may have profound consequences. For instance, when *RT1* was removed from the statistical model for gaze durations, the amount of variance explained by the fixed effects dropped by 1.3% percent. From a methodological perspective, bringing longitudinal and local effects in the course of the experiment may be crucial for coming to a proper understanding of the data (cf., De Vaan *et al.*, 2007; Kinoshita & Mozer, 2006; Taylor & Lupker, 2006).

General Discussion

This study primarily addressed the role of morphological structure in compound recognition. This section begins with a summary of findings, then we elaborate on the methodology of this study, and finally, we formulate requirements for a model of compound processing which would account for the present set of results.

To explore computation for multiply complex words, we considered a range of diagnostic measures traditionally interpreted as indicating decompositional processing. In our data, we observed facilitatory effects of the left and right constituent lemma frequencies, as well as the facilitatory effects of the left and right constituent family sizes. In addition, we found facilitatory effects of the compound lemma frequency, the traditional hallmark for non-decompositional processing.

The time-course of all these effects was tied to the time-course and direction of reading. Properties associated with the left constituent played a role in the early measures of eye movements, while the role of the right constituent emerged relatively late (cf., Hyönä *et al.*, 2004). Moreover, the effect sizes observed for the right constituent were considerably smaller.

The constituent frequency and family size effects may have arisen at the level of form processing, at the level of semantic processing, or possibly at both levels. At the level of form, the effect of a constituent's frequency may reflect the reader's experience with identifying that constituent's string of characters. The effect of morphological family may tap into a reader's more specific experience with parsing out and recognizing the constituent as part of a larger word. At the level of word meaning, a constituent's frequency may gauge the ease of access to its meaning. A constituent's family size would then estimate the resonance that

activation of a constituent morpheme gives rise to in its morphological family.

The effect of compound frequency emerged already at the first fixation duration, a point in time when most compounds have not yet been fully scanned. There are several ways in which this surprising effect can be interpreted. This full-form frequency effect may result from unstructured form processing in which the available visual input at the first fixation (the initial characters, the previewed characters in the middle of the word, as well as the word's length, cf., Pollatsek & Rayner, 1982; Rayner, Well, Pollatsek & Bertera, 1982) is matched against stored form representations. The more entrenched this full-form representation is, the earlier the benefits of its availability emerge in the eye-movement record. Importantly, this interpretation presupposes that full-form representations do not require full visual inspection of the input and may be accessed on the basis of partially matching information (cf., de Almeida & Libben, 2002). The fact that the effect of compound frequency is also visible in later measures implies that the full-form representation of a compound is actively involved in the process of compound recognition even when other sources of lexical information become available, possibly for checking the new input for consistency with the already activated full-form and/or deactivating other competitors in the morphological family.

It is unlikely, however, that unstructured form processing would fully account for the compound frequency effect and especially for its presence in the late eye movement measures. The compound frequency effect survives inclusion in the statistical model of the frequency of the initial quadrogram summed over words that match the target compound in length (model not shown). This indicates that it is unlikely that the compound frequency effect can

be reduced specifically to the earliest available visual information. Following Wurm, Aycock and Baayen (2008), it is conceivable that full-form frequency effects reflect, at least in part, memory traces of constituent morphemes having been combined together into one lexical unit. 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.

At the present stage of our knowledge, we cannot exclude that the compound frequency effect is also indicative of facilitation from semantic processing, given that semantic effects have been observed for very short initial time spans (cf., e.g., Diependaele, Grainger & Sandra, 2005; Hauk & Pulvermüller, 2004; Penolazzi *et al.*, 2007; cf., also Baayen, Feldman & Schreuder, 2006, for evidence concerning a strong semantic component to the word frequency effect).

In addition to constituent frequency and family size effects, and in addition to the compound frequency effect, we obtained ample evidence for a role of morphemes that are embedded inside the immediate constituents of compounds. Thus, embedded affixes that are more productive elicited shorter gaze durations, as expected given previous studies of bimorphemic derivations (cf., e.g., Bertram *et al.*, 1999). We also observed that compounds embedded in compounds require more reading time than derivations embedded in compounds. We have two possible explanations for that. First, compounds with three free-standing lexemes are more difficult to integrate semantically than those with two such lexemes. For instance,

readers need to determine whether a compound with three lexemes is left-branching (i.e., the first two constituents modify the third, as in *voet-bal+bond* "football association") or right-branching (i.e., the first constituent is a modifier of the two latter constituents, as in *zaal+voet-bal* "indoor football"). Second, the derivational morpheme may have served as a parsing cue to identification of immediate constituents, and using such cues allows faster access to morphological constituents and faster semantic wrap-up of the complex word (see Kuperman *et al.*, 2008, for a more detailed discussion on this issue).

Methodological considerations. A comparison of the results obtained with the visual lexical decision task and those obtained with the cumulative eye-movement measures (subgaze and gaze durations) show remarkable convergence. In the RTs, just like in eye movements, we observe facilitatory effects of constituent frequencies and family sizes, and also those of compound frequencies. We also find qualitatively similar interactions between morphological predictors (*WordFreq* by *LeftFreq*, and *ResidLeftFamSize* by *RightFreq*) in lexical decision latencies and eye-movement durational measures. Furthermore, embedded morphemes and experimental control variables give rise to very similar patterns of results in the two datasets, lexical decision latencies and eye-movements. What the analysis of the eye-movements adds is detailed information about the time-course of morphological processing, including the early and lingering compound frequency effect, the early left constituent family size effect, and the temporal sequence of the effects pertaining to the compounds' left and right constituents.

Our choice of investigating the processing of isolated existing and nonce compounds in visual lexical decision has offered us both advantages and disadvantages. The main advan-

tage of using isolated words is the ability to collect large numbers of data points from the same participant relatively quickly. As a result, our statistical analyses enjoy the benefit of enhanced power. In addition, combining lexical decision, the task that has been used most intensively to study morphological processing, with eye-tracking allows us to evaluate to what extent the two paradigms converge (cf., Grainger's (2003) program of investigating functional overlap between tasks). As noted above, there is indeed remarkable convergence in our data.

Our choice for using isolated words in lexical decision also comes with several disadvantages, most of which concern the issue of the ecological validity of our results. In single word reading, there is no parafoveal preview from the preceding word, and there is no natural spillover effect from the target word to the next word to be investigated. More importantly, lexical decision may induce rather different kinds of processing strategies than those used for the natural integration of word meaning into the sentence and discourse.

Another methodological decision that we had to make is whether to include a look away point on the screen, that is, whether to instruct participants to complete their lexical decision task by fixating either the word "Yes" or the word "No", which would be displayed in two different areas on the screen equally distant from the area where the target word was displayed (for the full description of this technique, see Hyönä, Laine & Niemi, 1995). For compatibility with the existing body of literature, we stayed as close to the conventional lexical decision paradigm as possible, and did not make use of such a look away point. Instead, we considered in our analyses only those fixations that were completed before the

button press registering a lexical decision. The price we pay is the possibility of some more noise in the eye-movements measures, especially in the gaze durations. Yet in our data, gaze durations and RTs are not that highly correlated: $R^2 = 0.46$ only. Thus, both gaze durations and RTs serve as dependent variables in their own right.

We also note that the presence of nonce compounds and of many low-frequency existing compounds in our experiment may have enhanced decompositional processing and inhibited full-form processing. In the light of this possibility, it is all the more surprising that an effect of compound frequency is observed at the very first fixation.

Whatever the disadvantages of our methodology may be, the pattern of results that we have obtained and reported either in the body of the paper or in Appendix 1, dovetails perfectly with many of the results obtained in the literature for sentential reading, such as visuo-oculomotor effects (cf., e.g., O'Regan *et al.*, 1994; Rayner, 1998; Vitu, McConkie, Kerr & O'Regan, 2001), effects of compound length and frequency, as well as of constituent frequencies (cf., e.g., Andrews *et al.*, 2004; Duñabeitia, Perea & Carreiras, 2007; Hyönä & Pollatsek, 1998; Hyönä *et al.*, 2004; Juhasz *et al.*, 2003; Taft & Forster, 1976), and effects of orthographic n-grams (reported in Appendix 1, cf., e.g., Lima & Inhoff, 1985). Furthermore, in a recent sentential reading study (Kuperman, Bertram & Baayen, 2008), in which Finnish compounds were embedded in context, a highly similar pattern of results was observed, including early effects of compound frequency, left constituent frequency and family size, later and weaker effects of right constituent frequency and family size, interactions between morphological predictors, as well as longitudinal experimental effects.

Towards a theory of compound processing. According to Libben (2006), readers and listeners maximize their opportunities for comprehension by the simultaneous use of all processing cues available to them, and all processing mechanisms that they have at their disposal, including retrieval from memory and compositional computation. The present study provides support for Libben's hypothesis of maximization of opportunity. All constituent morphemes, the whole compound itself and morphological families that share one of the compound's constituents play a noticeable role in lexical processing of compounds. This indicates that there are multiple routes at work in compound processing, and readers use these routes interactively, at different times and to a different extent, to efficiently and accurately recognize compounds. The early compound frequency effect shows that readers do not wait for all the characters of the word to be seen before making inferences about the word's identity. The early compound frequency effect also shows that readers do not gain access to compound representations only after having accessed its constituents. The interactions of morphological predictors (compound frequency by left constituent frequency and left constituent family size by right constituent frequency) show that the cues modulate each other, and that decompositional processes and full-form driven processes are not independent. Using one kind of morphological information for compound identification as if other sources of information do not exist amounts to missing out on the cumulative use of informations and on concomitant facilitation of performance.

In what follows, we take as the point of departure the basic assumption of parallel dual route models, given the evidence in our data for both processing routes. As the detailing of a

full-fledged model of morphological processing is beyond the scope of this study, we restrict ourselves to listing a number of requirements that are not satisfied by the current parallel dual route models proposed in the literature (e.g., Schreuder & Baayen, 1995). While our results were obtained in the visual domain, we believe that the requirements outlined below would equally hold for the models of the auditory processing of compounds.

First, current models of morphological processing almost always discuss complex words as if they are read with only one fixation. An example of a model that addresses the temporal dynamics of reading complex words is the one proposed by Pollatsek, Reichle and Rayner (2003), and they conclude that a parallel dual route architecture is unable to approximate the empirical data, unless the two routes of lexical processing are allowed to interact. It is clear, also from the present data, that the details of the time-course of information entering the system needs to be explicitly included in models of morphological processing in reading.

In the typical left-to-right reading of long compounds the very first opportunities for comprehension of the compound present themselves already during parafoveal preview, when information about the initial characters and word length becomes available (Rayner *et al.*, 1982). In single-word reading, this information is also available very early, during the low-level attentional scan of the word that occurs in the beginning of fixation, cf., Reichle, Rayner and Pollatsek, 2003. Following Clark and O'Regan (1999) and O'Regan (1979), word length may play a disambiguating role in word recognition (for the opposing view, see Inhoff & Eiter, 2003). For words embedded in the sentential context, additional information may come from contextual predictability (e.g., Ehrlich & Rayner, 1981), collocational strength

(e.g., McDonald & Shillcock, 2001) and constructional cues (e.g., Frazier *et al.*, 2006).

The next opportunities for restricting the range of possible interpretations for the visual input arise at the first fixation, where a range of properties of the first constituent come into play: not only the frequency of the left constituent, its length, and its morphological family, but also the combinatorial likelihood of morphemes within the whole compound, in conjunction with information about the compound's length. Later opportunities (at second and subsequent fixations) include properties of the right constituent. New information obtained at this stage is processed against the backdrop of the information already extracted about the word.

Second, models of morphological processing in reading need to allow for a simultaneous processing of information at different levels without requiring strict sequentiality of processing stages, as witnessed, for instance, by the simultaneous early effects in our data of compound frequency, left constituent frequency and family size, and orthographic n-gram effects[7]. Our results challenge sublexical models, which allow full-form access only after morphological constituents have been recognized (cf., Pinker, 1999; Taft, 2004; Taft & Forster, 1976; Taft, 1991). Our results also challenge supralexical models, which only allow constituents to come into play after the compound as a whole has been recognized (Giraudo &

[7]A modeling framework that may prove to be useful here is the hierarchical temporal memory framework proposed by Hawkins and George (2006), see also Hawkins and Blakeslee (2004). In the hierarchical temporal memory framework, the simultaneous processing would be accomplished by generation skip, i.e., lower-level detectors in the hierarchy propagating information about the input to higher levels, skipping intermediate levels.

Grainger, 2001).

Third, models of compound processing should allow for the modulation of the weight of one opportunity by the presence and strength of other opportunities, as witnessed by the interaction of compound frequency and left constituent frequency (for related discussion of cue trade-offs in speech processing see e.g., Mattys, White & Melhorn, 2005; McClelland & Elman, 1986). Current parallel dual route models tend to simplify morphological processing to activation of autonomous lexical representations that are blind to each other's activation (cf., Laudanna & Burani, 1985; Frauenfelder & Schreuder, 1991, and Schreuder & Baayen, 1995; see however Baayen & Schreuder, 2000). In general, the fact that we find, also in the parallel study, early constituent frequency effects and whole-word frequency effects at the same time, tells us that one cue or route is not cancelling out the other completely, a prediction that would directly derive from a strict dual route model. Depending on the strength of the available cues, the fine-tuning of this kind of co-operative system depends on the specific properties of the complex word.

Fourth, models of morphological processing should come to grips with fast activation of morphological paradigms (families) associated with a compound's constituents. One important constraint on morphological models is our finding that left constituent families are activated immediately upon access to those constituents, and not after full-form access.

Effectively, a model that meets these requirements is no longer a dual route model, but rather a multiple route model that, in morphological terms, allows access to full-forms, immediate constituents, embedded morphemes and morphological families. More generally,

such a model will have as its basic principle maximization of all opportunities, both morphological, orthographic, phonological, and contextual, for comprehension of the visual input. We believe that probabilistic and information-theoretical approaches to lexical processing developed recently in morphological and syntactic research (cf. e.g., Moscoso del Prado Martín *et al.*, 2004; Levy, 2008) hold promise for formalization of those opportunities and for computational implementation of the multiple-route model of compound recognition.

Acknowledgements

Thanks are due to Françoise Vitu, Ingo Plag and two anonymous reviewers for helpful comments on previous versions of this paper. We also thank Edwin Commandeur for his assistance with the experimental setup.

References

Allen, M. and Badecker, W. (2002). Inflectional regularity: Probing the nature of lexical representation in a cross-modal priming task. *Journal of Memory and Language*, 46:705–722.

Andrews, S. (1986). Morphological influences on lexical access: lexical or non-lexical effects? *Journal of Memory and Language*, 25:726–740.

Andrews, S., Miller, B., and Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, 16(1/2):285–311.

Baayen, R. H. (2007). Storage and computation in the mental lexicon. In Jarema, G. and Libben, G., editors, *The Mental Lexicon: Core Perspectives*, pages 81–104. Elsevier.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge University Press, Cambridge.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *In press*.

Baayen, R. H., Feldman, L. B., and Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55:290–313.

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

Baayen, R. H. and Schreuder, R. (1999). War and peace: morphemes and full forms in a non-interactive activation parallel dual route model. *Brain and Language*, 68:27–32.

Baayen, R. H. and 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.

Baayen, R. H., Wurm, L. H., and Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *Mental Lexicon*, 2:419–463.

Bates, D. M. and Sarkar, D. (2005). The lme4 library. *[On-line], Available: http://lib.stat.cmu.edu/R/CRAN/.*

Bertram, R. and 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.

Bertram, R., Laine, M., and Karvinen, K. (1999). The interplay of word formation type, affixal homonymy, and productivity in lexical processing: Evidence from a morphologically rich language. *Journal of Psycholinguistic Research*, 28:213–226.

Bertram, R., Pollatsek, A., and Hyönä, J. (2004). Morphological parsing and the use of segmentation cues in reading Finnish compounds. *Journal of Memory and Language*, 51:325–345.

Bertram, R., Schreuder, R., and Baayen, R. H. (2000). The balance of storage and computation in morphological processing: the role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26:489–511.

Clark, J. and O'Regan, J. (1999). Word ambiguity and the optimal viewing position in reading. *Vision Research*, 39:842–857.

Cleland, A., Gaskell, M., Quinlan, P., and Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology:Human Perception and Performance*, 32:104–119.

Cunnings, I. and Clahsen, H. (2007). The time-course of morphological constraints: Evidence from eye-movements during reading. *Cognition*, 104:476–494.

de Almeida, R. and Libben, G. (2002). Compound pre-access decomposition: Effects of constituent disruption. *Folia Linguistica*, 36:97–115.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., and Baayen, R. H. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects. *Brain and Language*, 81:555–567.

De Jong, N. H., Schreuder, R., and Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, 15:329–365.

De Vaan, L., Schreuder, R., and Baayen, R. H. (2007). Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon*, 2(1):1–24.

Diependaele, K., Sandra, D., and Grainger, J. (2005). Masked cross-modal morphological priming: unravelling morpho-orthographic and morpho-semantic inluences in early world recognition. *Language and Cognitive Processes*, 20:75–114.

Dijkstra, T., Moscoso del Prado Martín, F., Schulpen, B., Schreuder, R., and Baayen, R. (2005). A roommate in cream: Morphological family size effects on interlingual homograph recognition. *Language and Cognitive Processes*, 20:7–41.

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

Ehrlich, S. F. and Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20:641–655.

Frauenfelder, U. H. and Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In Booij, G. E. and Marle, J. v., editors, *Yearbook of Morphology 1991*, pages 165–183. Kluwer Academic Publishers, Dordrecht.

Frazier, L., Carminati, M., Cook, A., Majewski, H., and Rayner, K. (2006). Semantic evaluation of syntactic structure: Evidence from eye movements. *Cognition*, 99:B53–B62.

Giraudo, H. and Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review*, 8:127–131.

Grainger, J. (2003). Moving eyes and reading words: How can a computational model combine the two. In Hyönä, Y., Radach, R., and Deubel, H., editors, *The mind's eye: Cognitive and applied aspects of eye movement research*. Elsevier, Amsterdam.

Hauk, O. and Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, 115:1090–1103.

Hawkins, J. and Blakeslee, S. (2004). *On Intelligence*. Henry Holt and Co, New York.

Hawkins, J. and George, D. (2006). Hierarchical temporal memory. concepts, theory and terminology. *[On-line], Available: http://www.numenta.comtechnology.php*.

Hyönä, J., Bertram, R., and Pollatsek, A. (2004). Are long compound words identified

serially via their constituents? Evidence from an eye-movement-contingent display change study. *Memory and Cognition*, 32:523–532.

Hyönä, J., Laine, M., and Niemi, J. (1995). Effects of a word's morphological complexity on readers' eye fixation pattern. In Findlay, J. M., Kentridge, R., and Walker, R., editors, *Eye movement research: mechanisms, processes and applications*, pages 207–231. Elsevier, Amsterdam.

Hyönä, J. and 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.

Inhoff, A. and Eiter, B. (2003). Knowledge of word length does not constrain word identification. *Psychological Research*, 67:1–9.

Juhasz, B. (2007). The influence of semantic transparency on eye movements during english compound word recognition. In Van Gompel, R., Fischer, M., Murray, W., and Hill, R., editors, *Eye Movements: A Window on Mind and Brain*, pages 373–389. Elsevier, Amsterdam.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., and 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.

Kinoshita, S. and Mozer, M. (2006). How lexical decision is affected by recent experience:

Symmetric versus asymmetric frequency-blocking effects. *Memory and Cognition*, 34:726–742.

Krott, A., Baayen, R. H., and Schreuder, R. (2001). Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics*, 39(1):51–93.

Kuperman, V., Bertram, R., and Baayen, R. H. (2008). Morphological dynamics in compound processing. *In press, Language and Cognitive Processes*.

Laudanna, A. and Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 345–364. Lawrence Erlbaum Associates, Hillsdale, N. J.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177.

Libben, G. (2006). Why study compound processing? In Libben, G. and Jarema, G., editors, *The representation and processing of compound words*, pages 1–23. Oxford University Press, Oxford.

Lima, S. D. and Inhoff, A. (1985). Lexical access during eye fixations in reading: Effects of word-initial letter sequence. *Journal of Experimental Psychology: Human Perception and Performance*, 11:272–285.

Mattys, S., White, L., and Melhorn, J. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134:477–500.

McClelland, J. L. and Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18:1–86.

McDonald, S. and Shillcock, R. (2001). Rethinking the Word Frequency Effect: The Neglected Role of Distributional Information in Lexical Processing. *Language and Speech*, 44(3):295–323.

McDonald, S. and Shillcock, R. (2004). The potential contribution of preplanned refixations to the preferred viewing location. *Perception and Psychophysics*, 66(6):1033–1044.

Meeuwissen, M., Roelofs, A., and & Levelt, W. J. M. (2003). Planning levels in naming and reading complex numerals. *Memory & Cognition*, 31:1238–1248.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., and Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30:1271–1278.

Niswander-Klement, E. and Pollatsek, A. (2006). The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. *Memory and Cognition*, 34:685–702.

O'Regan, J. (1979). Saccade size control in reading: Evidence for the linguistic control hypothesis. *Perception and Psychophysics*, 25:501–509.

Penolazzi, B., Hauk, O., and Pulvermüller, F. (2007). Early semantic context integration

and lexical access as revealed by event-related brain potentials. *Biological Psychology*, 74:374–388.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Statistics and Computing. Springer, New York.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Weidenfeld and Nicolson, London.

Pollatsek, A. and Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, 20:261–290.

Pollatsek, A., Hyönä, J., and Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 26:820–833.

Pollatsek, A. and Rayner, K. (1982). Eye-movement control in reading - the role of word boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 8:817–833.

Pollatsek, A., Reichle, E., and Rayner, K. (2003). Modeling eye movements in reading: Extensions of the E-Z Reader model. In Hyönä, Y., Radach, R., and Deubel, H., editors, *The mind's eye: Cognitive and applied aspects of eye movement research*, pages 361–390. Elsevier, Amsterdam.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372–422.

Rayner, K., Well, A., Pollatsek, A., and Bertera, J. (1982). The availability of useful information to the right of fixation in reading. *Perception and Psychophysics*, 31:537–550.

Reichle, E., Reyner, K., and Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26:445–526.

Schreuder, R. and Baayen, R. H. (1995). Modeling morphological processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 131–154. Lawrence Erlbaum, Hillsdale, New Jersey.

Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96:523–568.

Sereno, S. C., Rayner, K., and Posner, M. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *Neuroreport*, 9:2195–2200.

Taft, M. (1991). *Reading and the mental lexicon*. Lawrence Erlbaum, Hove, U.K.

Taft, M. and Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14:638–647.

Taft, M. and Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, 15:607–620.

Taylor, T. E. and Lupker, S. J. (2006). Time perception and word recognition: An elaboration of the time-criterion account. *Perception and Psychophysics*, 68:933–945.

Vergilino-Perez, D., Collins, T., and Doré-Mazars, K. (2004). Decision and metrics of refixations in reading isolated words. *Vision Research*, 44:2009–2017.

Vitu, F., Lancelin, D., and Marrier d'Unienville, V. (2007). A perceptual-economy account for the inverted-optimal viewing position effect. *Joural of Experimental Psychology: Human Perception and Performance*, 33:1220–1249.

Vitu, F., McConkie, G., Kerr, P., and O'Regan, J. (2001). Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision Research*, 41:3513–3533.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, 9:341–368.

Appendix 1

Key to Table 3: Predictors of primary interest for this study are presented in the main body of paper. Additional control variables that show significant effects in our statistical models are as follows: *Correct1*, the binary indicator of whether the previous trial was a correct lexical decision; *FixPos* and *FixPos2*, first fixation position and its squared value; *FinTrigram*, frequency of the word-final trigram; and *Nomore*, indicator of whether the fixation is word-final. In addition to these, we have considered a large number of control variables that were not significant predictors of reading times, fixation probabilities or lexical decision latencies. These included variables listed in the subsection Dependent variables as well as initial trigram frequency, mean bigram frequency of the word, position of the minimal bigram, affix length, branching of triconstituents, and frequencies of deeply embedded stems in triconstituents.

Specifications of statistical models

Specifications include estimates of the regression coefficients; 95% highest posterior density intervals (HPDs), which are a Bayesian estimate of the most likely values of a parameter, roughly comparable to traditional 95% confidence intervals; p-values estimated by the Monte Carlo Markov chain (MCMC) method using 1000 simulations; and p-values obtained with the t-test for fixed effects using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom (see Pinheiro & Bates, 2000 for discussion of the method). We also report the estimated standard deviations for each random intercept (e.g., *Subject* or *Word*) and each random slope (e.g., *Subject* by

WordLength), together with the estimates based on the MCMC samples and HPD intervals, such as the MCMC mean and 95% HPDs (Table 9 for all models), see Pinheiro and Bates (2000) for detailed treatment of random effects in mixed-effects models.

Computation of effect sizes

Effect sizes were estimated as follows. For factors, for which we used contrast coding, effect size was calculated as the difference between (i) the sum of the intercept and the contrast coefficient, $\hat{\beta}$, and (ii) the intercept. For log-transformed dependent variables (fixation duration, gaze duration, RT), effect sizes were calculated for back-transformed values, so that effect sizes are reported in ms. Effect sizes for simple main effects of a covariate were calculated as the difference between the model's predictions for the minimum and maximum values of that covariate.

Comparison between effect sizes of numeric variables obtained in our study and in previous studies that set these variables to a discrete number of levels for factorial designs is not straightforward. Our estimates are defined over the entire range of values of the variable, while a factorial contrast is defined as a difference between group means, where groups are formed (in the simplest case) by dichotomization of a given predictor. The best approximation to factorial estimates is one half of our effect sizes, which is equivalent (for linear effects) to the factorial contrast where the variable of interest is dichotomized and where the group means are positioned at the first and the third quartiles. Obviously, factors do not pose such a problem and are directly comparable across reports.

Appendix 2

In the present study, we chose to present readers with a fixed order of items in the block and the fixed order of blocks, so that each reader saw the words in the same order (even though that one order was set randomly). We hypothesized that by using one list order we would have tighter experimental control, especially as we have the position of an item in the experimental list as a covariate in the model, so that longitudinal effects of practice or fatigue are modeled explicitly. By using the fixed list order we also attempted to avoid the increase in between-participant variance, which derives from the random ordering of items across participants. By that, we aimed at gaining increased statistical power. Using the fixed list order, however, goes against the common practice of counterbalancing (or otherwise randomizing) the presentation order of items across participants. The problem that is usually claimed to follow from using the fixed item order is that the variance that one attributes to a predictor of interest might in fact be due to the influence of the item order. In other words, the item order is a potential confound for estimates of other effects.

It turns out that the item order is no a priori reason for worry. Linear mixed effects models are very well able to disentangle the various sources of variance for a design such as we used. In a simulation study, we considered a repeated measures design with 20 participants, 1000 items, list position (the rank or trial number in the list) as a predictor, and five predictors specifying properties linked to the items (standing for word length, word frequency, left constituent family size, left constituent and right constituent frequencies). In other words, the simulated data have the same design as our experimental data, albeit with fewer predictors

and fewer items. The question of interest is whether the mixed-effects modeling algorithm can adequately separate the different sources of variance under two conditions, one in which each participant is exposed to the items in the same order (as in our manuscript) and one in which each participant is exposed to the items in a different random order. This simulation does not aim to assess the significance of our predictors or to validate our statistical models. Rather we simulate two types of experimental designs (with a fixed number of items, participants and predictors) to see whether, under these conditions, predictions of statistical models would differ across designs.

More formally, let i index participants, j index items, and k index trial number. Furthermore, let X_1 denote *Trial*, X_2 to X_6 item-bound properties, and β_{0-6} the intercept and regression coefficients, respectively. We further denote participant random effect as b_{S_i} (normally distributed with standard deviation σ_S) and item random effect as b_{W_j} (normally distributed with standard deviation σ_W), and we denote the error term as ϵ_{ijk} (normally distributed with standard deviation σ). Our simulated data have the general form

$$y_{ijk} = \beta_0 + \beta_1 X_{1_k} + \beta_2 X_{2_j} + \beta_3 X_{3_j} + \beta_4 X_{4_j} + \beta_5 X_{5_j} + \beta_6 X_{6_j} + b_{S_i} + b_{W_j} + \epsilon_{ijk}, \quad (1)$$

$$b_{S_i} \sim \mathcal{N}(0, \sigma_S), b_{W_i} \sim \mathcal{N}(0, \sigma_W), \epsilon_{ijk} \sim \mathcal{N}(0, \sigma), b_{S_i} \perp b_{W_j} \perp \epsilon_{ijk}.$$

In building a simulation there are many choices to be made. In this simulation we make a simplifying assumption that item-bound predictors X_2 to X_6 are uncorrelated, while the predictors in the empirical data show mild collinearity. Also, it is not necessarily the case that there is a unique value for a predictor for each trial. Say, if we take word length to range from 4 to 12 characters, there will not be 1000 different values of word length for 1000 trials,

rather integer values for 4 to 12 will be repeated multiple times, just like in the original data.

We distinguish between a model with fixed order for all participants, so $k = j$, henceforth $M_{k=j}$, and a model in which each participant has a different order, so $k \neq j$, henceforth $M_{k\neq j}$. We studied the behavior of both models for 20 participants and 1000 items, across 1000 simulation runs. Columns 1-3 in Table 10 specify which fixed and random predictors were used in the simulation, what their coding is in (1), and what the values were that we set for those predictors. We based the ranges of values for item-bound predictors on the actual ranges in the experimental data. Values of X_2 to X_6 varied randomly (uniformly) in the corresponding value ranges. For all predictors, only integer values were considered. Our estimates for regression coefficients and the intercept closely follow the output of the statistical model for lexical decision latencies (Table 4), with a few exceptions. We increased variance in data by setting higher values for random errors, and we diminished the influence of the strongest lexical predictors, word length and word frequency, by dividing their regression coefficients by 10. We increased noise and weakened some effects, because the simulation ran on the original data showed the almost perfect accuracy in estimating the coefficients, and it reported significance of predictors correctly in almost 100% of cases. Results of the simulation are summarized in Table 10.

Columns 4 and 6 in Table 10 show the means of the estimates of the coefficients for the fixed effects and for the standard deviations of the random effects, obtained with the model with fixed order of items and the model with the randomized order of items for each participant, correspondingly. Columns 5 and 7 show proportions of correctly reported

significance across simulation runs for both types of models. It is evident that for large data samples, such as we used in this study, there is no appreciable difference across presentation orders in the performance of statistical models, neither in the accuracy of estimates for model coefficients or for standard deviations of random effects, nor in the power to detect the effect of the item-bound predictors. For smaller samples, we have seen cases where a single experimental list (fixed order) comes with slightly reduced power than the list with the random presentation of items.

We have carried out more simulations with different values for the fixed and random effect parameters, time and again the pattern is like the one summarized in Table 10. Importantly, these simulations show no a priori reason to believe that sources of variance are confounded: at least, not with the number of items and the number of uncorrelated predictors that we used here. It is important to realize that the strength of linear mixed-effects models lies precisely in their ability to 'unconfound' different sources of variance.

We have double-checked whether there were interactions of this longitudinal effect with item-bound predictors, including lexical, distributional or orthographic characteristics of compounds as whole words or their morphemes, but there were none. This gives additional assurance that the morphological effects of our primary interest are not modulated by the longitudinal effects of the experimental list. We have also investigated whether other longitudinal effects might be present (ranging from priming effects due to constituents that appeared earlier in the list to effects of sharing onset or rhyme). None turned out to be significant. In other words, the morphological and orthographic effects that we report are

not artifacts nor confounds of experimental control variables, as can be demonstrated both in a simulation and in an experiment.

Table 2: Summary of morphological effects on durational measures

Predictor	RT	FirstDur	SubgazeLeft	SubgazeRight	GazeDur
LeftFreq	-32 ms (<0.001)	-35 ms (<0.001)	-48 ms (<0.001)	ns	-34 ms (<0.001)
- *interaction with*	*WordFreq* (0.006), Fig. 1	*WordFreq* (0.006)	*WordFreq* (0.001)		*WordFreq* (0.01)
ResidLeftFamSize	-43 ms (<0.001)	-47 ms (<0.001)	-83 ms (<0.001)	-24 ms (0.028)	-61 ms (<0.001)
- *interaction with*	*RightFreq* (0.004), Fig. 2	*WordLength* (0.008)			*RightFreq* (0.008)
RightFreq	-37 ms (0.002)	ns	ns	-27 ms (0.001)	-27 ms (<0.001)
- *interaction with*	*ResidLeftFamSize* (0.004), Fig. 2				*ResidLeftFamSize* (0.008)
ResidRightFamSize	-40 ms (0.001)	ns	ns	-54 ms (0.001)	ns
WordFreq	-96 ms (<0.001)	-24 ms (<0.001)	-42 ms (<0.001)	-47 ms (<0.001)	-73 ms (<0.001)
- *interaction with*	*LeftFreq* (0.006), Fig. 1	*LeftFreq* (0.01)	*LeftFreq* (0.001)		*LeftFreq* (0.006)

Numbers in columns 2-6 show sizes of statistically significant main effects. In the case where the main effect is qualified by the interaction (e.g., *WordFreq* by *LeftFreq*, we report the numerical estimate of the main effect size (*WordFreq*) for the median value of the interacting term (*LeftFreq*). The column *RT* provides references to figures illustrating the interactions, which are qualitatively similar for both lexical decision latencies and reading times. Numbers in brackets provide p-values for the main effects and interactions. "ns" stands for non-significant.

Figure 1: Interaction of compound frequency by left constituent frequency for lexical decision latencies. The lines plot the effect of compound frequency for the quantiles of left constituent frequency (quantile values provided at the right margin). Compound frequency comes with the strongest negative effect at the 1st quantile (solid line), the effect gradually levels off at the 2nd quantile (dashed line), the 3d quantile (dotted line) and the 4th quantile (dotdash line), and is weakest for the compounds with highest-frequency left constituents, the 5th quantile (longdash line).

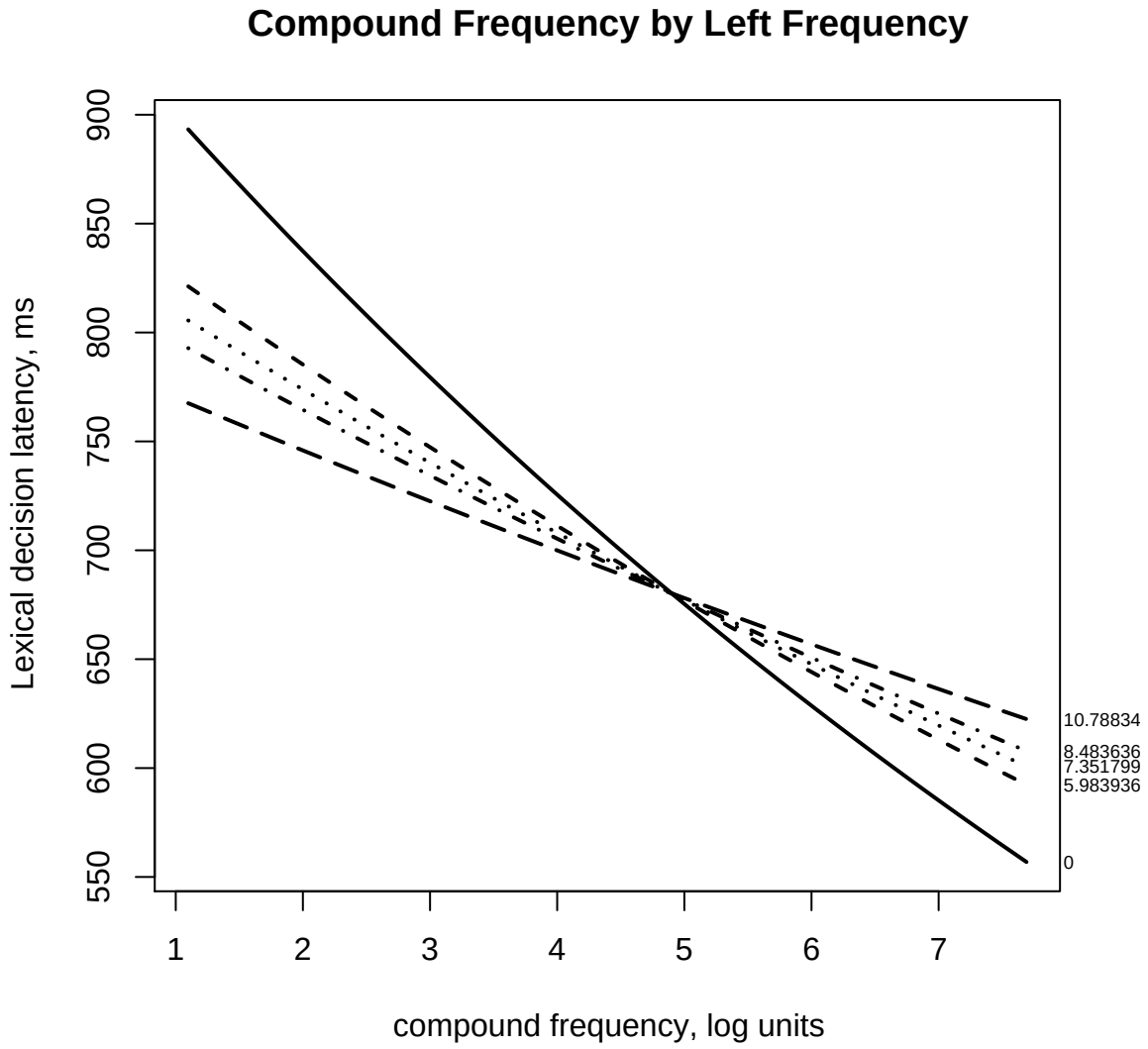

Figure 2: Interaction of right constituent frequency by (residualized) left constituent family size for lexical decision latencies. The lines plot the effect of right constituent frequency for the quantiles of left constituent family size (quantile values provided at the right margin). Right constituent frequency has no substantial effect for smallest left constituent families, represented as the 1st quantile (solid line). The effect gradually increases at the 2nd quantile (dashed line), the 3d quantile (dotted line) and the 4th quantile (dotdash line), and it is strongest for compounds with the largest left constituent families, the 5th quantile (longdash line).

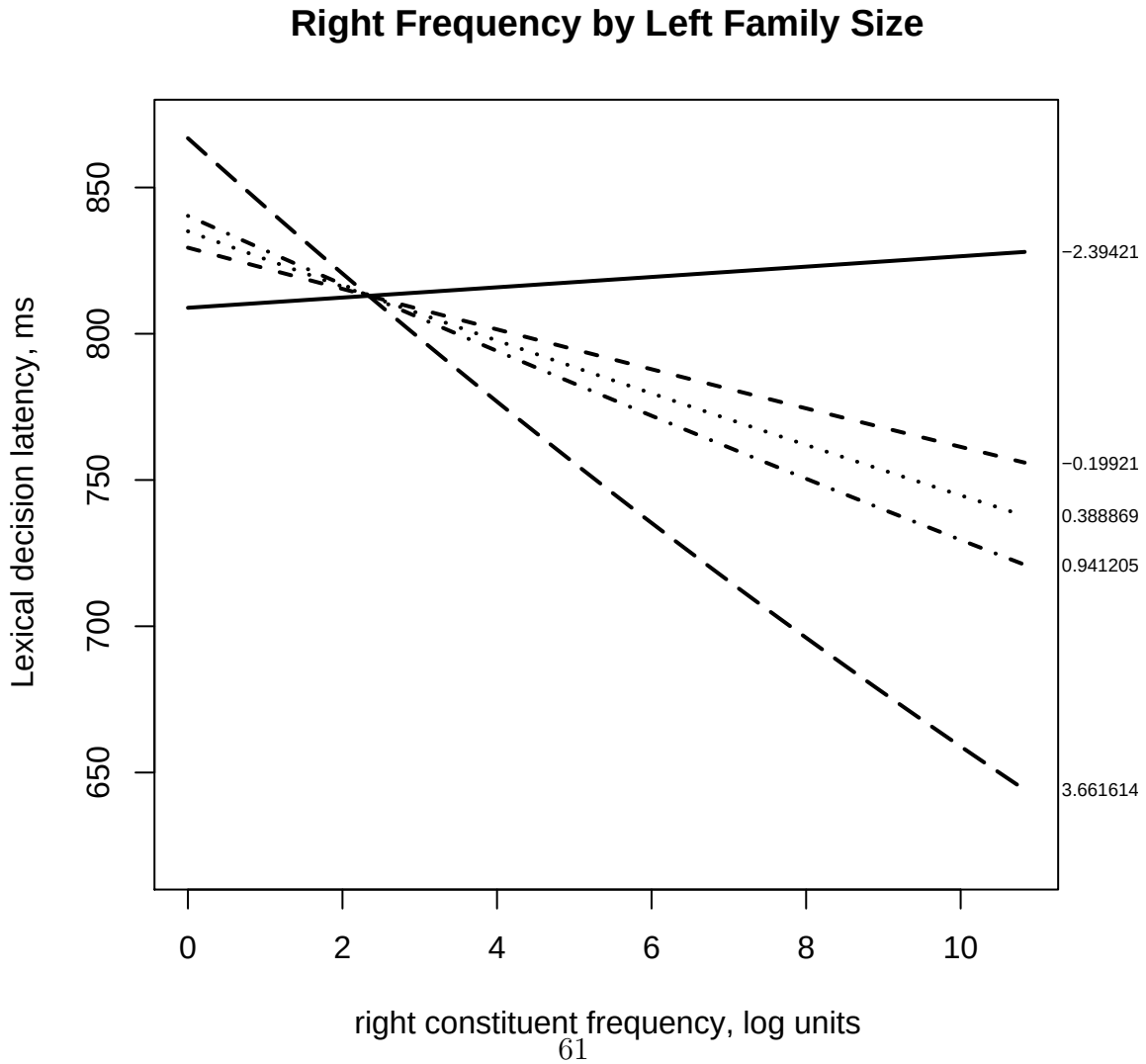

Table 3: *Summary of Continuous Variables*

Variable	Range (Adjusted Range)	Mean(SD)	Median
RT	270:2208 ms (5.6:7.7 log units)	6.7(0.3)	6.6
InitPos	0.1:11.9 characters (1:119 pixels)	40.7(8.4)	40
FirstDur	50:1200 ms (3.9:7.1 log units)	5.6(0.4)	5.6
SubgazeLeft	60:1808 ms (4.1:7.5 log units)	5.8(0.5)	5.7
SubgazeRight	82:1097 ms (4.4:7.0 log units)	5.6(0.4)	5.5
GazeDuration	50:2208 ms (3.9:8.2 log units)	6.5(0.5)	6.5
TrialNum	1:2500	12.0(7.2)	12
RT1	148:4023 ms (5.0:8.3 log units)	6.73(0.3)	6.7
WordLength	8:12 characters	11.6(0.7)	12
LeftLength	2:10 characters	5.4(1.6)	5
FinTrigram	1:984609 (0:13.8 log units)	9.6((2.6)	9.9
WordFreq	3:2207 (1.1:7.7 log units)	2.2(1.1)	1.9
LeftFreq	1:24343 (0.0:10.1 log units)	5.0(2.9)	5.4
RightFreq	1:49020 (0:10.8 log units)	4.5(3.0)	4.2
ResidLeftFamilySize	3:298 (-2.3:3.7)	0.0(1.0)	0.0
ResidRightFamilySize	3:270 (-3.5:7.4)	0.0(1.1)	-0.1
AffixProd	3:6002 (0.7:8.7 log units)	6.8(1.3)	6.99
Complexity	3:6 morphemes	3.2(0.4)	3

Numbers in the second column show original value ranges for predictors. If any transformations have been made with the original values for statistical reasons (i.e., natural log transformation, decorrelation with other predictors or scaling), the numbers in the brackets show the ranges actually used in statistical models. Means, standard deviations and median values refer to the predictor values used in the models. Values for frequency and family size measures are based on the corpus with 42 million word-forms.

Table 4: Fixed Effects of the Model for Lexical Decision RT for Existing Compounds

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.9740	5.9771	5.8176	6.1336	0.001	0.0000
WordLength	0.0148	0.0149	0.0083	0.0226	0.002	0.0000
LeftFreq	-0.0181	-0.0183	-0.0250	-0.0115	0.001	0.0000
RightFreq	-0.0095	-0.0096	-0.0129	-0.0060	0.001	0.0000
Complexity	0.0639	0.0634	0.0302	0.0953	0.001	0.0002
Trial	-0.0041	-0.0042	-0.0050	-0.0032	0.001	0.0000
RT1	0.1288	0.1286	0.1144	0.1413	0.001	0.0000
Correct1Y	-0.0160	-0.0159	-0.0285	-0.0031	0.012	0.0146
ResidLeftFamSize	0.0114	0.0111	-0.0106	0.0353	0.354	0.3557
ResidRightFamSize	-0.0122	-0.0121	-0.0194	-0.0049	0.001	0.0010
AffixFinal	-0.0527	-0.0526	-0.0796	-0.0295	0.001	0.0001
AffixInitial	-0.0178	-0.0169	-0.0613	0.0339	0.500	0.4801
AffixMedial	-0.0382	-0.0378	-0.0653	-0.0116	0.006	0.0062
AffixMultAffix	-0.0897	-0.0887	-0.1371	-0.0394	0.001	0.0004
WordFreq	-0.0717	-0.0722	-0.0904	-0.0533	0.001	0.0000
LeftFreq:WordFreq	0.0037	0.0037	0.0012	0.0062	0.002	0.0047
RightFreq:ResidLeftFamSize	-0.0049	-0.0048	-0.0079	-0.0017	0.001	0.0040

Table 5: Model for First Fixation Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.8489	5.8532	5.5993	6.1337	0.001	0.0000
NomoreTRUE	0.2345	0.2356	0.1773	0.3067	0.001	0.0000
WordLength	-0.0394	-0.0390	-0.0575	-0.0202	0.001	0.0000
LeftLength	-0.0261	-0.0260	-0.0304	-0.0209	0.001	0.0000
FixPos	0.0088	0.0088	0.0065	0.0113	0.001	0.0000
FixPos2	-0.0001	-0.0001	-0.0001	0.0000	0.001	0.0000
WordFreq	-0.0347	-0.0346	-0.0490	-0.0171	0.001	0.0001
LeftFreq	-0.0172	-0.0172	-0.0231	-0.0115	0.001	0.0000
ResidLeftFamSize	0.0728	0.0733	0.0001	0.1559	0.058	0.0690
Trial	0.0000	0.0000	0.0000	0.0000	0.001	0.0000
RT1	0.0352	0.0348	0.0168	0.0515	0.001	0.0001
WordFreq:LeftFreq	0.0031	0.0031	0.0010	0.0052	0.010	0.0058
WordLength:ResidLeftFamSize	-0.0085	-0.0086	-0.0156	-0.0019	0.018	0.0171

Table 6: Model for Subgaze Duration for the Left Constituent

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.9312	5.9380	5.7110	6.1295	0.001	0.0000
WordLength	-0.0628	-0.0627	-0.0729	-0.0529	0.001	0.0000
LeftLength	0.0777	0.0774	0.0687	0.0856	0.001	0.0000
WordFreq	-0.0591	-0.0598	-0.0846	-0.0341	0.001	0.0000
LeftFreq	-0.0272	-0.0275	-0.0361	-0.0175	0.001	0.0000
RightFreq	-0.0028	-0.0029	-0.0070	0.0015	0.184	0.2056
ResidLeftFamSize	-0.0378	-0.0382	-0.0472	-0.0280	0.001	0.0000
ResidRightFamSize	-0.0023	-0.0024	-0.0116	0.0062	0.624	0.6280
Trial	0.0000	0.0000	0.0000	0.0000	0.004	0.0026
RT1	0.0748	0.0747	0.0529	0.0988	0.001	0.0000
AffixMedial	-0.0472	-0.0468	-0.0831	-0.0151	0.008	0.0084
AffixFinal	-0.0216	-0.0218	-0.0588	0.0138	0.266	0.2556
AffixMultAffix	-0.0805	-0.0808	-0.1153	-0.0472	0.001	0.0000
WordFreq:LeftFreq	0.0052	0.0053	0.0017	0.0085	0.001	0.0019

Table 7: Model for Subgaze Duration for the Right Constituent

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.6285	5.6278	5.3264	5.9335	0.001	0.0000
WordLength	0.0289	0.0289	0.0139	0.0408	0.001	0.0000
LeftLength	-0.1105	-0.1105	-0.1211	-0.0994	0.001	0.0000
WordFreq	-0.0340	-0.0339	-0.0444	-0.0244	0.001	0.0000
LeftFreq	-0.0016	-0.0016	-0.0068	0.0037	0.550	0.5612
RightFreq	-0.0103	-0.0103	-0.0167	-0.0043	0.001	0.0009
ResidLeftFamSize	-0.0154	-0.0153	-0.0296	-0.0027	0.028	0.0292
ResidRightFamSize	-0.0188	-0.0188	-0.0317	-0.0074	0.001	0.0052
Trial	0.0000	0.0000	0.0000	0.0000	0.040	0.0528
RT1	0.1029	0.1032	0.0676	0.1364	0.001	0.0000
AffixMedial	-0.0598	-0.0594	-0.1149	-0.0120	0.020	0.0210
AffixFinal	-0.1022	-0.1016	-0.1513	-0.0529	0.001	0.0000
AffixMultAffix	-0.0005	-0.0010	-0.0520	0.0486	0.966	0.9852

Table 8: Gaze Duration Model

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.6415	5.6385	5.4218	5.8684	0.001	0.0000
WordLength	0.0173	0.0173	0.0032	0.0319	0.012	0.0174
LeftLength	-0.0173	-0.0173	-0.0291	-0.0061	0.002	0.0029
WordFreq	-0.0912	-0.0908	-0.1194	-0.0621	0.001	0.0000
LeftFreq	-0.0253	-0.0253	-0.0354	-0.0142	0.001	0.0000
RightFreq	-0.0080	-0.0081	-0.0132	-0.0026	0.008	0.0054
ResidLeftFamSize	0.0073	0.0075	-0.0291	0.0446	0.672	0.6975
ResidRightFamSize	-0.0070	-0.0072	-0.0177	0.0025	0.164	0.2102
Trial	0.0000	0.0000	-0.0001	0.0000	0.001	0.0000
RT1	0.1506	0.1509	0.1273	0.1752	0.001	0.0000
AffixMedial	-0.0812	-0.0798	-0.1201	-0.0400	0.001	0.0001
AffixFinal	-0.1001	-0.0985	-0.1413	-0.0585	0.001	0.0000
AffixMultAffix	-0.0834	-0.0826	-0.1250	-0.0470	0.001	0.0001
FinTrigram	-0.0070	-0.0071	-0.0124	-0.0015	0.012	0.0185
RightFreq:ResidLeftFamSize	-0.0068	-0.0068	-0.0119	-0.0018	0.008	0.0087
WordFreq:LeftFreq	0.0053	0.0053	0.0016	0.0091	0.008	0.0055

Table 9: Random effects for *RT*, *FirstDur*, *SubgazeLeft*, *SubgazeRight* and *GazeDur*

A. Lexical decision latency				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.095	0.095	0.090	0.101
Subject	0.151	0.155	0.110	0.215
Residual	0.241			
B. First fixation duration				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.049	0.050	0.042	0.057
Subject	0.42	0.415	0.286	0.608
Subject by Nomore	0.099	0.098	0.070	0.148
Subject by WordLength	0.035	0.035	0.024	0.051
Residual	0.289			
C. Subgaze duration for the left constituent				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.088	0.087	0.078	0.096
Subject	0.114	0.12	0.087	0.167
Residual	0.335			
D. Subgaze duration for the right constituent				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.010	0.097	0.075	0.116
Subject	0.107	0.110	0.079	0.158
Residual	0.456			
E. Gaze duration				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.014	0.122	0.113	0.132
Subject by LeftLength	0.015	0.014	0.008	0.023
Subject by WordLength	0.017	0.018	0.012	0.025
Subject	0.082	0.022	0.001	0.172
Residual	0.386			

Table 10: Parameters, estimates of the parameters, and power (for $\alpha = 0.05$) for the models without ($M_{k=j}$) and with ($M_{k\neq j}$) random orders of items for each participant. Averages over 1000 simulation runs.

predictor	parameter	value	$M_{k=j}$		$M_{k\neq j}$	
			estimate	power	estimate	power
Intercept	β_0	5.82	5.8074	1.00	5.8084	1.000
Trial	β_1	-0.0048	-0.0010	1.00	-0.0010	1.00
WordFreq	β_2	-0.0043	-0.0045	0.35	-0.0045	0.35
WordLength	β_3	0.0013	0.0010	0.05	0.0010	0.05
LeftFamSize	β_4	-0.0084	-0.0082	0.27	-0.0082	0.27
LeftFreq	β_5	-0.0026	-0.0027	0.21	-0.0027	0.22
RightFreq	β_6	-0.0072	-0.0072	0.88	-0.0072	0.88
Subj	σ_S	0.32	0.3100		0.3100	
Item	σ_W	0.20	0.1999		0.1999	
Resid	σ	0.60	0.5996		0.5996	