Morphological Dynamics in Compound Processing

Victor Kuperman*

Radboud University Nijmegen, The Netherlands

Raymond Bertram

University of Turku, Finland

R. Harald Baayen

University of Alberta, Canada

February 1, 2008

Running Head: Morphological Dynamics

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

Abstract

This paper explores the time-course of morphological processing of trimorphemic Finnish compounds. We find evidence for the parallel access to full-forms and morphological constituents diagnosed by the early effects of compound frequency, as well as early effects of left constituent frequency and family size. We further observe an interaction between compound frequency and both the left and the right constituent family sizes, which implies that full-form access and decompositional access are not independent. Furthermore, our data show that suffixes embedded in the derived left constituent of a compound are efficiently used for establishing the boundary between compounds' constituents. The success of segmentation of a compound is demonstrably modulated by the affixal salience of the embedded suffixes. We discuss implications of these findings for current models of morphological processing and propose a new model that views morphemes, combinations of morphemes and morphological paradigms as probabilistic sources of information that are interactively used in recognition of complex words.

Keywords: morphological structure; lexical processing; eye movements; segmentation cues; models

Current models of morphological processing vary widely in their assumptions about what morphological information is used, and in what order, to identify and interpret complex words, for instance *dish+wash-er* or *happi-ness*. For instance, sublexical and supralexical models advocate obligatory sequentiality: The former class of models posits that full-forms can only be accessed via morphological constituents (e.g., Taft & Forster, 1975; Taft, 1979; Taft, 1991), while the latter class claims that the activation of the full-form precedes the activation of constituents (e.g., Giraudo & Grainger, 2001). Some parallel dual-route models allow for simultaneous activation of both the full-forms of complex words and their morphological constituents, but assume that the two routes proceed independently of each other (e.g., Schreuder & Baayen, 1995; Baayen & Schreuder, 1999). The computational model MATCHEK (Baayen & Schreuder, 2000) implements the interaction between the two processing routes, but is silent about the time-course of visual information uptake, and assumes that all words are read with a single fixation. The present eye-tracking study adresses the temporal unfolding of visual recognition of trimorphemic Finnish compounds, in order to establish whether the requirements posed by current models (e.g., obligatory sequentiality or independence of processing stages) hold for reading of long words. We present evidence that more sources of morphological information are at work and interacting with each other in compound processing than previously reported.

The central research issue that this paper addresses is the hotly debated topic of the time-course of morphological effects in recognition of long compounds. It is a robust finding that full-form representations of compounds are involved in compound processing, as indicated by the effect of compound frequency (e.g., De Jong, Feldman, Schreuder, Pastizzo & Baayen, 2002; Hyönä & Olson, 1995; Van Jaarsveld & Rattink, 1988). The question that remains open, however, is how early this involvement shows up. Several studies of English and Finnish compounds found a weak non-significant effect of compound frequency as early as the first fixation on the compound (cf., Andrews, Miller & Rayner, 2004; Bertram & Hyönä, 2003; Pollatsek, Hyönä & Bertram, 2000). The presence or absence of compound frequency effects at the earliest stages of word identification may inform us about the order of activation of the full-forms of compounds and their morphological constituents. Specifically, an early effect of compound frequency may be problematic for obligatory decompositional models.

The role of constituents in compound processing is also controversial. Taft and Forster (1976) claimed that the left constituent of a compound serves as the point of access to the meaning of the compound, while Juhasz, Starr, Inhoff and Placke (2003) argued for the primacy of the right constituent see also Duñabeitia, Perea & Carreiras, 2007). Several studies of Finnish compounds established the involvement of both the left and the right constituent in reading of compounds (cf., e.g., Hyönä & Pollatsek, 1998; Pollatsek *et al.*, 2000). Moreover, Bertram and Hyönä (2003) argued on the grounds of visual acuity that the longer the compound, the more prominent the role of its morphological structure becomes.

An eye-tracking visual lexical decision study of 8-12 character-long isolated Dutch compounds by Kuperman, Schreuder, Bertram and Baayen (2007) established a significant effect of compound frequency emerging as early as the first fixation. Given the length of target words and constraints of visual acuity, the compound frequency effect at the first fixation is likely to precede the identification of all characters of the compound. This is supported by the fact that most compounds in their study elicited more than one fixation. The authors suggest that readers aim at identifying the compound on the basis of partial information obtained during the first fixation (e.g., initial characters, compound length and possibly an identified left constituent, see also the General Discussion). They also observed an interaction between compound frequency and left constituent frequency, suggesting that access to the full-form of a compound and access to its morphological constituents are not independent, contrary to the assumptions of e.g., Schreuder and Baayen (1995). Furthermore, they reported effects of frequency and family size for both the left and the right constituents of the compound[1].

Kuperman *et al.* (2007) explained their findings within the conceptual framework of maximization of opportunity (Libben, 2006). This framework argues that readers simultaneously use, as opportunities for compound recognition, multiple sources of information (as soon as those are available to them), and multiple processing mechanisms that they have at their disposal, including full-form retrieval from the mental storage and on-line computation. Kuperman *et al.* (2007) propose that an adequate model of compound processing needs to meet at least the following four requirements: (i) explicit consideration of the temporal order of information uptake, (ii) absence of strict sequentiality in the processing of information,

[1] The left (right) morphological family of a compound is the set of compounds that share the left (right) constituent with that compound (e.g., the left constituent family of *bankroll* includes *bankbill*, *bank holiday*, *bank draft*, etc.). The size of such family is the number of its members, while the family frequency is the cumulative frequency of family members.

i.e., simultaneous processing of information at different levels in representational hierarchies; (iii) the possibility for one processing cue to modulate the presence and strength of other cues; and (iv) fast activation of constituent families, along with activation of constituents and full-forms.

The present study explores the role of morphological structure in compound processing in a way that differs from the experiment with Dutch compounds by Kuperman *et al.* (2007) in several crucial respects. We use a different experimental technique (reading of compounds in sentential contexts, no lexical decisions on compounds presented in isolation), a different language (Finnish) and a different range of word lengths (10-18 characters, mean 15). We specifically address the following questions. Does the pattern of results obtained with the visual lexical decision paradigm generalize to a more natural task of sentential reading with words in normal context? Will compound frequency have an early effect in longer words, where more characters fall outside of the foveal area with high visual acuity? Will morphological families show the same facilitation in reading as they show in lexical decision? The effect of constituent family size may differ across tasks, since a more ”word-like” target with a large family may facilitate a positive lexical decision. In normal reading, however, the members of the family might function as competitors and hamper the integration of the word in the sentence, which would show as inhibition in the eye movement record. Finally, is there evidence in the eye movement record that different routes of lexical processing interact, when compounds are placed in sentential contexts? Another task that we set for ourselves is to formalize the specifications for a model of morphological processing outlined in Kuperman *et al.* (2007). We propose such a model in the General Discussion.

Additionally, we consider the processing of compounds with more than two morphemes. Current research on visual processing of morphologically complex words is largely constrained to bimorphemic words (for exceptions see e.g., De Almeida & Libben, 2005; Inhoff, Radach & Heller, 2000; Krott, Baayen & Schreuder, 2001; Krott, Libben, Jarema *et al.*, 2004; Kuperman *et al.*, 2007). At the same time, such complexity is anything but rare in many languages: In German, Dutch and Finnish words with three or more morphemes account for over 50% of word types. Similarly, words in the length range of 10-18 characters that we use in this study account for over 60% of word types and over 20% word tokens in Finnish. In the present experiment, we zoomed in on one type of morphological structure, where the left constituent is a derived word with a suffix and the right constituent is a simplex noun (e.g., *kirja-sto/kortti* "library card", where *kirja* is "book", *kirjasto* is "library" and *kortti* is "card").

We took into consideration two suffixes: the suffix -*stO*[2], which attaches to nouns forming collective nouns (e.g., *kirja*, "book", and *kirjasto*, "library"), and the suffix -*Us*, which attaches to verbs and forms nouns with the meaning of the act or the result of the verb (analogous to the English -*ing*, e.g., *aloittaa* "to begin" and *aloitus* "beginning"), cf., Järvikivi, Bertram and Niemi (2006). Bertram, Laine and Karvinen (1999) and Järvikivi *et al.* (2006) argue that these two suffixes differ in their affixal salience, defined as the likelihood of serv-

[2] The capital characters in suffixes refer to the archiphoneme of the vowel that has back and front allophones. Realization of Finnish suffixes alternates due to the vowel harmony with the vowels in the stem, e.g., -*stO* may be realized either as /sto/ or /stœ/, and -*Us* either as /us/ or /ys/.

ing as a processing unit in identification of the embedding complex form (cf., Laudanna & Burani, 1995). The suffix *-stO* is arguably more salient and less ambiguous than the suffix *-Us*. Järvikivi *et al.* (2006) attribute this difference in salience to the fact that the suffix *-stO* has no allomorphs (i.e., is structurally invariant across inflectional paradigms), nor homonyms. Conversely, the suffix *-Us* has a very rich allomorphic paradigm (cf., several inflectional variants of *räjähdys* "explosion": *-ysken, -yksien, -ysten, -ystä, -yksiä, -yksenä*, Table 2 in Järvikivi *et al.*, 2006) and is homonymous with the deadjectival suffix *-(U)Us*.

The difference in affixal salience has demonstrable consequences for the processing of derived words. In particular, Järvikivi *et al.* (2006) showed in a series of lexical decision experiments that Finnish derived words ending in relatively salient affixes, like *-stO*, show facilitatory effects of both the surface frequency of the derived form (e.g., *kirjasto*) and the base frequency of its stem (e.g., *kirja*). At the same time, complex words that carry less salient affixes, like *-Us*, show facilitation only for surface frequency. In other words, salient affixes tend to shift the balance towards decomposition of complex words into morphemes and towards subsequent computation of a word's meaning from these constituent morphemes (e.g., Baayen, 1994; Bertram, Schreuder & Baayen, 2000; Järvikivi *et al.*, 2006; Laudanna & Burani, 1995; Sereno & Jongman, 1997).

Crucially, in bimorphemic derivations, one of the affix boundaries is explicitly marked by a space, which makes easier the task of parsing morphemes out of the embedding word. Our goal was to determine the role of affixal salience for suffixes orthographically and morphologically embedded in larger words. We envisioned several possible states of affairs. First, the suffix may, depending on its salience, facilitate activation of the base of the derived left constituent of the compound (i.e., *kirja* "book" in *kirjastokortti* "library card"), as shown for bimorphemic derivations by Järvikivi *et al.* (2006). On this account, one expects an interaction of base frequency by suffix type. Specifically, compounds with a relatively salient suffix -*stO* would show effects of both the base and the surface frequency of the left immediate constituent, while for the less salient suffix -*Us*, we expect to only witness the effects of left constituent surface frequency, in line with findings by Järvikivi *et al.* (2006). Second, the suffix demarcates the boundary between the two immediate constituents of the compound (i.e., *kirjasto* "library" and *kortti* "card" in *kirjastokortti*). If so, it is plausible that a more salient affix serves as a better segmentation cue and facilitates decomposition of a compound into its major constituents (for the discussion of segmentation cues in compound processing, see e.g., Bertram, Pollatsek & Hyönä, 2004). The finding expected on this account is the interaction between characteristics of the compound's constituents and the suffix type. For instance, we would expect the effects of left constituent frequency or family size to interact with the salience of our suffixes. Third, suffixes might pave the way for both parsings (*kirja* in *kirjastokortti* and *kirjasto* in *kirjastokortti*), as they may demarcate both the boundary of the base in the derived left constituent and the boundary between the compound's major constituents. If this is the case, we would expect the frequencies (or other morphological characteristics) of both the base and the full-form of the left constituent to interact with the suffix type.

As the time-course of morphological effects is essential for this study, we opted for using the eye-tracking experimental paradigm, which allows for a good temporal resolution of cognitive processes as reflected in eye movements. Furthermore, multiple regression mixed-effects modeling with participants and items as crossed random effects satisfied our need to explore simultaneously many predictors, both factors and covariates, while accounting for between-participants and between-items variance (cf., Baayen, Davidson & Bates, 2007; Bates & Sarkar, 2005; Pinheiro & Bates, 2000).

Method

Participants

Twenty-seven students of the University of Turku (18 females and 9 males) participated in this experiment for partial course credit. All were native speakers of Finnish and had normal or corrected-to-normal vision.

Apparatus

Eye movements were recorded with an EyeLink II eye-tracker manufactured by SR Research Ltd. (Canada). The eyetracker is an infrared video-based tracking system combined with hyperacuity image processing. The eye movement cameras are mounted on a headband (one camera for each eye), but the recording was monocular (right eye) and in the pupil-only mode. There are also two infrared LEDs for illuminating the eye. The headband weighs 450 g in total. The cameras sample pupil location and pupil size at the rate of 250 Hz. Recording is performed by placing the camera and the two infrared light sources 4-6 cm away from the eye. Head position with respect to the computer screen is tracked with the help of a head-tracking camera mounted on the center of the headband at the level of the forehead. Four LEDs are attached to the corners of the computer screen, which are viewed by the head-tracking camera, once the participant sits directly facing the screen. Possible head motion is detected as movements of the four LEDs and is compensated for on-line from the eye position records. The average gaze position error of EYELINK II is $<0.5^o$, while its resolution is 0.01^o. The stimuli were presented on a 21 inch ViewSonic computer screen, which had a refresh rate of 150 Hz.

Stimuli

The set of target words included 50 noun-noun compounds with the derivational first constituent ending in the suffix *-stO* (e.g., *tykistötuli* "cannon fire"), 50 noun-noun compounds with the derivational first constituent ending in the suffix *-Us* (e.g., *hitsaustyö* "a piece of welding"), and 50 bimorphemic compounds with two noun stems (e.g., *palkkasotilas* "a soldier of fortune"). All target words were selected from an unpublished Finnish newspaper corpus of 22.7 million word forms with the help of the WordMill database program (Laine & Virtanen, 1999). Each target word in the nominative case was embedded in a separate sentence, and it never occupied the sentence-initial or sentence-final position. All critical sentences had semantically neutral initial parts up to the target word. In a separate rating task, we asked five participants (none of whom participated in the eye-tracking experiment) to rate how felicitous the target words (e.g., *perhetapahtuma* "family happening") were given the preceding context (*Iloinen ja jännittävä...* "The happy and exciting ...") using a scale from 1 (does not fit at all) to 5 (fits very well). The task included all target sentences from the eye-tracking experiment, as well as fillers. The mean rating for target words was 3.7, which shows that the target words were in general a good continuation of the preceding context. Compound-specific ratings were not significant predictors of reading times in our statistical models. Averages per suffix type were 3.8, 3.7 and 3.6 for bimorphemic compounds, compounds with -*stO* and compounds with -*Us*, respectively. Pairwise t-tests showed no difference in ratings between the different compound types.

Eighty filler sentences were added to the 150 target sentences. All sentences comprised 5-12 words and took up at most one line. The sentences were displayed one at a time starting at the central-left position on the computer screen. Stimuli were presented in fixed-width font Courier New size 12. With a viewing distance of about 65 cm, one character space subtended approximately 0.45^o of visual angle.

Sentences were presented in two blocks, while the order of sentences within the blocks was pseudo-randomized and the order of blocks was counterbalanced across participants. Approximately 14% of sentences were followed by a screen with a yes-no question pertaining to the content of the sentence. The experiment began with a practice session consisting of five filler sentences and two questions.

Procedure

Prior to the presentation of the stimuli, the eye-tracker was calibrated using a three-point grid that extended over the horizontal axis in the middle of the computer screen. Prior to each stimulus, correction of calibration was performed by displaying a fixation point in the central-left position. After calibration, a sentence was presented to the right of the fixation point.

Participants were instructed to read sentences for comprehension at their own pace and to press a "response" button on the button box. Upon presentation of a question, participants pressed either the "yes"-button or the "no"-button on the button box. If no response was registered after 3000 ms, the stimulus was removed from the screen and the next trial was initiated. Responses and response times of participants were recorded along with their eye movements. The experimental session lasted 50 minutes at most.

Dependent variables

In the analysis of the eye-tracking data, we considered as measures of early lexical processing the duration of the first fixation (*FirstDur*), as well as the subgaze duration for the left constituent of a compound (the summed duration of all fixations that landed on the left constituent of a compound before fixating away from that constituent, *SubgazeLeft*. As a measure of later lexical processing, we focused on the subgaze duration for the right constituent of a compound (the summed duration of all fixations that landed on the right constituent of a compound before fixating away from that constituent, *SubgazeRight*. As a global measure, we considered the gaze duration on the whole word (the summed duration of all fixations on the target word before fixating away from it, *GazeDur*). We obtained additional information from two other measures: the probability of a single fixation (*SingleFix*) and - in order to assess how smoothly compound processing proceeded - the probability of the second fixation landing to the left of the first fixation position (*Regress*)[3]. All durational

[3] Other considered dependent measures included the total number of fixations, durations of the second and third fixation, amplitude of the first and second within-word saccades, and the probability of eliciting more than two fixations. The measures did not provide additional insight into our research questions.

measures were log-transformed to reduce the influence of atypical outliers.

Predictors

Trials were uniquely identified by the participant code (*Subject*) and item (*Word*). The type of affix used in the target words was coded by the factor *SuffixType* with values "stO", "Us" and "none" (for bimorphemic compounds).

Lexical distributional properties of morphological structure. We considered compound lemma frequency, *WordFreq*, while lemma frequency was defined as the summed frequency of all inflectional variants of a word (e.g., the lemma frequency of *cat* is the sum of the frequencies of *cat*, *cats*, *cat's* and *cats'*). As frequencies of compounds' constituents have been shown to codetermine the reading times along with compound frequency (e.g., Andrews *et al.*, 2004; Hyönä & Pollatsek, 1998; Juhasz *et al.*, 2003), we included lemma frequencies of the compound's left and right constituents as isolated words, *LeftFreq* and *RightFreq*. Additionally, for each derivational left constituent (e.g., *kirjasto* "library" in *kirjastokortti* "library card") we included the lemma frequency of its base word (e.g., *kirja* "book"), *BaseFreq*, as a predictor. All frequency-based measures in this study, including the ones reported in the remainder of this section, were (natural) log-transformed to reduce the influence of outliers.

The morphological family sizes and family frequencies of a compound's constituents are known to codetermine the processing of compounds (cf., e.g., De Jong, Schreuder & Baayen, 2000; Juhasz *et al.*, 2003; Krott & Nicoladis, 2005; Kuperman *et al.*, 2007; Moscoso del Prado Martín, Bertram, Haikio *et al.*, 2004; Nicoladis & Krott, 2007; Pollatsek & Hyönä, 2005).

The larger the number of members in such a family or the larger their cumulative frequency, the faster the identification of the constituent and the embedding compound proceeds, as shown in lexical decision and eye-tracking studies. Since Moscoso del Prado Martín *et al.* (2004) have shown that it is only the subset of words directly derived from the complex word itself that codetermines the speed of lexical processing in Finnish morphological families, we restricted our families to compounds derived from the target compound. To give an example in English, we would consider *vanilla cream* and *shoe cream* as members of the right constituent family of *ice cream*, but not, say, *chocolate ice cream.* We collected counts of the family members for the left and the right constituent families (i.e., constituent family sizes) for our compounds, *LeftFamSize* and *RightFamSize*, where families were defined over compounds and did not include derived words. The related measure, the family frequency of the left (right) constituent, failed to reach statistical significance in our models (even when the respective family size was not included in the models) and will not be further discussed.

Other variables.

To reduce variance in our models, we controlled for several variables that are known to modulate visual processing. Among many other predictors (see Appendix for the full list), we considered compound length (*WordLength*) and the length of the left constituent *LeftLength.* We also included as a predictor the position of trial N in the experimental list as a measure of how far the participant has progressed into the experiment. This measure, *TrialNum*, allows us to bring under statistical control longitudinal task effects such as fatigue or habituation.

Statistical considerations

Several of our measures showed strong pair-wise correlations. Orthogonalization of such variables is crucial for the accuracy of predictions of multiple regression models. Teasing collinear variables apart is also advisable for analytical clarity, as it affords better assessment of the independent contributions of predictors to the model's estimate of the dependent variable (see Baayen, 2008: 198). We orthogonalized every pair of variables for which the Pearson correlation index r exceeded the threshold of 0.5. Decorrelation was achieved by fitting a regression model in which one of the variables in the correlated pair, e.g., *LeftLength*, was predicted by the other variable, e.g., *WordLength*. We considered the residuals of this model, *ResidLeftLength*, as an approximation of the left constituent length, from which the effects of compound length were partialled out. Using the same procedure, we obtained *ResidLeftFreq* (orthogonalized with *WordFreq* and *LeftLength*), *ResidLeftFamSize* (orthogonalized with *LeftFreq*), *ResidBaseFreq* (orthogonalized with *LeftFreq*), and *ResidRightFamSize* (orthogonalized with *RightFreq*). All orthogonalized measures were very strongly correlated with the measures, from which they were derived (rs > 0.9, p < 0.0001). The collinearity between the resulting set of numerical predictors was low, as indicated by $\kappa = 1.44$.

Additionally, some of the predictors were centered, so that the mean of their distribution was equal to zero. This procedure is crucial to avoid spurious correlations between random slopes and random intercepts in mixed-effects regression models (cf., Baayen, 2008: 276).

Table 3 in the Appendix lists the distributions of the continuous variables used in this study, including statistics on their original values and (if different from the original values) the values actually used in the models.

In this study we made use of mixed-effects multiple regression models with *Subject* and *Word* as random effects. For predicting binary variables (e.g., indicators of whether the given fixation is word-final or regressive), we used generalized mixed-effects multiple regression models with a logistic link function and binomial variance. We coded the "Yes" values as successes and "No" values as failures.

The distribution of durational dependent measures was skewed even after the log transformation of durations. Likewise, residuals of the mixed-effects models for durations were almost always skewed. To reduce skewness, we removed outliers from the respective datasets, i.e., points that fell outside the range of -2.5 to to 2.5 units of SD of the residual error of the model. Once outliers were removed, the models were refitted, and we reported statistics for these trimmed models. Unless noted otherwise, only those fixed effects are presented below that reached significance at the 5%-level in a backwards stepwise model selection procedure.

The random effects included in our models significantly improved the explanatory value of those models. Improvement was indicated by the 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).

Results and Discussion

The initial pool of data points comprised 13394 fixations. We log-transformed the fixation durations and removed from the dataset for each participant those fixations that exceeded

3.0 units of SD from that participant's mean log-transformed duration. The number of removed fixations was 397 (3%), and the resulting range of fixation durations was 60 to 892 ms. Subsequently, fixations that bordered microsaccades (fixations falling within the same letter) were removed (44 x 2 = 88 fixations, 0.6%). Finally, we only considered the fixations pertaining to the first-pass reading (i.e., the sequence of fixations made before the fixation is made outside of the word boundaries, 67% of the original dataset). As a result, we were left with a pool of 9023 valid fixations.

A negligible percent of the target words was skipped ($< 0.01\%$). Twenty-seven percent of the target words required only one fixation, 40% required exactly two fixations, 20% required exactly three fixations, and it took four or more fixations to read the remaining 13% of our compounds. The average number of fixations on a stimulus was 2.2 ($SD = 1.2$). Regressive fixations (i.e., fixations located to the left of the previous fixation within same word) constituted 14.2% of our data pool. The average fixation duration was 234 ms ($SD = 84$), and the average gaze duration was 455 ms ($SD = 263$).

We report in the Appendix full specifications of the models for the first fixation duration (3967 datapoints, Table 4), subgaze duration for the left constituent (3800 data points, Table 5), subgaze duration for the right constituent (2342 data points, Table 6), and gaze duration (3884 data points, Table 7).

Time-course of morphological effects

Table 1 summarizes effects of morphological predictors on reading of long, multiply complex Finnish compounds across statistical models for early and cumulative measures (see full specifications for the models in Appendix). The table provides effect sizes (see Appendix for the explanation as to how these were computed) and p-values for main effects, as well as indicates interactions between morphological and other predictors of interest. For clarity of exposition, we leave out in this section interactions between morphological predictors and the type of the suffix in the compound's left constituents: These interactions are presented in detail in the next section.

INSERT TABLE 1 HERE

Results presented in Table 1 reveal the temporal pattern of how effects of morphological structure unfold in complex word recognition. First, characteristics pertaining to the compound's left constituent, such as left constituent frequency and family size, show effects in both the early measures of reading times (first fixation duration, subgaze duration on the left constituent), and in the later measure (subgaze duration of the right constituent). Conversely, characteristics of the compound's right constituent are not significant predictors at early stages of lexical processing and only yield significant effects (always modulated by interactions with other predictors) in the measures of right constituent subgaze duration and gaze duration. This sequence of effects corroborates previous findings that both constituents are activated during processing of compounds (cf., Hyönä, Bertram & Pollatsek, 2004). Moreover, the order of their activation goes hand in hand with the typical sequence of the visual uptake in long compounds that was observed previously in Hyönä *et al.* (2004), Kuperman *et al.* (2007) and again in the present study, such that the first fixation tends to land on a compound's left constituent and the second fixation on its right constituent[4]. We also note that the influence of the frequency-based characteristics of the left constituent on the lexical processing of compounds is qualitatively stronger than the corresponding measures for the right constituent. Left constituent frequency and family size show main effects in the models for fixation durations and subgaze and gaze durations, whereas effects of the right constituent frequency and family size are qualified by the interaction with compound length and compound frequency, respectively. The dominant involvement of the left constituent in compound processing is in line with the findings of Taft and Forster (1976). It is at odds with the important role of the right constituent as the access code to the compound's meaning proposed by Juhasz *et al.* (2003).

Second, we observed effects of constituents' morphological families emerging simultaneously with the effects of the respective constituent frequencies. The early effect of the left constituent family size goes against the traditional interpretation, which holds that the semantic family size effect arises due to post-access spreading activation in the morphological family (cf., De Jong *et al.*, 2002). Surprisingly, the right constituent family (e.g., *vanilla cream*, *ice cream*, *shoe cream*) is activated even when the lexical processor might have begun identification of one member of that family (e.g., *vanilla cream*), the target compound itself

[4] The size of perceptual span in reading (3-4 characters to the left and 10-15 characters to the right of the fixation position, see e.g., Rayner, 1998) suggests that at least some characters from the compound's right constituent are very likely to be identified either foveally or parafoveally. The absence of early effects stemming from the compound's right constituent implies, however, that the available orthographic information is apparently not sufficient for early activation of that morpheme (cf., Hyönä *et al.*, 2004).

(the left constituent of which was processed at the preceding fixation). It may be that this effect is driven by the cases in which a compound's left constituent is particularly difficult to recognize (e.g., due to its lexical properties or non-optimal foveal view). In such cases identification of the left constituent may not be complete at the first fixation and may continue even as the eyes move to the right constituent. It may also be that activation of morphological families is automatic and happens even when not fully warranted by the processing demands: This is an empirical question that requires further investigation. More generally, we argue in the General Discussion that characteristics of the compound's right constituent may provide a valuable source of information that facilitates recognition of a complex word and its constituents, even when other such constituents have received sufficient activation and produced detectable effects on reading times.

Third, higher compound frequency came with a benefit in speed that was present as early as the first fixation, and extended over late measures of reading times. Given the lengths of our compounds (10-18 characters), it is very likely that not all the characters of the compounds are identified at the first fixation. In fact, for nearly three quarters of our compounds, visual uptake is not completed at the first fixation. Importantly, the effect of compound frequency on fixation duration is still present when single-fixation cases are removed from the statistical model. We outline possible reasons for the very early and lingering effect of compound frequency in the General Discussion.

Fourth, the effect of compound frequency on cumulative reading times was weaker in compounds that had constituents with large families. In the compounds with very large left or right constituent families the effect of compound frequency vanished (see Figs. 1 and 2).

INSERT FIGURES 1 and 2 HERE

The interactions of characteristics traditionally associated with the full-form representation (i.e., compound frequency) and characteristics of morphemes that imply decomposition (i.e., constituent family sizes) provides evidence against race models in which full-form access and morpheme-based access are presented as strictly independent (cf., Schreuder & Baayen, 1995). Additionally, we observe that higher right constituent frequency correlated with shorter *SubgazeRight*, and this effect was stronger in longer compounds. This implies that the strength of morphological effects can also be modulated by visual characteristics of the word, in line with the earlier report of Bertram & Hyönä (2003).

Differences across types of compounds

Recall that our data comprised three types of compounds: compounds with the left constituent ending in the relatively salient affix -*stO*, compounds with the left constituent ending in the less salient affix -*Us*, and bimorphemic compounds with two simplex constituents. *SuffixType* did not reveal a simple main effect in our statistical models, but it qualified the effects of several morphological predictors, summarized in Table 2 across several statistical models. Table 2 provides a comparative overview of morphological effects across suffix types, including effect sizes and associated p-values per suffix, as well as p-values for interactions.

INSERT TABLE 2 HERE

Measures of the early visual uptake (probability of a single fixation and probability of the regressive second fixation) suggest that bimorphemic compounds and especially compounds with the suffix -*Us* come with a higher processing load (i.e., require more fixations and elicit more regressive fixations) than words with the salient suffix -*stO*, which benefit most from the properties of the left constituent (i.e., require fewer fixations).

The cumulative measures of reading times demonstrate a straighforward pattern: Compounds with left constituents ending in the suffix -*stO* show much stronger effects of the left constituent frequency and family size than bimorphemic compounds and especially than compounds with the suffix -*Us*. We view this difference as evidence that this relatively salient suffix acts as a better segmentation cue for parsing out a compound's constituents than the suffix -*Us* with its many allomorphs, or the constituent boundary in bimorphemic compounds. Earlier identification of the left constituent ending in -*stO* may lead to easier recognition of that constituent and to earlier and larger effects of distributional characteristics pertaining to that constituent.

Surprisingly, bimorphemic compounds demonstrated stronger effects of the left constituent than compounds with the suffix -*Us* did. The three types of compounds can be ordered by the relative ease of processing (and, we argue, by the salience of their segmentation cues) as follows: (i) compounds with the suffix -*stO*, (ii) bimorphemic compounds and (iii) compounds with the suffix -*Us*. This finding is counterintuitive given that the bigram ”Us” has a very high frequency of occurrence and a high productivity as a suffix in Finnish (see Table 1 in Järvikivi *et al.*, 2006). It represents the nominative case of two suffixes with high-frequency and high-productivity, deadjectival -*Us*, which we focus on in this study, and a homonymous deverbal -*(U)Us* (cf., Järvikivi *et al.*, 2006). That is, the character string

"Us" would be a likely candidate for serving as a suffix and thus would be expected to perform as a better segmentation cue than the n-gram at the constituent boundary of a bimorphemic compound (we note that the frequency of a bigram straddling the constituent boundary was not a significant predictor in any of our models).

One explanation for this finding is offered by Järvikivi *et al.* (2006) who argue that the identification of the suffix -*Us*, and subsequent parsing of the derived word, is impeded by the rich allomorphic paradigm that comes with that suffix. The two-level version of the dual-route model (Allen & Badecker, 2002) would predict that activation of competing allomorphic variants takes place as soon as access is attempted to any of the variants due to the lateral links between the different allomorphs. The early allomorphic competition for a structurally variant suffix may explain the worse performance of the suffix -*Us* as a segmentation cue in comparison to bimorphemic words, which indeed is noticeable from the first fixation onwards.

Another dimension of salience that differs across our suffixes is homonymy. The deverbal suffix -*Us* (analogous to the English -*ing*) is homonymous with the highly frequent deadjectival suffix -*(U)Us* (analogous to the English -*ness*), while the suffix -*stO* has no homonyms. Bertram, Laine and Kalvinen (1999) and Bertram, Schreuder and Baayen (2000) found that the presence of homonymy may create ambiguity as to the semantic/syntactic role that the suffix performs in the given word (in our case, the left constituent of a compound). Resolving this ambiguity might then come with slower processing of the homonymous suffix. This is unlikely to happen in our case, though, since the homonymous suffixes -*Us* and -*(U)Us* are very close in their meaning and syntactic function (cf., Järvikivi *et al.*, 2006).

A more important factor may be that the phonotactic rules of Finnish are such that the trigram "stO" only occurs in a word-initial position in a small number of borrowed words (26 word types, e.g., *stockman*). Thus, when embedded in complex words, this trigram serves as a clear cue of the constituent boundary, since it is much more probable to occur at the end of the left consituent than in the beginning of the right one. On the other hand, a substantial number of Finnish words begin with the bigram "Us" (509 word types, including highly frequent words like *ystävä* "friend" or *uskoa* "to believe"). The high positional probability of the bigram "Us" at the word's beginning may pave the way for misparsings that attribute the suffix -*Us* to the final constituent, rather than to the initial constituent in which the suffix is actually embedded. Due to a higher likelihood of misparsings, the suffix -*Us* would then figure as a less salient affix than its counterpart -*stO* in the situation when suffixes occupy a compound-medial position.

We find no effects of the morphological base of a compound's left constituent for any type of compound that we considered. This is at odds with the results of Järvikivi *et al.* (2006), who show significant effects of the base frequency for derivations with the relatively salient suffix -*stO*, as opposed to derivations with -*Us*. Clearly, in their data the identification of the suffix makes available two morphological sources of information, one provided by the base of the left constituent (e.g., *kirja* in *kirjastokortti*) and the other provided by the major constituent boundary between the left constituent *kirjasto* and the right constituent *kortti*. Our data only provides support for the detection of the immediate constituents. At all appearance, in trimorphemic compounds left constituent bases do not offer much information in addition to what information is carried by a compound's immediate constituents, and so the contribution of left constituent bases is too weak to be detected in our experiment.

We also report an interaction of *SuffixType* with *TrialNum*, such that the reading times for the right constituent were shorter towards the end of the experiment only for compounds including the suffix -*stO*, and not for other types of compounds ($p = 0.0015$ as estimated via the Monte Carlo Markov chain (MCMC) random-walk method using 1000 simulations). The suffix -*stO* is not too frequent in Finnish, so its presence in 22% of our stimuli sentences may have led to overrepresentation and easier recognition of this sequence of characters towards the end of the experimental list, more so than for the high-frequency suffix -*Us*. We note, however, that the covariance-analytical technique implemented in multiple regression models ensures that all other effects predicted by those models are observed over and above the impact of overrepresentation on eye movements.

Below we offer a formal, model-based view of the role that affixes structurally and orthographically embedded in compounds play in activation of other morphological constituents.

General Discussion

The key issue that we investigated in this paper is the time-course of morphological effects in the lexical processing of long, multiply complex Finnish compounds.

We found evidence for the activation of most morphological cues (i.e., morphemes, sequences of morphemes and morphological paradigms) that build up our compounds. These cues create opportunities for recognition of complex words. Moreover, there is a temporal flow of morphological information during reading of our compounds, which is roughly as follows. Typically the first fixation on a compound lands on its left immediate constituent. As early as the first fixation, we observe simultaneous effects of compound frequency, compound length, left constituent frequency and left constituent family size. The second and subsequent fixations usually land further into the word, such that the right constituent comes under foveal inspection and a new source of morphological information becomes available for recognition of compounds. Consequently, the effects of right constituent frequency and right constituent family size emerge late, and their effects are weaker than those of the left constituent. Finally, we observe interactions between compound frequency and both the left and the right constituent family sizes.

Perhaps the most intriguing of our findings is that the early effect of compound frequency apparently precedes the complete identification of all characters and of the right constituents of our long compounds. This effect suggests that readers make inferences about the compound's identity as soon as they have available any (potentially incomplete) information about the word. Information about formal compound properties, such as its initial characters or length, may be available from the parafoveal preview and from the earliest stages of foveal inspection of the word (see Rayner, Well, Pollatsek & Bertera, 1982). Readers may match the visual pattern consisting of several initial characters in combination with word length against words stored in memory long before the compound as a whole is scanned. The more frequent matches to such patterns may boost the identification of that compound. Compound frequency may also be considered as the combinatorial strength of association between the morphemes of a compound and its full-form representation. Activation of one morpheme may then lead to activation of combinations with that morpheme, which will be stronger for higher-frequency combinations. Thus, identification of the left constituent, potentially enhanced by the information about word length, may also lead to early identification of compounds that embed that constituent (for the length constraint hypothesis, see O'Regan, 1979; Clark & O'Regan, 1999; for the opposing view, see Inhoff & Eiter, 2003). We note that the effect of compound frequency lingers on throughout the entire course of reading a compound, which implies that the full-form representation of a compound keeps being actively involved in the recognition process as other morphological and orthographic cues to identification become available to the reader.

Observed effects of left and right constituent frequency, like the effect of compound frequency, may gauge both the ease of access to the morpheme in the mental lexicon, and, at the level of form, the reader's experience with identifying a character string that represents the constituent as a word pattern within a larger word. Additionally, left and right constituent family sizes may be measures of the semantic resonance following activation of a constituent, but also a measure of experience that the reader has with parsing that constituent out of compound words.

We explain qualitatively stronger effects pertaining to the compound's left constituent (as compared to those pertaining to the compound's right constituent) by the time-course of visual uptake. As a result of its later availability for the visual system, identification of a compound's right constituent may proceed against the backdrop of existing knowledge gleaned from the left constituent. Since the informational value carried by a compound's right constituent is attenuated by the information obtained earlier, the contribution of that constituent to the comprehension of a compound is smaller than the contribution of the left constituent.

We note that most of the morphological measures that we have described so far can be argued to tap both into the formal properties of a compound or its morphemes, and into their semantic representations and semantic integration of morphemes in a whole: This duality is quite in line with recent findings that morphological effects imply at least two processing stages, that of form-based decomposition and that of semantic integration (e.g., Meunier & Longtin, 2007).

The present findings show remarkable convergence with the findings in Kuperman *et al.* (2007), which included the early effect of compound frequency, early effects of left constituent frequency and family size, late effects of right constituent frequency and family size, and interactions between compound frequency and frequency-based measures of the left constituent. In other words, the findings are robust to language (Dutch vs. Finnish), the experimental task (lexical decision vs. reading), the experimental technique (single word reading vs. sentential reading), or the range of word lengths (8-12 vs. 10-18 characters). Below we discuss implications of these findings for current models of morphological processing, and propose a formal model, the PRObabilistic Model of Information SourcEs (henceforth, PROMISE) to account for the present results and results of Kuperman *et al.* (2007).

Our set of findings has far-reaching consequences for current theories of morphological processing. While eye-movements (like any other known experimental paradigm) cannot exhaustively access the time course of compound processing in absolute terms, they certainly give us insight in some crucial aspects of the processing time-flow. The fact that we are using long compounds allows for naturalistic separation of information sources into those that are available (and used) early in the processing and those that come into play only relatively late. For instance, the early effect of compound frequency is problematic for approaches that require prelexical decomposition of full-forms prior to identification of complex words (e.g., Taft, 1991; Taft, 2004). A pure decompositional model proposed for inflections and derivations assumes access to both morphological constituents before full-form representations are activated. More specifically, Taft and Ardasinski (2006) argue that in the case of inflections, full-form representations are not activated at all, while in the case of derivations, full-form representations are activated at the lemma level after activation of both constituents. Our results go against these assumptions, since we find evidence for activation of the full-form representation before the activation of the right constituent. The kind of a decompositional feed-forward model, advanced by Taft and Forster (1976) for compounds, assumes that the compound's full-form is activated by and after access to the left constituent. It does not predict any effect of the right constituent at all, contrary to our results (see also Lima & Pollatsek, 1983).

For supralexical models, there is a logical possibility that the full-form representation of the compound is activated and, in sequence, this activation spreads to the compound's left constituent, such that the effects of both the compound as a whole and its left constituent are detectable within the short duration span of the first fixation. A problem for this class of models, however, is that activation of the right constituent of a compound is predicted to be simultaneous with that of the left constituent, but we observed no effect pertaining to characteristics of right constituents in either first nor second fixation measures.

Another finding that is not easy to reconcile with several current models of morphological processing is the interaction between the characteristics of a full-form (e.g, compound frequency) and the characteristics of a compound's constituents (left and right constituent family sizes), such that compound frequency has little or no effect on the reading time for the words with large constituent families. In the "horse race" models of dual-route parallel processing, the full-form route and the decompositional route of lexical access are assumed to be autonomous and thus the strength of the compound frequency effect and the strength of the constituent family size effect are not predicted to interact. In the strictly sublexical models and in supralexical models, activation of full-forms and that of morphemes are separated in time (i.e., are not parallel), so the effects of full-forms and of those morphemes are expected to fully develop on their own. Thus, the strength of effects pertaining to the full-form representation is not supposed to modulate, or be modulated by, the influence of morphemic properties.

Our results show that the patterns of morphological effects in compound processing are not captured in their entirety by current models of morphological processing. Moreover, with the exception of Pollatsek, Reichle and Rayner (2003), models of morphological processing make no provision about the temporal unfolding of reading, as if complete identification of the word would always require a single fixation. Kuperman *et al.* (2007) suggest that theoretical assumptions such as instant access to full visual information, obligatory sequentiality or independence of processing stages need to be reconsidered in order to account for the readers' interactive use of multiple morphological cues (see Libben, 2005; Libben, 2006). In fact, most current models have been developed on the basis of experiments with relatively short compounds, i.e., those where the visual uptake is not stretched over time and the order of activation of morphemes and full-forms is difficult to establish empirically. From this perspective, it is not surprising that their predictions do not generalize to long morphologically complex words. Below we present the model of morphological processing that is based on the reading data from long words, yet it makes explicit predictions about the patterns of morphological processing expected of short complex words.

Towards a Probabilistic Model of Information Sources

We have documented a broad range of lexical distributional properties of morphological structure that codetermine the uptake of information (as gauged by durational measures in the eye-movement record). In what follows, we sketch a framework for understanding and modeling these lexical effects.

The mental lexicon is a long-term memory store for lexical information. We view an incoming visual stimulus as a key for accessing this lexical information. The information load of a stimulus is defined by the lexical information in long-term memory. Without knowledge of English, words like *work* or *cat* carry no information for the reader. It is the accumulated knowledge of words and their paradigmatic and syntagmatic properties that define a word's information load, and hence the speed with which information can be retrieved from lexical memory.

Our Probabilistic Model of Information Sources (PROMISE) takes as its point of departure the perhaps most basic insight of information theory, that information (I) can be quantified as minus log probability (P):

$$I = -\log_2 P \tag{1}$$

As P decreases, I increases: less probable events are more informative. A fundamental assumption of our model is that the time spent by the eye on a constituent or word is proportional to the total amount of lexical information available in long-term memory for identification of that constituent or word at that timepoint (cf., Moscoso del Prado Martín, Kostić & Baayen, 2004). Events with small probability and hence a large information load require more processing resources and more processing time (see Levy, 2008 for a similar probabilistic approach to processing demands in online sentence comprehension)[5].

Seven lexical probabilities are fundamental to our model. First, we have the probability

[5] While most of the measures considered below are traditionally considered as semantic (e.g., degree of compatability of constituents in a compound, degree of connectivity in a morphological paradgim, etc.), we remain agnostic in the present paper to whether information originates from the level of form or the level of meaning. In all likelihood, formal properties of words reach the lexical processing system earlier than their semantic properties. Yet, as argued in e.g., Meunier and Longtin (2007) and in the present paper, most morphological effects take place at both the level of form and that of meaning. The model is able to capture informations originating at either level as long as they can be represented numerically: as frequency measures, as the Latent Semantic Analysis scores, or as a number of members in a morphological family, of words of a given length, of synonyms, of orthographic or phonological neighbors, etc.

of the compound itself. We construe this probability as a joint probability, the probability of the juxtaposition of two constituents, μ_1 and μ_2: $\Pr(\mu_1, \mu_2)$. In what follows, subscripts refer to the position in the complex word. We estimate this probability by the relative frequency of the complex word in a large corpus with N tokens. With F_{12} denoting the absolute frequency of the complex word in this corpus, we have that

$$\Pr(\mu_1, \mu_2) = \frac{F_{12}}{N}. \tag{2}$$

This is an unconditional probability, the likelihood of guessing the complex word without further contextual information from sentence or discourse. Two further unconditional probabilities that we need to consider are the probability of the left constituent and that of the right constituent:

$$\Pr(\mu_1) = \frac{F_1}{N} \tag{3}$$

$$\Pr(\mu_2) = \frac{F_2}{N}. \tag{4}$$

The remaining four probabilities are all conditional probabilities. The first of these is the probability of the right constituent (μ_2) given that the left constituent (μ_1) has been identified: $\Pr(\mu_2|\mu_1)$. Using Bayes' theorem, we rewrite this probability as

$$\Pr(\mu_2|\mu_1) = \frac{\Pr(\mu_1, \mu_2)}{\Pr(\mu_{1+})}, \tag{5}$$

where μ_{1+} denotes the set of all complex words that have μ_1 as left constituent. Hence, $\Pr(\mu_{1+})$ is the joint probability mass of all words starting with μ_1. We estimate $\Pr(\mu_2|\mu_1)$ with

$$\Pr(\mu_2|\mu_1) = \frac{\Pr(\mu_1, \mu_2)}{\Pr(\mu_{1+})} = \frac{\frac{F_{12}}{N}}{\frac{F_{1+}}{N}} = \frac{F_{12}}{F_{1+}}, \tag{6}$$

where F_{1+} denotes the summed frequencies in the corpus of all μ_1-initial words. This probability comes into play when the left constituent has been identified and the right constituent is anticipated, either by the end of the information uptake from the left constituent, or during the processing of the right constituent.

The next conditional probability mirrors the first: It addresses the likelihood of the left constituent given that the right constituent is known. Denoting the set of words ending in the right constituent μ_2 by μ_{+2}, the summed frequencies of these words by F_{+2}, and the corresponding probability mass by $\Pr(\mu_{+2})$, we have that

$$\Pr(\mu_1|\mu_2) = \frac{\Pr(\mu_1, \mu_2)}{\Pr(\mu_{+2})} = \frac{\frac{F_{12}}{N}}{\frac{F_{+2}}{N}} = \frac{F_{12}}{F_{+2}}. \tag{7}$$

This probability is relevant in any situation where the right constituent is identified before the left, for instance, because the left constituent was skipped or only partly processed[6].

The preceding two probabilities are conditioned on the full availability of the left or the right constituent. The final two probabilities are more general in the sense that they condition on the presence of some unspecified right or left constituent, without narrowing this constituent down to one specific morpheme. The unspecified left constituent stands for the subset of all morphemes or words in a language that can appear in the word-initial position.

[6] μ_{1+} and μ_{+2} denote the left and right constituent families. In the present formulation of the model, we estimate the corresponding probabilities and informations using the summed frequencies of these families. It may be more appropriate to estimate the amount of information in the morphological family using Shannon's entropy, the *average* amount of information (cf. e.g., Moscoso del Prado Martín, Kostić & Baayen, 2004), or, under the simplifying assumption of a uniform probability distribution for the family members, by $\log V$, with V the family size, which is the measure we used for our experimental data.

Essentially, this subset is equal to full vocabulary with the exception of suffixes (e.g., *-ness, -ity*) and of those compounds' constituents that can only occur word-finally. Suppose that the reader has an intuition that the word under inspection, say *blackberry*, as potentially morphologically complex (based, for example, on its length or the low probability of the bigram "kb"). While the left constituent of such a compound is unspecified, combinations like **nessberry* or **ityberry* will never be part of the lexical space, which needs to be considered for identification of the full compound. Likewise, the unspecified right constituent is the set of morphemes that excludes prefixes (e.g., *un-, anti-*) or compounds' constituents (e.g., *cran*) that can only occur word-initially.

Denoting the presence of such an unspecified left constituent by M_1 and that of such an unspecified right constituent by M_2, we denote these more general conditional probabilities as $\Pr(\mu_1|M_2)$ and $\Pr(\mu_2|M_1)$ respectively, and estimate them as follows:

$$\Pr(\mu_1|M_2) = \frac{\Pr(\mu_1, M_2)}{\Pr(M_2)} = \frac{\Pr(\mu_{1+})}{\Pr(M_2)} = \frac{F_{1+}}{F_{M_2}} \tag{8}$$

$$\Pr(\mu_2|M_1) = \frac{\Pr(M_1, \mu_2)}{\Pr(M_1)} = \frac{\Pr(\mu_{+2})}{\Pr(M_1)} = \frac{F_{+2}}{F_{M_1}} \tag{9}$$

In these equations, F_{M_2} denotes the summed frequencies of all words that can occur as a right constituent. Likewise, F_{M_1} denotes the summed frequencies of all words that can occur as a left constituent in a complex word. The probabilities $\Pr(M_1)$ and $\Pr(M_2)$ are independent of μ_1 and μ_2 and hence are constants in our model. $\Pr(\mu_2|M_1)$ comes into play when the left constituent is not fully processed and the likelihood of the right constituent is nevertheless evaluated. $\Pr(\mu_1|M_2)$ becomes relevant when length information or segmentation cues clarify that there is a right constituent, and this information is used to narrow down the set of candidates for the left constituent. To keep the presentation simple, here we build a model for compounds with only two morphemes: Extension to trimorphemic cases, however, is straightforward.

The basic model. We introduce our model with only three of the seven probabilities defined in the preceding section. For each of the probabilities

$$\begin{aligned}
\Pr(\mu_2|\mu_1) &= \frac{F_{12}}{F_{1+}} \qquad (10)\\
\Pr(\mu_1,\mu_2) &= \frac{F_{12}}{N}\\
\Pr(\mu_1|M_2) &= \frac{F_{1+}}{F_{M_2}}
\end{aligned}$$

we calculate the corresponding weighted information using (1),

$$\begin{aligned}
\mathrm{I}_{\mu_2|\mu_1} &= w_1(\log F_{1+} - \log F_{12}) \qquad (11)\\
\mathrm{I}_{\mu_1,\mu_2} &= w_2(\log N - \log F_{12})\\
\mathrm{I}_{\mu_1|M_2} &= w_3(\log F_{M_2} - \log F_{1+})
\end{aligned}$$

with positive weights $w_1, w_2, w_3 > 0$. A crucial assumption of our model is that the time t spent by the eye on a constituent or word is proportional to the total amount of information available at a given point in time:

$$\begin{aligned}
t &= I_{\mu_2|\mu_1} + I_{\mu_1,\mu_2} + I_{\mu_1|M_2} \qquad (12)\\
&= w_1(\log F_{1+} - \log F_{12}) + w_2(\log N - \log F_{12}) + w_3(\log F_{M_2} - \log F_{1+})\\
&= w_1 \log F_{1+} - w_1 \log F_{12} + w_2 \log N - w_2 \log F_{12} + w_3 \log F_{M_2} - w_3 \log F_{1+}\\
&= w_2 \log N + w_3 \log F_{M_2} - (w_1 + w_2) \log F_{12} - (w_3 - w_1) \log F_{1+}.
\end{aligned}$$

Equation (12) states that processing time linearly covaries with $\log F_{12}$ and $\log F_{1+}$, with facilitation for compound frequency and facilitation or inhibition for left constituent family frequency, depending on the relative magnitude of w_1 and w_3. In other words, starting from simple probabilities and using information theory, we have derived a model equation the parameters of which can be directly estimated from the data using multiple (linear) regression models. Note that these parameters are simple sums of our weights w.

We now bring the remaining probabilities

$$\begin{aligned} \Pr(\mu_1|\mu_2) &= \frac{F_{12}}{F_{+2}} \\ \Pr(\mu_2|M_1) &= \frac{F_{+2}}{F_{M_1}} \\ \Pr(\mu_1) &= \frac{F_1}{N} \\ \Pr(\mu_2) &= \frac{F_2}{N} \end{aligned} \tag{13}$$

into the model as well. For each of these probabilities we have a corresponding weighted amount of information, again with positive weights:

$$\begin{aligned} I_{\mu_1|\mu_2} &= w_4(\log F_{+2} - \log F_{12}) \\ I_{\mu_2|M_1} &= w_5(\log F_{M_1} - \log F_{+2}) \\ I_{\mu_1} &= w_6(\log N - \log F_1) \\ I_{\mu_2} &= w_7(\log N - \log F_2) \end{aligned} \tag{14}$$

We can now define the general model as

$$\begin{aligned} t = {} & (w_2 + w_6 + w_7)\log N + w_3 \log F_{M_2} + w_5 \log F_{M_1} - (w_1 + w_2 + w_4)\log F_{12} \\ & -(w_3 - w_1)\log F_{1+} - (w_5 - w_4)\log F_{+2} - w_6 \log F_1 - w_7 \log F_2. \end{aligned} \tag{15}$$

This equation, as well as equations in (11) and (14), sheds light on some of the intriguing findings reported above. Compound frequency contributes to probabilities (and respective amounts of information) that readers can start estimating even before all characters may be scanned: for instance, as a term in the conditional information of the right constituent $\mathrm{I}_{\mu_2|\mu_1}$ given the (partial) identification of the left constituent (first equation in (11)). Also recall that the property of the right constituent family plays a role even though activation of this family would seem dysfunctional given that the only relevant right constituent family member is the compound itself. This seemingly unwarranted contribution of the right constituent family originates, however, from the fact that the family contributes to the estimate of the conditional probability $\mathrm{I}_{\mu_2|M_1}$ of the right constituent and to the conditional probability $\mathrm{I}_{\mu_1|\mu_2}$ of the left constituent. In other words, the family is used to narrow down the lexical space from which both constituents are selected, and thus it offers a larger amount of information about the compound and its morphemes.

Equation (15) in its present form treats all information sources as if they are simultaneously available to the processing system. This describes cases when the visual uptake of the word is complete in one fixation (typical of shorter and more frequent words). The formulation, however, is easily adjustable to the cases where multiple fixations are required to read the word, like in the long compounds used in the current study and in Kuperman *et al.* (2007). Information sources that are available early in the time-course of the visual uptake are demonstrably more important in compound recognition (cf. the weaker role of right constituent measures as compared to properties of the left constituent). In the equation, weights

w for "early" information sources can be multiplied by a time-step coefficient α_1, such that $\alpha_1 > 1$. For "late" information sources, the value of α_2 is equal to or smaller than 1. As with weights w, the value of α can be directly estimated from comparing regression coefficients of a predictor in the models for early measures of the visual uptake (cf., *SubgazeLeft*) vs. the models for later measures (e.g., *SubgazeRight*). For the sake of exposition, we restrict our further discussion to a simpler, temporally indiscriminate, model (15).

There are several falsifiable predictions that follow straightforwardly from the properties of (15).

- The frequency of the whole compound, as well as the frequencies of its constituents as isolated words, have negative coefficients in the equation. This predicts that higher *a priori*, unconditional, frequencies of complex words and their morphemes always come with facilitation of processing (e.g., shorter reading times or lexical decision latencies).

- Three corpus constants contribute to the intercept: the token size of the corpus/lexicon (N), the number of tokens in the corpus/lexicon that can occur as a left constituent (F_{M_1}), and the number of tokens in the corpus/lexicon that can occur as a right constituent (F_{M_2}). The larger the size of a corpus/lexicon, the higher the values of all three constants and the higher the intercept. Given the positive weight coefficients, the model predicts a longer processing time for a word in a larger corpus/lexicon. This is hardly surprising, since we use absolute frequencies in (15). So a word with 100 occurrences per corpus would be recognized slower in a corpus of 100 million word forms that in a corpus of 1000 word forms.

- All coefficients, with the exception of w_1, occur in more than one term of equation (15). This expresses various trade-offs in lexical processing. For instance, w_3 appears with a positive sign for the intercept ($w_3 \log F_{M_2}$) and with a negative sign for the left constituent family frequency ($-w_3 \log F_{1+}$). We predict that the stronger facilitation compounds receive due to their higher family frequency, the higher the intercept (i.e., average processing time) across compounds is.

In the remainder of this section we apply PROMISE to the key statistical models that we fitted to our experimental data. Since most results of the model for first fixation duration are also found in the model for left subgaze duration, and most results of the model for gaze duration are also attested in the model for right subgaze duration, in what follows we concentrate on the two models for subgaze durations (cf., Tables 5 and 6 in Appendix).

Left subgaze duration. A comparison of the general model (15) with the regression model for the subgaze for the left constituent (Table 5) shows that $w_4 = w_5 = w_7 = 0$. The information sources requiring identification of the right constituent $\mathrm{I}_{\mu_1|\mu_2}$, I_{μ_2}, as well as the information source conditioned on the presence of some unspecified left constituent $\mathrm{I}_{\mu_2|M_1}$, play no role when the left constituent is being processed.

Given the regression coefficients listed in Table 5, we infer that $w_1 + w_2 = 0.0471$ and $w_3 - w_1 = 0.0431$, from which it follows that 0.0471 is an upper bound for w_1 and that 0.0431 is a lower bound for w_3. In other words, $I_{\mu_1|M_2}$ receives greater weight than $I_{\mu_2|\mu_1}$. Apparently, the identification of the left constituent given the knowledge that there is some right constituent plays a more important role at that timepoint than anticipating the right constituent given the identity of the left constituent. Anticipation of the right morpheme probably is a process that only starts up late in the uptake of information from the left morpheme.

Interestingly, the importance of the *a priori*, context-free probability of the left constituent (I_{μ_2}) is much smaller than the contribution of that constituent recognized as part of a compound: given that 0.0431 is a lower bound for w_3 and that $w_6 = 0.0219$, the weight of this *a priori* probability is at best roughly half of that of the contextual probability of the left constituent.

An important finding for the left subgaze durations is that the effects of the left constituent frequency and left constituent family size were greater for those left constituents ending in the suffix -*stO*, cf., Table 2. Within the present framework, this implies that the weights w_6 (for the left constituent frequency) and w_3 (for left family size) have to be greater for left constituents with -*stO* compared to left constituents with -*Us* or simplex left constituents. Greater values for w_6 and w_3 for -*stO* also imply, according to (15), that the intercept should be larger as well for left constituents with this suffix. As can be seen in Table 5, this is indeed the case: The main effect for -*stO* is positive (0.045) and is more than twice the main effect for -*Us* (0.024). This suggests that conditioning narrows down the set of candidates and hence affords better facilitation, but it always comes with a price, the price of 'spurious' lexical co-activation.

Right subgaze duration. A comparison of the general model (15) with the regression model for the subgaze for the right constituent indicates that $w_6 = 0$: the unconditional information source for the left constituent, I_{μ_1}, no longer plays a role. In the context of the right constituent, this probability has become irrelevant.

The regression model for the subgaze durations for the right constituent presents us with the familiar and expected facilitation for compound frequency. The facilitation for the right constituent frequency and family size are also in line with (15).

For left constituents in -*Us*, there is no effect of left constituent family size ($\hat{\beta} = -0.028; p = 0.18$), implying that here $w_1 \approx w_3$. For left constituents in -*stO*, by contrast, we have facilitation ($\hat{\beta} = -0.055; p = 0.035$), indicating that $w_1 > w_3$, while for simplex left constituents there is some evidence for inhibition ($\hat{\beta} = 0.025; p = 0.085$). It follows from our model that the intercept must be greatest for -*stO*, and Table 6 shows that this is indeed the case (5.44 log units for bimorphemic compounds and compounds with -*Us* and $5.44 + 0.12 = 5.56$ for compounds with -*stO*). Compared to the model for the left subgaze durations, this balance between increased intercept and increased facilitation emerges more clearly, with unambiguous support from the significance levels.

The right subgaze durations are characterized by interactions of compound frequency by left constituent family size and compound frequency by right constituent family size that are absent for the left subgaze durations (see Figs. 1 and 2). Within the present framework, an interaction such as that of compound frequency by left constituent family size implies a more complex evaluation of $I_{\mu_2|\mu_1}$, which we weighted above simply by a scalar weight w_1.

First note that

$$I_{\mu_2|\mu_1} = w_1(\log F_{1+} - \log F_{12}) \quad = \quad \log(\frac{F_{1+}}{F_{12}})^{w_1} \tag{16}$$

We have to revise information $I_{\mu_2|\mu_1}$ in such a way that the magnitude of one cue contributing to an information source modulates the extent to which other cues contribute to that information source (see also Kuperman *et al.*, 2007). We achieve this by assigning weights to one term in the equation (e.g., F_{12}) so that it is proportional to another term (e.g., F_{1+}):

$$I_{\mu_2|\mu_1} = \log \frac{F_{1+}^{w_1+C_1 \log F_{12}}}{F_{12}^{w_1+C_2 \log F_{1+}}} = w_1 \log F_{1+} - w_1 \log F_{12} + (C_1 - C_2) \log F_{12} \log F_{1+}, \tag{17}$$

$(w_1, w_2, C_1, C_2 > 0)$.

Notably, this new weighting of terms in the information source introduces into our model the desired multiplicative interaction between compound frequency and left constituent family size[7].

[7] Other estimates of weights are also possible. For instance, the amount of information I_{μ_1,μ_2} can be derived from probability equation (2) using the same weight, rather than different weights for the numerator and denominator: $log[F_{12}/N]^{w_2+logF_{12}} = w_2logN - logF_{12}(logN + w_2) + logF_{12}^2$. Note that I_{μ_1,μ_2} becomes a polynomial with F_{12} as a negative linear term and a positive quadratic term. This equation predicts the L-shape or the U-shape functional relationship between processing time and compound frequency. The L-shape frequency effect is indeed observed in comprehension (Baayen, Feldman & Schreuder, 2006) and the U-shape effect in production (Bien, Levelt & Baayen, 2005).

The interaction of compound frequency with right constituent family size can be modeled in terms of $I_{\mu_1|\mu_2}$ in the same way ($w_4, K_1, K_2 > 0$):

$$I_{\mu_1|\mu_2} = \log \frac{F_{+2}^{w_4+K_1 \log F_{12}}}{F_{12}^{w_4+K_2 \log F_{+2}}} = w_4 \log F_{+2} - w_4 \log F_{12} + (K_1 - K_2) \log F_{12} \log F_{+2}. \quad (18)$$

This leads to the following model for the right subgaze durations:

$$\begin{aligned} t \;=\; & (w_2 + w_7) \log N + w_3 \log F_{M_2} + w_5 \log F_{M_1} \\ & -(w_1 + w_2 + w_4) \log F_{12} \\ & -(w_3 - w_1) \log F_{1+} - (w_5 - w_4) \log F_{+2} - w_7 \log F_2 \\ & +(C_1 - C_2) \log F_{12} \log F_{1+} + (K_1 - K_2) \log F_{12} \log F_{+2}. \quad (19) \end{aligned}$$

Figure 3 illustrates the geometry of these interactions.

INSERT Figure 3 ABOUT HERE

The upper panels illustrate the difference between a model without (left) and with (right) an interaction with a positive coefficient ($C_1 > C_2$). The right panel illustrates how facilitation can be reversed into inhibition depending on the value of the other predictor. Crucially, the interactions predicted by our statistical model for right subgaze duration in Fig. 1 and Fig. 2 are two-dimensional representations of the shape shown in the right panel of Fig. 3.

The coefficients for the interactions listed in Table 6 are all positive, which implies that $C_1 > C_2$ and $K_1 > K_2$. Apparently, the left (and right) family measures receive greater weight from compound frequency than compound frequency from the family measures. In other words, the compound's own probability has priority. The more C_1 (or K_1) increases with respect to C_2 (or K_2), the greater the inhibitory force of the interaction. The bottom panels of Figure 3 visualize the interactions of of compound frequency by left constituent family size, for compounds with left constituents ending in *-stO* (lower left panel) and compounds with simplex left constituents (lower right panel). For the compounds in *-stO*, we effectively have a floor effect, with a maximum for the amount of facilitation that never exceeds the maximum for any of the marginal effects. For the bimorphemic compounds, maximum facilitation is obtained only when compound frequency is large and family size is **small**. In terms of morphological processing, the observed interaction may receive the following interpretation. There is a balance between the contributions of compound frequency and left constituent family size to the ease of compound recognition. The effect of the family size may differ from facilitatory (as in the compounds with *-stO*) to slightly inhibitory (as in the bimorphemic compounds), see the lower panels of Figure 3. We believe that this reflects the potentially dual impact of constituent families: A large family may raise the resting activation level of its members (thus making easier lexical access to the target compound), and at the same time it brings along a larger number of competitors (thus inhibiting the recognition of the actual target). Crucially, regardless of the direction of the left constituent family size effect, the larger the morphological family, the more processing resources are allocated to it and the less impact is elicited by compound frequency. Again, we witness how the magnitude of some processing cues modulates the utility of the cues for compound recognition.

We note that the model can also predict facilitation for one of predictors at large values of the other predictor: an interaction of this nature would be a symptom that C_1 (or K_1) is smaller than C_2 (or K_2). Equal or similar values of C_1 (or K_1) and C_2 (or K_2) imply that the two terms contribute to a similar extent to the information and render the interaction statistically weak.

Since we focus on lexical distributional predictors in this version of the model, our formulation in (15) leaves out the interaction of right constituent frequency by word length attested for the right subgaze duration. The effect of length might be brought into the model, however, by conditioning on lexical subsets of the appropriate length. Alternatively, or complementary, processes of visual uptake may be at stake here. We leave this issue to future research.

The PROMISE model is a formalization of the idea that readers and listeners maximize their opportunities for recognition of complex words (see Libben, 2006 and Kuperman *et al.* 2007). Parameters of PROMISE can be directly estimated from the regression coefficients of statistical models. As we have shown, estimated values of parameters do not only shed light on which sources of information are preferred over others, but also specify at what timesteps of the visual uptake and at what cost to the processing system. Importantly, PROMISE is not restricted to compounding as a type of morphological complexity, nor to long polymorphemic words. The model allows dealing with word length and morphological complexity (e.g., simplex, inflected, derived or compound words) in a principled probabilistic way. As a research perspective, a series of experiments involving a broad spectrum of languages and word lengths would be desirable to quantify the range of opportunities that morphological structure offers for efficient recognition of complex forms. We also believe that PROMISE can be easily incorporated into general models of eye-movement control in reading, such as E-Z Reader or SWIFT, extending the line of research of Pollatsek, Reichle and Rayner (2003). Consideration of parameters of PROMISE along with other visual and lexical parameters may improve predictions of such models for the processing of complex morphological structures.

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., 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. (1994). Productivity in language production. *Language and Cognitive Processes*, 9:447–469.

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. 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.

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.

Bien, H., Levelt, W., and Baayen, R. (2005). Frequency effects in compound production. *PNAS*, 102:17876–17881.

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

de Almeida, R. and Libben, G. (2005). Changing morphological structures: The effect of sentence context on the interpretation of structurally ambiguous English trimorphemic words. *Language and Cognitive Processes*, 20:373–394.

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.

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.

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

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. and Olson, R. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, memory and cognition*, 21:1430–1440.

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.

Inhoff, A. W., Radach, R., and Heller, D. (2000). Complex compounds in German: Interword spaces facilitate segmentation but hinder assignment of meaning. *Journal of Memory and Language*, 42:23–50.

Järvikivi, J., Bertram, R., and Niemi, R. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. *Language and Cognitive Processes*, 21:394–431.

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.

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.

Krott, A., Libben, G., Jarema, G., Dressler, W., Schreuder, R., and Baayen, R. H. (2004). Probability in the grammar of German and Dutch: Interfixation in tri-constituent compounds. *Language and Speech*, 47:83–106.

Krott, A. and Nicoladis, E. (2005). Large constituent families help chidren parse compounds. *Journal of Child Language*, 32(1):139–158.

Kuperman, V., Schreuder, R., Bertram, R., and Baayen, R. H. (2007). Reading of polymorphemic Dutch compounds: Towards a multiple route model of lexical processing. *Manuscript submitted for publication.*

Laine, M. and Virtanen, P. (1999). *WordMill Lexical Search Program.* Center for Cognitive Neuroscience, University of Turku, Finland.

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. (2005). Everything is psycholinguistics: Material and methodological considerations in the study of compound processing. *Canadian Journal of Linguistics*, 50:267–283.

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 Pollatsek, A. (1983). Lexical access via an orthographic code? The basic orthographic syllabic structure (BOSS) reconsidered. *Journal of Verbal Learning and Verbal Behavior*, 22:310–332.

Meunier, F. and Longtin, C. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, 56:457–471.

Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94:1–18.

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.

Nicoladis, E. and Krott, A. (2007). Word family size and French-speaking children's segmentation of existing compounds. *Language Learning*, 57(2):201–228.

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

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

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., 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.

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.

Sereno, J. and Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition*, 25:425–437.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7:263–272.

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

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology*, 57A:745–765.

Taft, M. and Ardasinski, S. (2006). Obligatory decomposition in reading prefixed words. *Mental Lexicon*, 1:183–189.

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.

Van Jaarsveld, H. J. and Rattink, G. E. (1988). Frequency effects in the processing of lexicalized and novel nominal compounds. *Journal of Psycholinguistic Research*, 17:447–473.

Virtanen, P. and Pajunen, A. (2000). *ContextMill Computer Software.* General Linguistics, University of Turku, Finland.

Appendix

INSERT TABLE 3 APPROXIMATELY HERE

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: *NextLength*, length of the word to the right of the target word; *NextSkipped*, indicator of whether the word following the target is skipped during reading; *LeftLength*, length of the compound's left constituent; *InitTrigramFreq*, token-based frequency of the word-initial trigram (based on 22.7 million corpus of written Finnish); *AverageBigramFreq*, average bigram frequency across the target word (based on 22.7 million corpus of written Finnish); *LastSaccade*, amplitude of the saccade preceding the fixation; *NextSaccade*, amplitude of the saccade following the fixation; *FixPos* and *FixPos2*, first fixation position and its squared value; *Nomore*, indicator of whether the fixation is word-final; and *Sex*, participants' gender. Table 1 summarizes continuous (dependent and independent) variables, which show significant effects in our statistical models. In addition to these, we have considered a large number of control variables that were not significant predictors of reading times or probabilities. These included: transitional probabilities of word pairs N-1 and N and words N and N+1 (computed with the help the ContextMill software, Virtanen & Pajunen, 2000); frequencies of words N-1 and N+1; length of word N-1; and the frequency of the word-final trigram.

INSERT TABLES 4-8 APPROXIMATELY HERE

Key to Tables 4-8 and to estimating effect sizes for the models' predictors: Throughout the tables, the second column shows estimates of the regression coefficients for the model's predictors. Columns 3-6 provide information on the distributions of those estimates obtained via the Monte Carlo Markov chain (MCMC) random-walk method using 1000 simulations: this information is useful for evaluating stability of the models' predictions. The third column shows the MCMC estimate of the mean for each predictor, while the fourth and the fifth columns show highest posterior density intervals, which are a Bayesian measure for the lower and upper bounds of the 95% confidence interval, respectively. The sixth column provides a p-value obtained with the help of MCMC simulations; and the final column provides less conservative p-values obtained with the t-test using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom.

For the predictors of primary interest for this study we report effect sizes, either in the body of the paper or in Tables 1 and 2. These were obtained as follows. Our models used contrast coding for discrete variables. Therefore, the effect size for factors was calculated as the difference between (i) the (exponentially-transformed) sum of the intercept value and the contrast regression coefficient, $\hat{\beta}$, and (ii) the (exponentially-transformed) intercept value. Exponential transformation was only applied, when the dependent variable had log-transformed values, i.e. fixation or gaze duration. For instance, the effect size of the indicator of whether the word after the target word is skipped (*NextSkipped*) on gaze duration, after log gaze duration is back-transformed to original values in milliseconds, is:

$\exp(\text{Intercept} + \hat{\beta}) - \exp(\text{Intercept}) = \exp(5.9 + 0.105) - \exp(5.9) = 40\text{ms},$

where *Intercept* is the intercept of the model for gaze duration (= 5.9) and $\hat{\beta}$ is the contrast coefficient for *NextSkipped* (= 0.105).

Effect sizes for simple main effects of numeric variables were calculated as the difference between the (exponentially-transformed) model's predictions for the minimum and maximum values of a given variable. For instance, the regression coefficient, $\hat{\beta}$, associated with compound frequency, *WordFreq*, in the model for first fixation duration is -0.0111, while the range of values, *Min*:*Max*, used in that model for *WordFreq* and obtained via the operation of centering, is $-2.2 : 3.6$, see Table 3. To compute the effect size for log-transformed dependent measures, like first fixation duration, we used the following formula:

$\exp(\text{Intercept} + \hat{\beta} * \text{Max}) - \exp(\text{Intercept} + \hat{\beta} * \text{Min})$,

The effect of *WordFreq* (i.e., the difference between the model's predictions for the lowest-frequency and the highest-frequency target words) on first fixation duration is then:

$\exp(5.2 + -0.0111 * 3.6) - \exp(5.2 + -0.0111 * -2.2) = -11.6\text{ms}$

Computation of effect sizes for interactions involved obtaining model predictions for the extreme values of one term in the interaction of interest, while holding all other terms in that model (and in that interaction) constant at their median values. Again, the estimate of the effect size for an interacting variable was calculated as a difference between the (exponentially-transformed) values of the regression function corresponding to the minimum and the maximum values of that variable. To estimate the effect sizes for interactions we also used conditioning plots that are not explained here (for detailed treatment, see Baayen, 2008).

Table 1: Summary of morphological effects on durational measures

Predictor	FirstDur	SubgazeLeft	SubgazeRight	GazeDur
ResidLeftFreq	-13 ms (0.001)	-72 ms (<0.001)	ns	-72 ms (0.006)
ResidLeftFamSize	-9 ms (0.02)	-80 ms (<0.001)		-120 ms (0.001)
			interaction with *WordFreq* (0.004), Fig. 1	
RightFreq	ns	ns	ns	ns
			interaction with *WordLength* (<0.001)	
ResidRightFamSize	ns	ns	ns	ns
			interaction with *WordFreq* (0.022), Fig. 2	interaction with *WordFreq* (0.002), Fig. 1
WordFreq	-12 ms (0.010)	-110 ms (<0.001)	-44 ms (<0.001)	-136 ms (<0.001)
			interaction with family sizes	interaction with *ResidRightFamSize* (0.002)
			(left: 0.004; right: 0.022), Figs. 1, 2	

Numbers in columns 2-5 show sizes of statistically significant effects. Numbers in brackets provide p-values for the effects, estimated based on the MCMC method with 1000 simulations. ”ns” stands for non-significant. Estimation of effect sizes is based on models that do not include interactions of morphological predictors by suffix type: those interactions are summarized below in Table 2

Table 2: Summary of interactions of morphological predictors with SuffixType

Predictor	Measure	-stO	-Us	None	p-value
ResidLeftFreq					
	single fixation probability	more likely single fixation	ns	ns	p = 0.004
		3.1 log odds units(<0.001)			
	probability of regressive fixation	ns	more likely regression	ns	p = 0.009
			0.19 log odds units (0.025)		
	left subgaze duration	shorter duration	ns	shorter duration	p = 0.004
		-148 ms (0.0001)		-48 ms (0.07)	
	gaze duration	shorter duration	ns	shorter duration	p = 0.005
		-120 ms (0.0001)		-15 ms (0.08)	
ResidLeftFamSize					
	left subgaze duration	shorter duration	ns	shorter duration	p = 0.0045
		-204 ms (0.0001)		-80 ms (0.03)	
	right subgaze duration	shorter duration	ns	ns	p = 0.0045
		-35 ms (0.0345)			
	gaze duration	shorter duration	ns	ns	p = 0.0004
		-246 ms (<0.0001)			

Numbers in columns 3-5 show sizes of statistically significant effects. Numbers in brackets provide p-values for the effects. ”ns” stands for non-significant. Column 6 provides the estimate of statistical significance for the interactions with *SuffixType* based on the MCMC method with 1000 simulations.

Table 3: Summary of Continuous Variables Reported in Statistical Models.

Variable	Range (Adjusted Range)	Mean(SD)	Median
FixPos	0.1:16 characters (1:160 pixels)	37.1(21.8)	35.1
FirstDuration	67:735 ms (4.2:6.6 log units)	5.4(0.3)	5.4
SubgazeLeft	60:1808 ms (4.1:7.5 log units)	5.8(0.5)	5.7
SubgazeRight	81:812 ms (4.4:6.7 log units)	5.5(0.4)	5.5
GazeDuration	60:1998 ms (4.2:7.6 log units)	6.1(0.6)	6.2
LastSaccade	1:15 characters (10:151 pixels)	70.8(27.9)	70.5
NextSaccade	-12:19 characters (-112:189 pixels)	46.3(55.2)	54.7
NextLength	2:13 characters	4.9(3.1)	4
WordLength	10:18 characters (-3.1:4.9)	0.0(1.7)	-0.12
LeftLength	4:14 characters	7.5(1.4)	8
InitTrigramFreq	3:601 (1.1:6.4 log units)	4.3(1.0)	4.5
AverageBigramFreq	2:151 (0.7:5.0 log units)	4.1(0.9)	4.3
WordFreq	2:665 (-2.2:3.6 log units)	0.1(1.4)	0.1
ResidLeftFreq	11:1.8*10^4 (-4.1:3.1 log units)	0.0(1.5)	0.1
RightFreq	33:8.1*10^4 (-4.5:3.3 log units)	0.0(1.4)	0.14
ResidLeftFamilySize	2:812 (-3.0:1.7)	0.0(0.9)	0.1
ResidRightFamilySize	3:1808 (-2.0:1.3)	0.0(0.6)	-0.1
ResidBaseFreq	49:3.3*10^4 (-2.8:4.0)	0.0(1.2)	-0.2
TrialNum	11:272	142.1(76.3)	143

Numbers in the second column show original value ranges for predictors. If any transformations have been made to the original values for statistical reasons (i.e., natural log transformation, decorrelation with other predictors or centering), 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 22.7 million word-forms.

Table 4: First Fixation Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.2048	5.2060	5.1153	5.3001	0.001	0.0000
SuffixTypeSt	-0.0131	-0.0131	-0.0500	0.0207	0.458	0.4269
SuffixTypeUs	0.0143	0.0137	-0.0204	0.0463	0.428	0.3549
ResidLeftLength	-0.0099	-0.0095	-0.0196	0.0016	0.088	0.0533
NextSaccade	0.0010	0.0010	0.0008	0.0013	0.001	0.0000
LastSaccade	0.0013	0.0013	0.0009	0.0017	0.001	0.0000
WordFreq	-0.0111	-0.0109	-0.0179	-0.0033	0.008	0.0019
TrialNum	-0.0001	-0.0001	-0.0002	0.0000	0.158	0.1303
FixPos	0.0025	0.0025	0.0014	0.0036	0.001	0.0000
FixPos2	0.0000	0.0000	0.0000	0.0000	0.001	0.0000
NomoreTRUE	0.1194	0.1173	0.0718	0.1633	0.001	0.0002
RightFreq	-0.0080	-0.0079	-0.0161	-0.0010	0.044	0.0286
WordLength	-0.0066	-0.0064	-0.0137	-0.0003	0.062	0.0316
InitTrigramFreq	0.0072	0.0069	-0.0035	0.0177	0.190	0.1276
NextLen	0.0010	0.0009	-0.0022	0.0041	0.602	0.5148
ResidLeftFreq	-0.0129	-0.0128	-0.0196	-0.0057	0.002	0.0001
ResidFamSizeL	-0.0138	-0.0142	-0.0262	-0.0043	0.012	0.0062
SubjectSexM	-0.0069	-0.0085	-0.1112	0.0916	0.876	0.8958
SuffixTypeSt:ResidLeftLength	0.0229	0.0223	-0.0008	0.0466	0.068	0.0356
SuffixTypeUs:ResidLeftLength	0.0007	0.0000	-0.0235	0.0260	0.962	0.9526
SuffixTypeSt:NextSaccade	0.0000	0.0000	-0.0004	0.0003	0.888	0.8410
SuffixTypeUs:NextSaccade	-0.0002	-0.0002	-0.0006	0.0002	0.276	0.2698
RightFreq:WordLength	0.0016	0.0015	-0.0026	0.0057	0.494	0.4475
NomoreTRUE:SubjectSexM	-0.0620	-0.0758	-0.1403	-0.0070	0.026	0.2254

Table 5: Model for for Subgaze Duration for the Left Constituent

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.7703	5.7719	5.6822	5.8638	0.001	0.0000
WordLength	0.0219	0.0221	0.0072	0.0376	0.004	0.0046
WordFreq	-0.0471	-0.0469	-0.0646	-0.0283	0.001	0.0000
ResidLeftLength	0.0594	0.0600	0.0406	0.0802	0.001	0.0000
ResidFamSizeL	-0.0431	-0.0431	-0.0887	-0.0016	0.044	0.0529
SuffixTypeSt	0.0456	0.0451	-0.0206	0.1095	0.188	0.1796
SuffixTypeUs	0.0247	0.0242	-0.0328	0.0788	0.426	0.4044
ResidLeftFreq	-0.0219	-0.0216	-0.0460	0.0037	0.096	0.0713
SuffixTypeSt:ResidLeftFreq	-0.0384	-0.0396	-0.0804	0.0033	0.068	0.0608
SuffixTypeUs:ResidLeftFreq	0.0152	0.0148	-0.0220	0.0484	0.408	0.3948
ResidFamSizeL:SuffixTypeSt	-0.0814	-0.0835	-0.1526	-0.0136	0.008	0.0227
ResidFamSizeL:SuffixTypeUs	0.0316	0.0321	-0.0308	0.0821	0.250	0.2792

Table 6: Model for Subgaze Duration for the Right Constituent

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.4395	5.4387	5.3463	5.5407	0.001	0.0000
WordLength	0.0187	0.0189	0.0082	0.0295	0.002	0.0005
WordFreq	-0.0230	-0.0225	-0.0347	-0.0084	0.001	0.0006
TrialNum	0.0000	0.0000	-0.0003	0.0004	0.798	0.8069
ResidLeftLength	-0.0489	-0.0490	-0.0653	-0.0330	0.001	0.0000
SuffixTypeSt	0.1177	0.1208	0.0420	0.2107	0.001	0.0063
SuffixTypeUs	-0.0040	-0.0023	-0.0783	0.0811	0.950	0.9232
ResidFamSizeL	0.0259	0.0257	-0.0023	0.0554	0.084	0.0850
RightFreq	-0.0439	-0.0435	-0.0653	-0.0213	0.001	0.0001
NextSkipped	0.0777	0.0782	0.0329	0.1226	0.001	0.0003
NextLen	0.0079	0.0079	0.0007	0.0146	0.020	0.0180
ResidFamSizeR	-0.0024	-0.0022	-0.0303	0.0257	0.886	0.8711
TrialNum:SuffixTypeSt	-0.0008	-0.0009	-0.0013	-0.0004	0.001	0.0007
TrialNum:SuffixTypeUs	-0.0003	-0.0003	-0.0008	0.0001	0.228	0.2583
SuffixTypeSt:ResidFamSizeL	-0.0545	-0.0538	-0.1023	-0.0009	0.044	0.0345
SuffixTypeUs:ResidFamSizeL	-0.0282	-0.0277	-0.0679	0.0135	0.180	0.1808
WordLength:RightFreq	-0.0155	-0.0156	-0.0220	-0.0081	0.001	0.0000
WordFreq:ResidFamSizeL	0.0210	0.0210	0.0076	0.0367	0.004	0.0055
RightFreq:NextLen	0.0085	0.0084	0.0042	0.0123	0.001	0.0000
WordFreq:ResidFamSizeR	0.0242	0.0244	0.0051	0.0478	0.028	0.0222

Table 7: Model for Gaze Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.8979	5.9073	5.6691	6.1598	0.001	0.0000
WordLength	0.0540	0.0538	0.0376	0.0687	0.001	0.0000
TrialNum	-0.0001	-0.0002	-0.0003	0.0001	0.140	0.1633
WordFreq	-0.0303	-0.0302	-0.0514	-0.0123	0.004	0.0018
ResidLeftFreq	-0.0130	-0.0133	-0.0355	0.0122	0.268	0.2833
ResidFamSizeL	-0.0201	-0.0198	-0.0633	0.0261	0.376	0.3745
SuffixTypeSt	0.3112	0.3046	0.0512	0.5812	0.018	0.0227
SuffixTypeUs	0.3682	0.3636	0.0781	0.6204	0.010	0.0077
AverageBigramFreq	0.0638	0.0616	0.0158	0.1056	0.006	0.0063
ResidFamSizeR	-0.0079	-0.0087	-0.0543	0.0271	0.708	0.7075
SubjectSexM	-0.0385	-0.0370	-0.2782	0.2251	0.778	0.7580
NextSkipped	0.1051	0.1047	0.0711	0.1362	0.001	0.0000
SuffixTypeSt:AverageBigramFreq	-0.0623	-0.0604	-0.1257	0.0029	0.066	0.0636
SuffixTypeUs:AverageBigramFreq	-0.0821	-0.0810	-0.1442	-0.0171	0.010	0.0114
ResidLeftFreq:SuffixTypeSt	-0.0538	-0.0538	-0.0896	-0.0109	0.006	0.0076
ResidLeftFreq:SuffixTypeUs	0.0230	0.0228	-0.0186	0.0575	0.228	0.2028
ResidFamSizeL:SuffixTypeSt	-0.1233	-0.1239	-0.1987	-0.0574	0.002	0.0007
ResidFamSizeL:SuffixTypeUs	0.0206	0.0206	-0.0419	0.0760	0.452	0.4881
WordFreq:ResidFamSizeR	0.0535	0.0533	0.0257	0.0854	0.002	0.0005
TrialNum:SubjectSexM	-0.0007	-0.0007	-0.0010	-0.0003	0.001	0.0001

Table 8: Random effects for *FirstFixDur*, *SubgazeLeft*, *SubgazeRight* and *GazeDur*

A. First fixation duration

Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.015	0.025	0.011	0.045
Subject	0.106	0.114	0.084	0.156
Subject by Nomore	0.068	0.025	0.083	0.156
Residual	0.265			

B. Subgaze duration for the left constituent

Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.104	0.104	0.085	0.130
Subject	0.195	0.198	0.151	0.271
Residual	0.446			

C. Subgaze duration for the right constituent

Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.009	0.012	0.003	0.044
Subject	0.168	0.171	0.129	0.227
Residual	0.368			

D. Gaze duration

Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.113	0.114	0.095	0.139
Subject	0.298	0.303	0.233	0.398
Residual	0.394			

Figure 1: Interaction of compound frequency by (residualized) left constituent family size for right subgaze duration. The lines plot the effect of compound frequency for the quantiles of left constituent family size (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 even reverses to the positive direction for the largest left constituent families, the 5th quantile (longdash line).

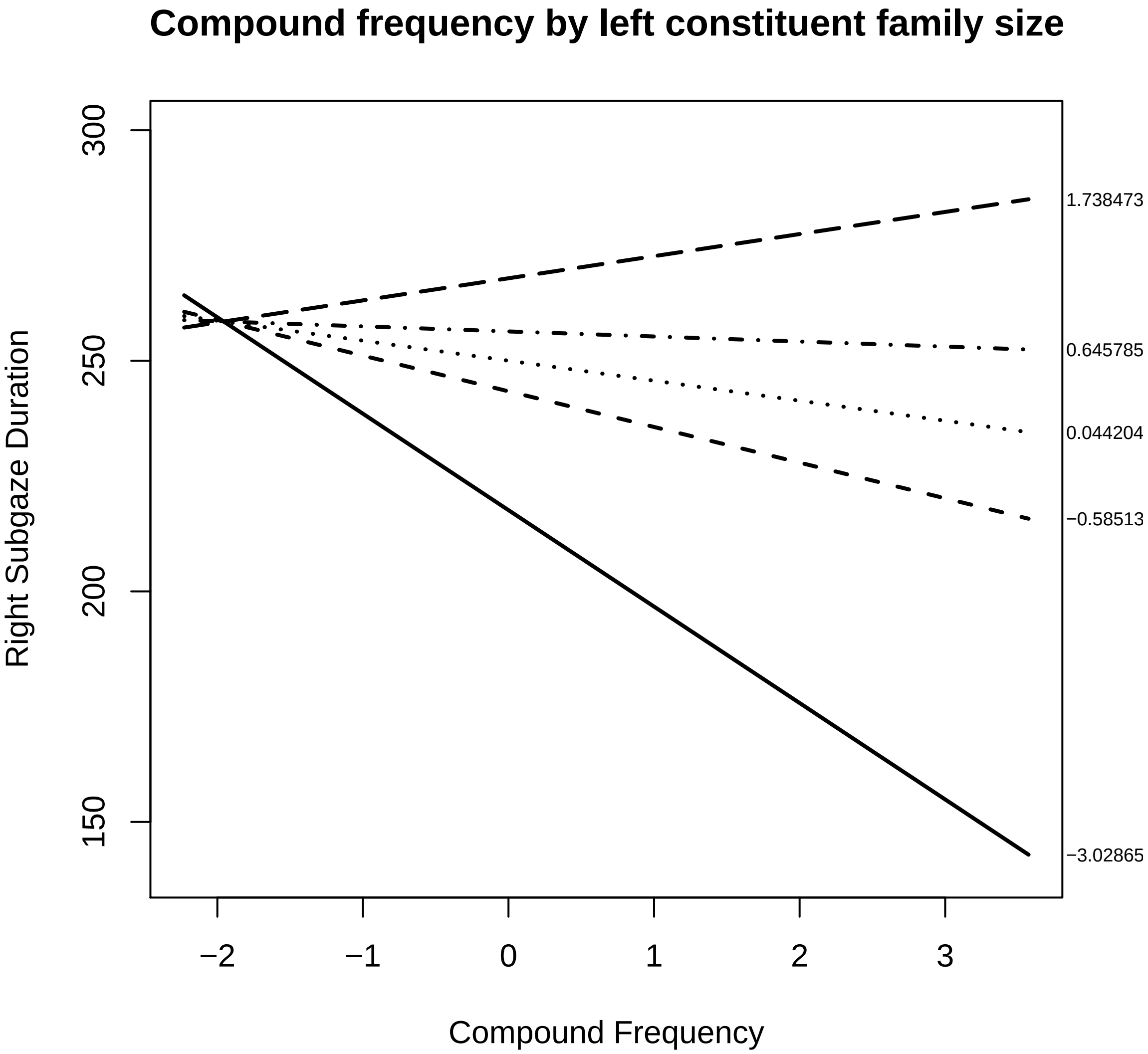
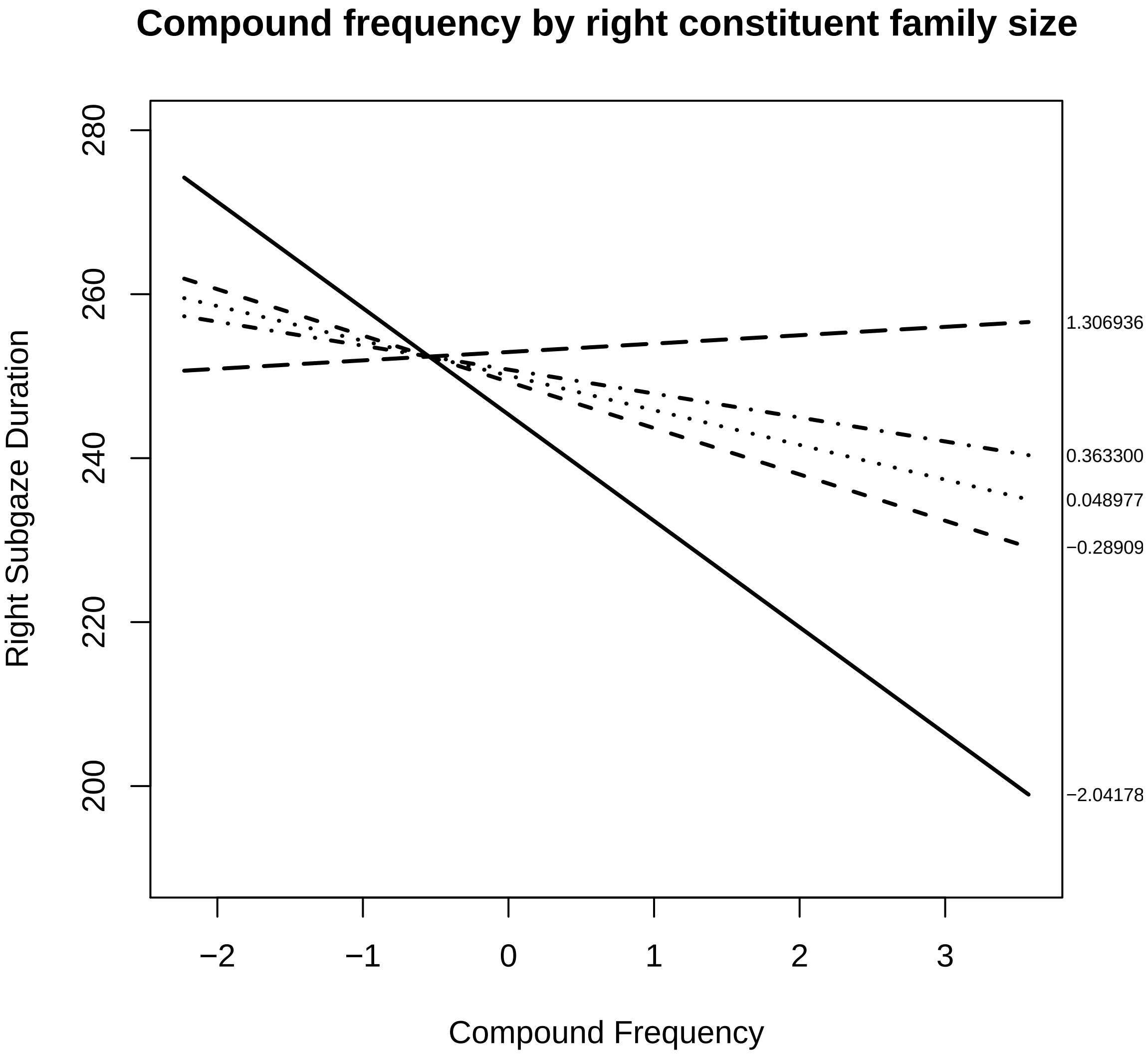


Figure 2: Interaction of compound frequency by (residualized) right constituent family size for right subgaze duration. The lines plot the effect of compound frequency for the quantiles of left constituent family size (quantile values provided at the right margin). Compound frequency comes with the strongest 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 even reverses to the positive direction for the largest left constituent families, the 5th quantile (longdash line).



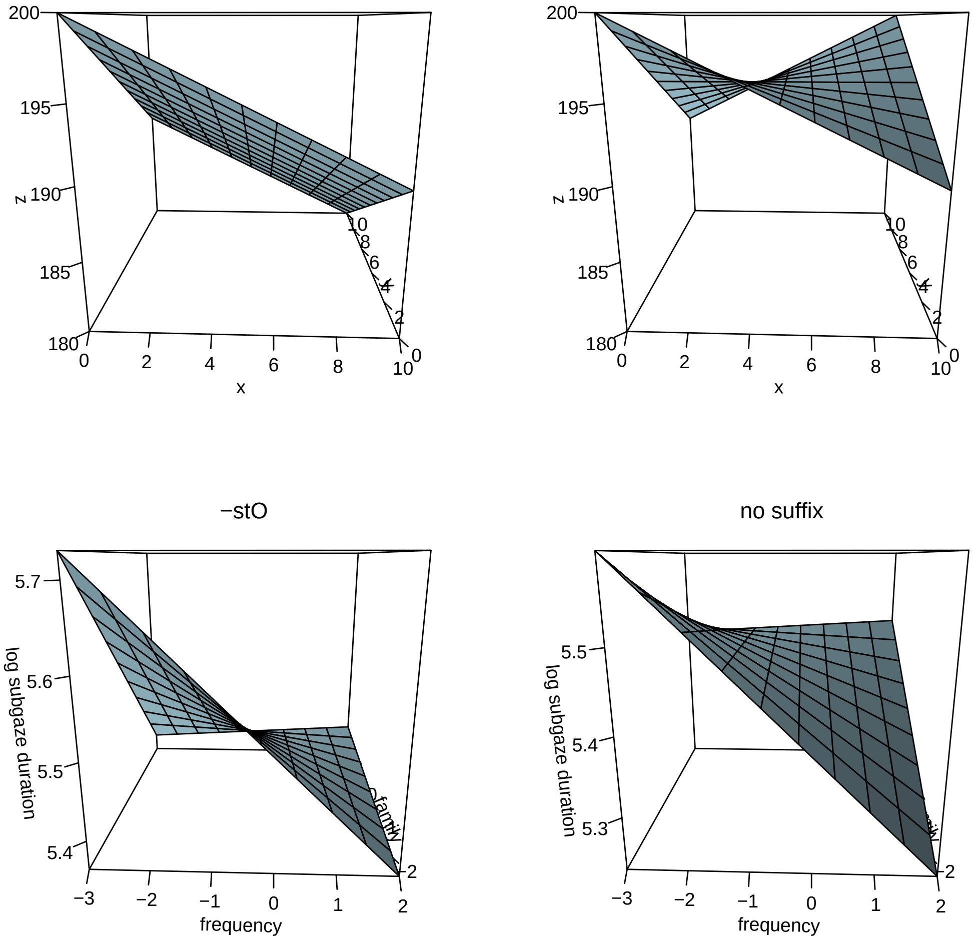


Figure 3: Perspective plots for (upper left panel) a linear model with additive main effects and no interaction and for (upper right panel) a linear model with a multiplicative interaction ($\beta_0 = 200, \beta_1 = -1, \beta_2 = -1$, for the left panel, $\beta_3 = 0$, for the right panel, $\beta_3 = 0.2$). The lower panels show the interaction of left constituent family size and compound frequency for the right subgaze durations for compounds with left constituents ending in the suffix *-stO* (left panel) and compounds with simplex left constituents (right panel).