Processing Trade-Offs in the Reading of Dutch Derived Words

Victor Kuperman*

Stanford University, USA

Raymond Bertram

University of Turku, Finland

R. Harald Baayen

University of Alberta, Canada

October 2, 2009

*Corresponding author: Victor Kuperman, Department of Linguistics, Stanford University, Stanford, CA 94305-2150 USA E-mail: vickup@stanford.edu. Phone: 1-619-5697220. Fax: 1-650-7990650

Abstract

This eye-tracking study explores visual recognition of Dutch suffixed words (e.g., *plaats+ing* "placing") embedded in sentential contexts, and provides new evidence on the interplay between storage and computation in morphological processing. We show that suffix length crucially moderates the use of morphological properties. In words with shorter suffixes, we observe a stronger effect of full-forms (derived word frequency) on reading times than in words with longer suffixes. Also, processing times increase if the base word (*plaats*) and the suffix (*-ing*) differ in the amount of information carried by their morphological families (sets of words that share the base or the suffix). We model this imbalance of informativeness in the morphological families with the information-theoretical measure of relative entropy and demonstrate its predictivity for the processing times. The observed processing trade-offs are discussed in the context of current models of morphological processing.

Keywords: lexical processing, eye movements, derived words, information theory, morphology

One of the objectives of psycholinguistic research on comprehension of morphologically complex words is to establish how the balance of storage (i.e., memorizing and recognizing complex words as unstructured units) and computation (i.e., decomposition of complex words into morphemes, from which the meanings of whole words are computed on-line) affects lexical processing. This research field develops in two major directions: (a) it explores what properties of complex words and their morphemes shift the balance between storage and computation towards one of these processing strategies, and (b) it models to what extent and in what relative order these strategies are employed in complex word recognition. The present paper aims at adding to the current knowledge on both topics by reporting an eye-tracking regression study of Dutch derived words embedded in sentential contexts.

There appears to be a consensus in the current literature that lexical processing of derived words is sensitive to characteristics of both derived words as wholes (e.g., *government*), and their morphological bases (e.g., *govern*) and affixes (e.g., *-ment*). To give only a few examples, Niswander, Pollatsek and Rayner (2000) observed the effects of suffixed word frequency and base frequency in the eye-movement record. Several chronometric and eye-tracking studies in several languages showed that (mostly, low-frequency) derived words elicit shorter processing times and come with lower error rates if they include higher-frequency bases (e.g., Beuavillain, 1996; Burani & Caramazza, 1987; Holmes & O'Regan, 1992; Schreuder, Burani & Baayen, 2003). Also, the family size and family frequency of the morphological base[1]

[1]A morphological family is a set of words that shares a constituent, for instance, the base family of *happiness* includes *happily*, *happier*, *happiest*, while the suffix family of *happiness* includes *goodness*, *madness* and *abruptness*. Family size is the count of members of the family, while family frequency is a cumulative

proved to co-determine processing costs of derived word recognition (e.g., Bertram, Baayen & Schreuder, 2000; Colé, Beauvillain & Segui, 1989), such that base families with many or more frequent members elicited shorter response latencies.

Furthermore, Laudanna and Burani (1995) proposed a conceptual framework for exploring the role of affixes by defining the notion of affixal salience, namely, the likelihood of recognizing the affix as a processing unit in its own right. The claim is that the more perceptual salience an affix has, the more it stands out of its embedding word, and the more biased lexical processing is towards morphological decomposition and towards using the properties of the base and the affix for the identification of the complex word. A wide range of studies have proposed dimensions that increase affixal salience and elicit different processing times for words with salient versus non-salient affixes, or give way to interactions of affixal salience with derived word frequency and base frequency. These dimensions include: orthographic properties of affixes (e.g., affix length, affixal confusability and transitional probabilities of n-grams near the morphemic boundary, e.g., Andrews & Davis, 1999; Laudanna & Burani, 1995), their phonological and phonotactic properties (e.g., co-occurrence probabilities of n-phones and of discontinuous patterns across the morphemic boundary, e.g., Bertram, Pollatsek & Hyönä, 2004; Hay & Baayen, 2003), and their lexical properties (e.g., word formation type of the affix, existence of inflectional allomorphs or homonyms for the affix, cf. Baayen, 1994; Bertram, Schreuder & Baayen, 2000; Bertram, Laine, Karvinen, 1999; Bertram, Laine, Baayen, Schreuder, & Hyönä, 2000; Järvikivi, Bertram & Niemi, 2006; Sereno & Jongman, 1997). Another important dimension of affixal salience is the distributional properties of

frequency of occurrence of family members.

affixes: Words that embed affixes which occur in a larger number of different, frequent or new words tend to be processed faster. For instance, Baayen, Wurm and Aycock (2007) and Plag and Baayen (2009) report the reduced lexical decision and naming latencies for English suffixed words with relatively productive suffixes, that is, suffixes occurring in a larger number of word types (see, however, Burani & Thornton, 2003 for a null effect of suffix frequency, the number of word tokens in which a suffix occurs).

While the accumulated knowledge on derived word processing is extensive, it predominantly originates from experiments that considered only a small number of morphemes and only a small number of predictors at a time (often experimenting on pairs of morphemes differing in only one dimension). That is, whether results obtained in many earlier studies generalize over the entire lexical space of suffixed words in a language is an open question. Furthermore, the vast majority of studies on derived words used the lexical decision paradigm, which imposes a number of task-specific limitations on the patterns of results. First, lexical decision operates on single words and may conceivably give rise to spurious lexical ambiguities that vanish when the sentential context is present. Second, classification of stimuli as words or nonwords is not part of the regular comprehension of a written text. The experimental task may induce rather different kinds of processing strategies than those used for the natural integration of word meaning into the sentence and discourse. Third, single word recognition and recognition of words within sentences differs in the time-course of visual uptake, which may affect the time-course and efficiency of lexical processing. For instance, in sentence reading there is a possibility of a parafoveal preview benefit, such that partial information about a word is available even as the previous word is fixated (for a survey see Rayner, 1998). Also, repeated visual inspection of the word after having moved further into the sentence is obviously not an option in tasks involving single word recognition, but such regressions occur frequently in natural reading, for a detailed discussion of relative merits of the lexical decision and eye-tracking experimental paradigms see e.g., Andrews, Miller and Rayner (2004) and Frisson, Niswander and Pollatsek (2008).

Given the possibility of methodological limitations in the state-of-the-art research on derived words, the first goal of the present study is to establish empirically which properties of complex words and their morphemes will emerge as significant contributors to recognition of derived words, when considered across a broad range of Dutch suffixed words and pitted against a variety of other such properties and control variables in a relatively naturalistic task of sentence reading for comprehension. We opt for a large number of suffixes and bases to ensure variability in the formal, semantic and distributional properties of our derived words and thus, to improve generalizability of our findings. We use step-wise multiple regression mixed-effects modeling with participants and items as crossed random effects (cf., Baayen, 2008; Bates & Sarkar, 2007; Pinheiro & Bates, 2000) as a statistical technique. These models allow one to consider many predictors simultaneously and to test the relative strength of their effects over and beyond the effects of other variables. We expect to find that some of the factors that earlier studies claim to affect the recognition of derived words will fail to elicit significant effects due to the disambuiguating role of the context and to the more ecologically valid time-course of visual information uptake. As argued in Gries

(2003), the burden of interpretation for this kind of research is two-fold: First, we need to explain how the important contributors to visual recognition of derived words affect the effort of processing of such words, and second, we need to show why some of the proposed predictors of complex word processing play no role in our study. By embedding derived words in sentences we aim at avoiding possible confounds of lexical decision as the experimental task. Also, we make use of eye-tracking as an experimental technique, which is ubiquitous in current research on word recognition, as the eye-movement record has good ecological validity as an approximation to the process of natural reading (cf. Richardson, Dale, & Spivey, 2007). Recent studies employing both lexical decision and eye-tracking techniques demonstrated that the latter paradigm is at least as sensitive to the morphological structure of complex words as lexical decision is (cf. Juhasz, Starr, Inhoff, & Placke, 2003; Inhoff, Starr, Solomon, & Placke, 2008; Kuperman et al., 2009).

The common assumption of the literature on morphological modeling is that effects of whole complex words (e.g., derived word frequency) serve as a diagnostic of full-form access, while effects related to morphemes point at the process of decomposition. Hence, the evidence for the simultaneous contribution of whole words and morphemes indicates that both storage and computation may be at stake in complex word recognition. The second, related, goal of this paper is to establish which current models of morphological processing make predictions compatible with the processes of storage and computation that are revealed in experimental measures of derived word recognition. Each of the options (pure storage, pure computation and the joint use of storage and computation) has been instantiated in models of morphological processing. Obligatory decomposition of complex words characterizes the sublexical models of morphological processing proposed, for instance, in Taft and Forster (1975) and Pinker (1999). These models advocate obligatory initial decomposition of derived words into bases (*govern*) and affixes (*-ment*) and subsequent lexical access to full-forms (*government*) via recombination of lemmas associated with morphemes, see also Fiorentino and Poeppel, 2007 and Taft, 2004. A word as a whole may have its own lemma if its meaning cannot be fully computed from the meanings of constituent morphemes. The ease of access to respective lemmas is modulated by frequencies of morphemes and the whole word. Conversely, obligatory full-form access is a hall-mark of supralexical models (e.g., Giraudo & Grainger, 2001). Supralexical models claim that full-forms are activated first, while the activation of morphemes is only attributed to the post-access processing stage. Furthermore, Diependaele, Sandra and Grainger (2005) come up with two distinct systems (morpho-orthographic and morpho-semantic) subserving both the sublexical and supralexical routes of lexical access. This is in the same spirit as dual-route parallel models (e.g., Baayen & Schreuder, 1999; Baayen & Schreuder, 2000; Schreuder & Baayen, 1995), which allow for morphological decomposition and full-form access to operate jointly. On these models, bases and affixes of the derived words are activated to the extent that lexical processing makes use of the decomposition route.

Recent eye-tracking studies of Dutch and Finnish polymorphemic compounds (Kuperman, Bertram & Baayen, 2008; Kuperman, Schreuder, Bertram & Baayen, 2009) have shown that the complexity of morphological processing may not be fully captured by single-route and most dual-route models proposed in the literature. For instance, in both studies an early effect of the compound's frequency of occurrence on reading times (e.g., Dutch compound *oorlogsverklaring* "declaration of war") preceded the inspection of the compound-final characters and of compounds' right constituents (e.g., *verklaring* "declaration"). This finding is at odds with strictly sublexical models of morphological processing. Moreover, Juhasz (2008) and Kuperman et al. (2008; 2009) report simultaneous effects of compound frequency and the compound's left constituent frequency in short English compounds, and long Finnish and Dutch compounds, respectively. That the effect of the compound's left constituent (e.g., *oorlog* "war") on eye-movement measures was simultaneous with, rather than followed in time, the effect of the compound frequency cannot be easily handled by supralexical models of morphological processing. Another piece of evidence that is problematic for single-route (sublexical or supralexical) models is the presence of interactions between properties of complex words and those of their constituents. For instance, the eye-tracking study of Niswander-Klement and Pollatsek (2006) reported interactions of prefixed word length with word frequency and base frequency in English. Similarly, Baayen, Wurm and Aycock (2007) observed an interaction between whole word frequency and base frequency in English (prefixed and suffixed) derived and inflected words, using the visual lexical decision data. Moreover, in an auditory lexical decision study, Winther Balling and Baayen (2008) found an interaction between derived word frequency and suffix frequency in Danish. Finally, even though interactions described above can be handled by dual-route models (as we discuss below), recent evidence on reading of novel prefixed words (Pollatsek, Slattery & Juhasz, 2008) also disconfirms the categorical "winner-takes-it-all" architecture implemented in "horse race" dual-route models of parallel morphological processing (e.g., Schreuder & Baayen, 1995).

To reiterate, the second goal of this paper is to establish whether predictions of single-, dual- and multiple-route models of morphological processing can adequately fit the range of effects that morphological structure shows in visual recognition of derived words in Dutch (e.g., *plaats+ing* "placing"). In particular, we focus on the kind of effects that emerged in the studies cited in the preceding paragraph and ran counter to the assumptions of sublexical, supralexical and "horse race" dual-route models, namely, interactions between properties of complex words and their constituents, and effects of constituents' morphological families. We expect to see that current models of morphological processing may be too restrictive in their assumptions of obligatory sequentiality and autonomy of processing routes, and that the balance between storage and computation in the processing of derived words is best captured by a multiple-route model, such as advocated in Kuperman et al. (2008; 2009).

Method

Participants

Twenty-eight students of the Radboud University Nijmegen (21 females and 7 males) participated in this experiment for the reward of 6 euros. All were native speakers of Dutch and had normal or corrected-to-normal vision.

Apparatus

Eye movements were recorded with an EyeLink II head-mounted eyetracker manufactured by SR Research Ltd. (Canada). The eye-tracker samples pupil location and pupil size at the rate of 500 Hz. 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 17-inch computer screen, which had a refresh rate of 60 Hz.

Materials

The set of target words included 156 Dutch bimorphemic words (e.g., *president+schap* "presidency") ending in one of the following derivational suffixes: *-achtig*, *-baar*, *-dom*, *-er*, *-erig*, *-erij*, *-es*, *-heid*, *-ig*, *-ing*, *-lijk*, *-loos*, *-nis*, *-schap*, *-sel*, *-ster*, *-te*, *-vol*, and, *-zaam* (3 to 12 words per suffix). These nineteen suffixes were selected for inclusion in our study since they are reasonably productive in modern Dutch and belong to the Germanic stratum[2]. To raise the likelihood that our target words were fixated, and not skipped, during reading, we set the minimum length of those words to 8 characters (range = 8-14, mean = 9, SD = 1.3). This range of word lengths is comparable to the ranges used across many studies of derived words (cf. 7-12 characters in Niswander et al., 2000; 6-11 characters in Burani & Thornton, 2003, etc.).

Each target word was embedded without further inflectional suffixes into a separate sentence, and it never occupied the sentence-initial or sentence-final position. The experimental list also included 136 filler sentences with a different experimental manipulation: Analyses of these sentences are not reported here. All sentences comprised 6-17 words (mean = 11.2 words, SD = 2.2) and took up at most one line on the screen. The sentences were displayed

[2] Latinate affixation is marginal in Dutch as compared to English, and has been argued to be unproductive (cf. Van Marle, 1985).

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 80 cm, one character space subtended approximately 0.36^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 15% of sentences were followed by 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.

Since sentential contexts were different for each target word, we conducted an additional plausibility rating experiment. Seventeen raters, none of whom participated in the main experiment, were presented with the fragment of the sentence preceding the word (e.g., *Ze luistert...* "She listened") and the target word (*ademloos* "breathless(ly)"), and they were asked to evaluate the plausibility of the target word in this context on the scale from 1 "completely implausible" to 7 "extremely plausible". The mean plausibility ratings ranged from 1.9 to 6.7. We considered these ratings as one of our contextual predictors in the regression models.

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

We used well-established measures of eye-movements to estimate the reading behavior of our participants. The duration of the single fixation landing on the target word (*SingleDur*) and gaze duration (the summed duration of all fixations on the target word before fixating away from it, *GazeDur*) provided most insight into the reading of derived words. Both measures serve as an index of the processing load at the first encounter with the derived word. Other dependent variables considered in this study included measures associated with early lexical processing (initial fixation position, duration of the first fixation and the amplitude of the first within-word saccade), measures associated with global processing costs (probability of a single fixation on the word, and the total number of fixations on the word in the first-pass), and finally, measures associated with integration of word meaning in the sentence (total reading time and total number of fixations on the word). All durational measures were log-transformed to reduce the influence of atypical outliers.

Predictors

The full list of predictors considered in this study is presented in Appendix 1, along with ranges, means and median values for numerical predictors (see Table 1). In what follows we describe predictors of primary interest for this study.

Distributional predictors: Previous research has shown that lexical processing of derived words is codetermined by the distributional properties of those words, as well as by the properties of their bases and affixes (cf., e.g., Bertram et al., 2000; Niswander-Klement et al., 2000; Plag & Baayen, 2009). There is a considerable number of similar lexical-statistical measures that attempt to operationalize the intuition that more productive suffixes are easier to parse out; these are surveyed in Hay and Baayen (2002), see also Plag and Baayen (2009) and Baayen, Wurm and Aycock (2007). In this study, we opted for simple measures of productivity, the morphological family size and family frequency (henceforth *SuffixProd* and *SuffixFreq*) of the suffix: the number of word types in which the suffix occurs, and their cumulated frequencies. The use of these productivity measures allows us to compare and evaluate base family size and suffix family size (productivity) along similar lines. All frequency-based measures described here and in the remainder of the section were transformed logarithmically (base e) to decrease the influence of atypical outliers.

Higher frequencies and larger morphological families of constituents in compounds and derived words tend to increase the speed of visual recognition (Andrews, Miller & Rayner, 2004; De Jong, Schreuder & Baayen, 2000; Hyönä, Bertram & Pollatsek, 2004; Juhasz, Starr, Inhoff & Placke, 2003; Pollatsek, Hyönä & Bertram, 2000). Distributional properties of base words and derivations as whole words were estimated using the following variables:

base word frequency, *BaseFreq* (lemma frequency of *president* in *presidentschap*), family size of the base, *BaseFamilySize* (the type-based count of derived words in which the base (*president*) occurs in the word-initial position)[3], and frequency of the whole derived word, *WordFreq* (e.g., lemma frequency of *presidentschap*). Computation of these distributional measures was based on the combined pool of roughly 120 million tokens, obtained from the CELEX lexical database (Baayen, Piepenbrock & Gulikers, 1995) and from the newspapers in the Twente News Corpus (Ordelman, 2002). All lemma frequency measures were collapsed over inflectional variants (*cat*, *cats*, *cat's* and *cats'*).

Other predictors: Word length is a robust predictor of reading times (cf. Rayner, 1998), while the length of the suffix is one of the proposed dimensions of affixal salience (Bertram, Laine & Karvinen, 1999; Laudanna & Burani, 1995). We took into consideration the lengths of derived words and suffixes, as measured in characters, phonemes and syllables. Since plausibility of words in context affects the time it takes to read those words (e.g., Rayner, Warren, Juhasz & Liversedge, 2004), we took into account plausibility ratings obtained from the separate norming experiment.

We also considered a broad range of predictors that were proposed in the literature as codeterminers of affixal salience, including homonymy (whether or not suffixes can serve multiple syntactic functions, as the English suffix *-er* in *warmer* and *builder*), confusability (the ratio of word types in which the character string functions as a suffix and all word types ending in that character string), structural invariance (whether or not suffixes change

[3] Base family frequency (the token-based count of derived words in which the base *president* occurs in the word-initial position) did not reach significance in any of our models.

their orthographic form across inflectional paradigms), as well as frequencies of bigrams preceding, following and straddling the morphemic boundary in the derived words. None of these latter predictors reached significance. We provide possible explanations for the discrepancies between our findings and the previously reported role of these predictors in the Results and Discussion section.

Statistical considerations:

In this study we made use of mixed-effects multiple regression models with participant and word as random effects (cf., Baayen, 2008; Baayen, Davidson & Bates, 2008; Bates & Sarkar, 2007; Pinheiro & Bates, 2000). Unless noted otherwise, only those fixed effects are presented below that reached significance at the 5%-level in a backwards stepwise model selection procedure. 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 -3.0 to to 3.0 units of SD of the residual error of the model. Once outliers were removed, the models were refitted.

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 log 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).

Several of the measures we considered showed strong pair-wise correlations. We attenuated collinearity in the data exploration phase using several methods, including standardization of continuous predictors (subtracting the mean and dividing by two standard deviations to enable the comparison of discrete and continuous predictors on the same scale; for benefits of this operation for models with interactions, see Gelman & Hill, 2007). Ultimately, collinearity was not a problem for the the set of (standardized) predictors that showed significant effects on reading times in our final statistical models: the values of the variance inflation factor were below 2 for each predictor in the single-fixation duration and gaze duration models. Also, the condition number κ in the final models was below 10, which indicates relatively low collinearity. We note that plots below are presented for non-standardized, raw values of predictors and dependent variables to ensure interpretability and comparability with earlier studies.

Results and Discussion

The initial pool of data points comprised 6672 fixations. We removed fixations that were shorter than 50 ms and longer than 1,000 ms (201 fixations, 3%). Subsequently, fixations that bordered microsaccades (fixations falling within the same letter) were removed (28 x 2 = 56 fixations, 0.8%). There were 4,916 valid fixations pertaining to the first-pass reading (i.e., the sequence of fixations made before the fixation is made outside of the word boundaries, 77% of the original dataset). A negligible percent of the target words was skipped (< 0.1%). Eighty-three percent of the target words elicited exactly one fixation, 16% elicited exactly two fixations, and only 1% elicited more than two fixations. The average number of fixations on a stimulus was 1.2 (SD = 0.4). The average fixation duration was 229 ms (SD = 64), and the average gaze duration was 262 ms (SD = 93). All participants responded correctly to at least 90% of the comprehension questions.

Since the majority of target words elicited exactly one fixation in the first-pass reading, the analyses for gaze duration, first fixation duration and single-fixation duration yielded very similar results. Analyses of the measures associated with global processing costs (e.g., total reading time) did not provide any additional insight into questions of our interest. We opted for providing in Appendix 1 full specifications of the models for single-fixation duration (3,267 data points, see Tables 2 and 3) and for gaze duration (3,950 datapoints, Table 4). Specifications for all models include estimates of the regression coefficients; highest posterior density intervals (HPDs), which are a Bayesian measure of confidence intervals; p-values estimated by the Monte Carlo Markov chain (MCMC) method using 10,000 samples; and p-values obtained with the t-test for fixed effects using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom (for a detailed treatment of the method, see Baayen, 2008; Baayen, Davidson & Bates, 2008; Pinheiro & Bates, 2000). For the effects reported in the body of the paper, rather than in Appendix 1, we provide beta coefficients and p-values estimated by the MCMC method using 10,000 samples.

Oculo-motor, orthographic and contextual predictors: Longer words preceding the fixated word, *PrecLength*, induced longer single-fixation durations (with a 30 ms difference between the shortest and the longest preceding word). Additionally, the initial fixation position (*InitFixPos*) shows a well-attested inverse-U shape relationship with single-fixation duration, such that reading times are longest when the initial fixation lands around the center of the word and reading times decrease if the initial fixation is closer to one of the word's extremes. Vitu, Lancelin & Marrier d'Unienville (2007) explain this so-called Inverted Optimal Viewing Position effect by claiming that fixations tend to be longer when the eyes are at locations in stimuli in which greater amounts of information are anticipated. Higher plausibility of the derived word given preceding context came with shorter single-fixations (a 20 ms decrease between the lowest and the highest plausibility ratings). This is in line with previous studies on the facilitatory role of contextual plausibility in reading (e.g., Rayner, Warren, Juhasz & Liversedge, 2004).

Main predictors of interest: Higher-frequency derived words elicited shorter single-fixations (a 20 ms decrease between the word with the lowest and the highest frequency) as a main effect. Moreover, *WordFreq* entered into a significant interaction with the length of the suffix in phonemes (*SuffixLength*) ($p = 0.046$, see Table 2), such that the effect of *WordFreq* on single-fixation durations was strongest for derived words with shortest suffixes, gradually weakened in derived words with phonologically longer suffixes and virtually vanished in words with the longest suffixes, see the conditioning plot in Figure 1A. Apparently, the more complexity there is in the phonological representation of the suffix, the more perceptually salient it is in the derived word and the less biased readers are towards using the properties of the complex word as a whole as processing cues. There was no main effect of suffix length (measured in phonemes or characters) on reading times.

[Figure 1 about here.]

The orthographic lengths of the base and the suffix in our target words correlated negatively and with a marginal significance at $r = -0.40, p = 0.06$. It might be that the observed effect of suffix length masked the effect of base length. We ruled out this possibility by including base length in the model for single fixation duration which did not have suffix length as one of its predictors, and – separately – by including residualized base length (base length from which the influence of suffix length was initially partialled out) along with suffix length as predictors in the model for single fixation duration. In neither model did base length elicit an effect below the 0.05-threshold of significance.

Neither the frequency nor the family frequency of the derivation's base word codetermined single fixation durations on derived words. However, base family size entered into a statistically significant interaction with the measure of suffix productivity, i.e., the type count of words in which the suffix occurs[4], i.e., *SuffixProd*, $p = 0.039$, see Figure 1B.

The interaction indicates that a large base family size came with longer reading times in those derived words that embedded low-productivity suffixes (i.e., suffixes that could only combine with a small number of bases). Furthermore, the base family size effect reversed in words with suffixes that were relatively productive (i.e., could combine with a large number of bases). In other words, we observed an interaction of two morphological families, the one reflecting combinability of the base word (e.g., *happy*) with suffixes (*-ly, -ness, -less, -lessness*) and the other reflecting the ability of a given suffix (e.g., *-able*) to attach to a range

[4]We use the count of word types (or suffix family size) as the measure of suffix productivity here. However, the interactions with base family size retain significance, even if we use — as alternative measures of suffix productivity — the count of hapax legomena in which the suffix occurs, or its growth rate (for detailed definitions, see Hay & Baayen, 2004). We opted for reporting here only one of alternative measures.

of base words (e.g., *love, dispense, expand*). When both families are similar in size (both are small or both are large), the processing costs are minimal. The costs increase, however, if there is a substantial discrepancy in the sizes of the two families. This interaction appears to reflect a conflict between base family size and suffix productivity (i.e., suffix family size). If the two constituents diverge in their family sizes, they also diverge in the amount of information they carry. This imbalance in the informativeness of morphological constituents may slow down the parsing (decompositional) route and result in inflated reading times. If this interpretation is correct, we should be able to replace three terms in our regression model (two main effects and a multiplicative interaction) by a single measure for the degree of imbalance between the family of the base and the family of the suffix. Moreover, this single measure should provide a better fit to the data. In what follows, we show that information theory provides us with exactly the right measure, relative entropy.

Before defining the measure of relative entropy, we note that the statistical model for the single fixation duration (Table 2) reveals a significant three-way interaction of base family size, *BaseFamilySize*, by suffix productivity, *SuffixProd*, by suffix length in phonemes, *SuffixLength*. This interaction is such that the two-way interaction of *BaseFamilySize* by *SuffixProd* presented in Figure 1B is observed for derived words with longer suffixes and is attenuated in words with shorter suffixes. If relative entropy is the single measure that can replace the two-way interaction of base family size by suffix family size, we expect relative entropy to also interact with suffix length.

Relative entropy

Relative Entropy, also known as Kullback-Leibler divergence and as information gain,

$$\mathrm{RE}(P||Q) = \sum_{i}^{n} p_i \log_2 \frac{p_i}{q_i}, \tag{1}$$

gauges how similar two probability distributions

$$\begin{aligned} P &= \{p_1, p_2, \ldots, p_n\} \\ Q &= \{q_1, q_2, \ldots, q_n\} \end{aligned}$$

are (for an overview of applications of relative entropy in morphological processing see Milin, Kuperman, Kostić & Baayen, 2009). Relative entropy equals zero when the two probability distributions P and Q are identical, and it increases as the two distributions become more and more dissimilar. The P and Q distributions that we need for modeling the family size interaction have two elements each ($n = 2$), namely, the type-based probability of a word belonging to the family, and its complement. Let $\mathcal{S}$ denote the family of the suffix, and let $\mathcal{B}$ denote the family of the base. The distributions P_w and Q_w for a word w can now be defined as

$$\begin{aligned} P_w &= \{p_1 = \mathrm{Pr}(\mathrm{w} \in \mathcal{S}), p_2 = 1 - p_1\}, \\ Q_w &= \{q_1 = \mathrm{Pr}(\mathrm{w} \in \mathcal{B}), q_2 = 1 - q_1\}. \end{aligned}$$

We estimate the probabilities p_1 (q_1) as the number of derived word types in which the suffix (base) occurs, divided by the total number of types in CELEX (roughly, 40000). For the suffix *-ing*, for instance, $p_1 = 233/40000 = 0.006$ and $p_2 = 1 - p_1 = 0.994$. For the base *plaats*, $q_1 = 120/40000 = 0.003$ and $q_2 = 0.997$. Hence the relative entropy for the two morphological families of the word *plaats+ing* equals

$$\begin{aligned} \mathrm{RE}(P||Q) &= 0.006 * \log_2\left(\frac{0.006}{0.003}\right) + 0.994 * \log_2\left(\frac{0.994}{0.997}\right) \\ &= 0.006 - 0.004 = 0.002. \end{aligned} \quad (2)$$

In our analysis, we multiplied the relative entropies calculated for the words in our data by 100 to bring them to a similar scale as the other predictors in the model.

Figure 2 illustrates the difference between the multiplicative interaction in our current regression model (left panel) and relative entropy (right panel).

[Figure 2 about here.]

The left panel plots, for each combination of suffix and base family size, the weighted contribution of the two main effects and their interaction to the estimate of the fixation (or gaze) duration. The minima of this joint function occur when the marginals are both at their minima or both at their maxima. The maxima of the joint function are found when one marginal is at its mimimum and the other marginal is at its maximum. In between we find a surface with intermediate values in the center. Note that the conditioning plot in Figure 1B is a two-dimensional representation of the left panel of Figure 2.

The right panel of Figure 2 illustrates the surface defined by relative entropy. The X-axis shows $p_1 = \mathrm{Pr}(\mathrm{w} \in \mathcal{S})$, abbreviated in the figure to 'p(suffix)'. Likewise, the Y-axis shows $q_1 = \mathrm{Pr}(\mathrm{w} \in \mathcal{B})$, abbreviated in the figure to 'p(base)'. The complement probabilities p_2 and q_2 are fully determined given p_1 and q_1, and are therefore not shown. We see that the minima and maxima of the relative entropy surface are located at the same combinations of minima and maxima of its marginal distributions as observed in the left panel for the surface defined by the multiplicative interaction. However, there is now a flat valley between the two minima, instead of a valley with a higher pass in between the two minima, as in the left panel.

The question that we now have to address is whether the entropy surface shown in the right panel is a better predictor for fixation durations than the multiplicative surface illustrated in the left panel. We therefore included relative entropy as a predictor in our statistical model for single fixation duration.

First, we compare a model for single fixation durations with *RE* as a predictor to the model with the two-way interaction *BaseFamilySize* by *SuffixProd* as predictors. We observe that *RE* is indeed a highly significant predictor of the reading time ($\hat{\beta} = 0.035, SE = 0.012; p = 0.003$), and its regression coefficient is positive, as we anticipated. Moreover, the interaction of base family size by suffix productivity loses its significance in the new model, and so do the main effects of base family size and suffix productivity (all ps > 0.1). This suggests that the relative entropy measure absorbs the variance in the data previously explained by the other predictors. Furthermore, *RE* retains significance when the non-significant interaction and the main effects are removed from the model. Importantly, the model with *RE* as a predictor shows a better performance than the model with the two-way interaction in that it has a better (greater) value of log-likelihood (which is a measure of the model's fit to the data), -848 vs. -858, whereas it uses less parameters. We conclude that *RE* allows us to fit superior models to the data and thus, as a measure of imbalance in the size (informativeness) of the families, relative entropy is preferrable to the multiplicative interaction of the base family size and suffix productivity.

Second, we test whether relative entropy interacts with suffix length: the possibility of this interaction is suggested by the significant three-way interaction of *BaseFamilySize* by *SuffixProd* by *SuffixLength*, Table 2. Indeed, we observe an interaction of relative entropy by suffix length, such that the effect of relative entropy is at its strongest in words with long suffixes and weakens in words with shorter suffixes, Figure 1C. Also, the interaction of derived word frequency by suffix length retains significance in our new model. Using the same criteria as described above, we come to the conclusion that the model with the interaction of *RE* by *SuffixLength* reported in Table 3 outperforms the model with the three-way interaction, Table 2. To summarize the statistical analyses, our final model for single fixation duration (Table 3) includes, among other significant predictors, crucial interactions of derived word frequency by suffix length (Figure 1A) and of relative entropy by suffix length (Figure 1C). These results are supported by the model for gaze duration, which replicates the significant interactions of both derived word frequency and relative entropy by suffix length, see Table 4. The superiority of relative entropy as a predictor illustrates the advantages of an information-theoretical approach to lexical processing, which motivated the measure that we hypothesized to tap into the imbalance of informativeness of the two morphological families. We discuss the implications of this imbalance for morphological processing in the General Discussion.

Discrepancies with earlier reports

Our results demonstrate that the processing of derived words in reading is codetermined by a constellation of interacting phonological, distributional, and orthographic properties of derivations and their morphological constituents. These findings allow us to delimit the large number of proposed predictors of lexical processing to only those that show robust effects when considered against the backdrop of multiple control variables and of multiple competing factors.

We see it as an important task to explain why we find no evidence for some of predictors of affixal salience proposed in literature (cf., Gries, 2003). Specifically, we address below such predictors as homonymy, confusability and structural invariance of suffixes. Several studies of derivation in English, Dutch and Finnish reported that derived words with homonymous suffixes (i.e., suffixes that can serve multiple syntactic or semantic functions, e.g., the English suffix -*er* in *warmer* and *builder*) tend to be processed as full-forms, rather than via their morphemes (e.g., Bertram, Laine, Baayen et al., 2000a; Bertram, Schreuder & Baayen, 2000b; Sereno & Jongman, 1997). In Dutch, two derivational suffixes from our list exhibit homonymy, -*er* and -*te*. Both suffixes form different word classes in their respective syntactic functions: adjectives in the comparative form versus agentive nouns, for the suffix -*er* (cf., *warmer* "warmer" and *werker* "worker"), and verbs in the past tense versus nouns, for the suffix -*te* (cf., *hoopte* "hoped" and *lengte* "length"). Experiments on Dutch derivations established that derived words ending in the homonymous suffixes -*er* and -*te* showed effects of whole word frequency, but no effects of base frequency on lexical decision latencies (Bertram, Schreuder & Baayen, 2000). Yet we found no interaction between homonymy and either whole word frequency or base frequency in our derived words. Possibly, the number of words ending in the suffixes *-er* and *-te* in our experimental list was too small to offer sufficient statistical power to the test. Another, perhaps more likely, explanation for the lack of a homonymy effect may arise from the presence of sentential context in our experimental stimuli, and its absence from the lexical decision studies. The sentential context preceding a complex word may offer strong syntactic cues as to what the expected class is for the word under identification (a noun or an adjective in the comparative; a noun or a past-tense verb) and, consequently, may allow the reader to anticipate the morphosyntactic function of the suffix, so that it does not cause problems in parsing. No such disambiguating cues are available in experimental paradigms where words are presented in isolation. This lack of contextual constraint may have given rise to ambiguities in word identification thus providing an advantage to full-form access. This advantage may in turn lead to task-specific differences in frequency effects reported for complex words with homonymous and non-homonymous suffixes.

Affix confusability (the ratio of word types in which the character string functions as an affix and all word types ending in that character string) has been argued to affect the balance between storage and computation in complex words, such that more confusable affixes are less salient and their processing is biased towards storage (proposed for prefixes by Laudanna & Burani, 1995). We observe no effect of confusability in our suffixed words and argue that previously reported effects may also be artefacts of the experimental presentation of words in isolation. Syntactic cues provided by the sentential context preceding the target word (for instance, word class) may greatly reduce the ambiguity of whether the word-final characters represent a suffix or not. To test this hypothesis, we considered the four Dutch suffixes with the largest confusability ratios (*-es*, *-te*, *-er* and *-nis*). We conditioned by word class the number of word types in which those character strings occurred in the word-final position. The resulting confusability ratios were reduced on average by a factor of 6.5. If readers anticipate word class given the preceding sentence fragment, the chances of confusing suffixes with non-morphological word endings are drastically reduced. This may explain the lack of the earlier reported effect in our data.

Finally, structural (in)variance, i.e., whether or not affixes change their orthographic form across inflectional paradigms, has been shown to influence reading of derived words in Finnish. The more allomorphs the suffix has, the slower its recognition proceeds (Järvikivi, Bertram & Niemi, 2006). The inflectional system of Dutch is much simpler than that of Finnish, and for nouns and adjectives considered here the main inflectional categories are Number (for nouns, e.g., singular *werk-er* "worker" vs. plural *werk-er-s* "workers") and Gender (for adjectives, common *bereikbare* vs. neuter *bereikbaar*). The most common change in the spelling of suffixes across inflectional forms is fully determined by regular spelling conventions for representing short and long vowels in different syllable types (cf. doubling of consonants in common gender *succesvolle* vs. neuter *succesvol*, or a vowel loss in the example with *bereikbare* above). The only two suffixes that are structurally variable are *-heid* (pl. *-heden*) and *-loos* (pl. *-lozen*). Since words with these suffixes do not come with slower reading times, we conclude that inflectional paradigms and the number of structurally variant suffixes are too small to elicit the effect observed in Finnish.

To sum up, using the distributional properties of the full-form, such as derived word frequency, can make the process of complex word recognition easier (faster) for the reader, and so can using the distributional properties of the word's morphemes, such as the morphological family sizes of the base word and the suffix. Relative entropy represents a difficulty that the lexical parser encounters under the imbalance in the informativeness of the word's morphemes. The relative entropy effect emerges side by side but independent of the effect of derived frequency (as a measure of how easily the full-form can be retrieved from the mental lexicon). Hence, we interpret relative entropy as a separate dimension of processing complexity that emerges during decomposition of words into morphemes.

Crucially, all the effects of morphological structure that we observe in the present study are significantly qualified by the measure of suffix length see Tables 2-4, and Figure 1A and C. That is, suffix length appears to serve as a key parameter that regulates the allocation of cognitive resources over available processing routes. Apparently, it fine-tunes the share of storage and computation in the processing of derived words. Words with extremely short and hence perceptually not salient suffixes do not provide a clear pointer to the reader that a complex word is at stake. This makes the full-form processing a preferred recognition route, which may be why the effect of derived word frequency on reading times is at its strongest in such words, while the effect of relative entropy is only weak. As affixal salience increases in words with longer suffixes (Laudanna & Burani, 1995), the effect of derived word frequency is attenuated and virtually vanishes, while the effect pertaining to the parsing of derived word's morphemes (as reflected in relative entropy) increase in size. In what follows we discuss observed effects of morphological structure and the modulating role of suffix length in shifting the balance between storage and computation in the processing of derived words.

General Discussion

The present data give rise to two insights into the lexical processing of morphologically complex words. First, access to the full-form (diagnosed by the derived word frequency effect) is modulated by suffix length in phonemes (see Figure 1A). Words with longer suffixes show a weaker effect of word frequency than those with shorter suffixes. This finding ties in with a number of reports that point at attenuated effects of whole word frequency in words in which morphemes are either of high-frequency (see interactions of compound frequency by left constituent frequency in Dutch, Kuperman et al., 2009; derived word frequency by base frequency in English, Baayen et al. 2007; and derived word frequency by suffix frequency in Danish, Winther Balling & Baayen, 2008), or have a large morphological family (see interactions of compound frequency with both left and right constituent family sizes in Finnish, Kuperman et al., 2008). This set of results strongly suggests that an easier recognition of one or both of the word's morphemes – due to their distributional or formal properties – leads to a stronger effect of these properties on the processing effort, and also diminishes the role of the complex word as a whole. We argue that the common cognitive underpinning for such an impact of a higher frequency, a larger family size or an increased phonological or orthographic length of a morpheme is that these morphemic properties increase the salience of the morpheme for the purposes of processing, see Laudanna and Burani (1995). To rephrase, a relatively salient substring in the word is a substring that the reader has more experience with identifying within this word and within many other words; it is also a substring that receives more perceptual bottom-up support. In the present study, the morpheme that shows a stronger contribution to morphemic salience and derived word recognition is the suffix, while other tasks, other languages and other types of morphological complexity (see papers cited in this paragraph) also emphasize the role of the base word.

The notion that morphemes and complex words as wholes interact in visual word recognition, supported here and in the studies cited above, is not easy to reconcile with any model that requires an obligatory temporary order in accessing the full-forms and morphemes of complex words. The single full-form route postulated by supralexical models (cf. Giraudo & Grainger, 2001) suggests that readers directly access the whole-word representation and this should be independent from the properties of constituent morphemes. Post-access activation of morphemes would not be able to modulate the derived word frequency effect. On the other hand, sublexical models require pre-access obligatory decomposition of the complex words into morphemes (cf. Taft, 2004, Taft & Ardasinsky, 2006). Decomposed morphemes activate their lexical representations with the speed proportional to morpheme frequencies. This phase is followed by a recombination process involving recognition of the complex words via its morphemes. If such recognition fails, a lemma that is associated with the whole word must be activated, and lexical information must be obtained from this lemma, with the speed proportional to the whole word frequency. On this account, a balance between storage and computation may also be present, but it is expected to occur at a later, post-decomposition phase. The obligatory decomposition account, however, makes predictions that are not compatible with present results. First, it predicts an effect of base frequency (or base length), which does not reach significance in any model in the present study. Second, it requires additional assumptions to explain the effect of suffix length in interaction with word frequency on the processing effort, since the processes involved in combining information associated with lemmas are not expected to be influenced by the length of the suffix. One such assumption compatible with the obligatory parsing framework might be that suffix length is confounded with semantic transparency of derived words, with longer suffixes being more transparent than shorter ones (for etymological or other reasons). Recognition of transparent derived words would allow skipping the recombination phase and relying solely on the outcome of the initial decomposition process. Conversely, opaque words would trigger recombination of lemmas associated with morphemes: since recombination is co-determined by whole word frequency, one might expect this frequency to modulate the speed of recombination for those words. Whether semantic transparency correlates with the length of Dutch suffixes and whether it can account for the behavioral patterns observed in this study is a question for further investigation[5].

In contrast to single-route models, the observed interaction of derived word frequency

[5]The link between suffix length and transparency is not confirmed in English. We collected measures on semantic distance between base words and whole derived words for over 800 English suffixed words. The distance was assessed using the Latent Semantic Analysis technique (Landauer & Dumais, 1997). Crucially, the correlation between semantic distances for derived words and their bases, on the one hand, and the lengths of suffixes in these words, on the other hand, was very weak and not significant ($r = 0.006, p = 0.70$).

and suffix length follows quite naturally from the premises of interactive dual- or multiple-route models of parallel processing. The longer (hence, more salient) suffix shifts the balance between storage and computation towards computation and makes the full-form route less beneficial for the lexical processor than decomposition. The stronger the bias towards decomposition is, the less use readers make of the full-form properties in the recognition of derived words (cf., Bertram & Hyönä, 2003, for compounds).

What is the cognitive mechanism that underlies the role that suffix length plays in the storage-computation trade-off? One possibility, well developed in the current literature, is that perceptual salience of longer suffixes makes them easier targets for detection during visual information uptake. This early segmentation may lead to a head-start for the parsing route and boost computation as a processing strategy, as argued, for instance, for Finnish compounds with vowel disharmony across the constituent boundary (Bertram, Pollatsek & Hyönä, 2004) and Finnish compounds with left constituents ending in derived suffixes (Kuperman et al., 2008). At the same time, the bias for computation suppresses the retrieval of the whole word: either because full-form access is not completed (and does not affect comprehension to the full extent) by the time the complex word is recognized via morphemes, or because early activation of morphemes interferes with the retrieval of full-forms due to a strong formal overlap, or because storage and computation compete with each other for the limited capacity of the cognitive resources required for word recognition. Given the present data, we cannot adjudicate between the reasons for suppression of full-form access and leave it to future research. We note here that the visual salience of the morphological base does not appear to have any influence on derived word processing in Dutch, even though such salience might give rise to the same kind of early segmentation, as suffix salience does. This may be due to the fact that the probability of encountering a morphological base in any word is 1: Each word consists of at least one morpheme, the base. Whether other processing units emerge out of the letter string depends on their salience, but the base does not need to be salient for comprehenders to know that it is realized in the word. Also, the base of most complex words can occur independently, which may imply that recognition of this morpheme, and its parsing out of the complex word, relies less on its visual salience. Moreover, suffixes tend to be much more frequent token-wise than bases are: This may give a head-start to the recognition of the suffix relative to the recognition of the base.

The second insight from the data is the interaction between morphological families of the base and suffix of the derived word, which we explain in terms of the relative entropy of the probability distributions of the base and suffix families. Figure 1B suggests that the time spent fixating derived words is minimal when the two families are of a similar size, or equivalently, when the relative entropy is minimal (see Figure 1C). Not only are both the base and the suffix accessed simultaneously, but also distributional properties of one morpheme modulate the magnitude of the contribution the other morpheme makes towards complex word recognition. At present, we can only speculate about why this might be so. Possibly, activation of morphological families gives rise to two subprocesses, recognition of one of the family members (the constituent actually realized in the complex word) and inhibition of competing members of the family. The first subprocess ties in with the experience of segmenting constituents out of the embedding word, and the larger the families the easier such segmentation proceeds. Yet there are fewer competitors to inhibit when families are small, so the inhibition may come with less processing costs when the families are small. We speculate that optimal processing coincides with either the easiest segmentation task (when both families are large) or with the easiest inhibition task (when both families are small).

One may also consider the effect of mismatch between base family size and suffix productivity in the framework of the computational Morphological Family Resonance Model (De Jong, Schreuder & Baayen, 2003). This model postulates that family size effects arise as a result of semantic resonance, defined as a joint action of activation spreading from central representations of a complex word to lemmas of contextually restricted subsets of morphological family members and the feedback activation spreading from those lemmas to central representations, including those shared by the word under identification. Possibly, the relative entropy effect bears witness to a competition between morphological families, such that the resonance between the family members of the morpheme with the larger family swamps the resonance of the family of the morpheme with the smaller family. That is, the activation threshold for one morpheme might be reached much faster than that for the other. This imbalance in the amount of lexical support would then delay the integration of the two morphemes into a coherent representation of the derived word as a whole. We leave to future research the further specification of how the imbalance in informativeness of morphological families affects lexical processing. Yet we note that, regardless of its interpretation, the effect of imbalance in informativeness of family sizes is problematic for dual-route models of parallel morphological processing, including those that advocate the interactive, rather than independent, use of processing routes, like MATCHEK (Baayen & Schreuder, 2000). The MATCHEK model predicts greater competition if representations of both constituents have the same strength, so these conditions should come with slower processing times, contrary to fact.

Evidently, a broad spectrum of single- and dual-route models of morphological processing cannot fully account for the present results. A multiple-route model that advocates a probabilistic approach to morphological complexity and is implemented as the Probabilistic Model of Information Sources (PROMISE) in Kuperman et al. (2009) fares better in accounting for the present set of data. The conceptual background for the model is the view that the mental lexicon is a long-term memory storage for lexical information. The visual uptake of a stimulus triggers access of this lexical information. The ease of access, and generally of lexical processing, depends in part on the amounts of information carried by words, which are defined by the accumulated knowledge of words and their paradigmatic and syntagmatic connectivity in the mental lexicon. The multiple-route model considers morphological structure as a conglomerate of sources of information, which contribute - to a different extent - to the recognition of polymorphemic words. Specifications for the model are as follows. First, the model does not require strict sequentiality of processing stages (as in sub- or supralexical models), but rather it allows for simultaneous processing of information at different levels (characters, immediate and deeply embedded morphological constituents, morphological paradigms and whole words). Second, the model allows for the contribution of one source of morphological information to modulate the presence and the strength of contributions made by other such sources towards complex word identification. This requirement runs counter to the implicit assumption of single- and some dual-route models that lexical representations are blind to each other's activation, and builds on the premises of the dual-route model of interactive parallel processing (Baayen & Schreuder, 2000). Third, the model takes into explicit consideration morphological families of constituents in complex words, and not just consituents as isolated words, since the effects of families on reading times in compounds are ubiquitous, early and strong. To sum up, the required model is a multiple-route model of morphological processing, which considers morphemes, combinations of morphemes, morphological paradigms and structurally complex words as sources of morphological information.

In consideration of space we do not present the mathematical treatment of the interactive patterns observed in the present data (for similar treatments in compound words see Kuperman et al., 2008). Instead we illustrate the complex dynamics of how multiple, interdependent, sources of information contribute to word recognition by way of example. Consider two hypothetical words with extreme distributional properties. Suppose word A has the highest frequency and the largest value of the relative entropy measure among all words in our experimental list. Our word B is the opposite of word A, with the lowest word frequency, and identical suffix productivity and base family size (i.e., zero relative entropy). We further assume that words A and B carry suffixes of the same length. We can now estimate the difference in processing times between A and B by visually inspecting interaction plots in Figure 1A and C, and estimating differences for the extreme values of predictors of interest. If both A and B have the shortest suffixes (two-phoneme long), our model for single-fixation duration makes the following predictions, as reflected in the interaction plots. Word A will come with a 40 ms reduction of reading time due to higher word frequency, and a 5 ms reduction due to greater entropy, as compared to word B. In total, word A would be processed approximately 45 ms faster than word B, given that suffixes in A and B are short and hence full-form access is likely. If A and B are words with a median suffix length of 3 phonemes, a similar calculation indicates that, in total, the processing advantage for word A over word B is reduced to 18 ms. Finally, if words A and B both carry the longest suffixes (5 phonemes), which prompt morphological decomposition, the processing time for word A is predicted to be 45 ms longer than that for word B.

Naturally, most words are not as extreme in their distributional properties as the words in our example. Still, this example highlights the dynamic nature of lexical processing for complex words. Apparently, the effect of virtually any single information source on the speed of word recognition can range from facilitatory to negligibly small to inhibitory depending on the effects of other such sources and the likelihoods that those other sources are available for processing. Methodologically, this implies that considering any one information source in isolation from others (by, say, keeping the values of other information sources constant through matching of stimuli) is bound to miss the essentially interactive use of bits and pieces of morphological structure in complex word recognition.

Our findings also raise a general question of whether using morphological sources of information is a viable alternative to the recognition of words as unstructured units, which is the option proposed in the full-listing models (cf. Butterworth, 1983; Janssen, Bi & Caramazza, 2008). The fact that we, along with a long tradition of morphological research, observe readers making use of morphemic properties does not necessarily imply that on average readers benefit from such use in terms of their processing speed. Morphological cues may merely impose themselves on the recognition system and be followed automatically, even to the disadvantage of the reader. However, there is a growing evidence from word comprehension studies that on average complex words are processed faster than simplex words with similar values of frequency, length and several other characteristics. Thus, Bertram et al. (1999) observed that Finnish derived words elicited shorter visual lexical decision latencies than monomorphemic words, Burani and Thornton (2003) reported shorter lexical decision latencies for Italian derived words with a high-frequency root as compared to simplex words matched on whole-word frequency. Winther Balling and Baayen (2008) reported shorter response times for Danish derived and inflected words in the auditory lexical decision task, while Fiorentino and Poeppel (2007) replicated this finding comparing English compounds and simplex words in a visual lexical decision task and Inhoff, Briihl and Schwartz (1996) found shorter naming latencies (but longer first fixation durations) for English compounds as compared to suffixed and simplex controls. That is, the fine-tuning of the balance between multiple processing routes may impose conditions on what counts as the optimal processing strategy and what the costs of suboptimality are, but it also offers an overall processing advantage unavailable to simpler recognition systems.

Acknowledgments

We wish to thank Christina Burani, Dominiek Sandra, Marcus Taft and two anonymous reviewers for their insightful comments on the earlier drafts of this paper, and Vitória Piai for her help with data collection. This research was conducted while the first author was at the Radboud University of Nijmegen. During the last stages of the research the first author was supported by a Rubicon grant of the Netherlands Organization for Scientific Research (NWO).

Appendix 1

Key to Table 1:

Table 1 provides statistics on continuous variables, which show significant effects in our statistical models, including ranges of their original values, and (where applicable) ranges of the values after (logarithmic or standardization) transformations. Column Variable lists predictors of interest. Numbers in the second column show original value ranges for predictors. If any transformations have been applied to the original values for statistical reasons (i.e., log transformation, standardization or scaling), the numbers in the brackets show the ranges actually used in statistical models. Means, standard deviations (Column 3) and median values (Column 4) refer to the predictor values used in the models. Computation of these distributional measures was based on the combined pool of roughly 120 million tokens, obtained from the CELEX lexical database and from the newspapers in the Twente News Corpus.

Predictors of primary interest for this study are presented in the main body of paper. In addition to the variables reported in Table 1, we considered a large number of control variables that were not significant predictors of reading times or probabilities. These included such distributional predictors as complexity-based ranking of suffixes (cf. Hay & Plag, 2004); number of word types in which the ratio of base frequency and whole word frequency is above/below the mean ratio for the suffix, and their ratio; number of hapax legomena in which the suffix occurs; growth rate and type/token ratio for the suffix; cumulative base frequency, the relative frequency of the base, average relative base frequency, and the ratio of word types in which whole word frequency exceeds base frequency and word types in which whole word frequency is lower than base frequency (cf. Hay, 2001). Orthographic and phonological factors included: whether or not the first or the last syllable of the word was stressed; whether stress falls on any of the suffix's syllables; whether or not a suffix began with a vowel; as well as type-based frequencies of the word-initial and word-final trigram and frequency of occurrence of the bigrams straddling, preceding and following the morphemic boundary. Lexical predictors included: whether or not suffixes change their orthographic form across the inflectional paradigm; whether the word class of the derivation as a whole differs from the word class of the base word; and whether suffixes in target words were homonymous with Dutch inflectional suffixes; and word class of the target word. We also considered the number of word types in which the character string occurs as a suffix and the number of word types in which it occurs in the word-final position in any other non-morphemic capacity. Contextual control variables included: joint probabilities of words N-1 and N, and of words N and N+1; lengths, frequencies and word classes of words N-1 and N+1; and word position in a sentence. Visual control variables included: amplitudes of saccades preceding and following the fixation. Also, including suffix as a random effect did not significantly improve the statistical models.

[Table 1 about here.]

Key to Tables 2-4:

Across tables 2-4, column *Variable* lists the intercept as well as all predictors and interactions that reached significance in the model. Column *Estimate* shows estimates of the regression coefficients for the model's predictors. Columns *MCMCmean*, *HPD95lower* and *HPD95upper* provide the estimate of the mean value for each predictor, as well as the lower and the upper boundaries of highest posterior density intervals (which are a Bayesian measure of the 95% confidence interval) obtained via the Monte Carlo Markov chain (MCMC) random-walk method using 10,000 simulations: this information is useful for evaluating stability of the model's predictions. Column *pMCMC* provides a p-value obtained with the help of MCMC simulations; and column *Pr(>|t|)* 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.

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

References

Andrews, S. and Davis, C. (1999). Interactive activation accounts of morphological decomposition: Finding the trap in mousetrap? *Brain and Language*, 68:355–361.

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. *Journal of Memory and Language*, 59:390–412.

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, 37:94–117.

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

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

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

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

Bates, D. M. and Sarkar, D. (2007). The lme4 library. *[On-line], Available: http://lib.stat.cmu.edu/R/CRAN/*.

Beauvillain, C. (1996). The integration of morphological and whole-word information during eye-fixations on prefixed and suffixed words. *Journal of Memory and Language*, 35:801–820.

Bertram, R., Baayen, R. H., and Schreuder, R. (2000a). Effects of family size for complex words. *Journal of Memory and Language*, 42:390–405.

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., Baayen, R. H., Schreuder, R., and Hyönä, J. (2000b). Affixal homonymy triggers full-form storage even with inflected words, even in a morphologically rich language. *Cognition*, 74:B13–B25.

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. (2000c). 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.

Burani, C. and Caramazza, A. (1987). Representation and processing of derived words. *Language and Cognitive Processes*, 2:217–227.

Burani, C. and Thornton, A. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. In Baayen, R. H. and Schreuder, R., editors, *Morphological structure in language processing.*, pages 157–208. Mouton de Gruyter, Berlin.

Butterworth, B. (1983). Lexical representation. In Butterworth, B., editor, *Language production (Vol. II): Development, Writing and Other language Processes*, pages 257–294. Academic Press, London.

Colé, P., Beauvillain, C., and Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and Language*, 28:1–13.

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

De Jong, N. H., Schreuder, R., and Baayen, R. H. (2003). Morphological resonance in the mental lexicon. In Baayen, R. H. and Schreuder, R., editors, *Morphological structure in language processing*, pages 65–88. Mouton de Gruyter, Berlin.

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

Fiorentino, R. and Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, 22:953–1000.

Frisson, S., Niswander-Klement, E., and Pollatsek, A. (2008). The role of semantic transparency in the processing of English compound words. *British Journal of Psychology*, 99:87–107.

Gelman, A. and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, Cambridge.

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

Gries, S. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement.* Continuum International Publishing Group Ltd., New York.

Hay, J. (2001). Lexical Frequency in Morphology: Is Everything Relative? *Linguistics*, 39:1041–1070.

Hay, J. (2003). *Causes and Consequences of Word Structure*. Routledge, New York and London.

Hay, J. and Baayen, H. (2002). Parsing and productivity. In Booij, G. and Van Marle, J., editors, *Yearbook of Morphology 2001*, pages 203–235. Kluwer Academic Publishers, Dordrecht.

Hay, J. and Baayen, H. (2003). Phonotactics, parsing and productivity. *Italian Journal of Linguistics*, 1:99–130.

Hay, J. B. and Plag, I. (2004). What constrains possible suffix combinations? on the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory*, 22:565–596.

Holmes, V. and O'Regan, J. (1992). Reading derivationally affixed French words. *Language and Cognitive Processes*, 7:163–192.

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.

Inhoff, A., Starr, M., Solomon, M., and Placke, L. (2008). Eye movements during the reading of compound words and the influence of lexeme meaning. *Memory & Cognition*, 36(3):675–687.

Inhoff, A. W., Briihl, D., and Schwartz, J. (1996). Compound word effects differ in reading, on-line naming, and delayed naming tasks. *Memory and Cognition*, 24:466–476.

Janssen, N., Bi, Y., and Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes*, 23:1191–1223.

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. (2008). The processing of compound words in English: Effects of word length on eye movements during reading. *Language and Cognitive Processes*, 23:1057–1088.

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.

Kuperman, V., Bertram, R., and Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, 23:1089–1132.

Kuperman, V., Schreuder, R., Bertram, R., and Baayen, R. H. (2009). Reading of polymorphemic Dutch compounds: Towards a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, 35:876–895.

Landauer, T. and Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2):211–240.

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.

Milin, P., Kuperman, V., Kostić, A., and Baayen, R. (2009). Words and paradigms bit by bit: An information-theoretic approach to the processing of inflection and derivation. In Blevins, J. and Blevins, J., editors, *Analogy in Grammar: Form and Acquisition*, pages 214–252. Oxford: Oxford University Press.

Niswander, E., Pollatsek, A., and Rayner, K. (2000). The processing of derived and inflected suffixed words during reading. *Language and Cognitive Processes*, 15:389–420.

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

Ordelman, R. (2002). Twente Nieuws Corpus (TwNC). Report 25, Parlevink Language Techonology Group. University of Twente.

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

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

Plag, I. and Baayen, R. (2009). Suffix ordering and morphological processing. *Language*, 85:109–152.

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.

Pollatsek, A., Slattery, T., and Juhasz, B. J. (2008). The processing of novel and lexicalized prefixed words in reading. *Language and Cognitive Processes*, 23:1133–1158.

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

Rayner, K., Warren, T., Juhasz, B., and Liversedge, S. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30:1290–1301.

Richardson, D., Dale, R., and Spivey, M. (2007). Eye movements in language and cognition: A brief introduction. In Gonzalez-Marquez, M., Mittelberg, I., Coulson, A., and Spivey, M., editors, *Methods in Cognitive Linguistics*. John Benjamins.

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.

Schreuder, R., Burani, C., and Baayen, R. H. (2003). Parsing and semantic opacity. In Assink, E. and Sandra, D., editors, *Reading Complex Words. Cross-language studies*, pages 159–189. Kluwer, Dordrecht.

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

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.

Van Marle, J. (1985). *On the Paradigmatic Dimension of Morphological Creativity*. Foris, Dordrecht.

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

Winther Balling, L. and Baayen, R. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*, 23:1159–1190.

List of Tables

Table 1: Summary of Continuous Variables

Variable	Range (units)	Mean(SD)	Median
InitFixPos	0.0:13.68 characters	3.56(2.4)	3.81
SingleDur	77:837 ms (4.3:6.7 log units)	5.5(0.3)	5.5
GazeDur	77:880 ms (4.3:6.8 log units)	5.6(0.4)	5.5
WordLength	8:14 characters (-0.39:2.08 standardized)	0.0 (0.5)	0.0
SuffixLength	2:5 phonemes (-0.67:1.22 standardized)	0.0 (0.5)	0.01
WordFreq	12:2607 (2.5:7.9 log units; -0.88:1.35 standardized)	0.0 (0.5)	0.0
SuffixProd	10:812 (-1.22:0.78 standardized)	0.0 (0.5)	0.0
RE	0.0:10.1 bits (-0.41:1.76 standardized)	0.0 (0.5)	0.0
BaseFamilySize	2:300 (-0.94:1.13 standardized)	0.0 (0.5)	0.0
PrecLength	2:17 characters (-0.43:2.51 standardized)	0.0 (0.5)	0.0
Plausibility	1.94:6.76 (-1.6:0.74 standardized)	0.0 (0.5)	0.0

Table 2: Original Model for Single-Fixation Duration. The names of standardized variables in Column 1 are preceded by the lower-case 's'.

Variable	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
Intercept	5.1365	5.1366	4.9381	5.3389	0.0002	0.0000
sPrecLength	0.0421	0.0422	0.0206	0.0642	0.0012	0.0003
InitFixPos	-1.4051	-1.4050	-2.0260	-0.8470	0.0002	0.0000
$InitFixPos^2$	-0.8589	-0.8534	-1.4233	-0.2765	0.0040	0.0039
sWordFreq	-0.0514	-0.0512	-0.0761	-0.0291	0.0002	0.0000
sSuffixLength	-0.6847	-0.6799	-1.1707	-0.2117	0.0072	0.0076
sWordLength	0.0101	0.0095	-0.0161	0.0342	0.4416	0.4431
sPlausibility	-0.0236	-0.0242	-0.0482	-0.0028	0.0392	0.0522
sBaseFamilySize	0.0949	0.0953	0.0335	0.1538	0.0044	0.0032
sSuffixProd	0.0747	0.0747	0.0345	0.1146	0.0004	0.0005
BaseFamilySize:AffixProd	-0.0219	-0.0220	-0.0344	-0.0097	0.0008	0.0009
sSuffixLength:BaseFamilySize	0.1763	0.1762	0.0273	0.3218	0.0208	0.0242
sSuffixLength:SuffixProd	0.1449	0.1442	0.0503	0.2304	0.0032	0.0026
sWordFreq:sSuffixLength	0.0406	0.0413	0.0000	0.0815	0.0456	0.0656
sSuffixLength:sBaseFamilySize:sSuffixProd	-0.0372	-0.0372	-0.0639	-0.0090	0.0104	0.0114

Table 3: Final Model for Single-Fixation Duration. The names of standardized variables in Column 1 are preceded by the lower-case 's'.

Variable	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
Intercept	5.4550	5.4544	5.4156	5.4904	0.0002	0.0000
sPrecLength	0.0400	0.0402	0.0185	0.0620	0.0002	0.0005
InitFixPos	-1.3891	-1.3896	-2.0277	-0.8451	0.0002	0.0000
InitFixPos^2	-0.8734	-0.8745	-1.4341	-0.2706	0.0040	0.0032
sPlausibility	-0.0311	-0.0312	-0.0533	-0.0086	0.0100	0.0086
sWordFreq	-0.0549	-0.0549	-0.0792	-0.0325	0.0002	0.0000
sSuffixLength	-0.0108	-0.0109	-0.0369	0.0138	0.4104	0.4218
sRE	0.0212	0.0213	-0.0037	0.0450	0.0936	0.1105
sWordFreq:sSuffixLength	0.0423	0.0426	0.0028	0.0838	0.0416	0.0487
sRE:sSuffixLength	0.0527	0.0528	0.0128	0.0945	0.0148	0.0134

Table 4: Final Model for Gaze Duration. The names of standardized variables in Column 1 are preceded by the lower-case 's'.

Variable	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
Intercept	5.5613	5.5619	5.5194	5.6055	0.0002	0.0000
sPrecLength	0.0407	0.0406	0.0088	0.0732	0.0100	0.0169
InitFixPos	-6.4198	-6.4084	-7.1054	-5.6941	0.0002	0.0000
InitFixPos^2	1.1956	1.2044	0.5209	1.9175	0.0012	0.0013
sWordFreq	-0.0742	-0.0747	-0.1081	-0.0401	0.0002	0.0000
sSuffixLength	0.0453	0.0453	0.0062	0.0818	0.0156	0.0261
sRE	-0.0087	-0.0091	-0.0477	0.0290	0.6348	0.6654
sPlausibility	-0.0342	-0.0341	-0.0661	-0.0004	0.0428	0.0535
sWordFreq:sSuffixLength	0.0632	0.0630	0.0010	0.1242	0.0492	0.0516
sSuffixLength:sRE	0.0853	0.0856	0.0256	0.1460	0.0084	0.0084

List of Figures

Figure 1: A. Interaction of derived word frequency by suffix length for single-fixation duration, measured in phonemes. The effect of derived word frequency is plotted separately for all suffix lengths. B. Interaction of base family size by suffix productivity (suffix family size). The effect of base family size is plotted for quartiles of suffix productivity. C. Interaction of relative entropy by suffix length. The effect of relative entropy is plotted separately for all suffix lengths.

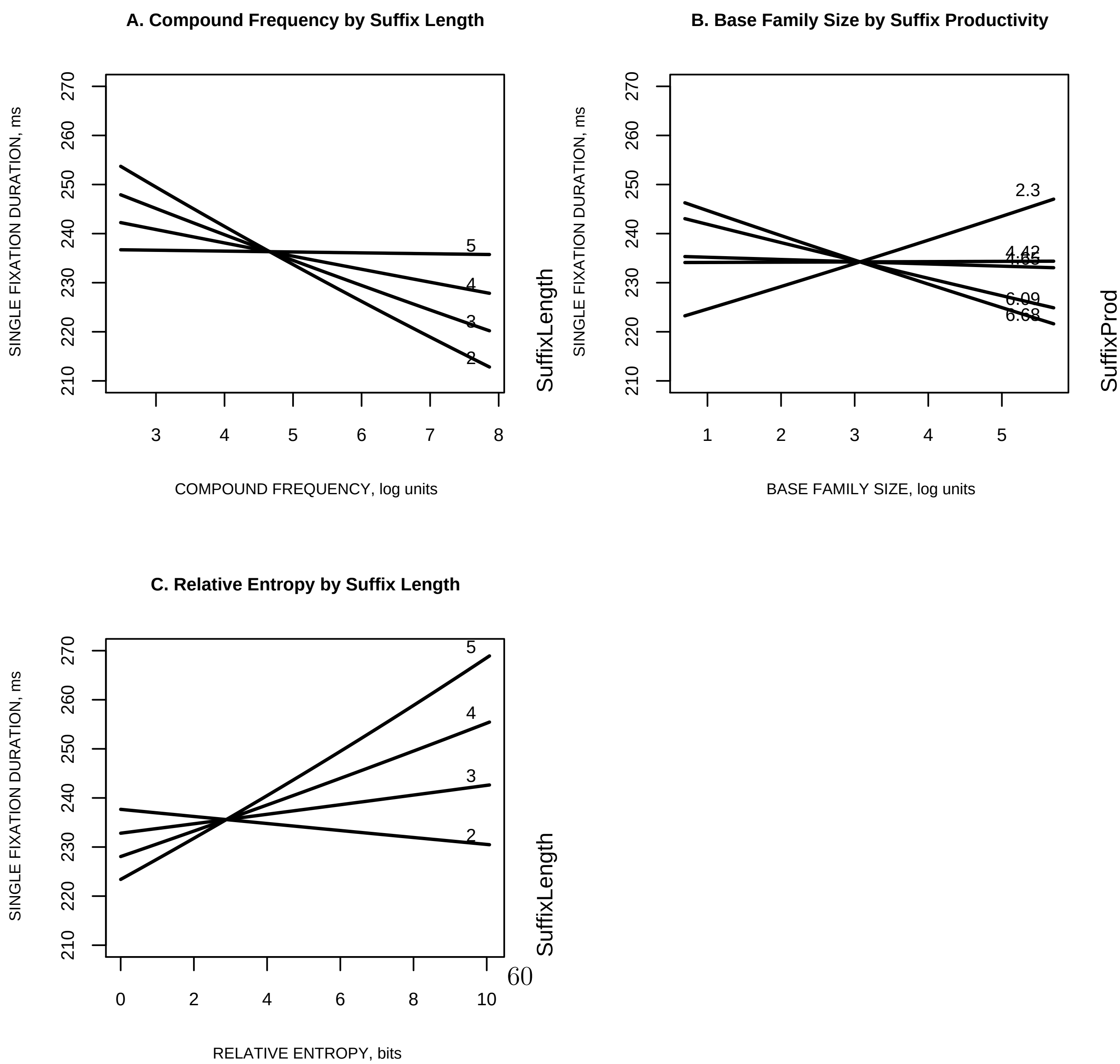

Figure 2: Left: The multiplicative interaction as a function of suffix productivity by base family size. Right: relative entropy as a function of the family-wise probabilities (and their complements, not shown) of suffix and base.