Journal of Experimental Psychology:
Human Perception and Performance
2000, Vol. 26, No. 2, 820–833

0096-1523/00/$5.00 DOI: 10.1037//0096-1523.26.2.820

# The Role of Morphological Constituents in Reading Finnish Compound Words

Alexander Pollatsek
University of Massachusetts at Amherst

Jukka Hyönä and Raymond Bertram
University of Turku

The processing of transparent Finnish compound words was investigated in 2 experiments in which eye movements were recorded while sentences were read silently. The frequency of the second constituent had a large influence (95 ms) on gaze duration on the target words, but its influence was relatively late in processing: A clear effect only occurred on the probability of a third fixation. The frequency of the whole compound word had a similar influence on gaze duration (82 ms) and influenced eye movements at least as rapidly as did the frequency of the second constituent. These results, together with an earlier finding that the frequency of the first constituent affected the first fixation duration, indicate that the identification of these compound words involves parallel processing of both morphological constituents and whole-word representations.

Although the mental processes by which printed words are recognized have received extensive attention from cognitive psychologists, the vast majority of studies have concentrated on the identification of short words that can be identified with a single eye fixation on the word. The general conclusion from the identification of short words is that it is a one-stage parallel process (e.g., McClelland & Rumelhart, 1981; Paap, Newsome, McDonald, & Schvaneveldt, 1982). Longer words (eight or more characters), however, typically require at least one refixation to ensure complete identification, which suggests that processing of longer words may involve more than one stage of processing.

A common type of long word in most languages is compound words, which consist of more than one morphological constituent, each of which is a word in the language. In Finnish, compound words are extremely common, and word compounding is a productive way to construct novel words. Typically, two or more nouns are attached to each other to form a compound word (e.g., *lumi* = snow, *lumipallo* = snowball, *lumipallosota* = snowball fight, *lumipallosotatantere* = snowball fight field). The purpose of the present study, which is a follow-up of Hyönä and Pollatsek (1998), is to examine the processing of two-noun compound words when they are read as part of a meaningful sentence.

Hyönä and Pollatsek (1998) used readers' eye fixations on transparent compound words to assess how they identified them in reading. In Experiment 2, they manipulated the frequency of the first constituent (i.e., its frequency as a separate word) and held the frequency of the whole word constant. The motivation for this experiment was to explore a simple decomposition hypothesis in which such long, two-noun compound words are recognized in successive processing stages: (a) the initial constituent is identified and (b) the second constituent is identified. This sequential decomposition model makes several predictions about the pattern of eye fixations that occur when reading these two-noun compound words.

First, the frequency of the initial constituent should affect the time to complete the first stage of processing the word. Because this is the first stage of processing, it should be reflected in eye movement indexes that tap early processing, such as the duration of the initial fixation on the word. However, it is unrealistic to think that cognitive stages of processing will have a simple one-to-one association with eye movement indexes, and thus it is likely that the initial stage of processing the compound word will affect subsequent processing on the word, such as the duration of subsequent fixations. In fact, in the Hyönä and Pollatsek (1998) data, the *first fixation duration* (the duration of the first or only fixation on a word) on the compound word was influenced by the frequency of the initial constituent. The effect of the frequency of the initial constituent on the *gaze duration* (the sum of all fixations landing on the word before exiting it), however, was substantially larger than its effect on the first fixation duration.[1] This indicates that the

---

Alexander Pollatsek, Department of Psychology, University of Massachusetts at Amherst; Jukka Hyönä and Raymond Bertram, Department of Psychology, University of Turku, Turku, Finland.

This research was supported by Grant HD26765 from the National Institutes of Health and by a grant from the Turun Yliopistosäätiö (Turku University Foundation). Alexander Pollatsek wishes to thank the Medical Research Council, Cognitive and Brain Sciences Unit, Cambridge, England, for the use of their facilities during the preparation of the manuscript.

Correspondence concerning this article should be addressed to Alexander Pollatsek, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003 or to Jukka Hyönä, Department of Psychology, University of Turku, FIN-20014 Turku, Finland. Electronic mail may be sent to pollatsek@psych.umass.edu or to hyona@utu.fi.

[1] More formally, the definitions of these two measures are as follows: Both are conditional on the first fixation on the word being on the first pass through the sentence (i.e., that the first fixation on the word was not preceded by a fixation on a word to the right of it). The *first fixation duration* is the duration of the first fixation on the

frequency of the first constituent had effects on processing after the initial fixation as well and that the putative first stage of processing must often extend past the first fixation.

Second, the model predicts that variables that affect the second stage of processing, such as the frequency of the second constituent, should only affect processing after the initial stage of processing is completed, which would presumably be after the initial fixation on the word. In fact, the results of post hoc regression analyses in Hyönä and Pollatsek (1998) were consistent with this: Gaze duration (but not first fixation duration) decreased as the frequency of the second constituent increased.

The question naturally arises as to whether the frequency of the compound word (independent of the frequency of its constituents) would have any effect on processing. Hyönä and Pollatsek (1998) hypothesized that, consistent with the decomposition model, either there would be no whole-word effects (if no processing is needed to process the meaning of the whole word beyond that needed to process the meaning of its parts) or, if there were such integrative effects (e.g., determining that a snowball fight is a fight using snowballs rather than a fight between snowballs), whole-word effects should appear only in a third stage of processing. If one accepts this hypothesis, one would predict that the frequency of the whole compound word would only affect processing very late in the eye movement record—certainly, one would predict that the frequency of the compound should not affect processing before the frequency of the second morpheme did. Of course, the "transparency" of the compound (i.e., how transparent its meaning is from the components) could influence whether a third stage is needed. Presumably, it might always be needed for quite opaque compounds such as *deadbeat.* However, as the *snowball fight* example indicates, even compounds that are usually classified as transparent may take some additional processing to determine the meaning.

The present experiments were designed to explore the Hyönä and Pollatsek decomposition model more fully. In Experiment 1, the frequency of the second constituent was manipulated while matching for both the frequency of the whole word and the frequency of the first constituent, whereas in Experiment 2, the frequency of the whole word was manipulated while matching for the frequencies of the first and second constituents.

One alternative model to a decomposition model is that compound words are identified solely by a direct lookup of the lexical entry. Such a model would appear to be ruled out by data, such as the above, which show that morphemic constituents play a role in the speed of word identification. The model which is closest in spirit to a direct lookup model, but which posits a role for constituents, is the sequential search model of Taft and Forster (1976), which makes use of a "file drawer" metaphor. According to this model (see also Cole, Beauvillain, & Segui, 1989, for a similar model for suffixed words), a word is looked up in a prestored location in the lexicon in two stages. The first stage is accessing the file drawer indexed by an "access code," such as the initial compound word constituent. The duration of this stage is presumably influenced by the frequency of this access code. The second stage is a frequency-ordered serial search for the lexical item in the file drawer, so that the duration of the second stage should be influenced by the frequency of the word in the file drawer relative to the frequencies of the other words stored in the same file drawer.

The decomposition model and the search model converge on the prediction that higher-frequency initial constituents will speed the initial processing stage, but they diverge on their predictions about subsequent processing. The decomposition model predicts that increasing the frequency of the second constituent should facilitate processing, whereas the sequential search model does not assume that the second morpheme has any direct influence on subsequent processing—it is only the location of the lexical item relative to the other lexical items in the file drawer that determines the processing duration of the second stage. There is a similar difference between the two models on their predictions of the effect of whole-word frequency. The decompositional model predicts that higher-frequency words should be processed more rapidly in the third stage, whereas the sequential search model makes no direct prediction about the effects of whole-word frequency—only the location in the file drawer should matter.

The Hyönä and Pollatsek (1998) study attempted to distinguish between the models in their regression analyses, because their words with infrequent first constituents were either the only compound beginning with that constituent or there was one other compound beginning with that constituent. Hence, the words with low frequency initial constituents were quite unique and would be first and second in their file drawer. In contrast, the words with higher frequency initial constituents were much lower down in their file drawers. Thus, the sequential search model makes the prediction that later processing would actually be longer for the higher frequency initial morpheme words in their experiment. However, there was no evidence for this; later processing was substantially facilitated by increasing the frequency of the first constituent.

Another alternative to a purely decompositional model is a hybrid model (Caramazza, Laudanna, & Romani, 1988; Frauenfelder & Schreuder, 1992; Niemi, Laine, & Tuominen, 1994; Schreuder & Baayen, 1995; Taft, 1994). Although most of these models have been constructed to explain processing of inflected and derived word forms—and most of the data come from tasks such as naming, lexical decision, or progressive demasking—they are plausible alternate hypotheses about how compound words are processed. In these models, there are two possible routes to identifying complex words: a decompositional route similar to that posited by Hyönä and Pollatsek and a direct route in

---

target word (the only fixation if there is only one fixation on the word). The *gaze duration* is the sum of the durations of all fixations on a word prior to another word being fixated. Hence, gaze duration does not include second pass fixations, that is, fixations after further text has been read and the word is refixated. In addition, both measures are conditional on the word being initially fixated. Hence, trials in which the word is initially skipped are not counted, rather than being counted as a zero fixation duration trial.

which the word as a whole is directly accessed by a lexical entry. Within these hybrid models, a distinction should be made between the either/or models and the and/and models. The either/or type of model claims that some polymorphemic words are processed via a direct route by means of whole-word representations, whereas others are processed solely by their constituents. Niemi et al., for example, claim that derived words are accessed by the former (direct) route, whereas inflected words are accessed by the latter route. In contrast, the and/and type of models are "race models," where decompositional and direct routes operate in parallel and are each, in principle, capable of accessing a given word. However, factors such as frequency and formal and semantic transparency will influence which of the routes is likely to be the winner of the race. For low-frequency transparent words, the decompositional route would be the most likely candidate to win the race, but for high-frequency and opaque words, the direct route may be the more likely winner. The morphological race model of Frauenfelder and Schreuder and the interactive activation network model of Schreuder and Baayen are instances of this type of model.

Both of these types of hybrid models are consistent with the Hyönä and Pollatsek (1998) data discussed earlier. (The either/or type of model, however, is consistent only if compounds with lower-frequency initial constituents are posited to be processed by the decompositional route, in which case it reduces to the decompositional model.) The race model version of the hybrid model is consistent with the Hyönä and Pollatsek data to the extent that the whole-word route does not always win the race, and hence the processing of constituents is relevant to identifying the compound at least some of the time. One clear prediction of such models that is different from a pure decompositional model is that whole-word effects (such as whole-word frequency) can influence processing at virtually any point of processing and certainly do not have to wait until after the second morpheme effects have surfaced. We will return to these issues when we discuss the experimental results.

The question of how compound words are processed has also been addressed using the lexical-decision paradigm. The seminal study of Taft and Forster (1976) employed nonword compounds, which were constructed from existing words but did not exist as whole words (e.g., *brieftax* or *toastpull*). Taft and Forster found that compound nonwords whose first constituent is a word took longer to classify as nonwords than those whose first constituent is not a word. This was taken as evidence to suggest that the first constituent plays a relevant role in the recognition of compound words. In contrast, they concluded that the second constituent was irrelevant in accessing these compound nonwords, because the lexical decision times did not differ as a function of the lexical status (word vs. nonword) of the second constituent. As a whole, these results were interpreted to support the sequential (i.e., two-stage) search model summarized above.

Using compound nonwords similar to those of Taft and Forster (1976), Lima and Pollatsek (1983) obtained evidence supporting the relevance of the first constituent, but in contrast to Taft and Forster they also found an effect for the lexical status of the second constituent, which stands in disagreement with the two-stage search model. A similar pattern of results was observed by Andrews (1986) for real compound words. Methodologically, her study differed from that of Taft and Forster in that she varied the frequency of compound word constituents as separate words instead of manipulating the lexical status of the constituents.

Another group of lexical-decision studies employed a priming paradigm (Monsell, 1985; Sandra, 1990; Zwitserlood, 1994). These studies indicated that a priming effect can be obtained for both the initial and the second constituent, especially with semantically transparent compound words, but not necessarily with semantically opaque compounds such as *deadbeat* (see Sandra, 1990; Zwitserlood, 1994). The priming effects for transparent compounds are consistent with any of the decompositional models cited above and suggest that when accessing transparent compound words, the word is decomposed into its components, which are then both used to derive the meaning of the word as a whole. The lack of consistent priming with the opaque words is perhaps easiest for an either/or model to explain if opaque transparent words are accessed by a direct route. However, what would need to be explained is how the system knows to turn off the compositional route when it encounters an opaque word. The either/or race model would have to posit that access of the whole word inhibits activation of parts that are not semantically congruent with the whole meaning. A pure decomposition model would similarly have to posit that after the nontransparent meaning in the third stage has been constructed, it inhibits the component morphemes.

In sum, the evidence from lexical-decision tasks on the recognition of compound words is more in favor of a decompositional model or a race model, rather than the sequential search model proposed by Taft and Forster (1976). However, the results from the lexical-decision task may not generalize to normal reading, because lexical-decision latencies often tap postaccess decision making that is likely to be unrelated to postaccess processes in reading comprehension. Moreover, strategic factors having to do with the stimulus environment have been shown to modulate the pattern of results in lexical decision (see Andrews, 1986). Thus, we feel that silent reading is a better paradigm for investigating how compound words are processed when people are processing for meaning.

## Experiment 1

### Method

*Participants.* Twenty-four students from the introductory psychology course at the University of Turku (Turku, Finland), all of whom were native speakers of Finnish, participated in the experiment as part of the course requirement.

*Materials and design.* Two sets of two-noun compound words (12–14 characters long) were used as target items: One set had a frequent second constituent and the other set had an infrequent second constituent. There were 32 words of each kind. The constituent frequency refers to the frequency the constituent has as a separate word in Finnish. The frequency of the compound words was fixed: All targets were infrequent, with an absolute frequency

Table 1
*Characteristics of the Target Words Used in Experiment 1*

| Stimulus characteristic | More-frequent 2nd constituent *M* | More-frequent 2nd constituent *SD* | Less-frequent 2nd constituent *M* | Less-frequent 2nd constituent *SD* |
|---|---|---|---|---|
| Frequency of compound word[a,b] | 0.2 | 0 | 0.2 | 0 |
| Frequency of first constituent[a,c] | 169 | 219 | 169 | 216 |
| Frequency of second constituent[a,c] | 359 | 274 | 1.6 | 1.3 |
| Average bigram frequency[a] | 6,320 | 1,616 | 7,161 | 2,181 |
| Beginning trigram frequency[a] | 719 | 425 | 796 | 788 |
| Final trigram frequency[a] | 819 | 719 | 858 | 719 |
| Length of compound word (in characters) | 12.7 | 0.6 | 12.8 | 0.8 |
| Length of first constituent (in characters) | 7.4 | 1.3 | 6.6 | 1.1 |

[a] These frequencies have been converted to frequencies per million to make them more comparable to the norms in English. [b] The frequencies are of the compound words in all their inflectional forms, including the nominative case. [c] The frequencies are of the constituents as separate words in all their inflectional forms, including the nominative case.

of 4 (i.e., 0.2 per million) in an unpublished 22.7 million-word newspaper corpus of Laine and Virtanen (1996).

Characteristics of the target words are presented in Table 1. There was a large difference in the frequency of the second constituent between the two word sets, whereas the frequency of the first constituents was closely matched. The target words were also matched for average bigram frequency, for beginning and final trigram frequency, for word length, and for first constituent length. Bigram and trigram frequencies were computed on the basis of the Laine and Virtanen (1996) corpus.

The target words, which always appeared in the nominative case, were near the beginning of sentences but never in the initial word position. Two words, one from each compound word type, were paired, and the sentence frames for each word pair were identical through the word following the target word.[2] A typical sentence pair follows, with the target word in italics:

1. Low-frequency second constituent condition: Tukholmassa *pommihysteria* onnistui valtaamaan tavallisten ihmisten mielet. (In Stockholm the bomb hysteria managed to overwhelm the minds of ordinary people.)
2. High-frequency second constituent condition: Tukholmassa *kuorokonsertti* onnistui kohtuullisesti, vaikka pitkä konserttikiertue oli uuvuttanut kuorolaisia. (In Stockholm the choral concert succeeded moderately well, although the long concert tour had tired the singers.)

There were two blocks of sentences, with the two sentences containing each word pair appearing in separate blocks. The order of the blocks was counterbalanced across participants. Thus, each participant saw all the critical target words. Within each block, the order of target sentences was randomized. There were 60 filler sentences among the critical sentences.

*Apparatus.* Eye movements were collected by the EYELINK eye-tracker, manufactured by SR Research Ltd. (Canada). The eye-tracker is an infrared video-based tracking system combined with hyperacuity image processing. There are two cameras mounted on a headband (one for each eye) including two infrared LEDs for illuminating each eye. (The headband weighs 450 g). The cameras sample pupil location and pupil size at the rate of 250 Hz. Registration is binocular and is performed for the selected eye(s) by placing the camera(s) and the two infrared light sources 4–6 cm away from the eye. In the present study, registration was done monocularly using the right eye. The resolution of eye position is 15 s of arc, and the average spatial accuracy is approximately 0.5 degrees. Head position with respect to the computer screen was 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 system allows free head motion within a 100-cm$^3$ cube. The compensation is better than 1 degree over the acceptable range of head motion.

*Procedure.* The eye-tracker was first calibrated using a 9-point calibration grid that spanned the entire screen area. Ten practice sentences were presented to participants before the actual experiment. Participants were instructed to read each sentence for comprehension. They were told that they would be occasionally asked to paraphrase the sentence they had just finished reading. Before the presentation of each sentence, they were required to gaze at a fixation point in the top left corner of the screen. In case of a calibration drift, the calibration was automatically corrected. Immediately after the fixation point was erased from the screen, a sentence, entailing 1–3 lines of text, was presented left-justified and centered vertically on the screen. The experimental session took a maximum of 30 min.

### *Results*

The data indicate that the frequency of the second constituent influenced processing but that its influence was not immediate. As shown in Table 2, the mean gaze duration

---

[2] A separate set of 9 participants was given the sentence pairs (each pair on a separate sheet of paper) and was asked to rate the relative plausibility of the two sentences with three options: (a) Sentence A is more plausible, (b) sentence B is more plausible, or (c) sentences A and B are equally plausible. (The sentences were presented in their entirety.) The sentence frames that the majority of these raters gave a rating favoring one sentence over the other were improved on the basis of their comments (we also asked them to write any comments they might have about the sentences). Changes were made to six pairs of sentences and almost all changes dealt with a fragment following the target word. The same procedure was used in Experiment 2 with seven raters, and five sentence pairs were changed on the basis of the ratings and comments.

Table 2
*Experiment 1: Various Reading Indexes as a Function of Second Constituent Frequency*

| Eye movement measure | Frequent 2nd constituent | Infrequent 2nd constituent |
|---|---|---|
| Gaze duration (in milliseconds) | 531 | 626 |
| First fixation duration (in milliseconds) | 214 | 215 |
| Second fixation duration (in milliseconds) | 204 | 212 |
| Probability of a first fixation[a] | .995 | .992 |
| Probability of a second fixation[a] | .880 | .905 |
| Probability of a third fixation[a] | .436 | .579 |
| Location of first fixation[b] | 4.43 | 4.34 |
| Location of second fixation[b] | 8.04 | 7.81 |

[a]These are the conditional probabilities of making a first, second, or third fixation on a word, given a valid trial. [b]Fixation location is in character spaces from beginning of the word.

on the target word with the more-frequent second constituent was 95 ms less than the mean gaze duration on the target word with the less-frequent second constituent, $t_1(23) = 7.40, p < .001, t_2(31) = 4.05, p < .001$.[3] Because the word frequencies were exactly controlled and the first constituent frequencies and other word characteristics were almost perfectly controlled (see Table 1), this indicates that the frequency of the second constituent had a large effect on the time to access the target word. To simplify exposition below, we will refer to differences between the two second constituent frequency conditions as "frequency effects."

Even though there was a large frequency effect on gaze duration, there was no frequency effect on the first fixation: There was only a 1-ms difference between the first fixation durations of the two conditions, and there was virtually no difference in the location of the first fixation duration ($ts < 1$). There were indications of a frequency effect on the second fixation, however. There was an 8-ms frequency effect on the second fixation duration, although it was not significant over items, $t_1(23) = 2.64, p = .015, t_2(31) = 1.31, p = .20$; a 2.5% difference in the probability of a second fixation, $t_1(23) = 1.70, p = .10, t_2(31) = 1.74, p = .09$; and about a quarter of a character difference in the location of the second fixation, $t_1(23) = 1.37, p = .18, t_2(31) = 1.08, p > .20$. The clearest frequency effect, however, was on the probability that the word needed to be fixated a third time: The words with more-frequent second constituents were fixated 14% less often than the words with the less-frequent constituent, $t_1(23) = 6.61, p < .001, t_2(31) = 3.81, p < .001$. Because there were many fewer observations on third fixations (with many missing cells in both the participant and item analyses), the fixation duration and fixation location measures were noisy and nonsignificant (all $ts < 1$). In fact, the differences in the third fixation duration went in opposite directions in the participant and item analyses. We will not discuss either of these third fixation measures below.

*Block analyses.* We used a design in which the part of the sentence preceding the target word was used in both blocks. Hence, there is some concern that some of the above effects might have only appeared in the second block and thus would be an artifact of this repetition. This hypothesis, however, is unlikely because the repeated part of the sentence was only about two to three words in length and fairly devoid of content. However, to ascertain whether this was a problem, subsidiary analyses were conducted in which block was an explicit factor in the analysis in addition to frequency condition. Fortunately, our design did not appear to be a problem as the interaction between block and second constituent frequency did not approach significance for any of the variables reported in Table 2 (the smallest $p$ value was .15).

*Lexicality and transparency analyses.* Two concerns about the generality of the above findings are (a) whether the stimuli were really perceived as words by the participants or merely thought of as novel expressions coined by the experimenters and (b) whether the words were semantically transparent. To assess each of these concerns, we had other native Finnish speakers rate the stimuli. For the novelty rating, we prepared a word list containing all the target words together with 32 high-frequency and 32 low-frequency (i.e., only one occurrence in the Laine and Virtanen, 1996, word corpus) filler compounds. Eighteen psychology students, who did not participate in the experiment proper, gave a familiarity rating for these words by merely marking the words they have seen before and the words that were novel to them. To assess the semantic transparency of the target words, we asked five linguistically experienced faculty members to rate each target word as transparent, opaque, or something in between. We converted the ratings into numbers by giving 2 points for a transparent rating, 1 point for an in-between rating, and 0 points for an opaque rating. We summed the ratings, so that a score of 10 was the highest transparency rating.

Fortunately, neither of the concerns appears to be serious. First, most of the words in the experiment were rated as quite familiar by the raters. More than half of the words were rated as familiar by all 18 raters, and the mean familiarity ratings were about the same for the two sets of words: 0.83 and 0.81 for the high-frequency and low-frequency second constitu-

---

[3] Analyses were also performed in both Experiments 1 and 2 in which all trials on which any individual fixation duration on the target words was less than 50 ms or more than 400 ms were excluded. As there were few such trials, the analyses were virtually the same as those presented.

Table 3

*Experiment 1: Various Reading Indexes as a Function of Second Constituent Frequency and Whether the Initial Fixation (Fix1) Landed on Letters 0–4 or Letters 5–8 of the Target Word*

| | Region 1: Fix1 on 0–4 | | Region 2: Fix1 on 5–8 | |
|---|---|---|---|---|
| Eye movement measure | Frequent 2nd constituent | Infrequent 2nd constituent | Frequent 2nd constituent | Infrequent 2nd constituent |
| Number of first fixations ($N$ = 32) | 19.8 | 20.2 | 12.0 | 11.5 |
| Gaze duration (in milliseconds) | 543 | 628 | 517 | 657 |
| First fixation duration (in milliseconds) | 203 | 203 | 233 | 233 |
| Second fixation duration (in milliseconds) | 218 | 219 | 191 | 202 |
| Probability of a second fixation given a first fixation | .904 | .925 | .814 | .912 |
| Probability of a third fixation given a first fixation | .478 | .608 | .357 | .566 |
| Mean location of second fixation[a] | 8.13 | 7.91 | 8.12 | 7.89 |

[a] Fixation location is in character spaces from beginning of the word.

ent sets, respectively. Second, virtually all the words in both sets were rated as transparent by all the reviewers (the mean ratings were 9.96 and 9.59 of 10 for the high-frequency and low-frequency second constituent sets, respectively).

To assess whether these small differences between the two sets could have artifactually produced the second-constituent frequency differences reported above, we used the differences between item pairs on familiarity and transparency as predictors in regression equations and differences in dependent variables on an item pair (e.g., difference in gaze duration) as the dependent variables. In these analyses, the intercept is the best estimate of the difference on the dependent variable when the confounding on the predictor variable is controlled for. Moreover, one can assess whether the intercept is significantly different from zero. Briefly, these analyses revealed the same general pattern as in the item analyses above. The sizes of the interesting effects (gaze duration, second fixation duration, and the probability of a second and third fixation) were virtually identical to those reported in the main analysis and the significance values were quite similar to those in the main analysis.[4] What was perhaps most surprising was that the difference in familiarity rating (which did vary appreciably over item pairs) had very little predictive power on any of the dependent variables (all $t$s $< 1$, and the direction of the effect was opposite to what one would expect for gaze duration and the percentage of second fixations). This suggests that whether these words are perceived as familiar or not played very little role in their processing.

*Assessing the effect of the initial landing position.* One possible explanation for the marginal effects of the second morpheme frequency on the second fixation was that, on many of these trials, the reader had initially landed near the beginning of the word, which may be a "bad location" for processing the target word (e.g., Vitu, O'Regan, & Mittau, 1990). In Vitu et al.'s analysis, when the initial fixation was near the beginning of a word, there is an immediate refixation programmed to the middle of the word, resulting in little lexical processing on the first fixation and brief fixations whose durations are not influenced by lexical factors. In contrast, when the initial fixation is near the middle of the word, there is presumably no need to make a quick refixation, so that initial lexical processing on the first fixation should be more extensive. Hence, first fixation durations should be longer than those on the beginning of the word and should be influenced by lexical factors. Moreover, when the initial fixation location is in a more optimal location, processing should be more advanced when the second fixation begins and thus it should be more likely for characteristics of the second morpheme to influence processing on it.

To assess the influences of initial landing position, we conducted subsidiary analyses in which the initial landing position on the target word was also a factor in the analysis. To simplify this analysis, we divided the trials into those in which the participant's first fixation landed on Locations 0–4 (Location 0 is the space preceding the word) and those in which the participant's first fixation landed on Locations 5–9 (the latter locations should be close to optimal). Approximately two thirds of the data were in the former condition and one third in the latter, with a negligible number of trials in which the initial landing position was further into the word than Letter Position 9 (see Table 3). Consistent with the above analysis, the frequency effects tended to be bigger and earlier when the initial landing position was in the middle of the word, although the interaction of second constituent frequency and initial landing position was marginal in most of the analyses. First, although there was no overall difference in gaze duration as a function of the initial fixation location on the target word ($F$s $< 1$), the frequency effect on gaze duration appeared to be a bit bigger when the initial landing position was in the middle (140 ms vs. 85 ms), although the interaction of frequency with initial fixation location was not significant, $F_1(1, 23) = 2.34, p = .14, F_2(1, 31) = 2.40, p = .13$. Also consistent with Vitu et al.'s (1990) analysis, first fixation durations were 30 ms shorter when the initial fixation was on the beginning of the word; however, as can be seen in Table 3, there was clearly no frequency effect for either initial fixation location condition. Thus,

---

[4] The intercept values and significance levels for the various measures are as follows: gaze duration, 76 ms, $t(29) = 2.95, p = .006$; second fixation duration, 8 ms, $t(29) = 1.59, p = .12$; probability of a second fixation, 2.8%, $t(29) = 1.74, p = .09$; and probability of a third fixation, 12.9%, $t(29) = 2.99, p = .006$.

even when the reader was fixated in a good place to process the second constituent, it had no effect on the first fixation duration. The initial fixation location appeared to modulate the frequency effect on the second fixation, as the frequency effect was only 1 ms when the initial fixation location was at the beginning of the word, but 11 ms when it was in the middle. However, the interaction of frequency and initial fixation location was not significant, $F_1(1, 23) = 2.05, p = .17, F_2(1, 31) = 1.28, p > .20$, and the 11-ms effect when the initial fixation was in the middle of the word was only marginally significant over items, $t_1(23) = 2.23, p < .05$, $t_2(30) = 1.63, p = .11$.

The initial landing position had similar effects on the probability of refixation. First, when the initial fixation location was in the middle of the word, the probability of a second fixation marginally decreased, $F_1(1, 23) = 3.68, p = .07, F_2(1, 31) = 9.67, p = .004$, and the probability of a third fixation clearly decreased, $F_1(1, 23) = 5.57, p = .027, F_2(1, 31) = 57.1, p < .001$. In addition, the frequency effects on the probability of a second fixation and the probability of a third fixation appeared to be bigger when the initial location was in the middle, although the interaction effects were not reliable, $F_1(1, 23) = 3.16, p = .09$; $F_2(1, 31) = 2.01, p = .17$; $F_1(1, 23) = 5.41, p = .029$; and $F_2(1, 31) = 3.27, p = .08$, respectively. The initial landing position, somewhat surprisingly, had virtually no effect on the location of the second fixation ($F$s $< 1$). The second fixation appeared to be targeted to the center of the word regardless of initial landing position.

To summarize, when readers' initial landing position was closer to the middle of the word and in a better position to encode the second constituent, there were some suggestions that the frequency of the second morpheme had a bigger effect on processing. However, most of these effects were marginally significant.

*Launch site analyses.* Another possible variable that may mediate processing is the *launch site* of the saccade that leads to the first fixation on the target word (i.e., the location of the fixation prior to the first fixation on the target word). When the launch site is far from the target word, lexical effects may become attenuated or delayed, or both, because (a) there may be less parafoveal processing of the target word and (b) the initial fixation location on the target word would tend to be in a less advantageous location (see above). To determine whether some of our effects were modulated by this variable, we conducted an analysis in which trials were divided up between those in which the launch site was four or fewer characters from the space preceding the target word and those in which the launch site was five or more characters from the space preceding the target word. In fact, these analyses did not yield much information of interest. Although there was a slight attenuation of effects when the launch site was farther away, the only significant effect was a 1.5-character main effect of launch site on the location of the first fixation, $F_1(1, 23) = 103, p < .001, F_2(1, 29) = 56.0, p < .001$. This led, as might be expected from the above landing location analyses, to somewhat longer gaze durations, somewhat shorter first fixation durations, and somewhat longer second fixation durations, for trials when the launch site was farther from the target word. No other main effect of launch site and none of the interactions of launch site with second constituent frequency approached significance (most $F$s were less than one).

## *Discussion*

As indicated earlier, the time to identify the target words was strongly affected by the frequency of the second constituent, but the effect surfaced relatively late in the eye movement record: There was no effect on the first fixation, suggestive effects on the second fixation, and a clear effect only on the probability of making a third fixation on the word. This contrasts with our earlier finding (discussed above) that the frequency of the initial constituent had a significant effect on the first fixation duration on the target word, marginal effects on the second fixation duration, and large effects on both the probability of a second fixation and the location of that second fixation (i.e., the second fixation was farther into the word for the words with a more-frequent constituent). Hence, the data of the two experiments indicate that the two constituents each have an effect but differ in the time course. The data are thus consistent with the decomposition model, which posits that the two constituents are processed in series and then the meaning of the word is composed from the two. (Other, less serial, explanations are, of course, also consistent with the data.)

These results are clear evidence against the sequential search model, which predicts that the frequency of the second constituent should not affect lexical processing. However, it might have been the case that the frequency manipulation of the second constituent in our experiment was confounded with position in the file drawer. To assess the impact of position in the file drawer, all words beginning with the first constituent were counted as being in the file drawer and the rank of each word was computed (i.e., the number of stimuli in the file drawer that had higher frequencies than the target word plus one). In fact, there were small differences in the mean ranks for the two types of words: 36 for the words with higher-frequency second constituents and 52 for the words with lower-frequency second constituents.

To determine whether this difference in ranks could have produced the observed differences between the two second-constituent frequency conditions in the experiment, regression analyses were run, similar to those described above examining transparency and familiarity effects. In fact, difference in rank in the file drawer had no significant effect in any of the analyses. Moreover, the intercepts of the regression lines (which would be the best estimate of the dependent variable when the difference in ranks is equal to zero) were similar to those reported in Table 2.[5] It is also worth noting that the file drawer variable, difference in rank,

---

[5] The intercept values on the variables of interest are as follows: gaze duration, 89 ms, $t(30) = 3.60, p < .001$; second fixation duration, 7 ms, $t(30) = 1.35, p < .20$; probability of a second fixation, 2.6%, $t(30) = 1.65, p < .20$; and probability of a third fixation, 15.2%, $t(30) = 3.79, p < .001$.

had virtually no predictive power: None of the $t$ values in any of the regression analyses for this variable exceeded 1 and most were below 0.5. Moreover, this was probably not due to a restricted range because the difference in rank varied over pairs from 149 to −119 ($SD = 47$). Most notably, the fact that difference in rank had little power in predicting the probability of a third fixation (presumably indexing a late stage of lexical processing) suggests that the file drawer model is not applicable to processing these Finnish compounds.

We were also interested in whether the frequency of the second morpheme affected the location of fixations in a word, because our earlier experiment (Hyönä & Pollatsek, 1998, Experiment 2) found significant effects of first constituent frequency on where the eye lands. That is, when the first constituent was common, the first fixation location was about 0.2 characters farther into the word and the second fixation location was about 0.6 characters farther into the word (both effects were significant). These findings are consistent with the processing-difficulty hypothesis, which posits that with increasing difficulty in parafoveal or foveal processing, the perceptual span around the fixation from where useful information is picked up is narrowed (cf. Henderson & Ferreira, 1990), which should lead to a shorter saccade into a new, still unprocessed text segment. The results of the current experiment offer some support to that hypothesis, as the second fixation location was about 0.2 characters farther into the word when the second constituent was frequent; however, this difference was not significant in the item analysis. (Given that the other effects of the second constituent frequency were relatively late in processing, it would have been unlikely for it to have had an effect on the first fixation location.)

## Experiment 2

The data of Experiment 1 are consistent with a completely decompositional model, but there are clearly other possible explanations for these data. A further test of such a model is to manipulate the frequency of the entire word. If processing of compounds has three serial stages, with only the last stage affected by properties of the whole word, then one would expect the frequency of the compound word to emerge only after the properties of the components manifested themselves. In the present case, it means that properties of the whole compound should emerge only after the properties of the second constituent emerged, which would be no earlier than the second fixation duration and probably later. In contrast, a model where there is a "race" between a compositional access of the meaning of the compound and a direct lexical lookup of the compound predicts that whole-word effects could occur at least as rapidly as second constituent effects.

A second way to assess the whole-word frequency effect for compound words is to compare it to the frequency effect for monomorphemic words of similar frequency and length. If, as argued above, whole-word effects for compounds only occur during the third of three serial stages, then one might expect that frequency effects would occur earlier for the monomorphemic words than for the compounds. In contrast, if compound-word processing is not such a serial process (e.g., if direct lexical lookup of the compound occurs in parallel with a decompositional process), one would expect significant effects of whole-word frequency for compounds and might expect them to surface as early as for matched noncompounds.

### Method

*Participants.* Twenty-four students from University of Turku, all of whom were native speakers of Finnish, participated in the experiment as part of the course requirement.

*Materials and design.* Four sets of words (20 words in each set) were used as target items. Two sets were two-noun compound words (11–15 characters long), with the members of one set having whole-word frequencies substantially greater than the other set (see Table 4). Similarly, there were two sets of monomorphemic words (11–15 characters long), with the members of one set having substantially greater whole-word frequencies than the other set. As seen in Table 4, all four sets were fairly closely matched for average bigram frequency, for beginning and final trigram frequency, and for word length. In addition, the two high-frequency sets and two low-frequency sets were each approximately equated on mean

Table 4
*Characteristics of the Target Words Used in Experiment 2*

| Stimulus characteristic | More-frequent compound | | Less-frequent compound | | More-frequent monomorpheme | | Less-frequent monomorpheme | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Frequency of compound word[a,b] | 41 | 31 | 0.4 | 0.4 | 45 | 62 | 0.7 | 0.4 |
| Frequency of first constituent[a,c] | 78 | 88 | 78 | 95 | — | | — | |
| Frequency of second constituent[a,c] | 117 | 112 | 116 | 110 | — | | — | |
| Average bigram frequency | 6,958 | 2,030 | 7,470 | 1,531 | 8,213 | 1,754 | 7,450 | 1,820 |
| Beginning trigram frequency | 796 | 533 | 703 | 595 | 897 | 657 | 982 | 897 |
| Final trigram frequency | 1,152 | 1,051 | 1,036 | 896 | 1,507 | 1,136 | 1,120 | 928 |
| Length of compound word (in characters) | 12.1 | 1.2 | 12.1 | 1.2 | 11.7 | 1.1 | 11.7 | 1.2 |
| Length of first constituent (in characters) | 6.5 | 1.3 | 6.5 | 1.3 | — | | — | |

[a] These frequencies have been converted to frequencies per million to make them more comparable to the norms in English. [b] The frequencies are of the compound words in all their inflectional forms. [c] The frequencies are of the constituents as separate words in all their inflectional forms, including the nominative case.

frequency. Perhaps most critically, the two sets of compounds were quite closely matched on both first constituent frequency and length as well as second constituent frequency and length. The stimulus, constituent, bigram, and trigram frequencies were computed using the same Laine and Virtanen (1996) corpus used in Experiment 1.

As in Experiment 1, the target words were near the beginning of sentences but never in the initial word position. Two words, one from each compound-word type, were paired, and the sentence frames for each word pair were identical through the word following the target word. The target words always appeared in the nominative case. There were two blocks of sentences in which the two sentences containing each word pair appeared in separate blocks. The order of the blocks was counterbalanced across participants. Thus, each participant saw all the critical target words. Within each block, the order of target sentences was randomized. There were 40 filler sentences among the critical sentences.

*Apparatus and procedure.* The apparatus and procedure were the same as in Experiment 1.

### Results

The data indicate that (a) whole-word frequency had a sizable effect on processing these compound words and (b) this effect was not particularly late, because it affected processing at least as early as the second constituent did. First, the whole-word frequency had an 82-ms effect on the gaze duration on the compound target words, $t_1(23) = 4.05$, $p < .001$, $t_2(19) = 3.60$, $p = .002$. The 34-ms frequency effect for the monomorphemic words, although significant, $t_1(23) = 4.05, p < .001, t_2(19) = 2.71, p = .014$, was quite a bit smaller than for the compounds; however, the difference between the frequency effects for the two types of words was only marginally reliable over items, $F_1(1, 23) = 17.7, p < .001, F_2(1, 38) = 3.677, p = 0.06$ (see Table 5).

The fact that the frequency effect for gaze duration was greater for the compounds than for the monomorphemes was initially somewhat surprising. However, the monomorphemic words were mostly loan words and thus had little resemblance to other words of similar length in Finnish. Hence, it may have been the case that fewer letters in these words needed to be identified in order for the word to be identified. Whatever the reason, the difference between the two frequency effects makes it a bit problematic to assess when whole-word frequency affects processing for the two types of words. Nonetheless, even though the pattern of data below is not identical for the two types of words, it seems to indicate that whole-word frequency effects (henceforth referred to as *frequency effects*) occurred at roughly the same time for the two types of words.

First, consider individual fixation durations. As seen in Table 5, first fixation durations on the higher-frequency words were 5 ms shorter for the compounds and 6 ms shorter for the monomorphemes. Whereas the difference for the monomorphemes was not close to being significant, $t_1(23) = 1.37, p = .18, t_2(19) = 1.19, p > .20$, the difference for the compounds was actually significant in the participant analysis, $t_1(23) = 2.11$, $p = .05$, $t_2(19) = 1.66$, $p = .11$. For second fixation duration, the frequency effect was substantially larger for the compound words. The 16-ms difference for the compound words was quite reliable, $t_1(23) = 3.66$, $p = .001$, $t_2(19) = 4.10$, $p = .001$, whereas the 9-ms difference for the monomorphemes was not, $t_1(23) = 1.54$, $p = .14$, $t_2(19) = 1.58$, $p = .13$; however, the difference between the two frequency effects was not close to significant, $F_1(1, 23) < 1, F_2(1, 38) = 2.08, p = .16$.

The pattern for fixation durations indicates that whole-word frequency was clearly affecting processing of the compound words at the time of the second fixation, with a hint that it may even have affected processing on the first fixation (but see below); the effects for the monomorphemes were somewhat smaller and unreliable. When we look at the number of refixations, however, the pattern is somewhat reversed. That is, the probability of refixating a first time was modulated by frequency for both classes of words, but the effect was bigger and more reliable for the monomorphemes. The difference in the probability of making a second fixation was 4.6% for the compounds, $t_1(23) = 2.20$, $p = .038$, $t_2(19) = 1.28$, $p > .20$, and 7.1% for the monomorphemes, $t_1(23) = 2.34$, $p = .028$, $t_2(19) = 2.11$, $p = .048$. Frequency significantly modulated the probability of making a third fixation for both types of words. There was a 17.3% difference for the compounds, $t_1(23) = 6.39, p < .001, t_2(19) = 4.28, p < .001$, and a 5.0% difference for the monomorphemes, $t_1(23) = 2.56$, $p = .017$, $t_2(19) = 2.30$, $p = .033$. (There were so few third fixations that it was not

Table 5
*Experiment 2: Various Reading Indexes as a Function of Whole-Word Frequency*

| Eye movement measure | Frequent compound | Infrequent compound | Frequent monomorpheme | Infrequent monomorpheme |
|---|---|---|---|---|
| Gaze duration (in milliseconds) | 345 | 427 | 310 | 344 |
| First fixation duration (in milliseconds) | 195 | 200 | 202 | 208 |
| Second fixation duration (in milliseconds) | 164 | 180 | 159 | 168 |
| Probability of a first fixation[a] | .971 | .973 | .988 | .979 |
| Probability of a second fixation[a] | .652 | .698 | .525 | .596 |
| Probability of a third fixation[a] | .146 | .319 | .092 | .142 |
| Location of first fixation[b] | 3.96 | 4.03 | 4.15 | 4.18 |
| Location of second fixation[b] | 7.48 | 7.67 | 7.16 | 7.34 |

[a] These are the conditional probabilities of making a first, second, or third fixation on a word given a valid trial. [b] Fixation location is in character spaces from beginning of the word.

worth analyzing any of their characteristics, such as their mean duration or mean location.)

Fixation locations were slightly farther into the word for the infrequent words, but none of the whole-word frequency effects on mean fixation location was close to reliable (all *t*s $< 1$). Because the direction of the effect was opposite of what one would expect in terms of processing difficulty (i.e., longer saccades for more frequent words), it seems safe to conclude that there was no effect of whole-word frequency on the location of fixations.

To summarize, the effect of whole-word frequency on gaze duration was substantial for the compound words and was about the same as the effect of second constituent frequency observed in Experiment 1. Moreover, the pattern of the whole-word frequency effects indicated that whole-word frequency was influencing processing at least as early as the second constituent in Experiment 2. The big difference in frequency effect on gaze duration for the monomorphemes and compounds made it hard to assess the relative time course of the effect of whole-word frequency on the two types of words. However, the pattern of data seems to indicate that the time course was about the same for the two. The gaze durations were shorter for the monomorphemic words than for the compound words (a finding also reported by Inhoff, Brühl, & Schwartz, 1996, for English compounds and monomorphemes); however, because these words undoubtedly differed on all sorts of key indexes, such as the pattern of bigram and trigram frequencies and neighborhood properties, we think it is hard to draw any definite conclusions from this fact.

*Block analyses.* As in Experiment 1, we employed a design in which the part of the sentence preceding the target word was used in both blocks, and therefore some of the above effects could have been due to effects that only occurred in the second block and were thus an artifact of this repetition. Accordingly, subsidiary analyses were conducted in which block was an explicit factor in the analysis in addition to frequency condition. The only frequency-modulated effect that was significant (most of the interactions of block with other variables had *F*s $< 1$) was the interaction of frequency and block on first fixation duration for the compounds, $F_1(1, 23) = 4.63, p = .04, F_2(1, 38) = 5.32, p = .03$. This interaction reflected the fact that the frequency effect displayed in Table 5 was actually about 2 ms in the opposite direction in Block 1. This indicates that one should be skeptical about whether whole-word frequency modulated the first fixation duration for compound words.

*Transparency analysis.* As in Experiment 1, we assessed the compound words to determine whether they were in fact perceived as transparent. (The same raters were employed as those assessing the Experiment 1 stimuli.) In fact, the compounds in this experiment were rated as a bit less transparent than those in Experiment 1, with mean scores (of a possible maximum of 10) of 8.8 for the high-frequency words and 9.3 for the low-frequency words. To make sure that this small transparency difference wasn't the cause of any of the above differences attributed to frequency differences, we conducted regression analyses similar to those in Experiment 1. In fact, the difference in transparency had very little effect on all the measures (all *t*s $< 1$), and the intercepts in the analyses were quite similar to those in the main analysis.[6]

*Assessing the effect of the initial landing position.* As in Experiment 1, we wanted to assess whether any marginal frequency effects would become more robust if one eliminated trials on which the reader initially fixated near the beginning of the word. In this analysis, we divided the initial landing position into two groups: (a) those on which the first fixation was on the space before the target words up to the boundary between the third and fourth letters and (b) those in which the first fixation was to the right of this boundary. We partitioned it this way, which was a bit different from Experiment 1, because there were substantially fewer second fixations than in Experiment 1 (probably because of the higher frequency of the words) and thus some missing data. However, the data looked quite similar when the data were partitioned as in Experiment 1 (see Table 6).

Roughly half of the initial landing locations were in each of the two groups (see Table 6). As in Experiment 1, the initial landing location appeared to modulate some of the frequency effects, although the interaction of initial landing position with almost all the effects was quite marginal. For gaze duration, the initial landing position had virtually no effect on the frequency effect for the monomorphemes, but the frequency effect appeared to be less for the compounds when the initial fixation was in the middle (however, *F*s $< 1$). This effect is essentially opposite to the one observed in Experiment 1. As in Experiment 1, there was an overall initial-landing-position effect on first and second fixation durations: When the initial fixation was near the beginning of the word, first fixation durations were shorter, $F_1(1, 23) = 69.5, p < .001, F_2(1, 38) = 49.8, p < .001$, and second fixation durations were longer, $F_1(1, 23) = 10.5, p = .004$, $F_2(1, 38) = 4.85$, $p = .034$. The pattern for the probability of refixating was similar. Although there were markedly fewer second or third refixations when the initial landing position was away from the beginning of the word—$F_1(1, 23) = 33.1, p < .001, F_2(1, 38) = 155, p < .001, F_1(1, 23) = 9.88, p = .002$, and $F_2(1, 38) = 81.7, p < .001$, respectively—there was virtually no modulation of the size of frequency effects by the initial landing location (all interactions had an $F < 1$ in both participant and item analyses). There was also virtually no interaction of initial landing location with frequency on the location of the second fixation (all *F*s $< 1$). Given that there was so little of interest with respect to frequency effects in this analysis, we didn't do a launch site analysis for Experiment 2.

*Assessment of file drawer effects.* Unlike in Experiment 1, there was a clear relationship between the frequency manipulation for compound words and the location in the file drawer. That is, if one holds first morpheme frequency

---

[6] The intercept values on the variables of interest are as follows: gaze duration, 83 ms, $t(18) = 3.48$, $p = .003$; second fixation duration, 18 ms, $t(18) = 3.79, p = .001$; probability of a second fixation, 5.6%, $t(18) = 1.53$, $p = .14$; and probability of a third fixation, 17.7%, $t(18) = 4.16, p = .001$.

Table 6

*Experiment 2: Various Reading Indexes as a Function of Whole-Word Frequency and Whether the Initial Fixation (Fix1) Landed on Letters 0–3 or Letters 4–End of the Target Word*

| | Region 1: Fix1 on 0–3 | | | | Region 2: Fix1 on 4–end | | | |
|---|---|---|---|---|---|---|---|---|
| Eye movement measure | Frequent compound | Infrequent compound | Frequent monomorpheme | Infrequent monomorpheme | Frequent compound | Infrequent compound | Frequent monomorpheme | Infrequent monomorpheme |
| Number of first fixations ($N = 24$) | 8.8 | 8.8 | 8.6 | 8.1 | 10.6 (10.4) | 10.7 (10.5) | 11.2 (11.0) | 11.5 (10.9) |
| Gaze duration (in milliseconds) | 359 | 455 | 323 | 357 | 325 (326) | 405 (406) | 359 (326) | 455 (406) |
| First fixation duration (in milliseconds) | 182 | 185 | 186 | 193 | 207 (207) | 208 (208) | 215 (212) | 222 (220) |
| Second fixation duration (in milliseconds) | 176 | 196 | 169 | 174 | 152 (151) | 180 (173) | 160 (156) | 168 (168) |
| Probability of a second fixation given a first | .814 | .847 | .677 | .719 | .542 (.540) | .654 (.659) | .436 (.451) | .554 (.551) |
| Probability of a third fixation given a first | .160 | .389 | .104 | .150 | .119 (.119) | .254 (.255) | .072 (.083) | .131 (.126) |
| Mean location of second fixation | 7.27 | 7.37 | 6.95 | 6.98 | 8.23 (8.24) | 7.93 (7.98) | 7.89 (7.96) | 7.56 (7.60) |

*Note.* Data in parentheses indicate data used only when the initial landing position was from the beginning of the fourth letter to the end of the seventh letter. Fixation location is in character spaces from beginning of the word.

constant and varies the frequency of the compound, the position of the word in the file drawer should also vary considerably. In fact, for the more-frequent compounds, the rank in the file drawer was almost always 1 or 2 (mean = 1.3), whereas for the less frequent compounds, it was usually considerably lower (mean = 26). However, there was considerable variability for the less-frequent compounds, so that the differences in rank between the less-frequent and more-frequent compound word of a matched pair of compounds varied between 1 and 82 ($SD = 22$). As a result of this variability in the rank difference scores (even though they all had the same sign), it seemed feasible to do the same regression analyses as those in Experiment 1, with the difference in file drawer rank as the predictor.

The results of these analyses were somewhat different from those in Experiment 1, because rank in the file drawer did have predictive power for one of the dependent variables, second fixation duration, $t(16) = 3.04$, $p = .008$. However, the intercepts (reflecting the best prediction for each of the dependent variable effects when difference in rank is controlled for) were pretty much the same as in the main analyses: gaze duration, 79 ms, $t(18) = 2.26, p = .037$; first fixation duration, 2 ms, $t < 1$; second fixation duration, 9 ms, $t(18) = 1.48, p = .18$; percentage of second fixations, 6.6%, $t(18) = 1.19$, $p > .20$; and percentage of third fixations, 13.1%, $t(18) = 2.15$, $p = .045$. The regression analyses thus indicate that word frequency is having its effect pretty much independently of file drawer location; however, the second fixation duration data suggest that the location in the file drawer may have some effect on later processing for the words in this experiment.

## General Discussion

Experiment 1 demonstrated that the frequency of the second constituent of a compound word had a substantial effect on the time to identify the compound word, because it had almost a 100-ms effect on gaze duration. However, this effect occurred relatively late in processing the target word, because there was no effect on the first fixation probability, duration, or location, and the effects on second fixation duration, probability, and location were relatively modest and not reliable. Instead, almost all of the effect on gaze duration was produced by a modulation of the probability of making a third fixation on these words. The finding that the second constituent frequency had a significant influence on the compound-word processing is consistent with previous lexical-decision and priming studies (Andrews, 1986; Lima & Pollatsek, 1983; Monsell, 1985; Sandra, 1990; Zwitserlood, 1994), except for that of Taft and Forster (1976), who employed nonwords as constituents of compounds.

Experiment 2 demonstrated that the frequency of the compound word had about the same effect on the gaze duration on the target compound words as the frequency of the second constituent did in Experiment 1. In addition, contrary to a completely serial model of constituent processing, the effect occurred no later than the second constituent effect, and probably earlier. That is, there was a relatively large and reliable effect of whole-word frequency for compounds on the second fixation duration and a suggestion that it affected the probability of the second fixation, and possibly even the first fixation duration under certain circumstances. Moreover, a comparison of the pattern of data for compounds and the matched monomorphemic words in Experiment 2 indicated that the influence of whole-word frequency was certainly not delayed for the compounds relative to the monomorphemes.

This pattern of data indicates that compositional processing is important in processing these Finnish compound words, but it also indicates that whole-word access can occur before access of the second constituent as a linguistic unit. This rules out Hyönä and Pollatsek's (1998) hypothesis that identification of transparent Finnish compounds is solely compositional and suggests that some sort of hybrid model is the best explanation. Let's consider the two types of hybrid models in turn.

The either/or version of a dual-route model claims that some compounds are always identified by a whole-word process, whereas others are always identified by a compositional process. For compound words, two candidates for how such a distinction may be made are differences in transparency and frequency. Because most of the high-frequency compounds in Experiment 2 were rated as quite transparent and because transparency differences had no predictive power in the analyses, this distinction can't explain the whole-word frequency effects observed in Experiment 2. Thus, if one were to champion an either/or model for Finnish compounds, one would have to claim that higher-frequency compounds were only identified by a whole-word route, whereas lower-frequency compounds were always identified by a compositional route.

There are basically two types of either/or models. The first posits that there is no a priori fixed order, but that an early "preprocessing" mechanism provides an early selection between the routes, leading to activation of the selected route and suppression or blocking of the other. The second posits that there is a fixed order. For compound words, it would be that the direct route is always the default, and the decomposition route comes to play only if the direct route fails. For the former type of model to make any sense, the preprocessing mechanism would have to decide very rapidly which route to take on the basis of some quick assessment of familiarity. The existence of such a fast-acting sensing mechanism may not be totally out of the question (cf. Reichle, Pollatsek, Fisher, & Rayner, 1998), but it doesn't seem very plausible. In addition, it isn't clear what benefit would accrue from shutting off the compositional route for higher-frequency compounds.

The second alternative appears to be functionally more plausible. However, it is at variance with the data. First, consider what our data indicate about the direct-access process. The earliest clear indications of direct-access processes occur on the second fixation, but it is not until the decision is made regarding whether to make a third fixation that whole-word frequency is having the bulk of its effect. It thus seems most plausible that the decision that "the direct route fails" would occur no earlier than at this time (i.e., relatively late into the second fixation). Given that, it would be difficult for such a model to explain how first constituent frequency could affect the duration of the first fixation (Hyönä & Pollatsek, 1998, Experiment 2) and how second constituent frequency could affect processes at more or less the same time as whole-word frequency. That is, if the sequence of operations assumed is "decide whole-word access fails—then access first constituent—then access second constituent," one would infer that first-constituent frequency effects would occur only after whole-word frequency effects and that second-constituent frequency effects would occur considerably after whole-word frequency effects. It thus seems that such an either/or model would be forced to deem the early constituent frequency effects observed in Experiment 2 of Hyönä and Pollatsek (1998) and Experiment 1 of the present study to be epiphenomenal (i.e., indicating that the eye movement system is influenced by the familiarity of the subcomponent but that these familiarity effects are irrelevant to accessing the compound). Such an explanation is possible but seems implausible, especially given the large effects on gaze duration produced by varying the frequencies of either the first or second constituent.

To summarize, an either/or model seems quite implausible. In contrast, the and/and or parallel version of a dual-route model gives a relatively straightforward account of all the data (see Frauenfelder & Schreuder, 1992, and Schreuder & Baayen, 1995, for a similar type of model suggested for the processing of inflected and derived words).[7] That is, if the two processes—the direct lookup and the compositional processes—occur in parallel, then it seems quite natural that access of the initial constituent would generally start to occur at least as rapidly as access of the whole word (i.e., that initial constituent effects occur before whole-word effects) and that second constituent effects would not necessarily have to occur after whole-word effects but that the two could overlap in time.

This leaves open the question about the effective finishing times of the two processes. That is, could it be that the processes occur in parallel but that the whole-word processes win the race for word identification almost all the time except for low-frequency words? In such a case, the and/and model would tend to look pretty much like an either/or model in terms of end result, if not in process. We think it unlikely, however, that the compositional process is restricted to an uninteresting set of essentially novel compound words. First, the whole-word frequencies of the compounds in Experiment 2 of Hyönä and Pollatsek (1998) that obtained constituent frequency effects weren't very low (5 per million).[8] Second, our ratings showed that, even for

---

[7] This type of model assumes that the race between the two routes starts right from the beginning of encountering a word, not taking into consideration that longer words (for example, eight letters or more) typically require more than one fixation, which may give the decomposition route a head start (this may explain why the whole-word frequency did not significantly influence the duration of the first fixation).

[8] In Hyönä and Pollatsek (1998), the Laine and Virtanen (1996) database was not available, and as a result, familiarity ratings were

the low-frequency compound words of Experiment 1, few Finnish college students viewed them as "novel." Moreover, in most languages, a large majority of compound words are fairly low in frequency, and these whole-word frequencies are even typical for a large majority of familiar compound words in English. We thus think that the compositional route is likely to play a significant role in the access of a large majority of compound words.

## Conclusions

One of our major motivations when we embarked on this research was to determine whether one could use eye movements to uncover compositional effects in processing compound words. A study of Finnish compounds appeared to be an ideal place to start, because it seemed unlikely that the entire word could be processed in a single fixation, so that the unfolding of subprocesses had a chance of playing out in time in the eye movement record. Our experiments have, in fact, demonstrated such an unfolding, indicating that eye movements can be used to study morphemic decomposition in natural reading. However, in other languages, where compounding is less productive and compounds tend to be shorter, it may be the case that one will obtain no evidence for morphemic decomposition in the eye movement record. This could either be because there is, in fact, no such morphemic decomposition in these languages or because there is morphemic decomposition but whole word processes are efficient enough to prevent the processes of morphemic decomposition from controlling the pattern of eye movements. What is perhaps more surprising in the present studies is that whole-word factors also play a part, even though (a) the whole word is unlikely to be encoded in a single fixation and (b) almost all the words were transparent.

A second final point needs to be made. Although we have tentatively accepted a model that posits a simple race between a morphemic decomposition process and a whole-word direct-access process, we think that the process of accessing Finnish compounds is probably more complex. The race model is consistent with the pattern of onsets of effects (i.e., when various frequency effects first appear in the eye movement record). However, the pattern of how long these effects last seems less readily explained by such a simple race model. For example, although there were early effects of first-morpheme frequency in Hyönä and Pollatsek (1998), a sizable fraction of the first-morpheme frequency effect on gaze duration was explained by the probability of making a third fixation. The latter effect seems quite unlikely in terms of a simple race model, because in such a model, the second morpheme is processed only after the first morpheme, and thus it is not clear how first-morpheme frequency effects can surface so late and can overlap so much in time with second-morpheme frequency effects and whole-word frequency effects. Instead, such overlap of effects in time seems more naturally explained by a model that posits fairly massive interaction between compositional and direct-access processes, analogous to a cooperative dual-route architecture posited by Carr and Pollatsek (1985; see Zorzi, Houghton, & Butterworth, 1998, for an interesting implementation of this idea). Another possibility is a "single-route" parallel distributed processing structure that posits lexemes or morphemes, or both, among the elements, such as an expanded version of a model like that of Plaut, McClelland, Seidenberg, and Patterson (1996).

## References

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

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28,* 297–332.

Carr, T. H., & Pollatsek, A. (1985). Models of word recognition. In D. Besner, T. G. Waller, & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 2–76). New York: Academic Press.

Cole, P., Beauvillain, C., & 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.

Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In G. Booij & J. van Merle (Eds.), *Yearbook of morphology 1991* (pp. 165–183). Dordrecht, Netherlands: Kluwer.

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 417–429.

Hyönä, J., & Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1612–1627.

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

Laine, M., & Virtanen, P. (1996). *Turun Sanomat* [a Turku newspaper] computerized lexical database [unpublished database and program].

Lima, S. D., & Pollatsek, A. (1983). Lexical access via an

---

used instead of whole-word frequencies. Given that the database is now available, we examined the frequencies of the two sets of words used in that experiment, and the average frequency was 5 per million. However, unfortunately, the whole-word frequencies of the two sets were not exactly the same, and the mean frequency of the words with higher-frequency initial constituents was somewhat higher than for the words with lower-frequency initial constituents (7 vs. 3 per million). To ensure that this relatively small uncontrolled difference in frequency was not responsible for the effects in that experiment that were putatively due to the frequency of the initial constituent, we did regression analyses similar to those outlined earlier. In fact, the intercepts for first fixation duration, second fixation duration, and gaze duration were 9 ms, $t(26) = 3.32$, $p = .003$; 8 ms, $t(26) = 1.93$, $p = .06$; and 76 ms, $t(26) = 3.42$, $p = .002$, respectively. These values (and the significance levels) are virtually identical to those reported in Hyönä and Pollatsek for first and second fixation duration (9 ms and 9 ms) and the difference in gaze duration is only slightly smaller than that reported in the original analysis (87 ms). Consistent with the above, there was virtually no predictive effect of difference in word frequency on either first or second fixation duration ($ts < .4$) but a marginal effect on gaze duration, $t(26) = 1.70$, $p = .10$.

orthographic code? The basic orthographic syllabic structure (BOSS) reconsidered. *Journal of Verbal Learning and Verbal Behavior, 22,* 310–332.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

Monsell, S. (1985). Repetition and the lexicon. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 2, pp. 147–196). London: Erlbaum.

Niemi, J., Laine, M., & Tuominen, J. (1994). Cognitive morphology in Finnish: Foundations of a new model. *Language and Cognitive Processes, 9,* 423–446.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation–verification model for letter and word recognition: The word superiority effect. *Psychological Review, 89,* 573–594.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Towards a model of eye movement control in reading. *Psychological Review, 105,* 125–157.

Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology, 42A,* 529–567.

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

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes, 9,* 271–294.

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

Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception & Psychophysics, 47,* 583–600.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1131–1161.

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

Received September 22, 1998
Revision received March 12, 1999
Accepted May 25, 1999 ■