CHAPTER 4

Processing of Finnish Compound Words in Reading

Jukka Hyönä
University of Turku

and

Alexander Pollatsek
University of Massachusetts

Abstract

In this chapter, we report data from four experiments on the identification of Finnish two-noun compound words during reading. In the experiments, we independently varied the frequency of the first and second compound word constituents and the frequency of the whole word; we also varied the constituent lengths while holding the length of the word and the frequencies constant. The primary processing measures were the durations and locations of eye fixations landing on the compound word. The data showed (a) that the frequency of the initial constituent influenced the duration of first fixation on the target word as well as later processing, (b) effects of second constituent frequency did not show up until the second fixation, whereas (c) the effects of word frequency emerged as early as the second constituent effect and perhaps even earlier. This pattern of data is consistent with a parallel dual-route model that assumes that the identification of compound words occurs via a direct look-up route and a decomposition route operating in parallel. The locations of fixations were affected both by the constituent length and constituent frequency. A refixation landed further into the word (a) the longer the initial constituent, and (b) the higher the frequency of either the first or second constituent. The frequency of the first constituent also produced a small but reliable effect on the incoming saccade. These frequency effects are largely consistent with a processing difficulty hypothesis, which posits a mechanism that is capable of fine-tuning the saccade length as a function of moment-to-moment fluctuations in the difficulty of carrying out lexical access. Finally, we also report data on the pupil size, which indicate
that the pupil size on the first fixation on the target word is larger than on the second fixation, possibly indicating that the identification process of long words requires more effort during the initial than later stages.

Introduction

In this chapter, we report and summarize data bearing on the question of how compound words are identified in reading, as revealed by readers’ eye fixation patterns. In cognitive psychology, a wealth of studies have been conducted to study the process of identifying printed words, but the majority of them has suffered from two shortcomings: the studies have focused on (a) the identification of short words and (b) words presented in isolation. Surprisingly little is known of how longer words, such as compounds, are processed during normal, continuous reading.

The registration of readers’ eye movements has proved to be a very fruitful way to study word identification during reading. Readers’ eye behavior consists of two components, fixations and saccades, that are assumed to be governed by relatively independent mental mechanisms (Rayner and McConkie, 1976; Rayner and Pollatsek, 1981). The visual intake of information takes place during eye fixations, and saccades bring new information to the foveal vision for scrutiny. The when mechanism that governs the duration of fixations decides when to terminate a fixation to move on in text, while the where mechanism that controls for the saccadic amplitude guides the eye to a specific location in text.

According to a widely held view (see e.g., Pollatsek and Rayner, 1990), the duration of fixations in reading is under the control of cognitive factors. It has been demonstrated that the fixation duration on a word reflects the speed with which the fixated word is accessed in the mental lexicon. Moreover, higher-order comprehension processes are capable of influencing the duration of individual fixations (see Rayner, 1998; Rayner and Sereno, 1994, for a review). In contrast, there are models that assign only a minor role to cognitive factors in controlling fixation durations. For example, in the strategy-tactics model proposed by O’Regan (1992), fixation durations are largely controlled by the location of a fixation on a word. That is, it is assumed that if a fixation is positioned non-optimally for word identification (e.g., in the beginning of the word), a short fixation will ensue that is not influenced by cognitive processing. On the other hand, it is conceded that a fixation located optimally near the center of the word will be longer and may be

---

1 However, according to a recent computational model of readers’ eye behavior (Reichle, Pollatsek, Fisher and Rayner, 1998; see also Chapter 27), this may not be due to completion of lexical access, but instead to a ‘familiarity check’ of the word that is also influenced by lexical variables such as word frequency.
affected by cognitive factors. (However, see O’Regan, Vitu, Radach and Kerr, 1994, for a less extreme position.)

The computation of where a saccade should be targeted is generally assumed to be carried out primarily under the guidance of low-level visual features of the text. Specifically, the *where* decisions in reading are believed to be made on the basis of word length and spacing information so that saccades are targeted to the center of the word to yield an optimal perception of the word. It has been shown that the eyes typically fixate somewhat to the left of the center of a word (the so-called *preferred landing position*). To achieve this, a saccade to a long word will be longer than to a short word, and a saccade leaving a long word will be longer than from a short word. However, there is some evidence, although disputed, that other, more linguistic, features of a word may be capable of influencing where the eyes fixate it. Hyönä, Niemi and Underwood (1989) showed that the incoming saccade lands further into the word for words in which the crucial information for identification is at the end. The generality of the finding was challenged by Rayner and Morris (1992), who could not replicate it using stimuli adopted from Underwood, Clews and Everatt (1990). There is also evidence, still undisputed, that orthographic regularity has a potential to influence the saccade trajectory. Hyönä (1995) and Beauvillain, Doré and Baudouin (1996) showed that a fixation lands closer to the word beginning when the word has an irregular letter cluster in the beginning (see also Chapter 11).

We think that this *when—where* distinction needs to be refined to be word-based (McConkie and Zola, 1984), or possibly constituent-based (see below). That is, there are abundant data that cognitive variables such as the word’s frequency and its predictability from prior context influence both (a) the probability of fixating a word and (b) the number of fixations it receives given that it is fixated (see Rayner and Pollatsek, 1989, for a review). Hence, we think it is more appropriate to make the distinction between the decision of which word (or other meaningful unit) to fixate and *where* on the word to fixate (see also Radach and McConkie, 1998). The former can be viewed as similar to a *when* decision as it is a decision of whether it is time to move on to a subsequent word and is heavily affected by cognitive variables, and the latter is a decision about *where* to land on a word, and seems largely uninfluenced by cognitive factors. Of course, a key question of interest with compound words is whether aspects of the constituents (such as their frequencies) have an influence on where readers fixate within a word.

In the following, we report data on the processing of long compound words as reflected in factors that affect the *when* decision (i.e., is it time to move on to the next word or constituent?) and factors that affect the *where* decision (i.e., where to target a saccade within a word or constituent). To date, very little is known about how long compound words are identified during reading. We started out with a working hypothesis that the identification of long compounds occurs in three
sequential stages: (1) the initial constituent is identified, (2) the second constituent is identified, and (3) the constituents are ‘glued’ together to form a meaning for the whole word. We suspected that this simple model of ours would not provide the ultimate description of the identification process, but we wanted to see how far we can get with such a straightforward stage model.

The experiments

In the experiments, we employed long (12–14 characters) two-noun compound words that typically receive more than one fixation. The compound words were embedded in single sentences, and readers were instructed to read the sentences for comprehension. There were 24 or 25 participants in each experiment. In all experiments, there were two compound word conditions and the number of words per condition varied between 20 and 51. Two target words, one from each compound word type, were paired, and the sentence frames for each word pair were identical up through the word following the target word. A typical sentence pair taken from Experiment 2) is as follows, with the target word in italics. The target words appeared near the beginning of sentences, but never as either the first word of the sentence nor in the initial or final position of a line. Eye movements were recorded using the EYELINK system (SR Research Ltd.).

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.)

High-frequency second constituent condition

“Tukholmassa kuorokonsertti onnistui kohtuullisesti, vaikka pitkä konserttiikertue oli uuvutanut kuoroolaisia.” (In Stockholm the choral concert succeeded moderately well, although the long concert tour had tired the singers.)

Our three-stage model assumes that properties of the initial constituent (e.g., constituent frequency and constituent length) affect an early processing stage, properties of the second constituent affect a relatively later processing stage, and properties of the whole word (while constituent properties are matched) affect a still later processing stage. Although there is no one-to-one correspondence between different processing stages and eye fixation measures, a processing sequence can be discerned from readers’ eye movement protocols: the initial fixation on the compound word reflecting the initial processing stages, the second fixation reflecting later processing stages as well, and a possible third fixation reflecting

3 The experiments were conducted using the EYELINK system (SR Research Ltd.).

4 The experiments are Experiment 1 and Experiment 2.

5 The experiments are Experiment 1 and Experiment 2.

2 In Experiment 2, only the length of the next word was equated.
still later stages. Thus, to the extent our model is correct, the properties of the initial constituent should have an effect on the duration and/or location of the initial fixation made on the compound word, although they may also have an effect on the duration and/or location of subsequent fixations as well. But what is crucial for our model, the effect of the initial constituent should begin to be seen at the earliest possible point, whereas the properties of the second constituent are not likely to influence the initial fixation. Instead, the second constituent properties would probably affect the location and/or duration of the second fixation made on the word. Finally, the properties of the whole word should not affect the eye movement record until the properties of the second constituent do.

In the following, we summarize the results of four experiments, in which different compound word features were manipulated (for a more detailed report of the experiments, see Hyönä and Pollatsek, 1998, for Experiments 1 and 4; and Pollatsek, Hyönä and Bertram, 2000, for Experiments 2 and 3). We also report some new results from follow-up analyses conducted on a subset of the complete data set. For the new analyses, detailed statistics will be reported (for the other results, the reader is asked to consult Hyönä and Pollatsek and Pollatsek et al., for more detail).

In Experiment 1, the frequency of the first constituent was manipulated while controlling for the familiarity of the whole word, the frequency of the second constituent and the length of both constituents. In Experiment 2, the frequency of the second constituent was varied while controlling for the frequencies of the first constituent, and the whole word and the lengths of both constituents. Finally, in Experiment 3, the frequency of the whole word was manipulated while matching the frequency and length of both constituents. In Experiment 3, also a set of frequency-matched and length-matched high and low frequency monomorphemic words were included against which the frequency effects for compound words could be compared. We present the data separately for the duration of fixations (i.e., the when decisions) and for the location of fixations (i.e., the where decisions). In Experiment 4, the length of the two constituents was varied (but the length of the whole word and the frequencies of the constituents and the whole word were

---

3 The experiment numbering we have adopted here is for ease of exposition and is not the chronological ordering of the experiments. What we are referring to Experiment 1 here is Experiment 2 of Hyönä and Pollatsek (1998) and what we are referring to as Experiment 4 here is Experiment 1 of Hyönä and Pollatsek. What we are referring to as Experiments 2 and 3 here are Experiments 1 and 2, respectively, of Pollatsek et al. (2000).

4 At the time of study, we did not have a representative word corpus available so that the words were matched using a familiarity rating instead of frequency counts. However, we subsequently did get the frequencies; follow-up analyses indicated that the small whole-word frequency difference apparent for the two types of compound words did not modulate the pattern of results.
controlled for): in one condition, the initial constituent was short (3–5 characters) and the second long (8–11 characters), and in the other condition, the initial constituent was long (8–11 characters) and the second short (3–5 characters).

In the final section, we report analyses in which pupil size was employed as a putative measure of cognitive effort (see Kahneman, 1973) during word identification. There is evidence reported in the literature suggesting that fluctuations in the processing load during language comprehension and production tasks are reflected in the pupil size (see Beatty, 1982, for a review of the earlier studies). That is, the pupil has shown to dilate with increasing task difficulty. For example, Hyönä, Tommola and Alaja (1995) showed that the relative difficulty of repeating back and translating auditorily presented English and Finnish words was reliably reflected in pupil size, and Just and Carpenter (1993) demonstrated that complexity of syntactic processing during reading is reflected in pupil size. However, to our knowledge no previous study has tried to employ pupillometry to study lexical access during reading.

**When decisions**

In this section, we begin by summarizing data on durations of individual fixations. The central issue here is to examine the validity of our three-stage model outlined above, which assumes that the frequency of first and second constituents as well as the whole-word frequency all affect the identification process, but at different points in time.

In Experiment 1, the first fixation duration on the compound words was significantly longer when the initial constituent was low-frequency than when it was high-frequency. This early effect of initial constituent frequency is consistent with the model’s prediction and indicates that morphemic components of words are playing an active role early in processing. The effect of first constituent frequency also spilled over to the second and third fixations, the second fixations being significantly longer and the probability of making a third fixation significantly greater for low-frequency initial constituents. But what is crucial for the model, the effect emerged early (i.e., during the first fixation). Also consistent with the model, the frequency of the second constituent influenced processing in Experiment 2, but later in the time-course than the effects of the first constituent. Although the effect showed up in the duration of the second fixation, the duration of the initial fixation remained unaffected. The strongest effect of second constituent frequency, however, was on the probability of making a third fixation on the word, which was clearly greater for compounds with a low-frequency second constituent.

On the other hand, the pattern of results of Experiment 3, where the whole-word frequency was manipulated, was not completely in line with the simple three-stage model. The model predicts that whole word frequency effects should appear
later than those of the second constituent frequency. However, in the data, word frequency effects emerged no later, and possibly earlier, than the second constituent frequency effect. The whole-word frequency affected the second fixation duration and the probability of making a third fixation on the word at least as strongly as the second constituent frequency. There was even an indication that it affected the first fixation duration: there was a 6-ms difference that proved significant in the subject analysis (but it was non-significant in the item analysis). These data are thus problematic for the three-stage model, which predicts a later effect of whole-word frequency than for second constituent frequency (see below for further discussion).

The results from Experiments 1 and 2 suggest that the frequencies of the constituents of compound words affect the identification process pretty much the same way they would if they were separate words. The first fixation (which typically landed on the initial constituent) was influenced by the frequency of the initial constituent, while the second fixation (which typically landed on the second constituent) was influenced by the frequency of the second constituent (see Hyönnä and Pollatsek, 1998; Pollatsek et al., 2000). To examine further whether this is the case, we conducted a follow-up analysis of the data of Experiments 1 and 2. In these analyses, constituents were treated as if they were separate words, and fixation durations were analyzed accordingly. We employed four eye fixation measures: (a) the duration of the first fixation on a constituent; (b) the gaze duration on a constituent (i.e., the sum of fixations before exiting a constituent); (c) the probability of fixating a constituent during the 'first-pass reading' (if the initial fixation on the word landed on the second constituent, the first constituent was considered to be skipped or if the second constituent was not fixated, it was considered to be skipped); and (d) the probability of fixating a constituent during the second-pass reading. (To initiate a second pass, the reader needed to regress back from the second to the first constituent; all fixations that followed such a regression were considered second-pass fixations).

As indicated above, in Experiment 1, the frequency of the initial constituent was manipulated. Perhaps not surprisingly, the when fixation measures on the first constituent were similar to what they would have been if it was a separate word (see Table 1). That is, when the initial constituent was low-frequency, the duration of the first fixation on it was 14 ms longer \((F_1(1, 23) = 14.83, p = 0.001, F_2(1, 27) = 29.68, p < 0.001)\), the ‘gaze duration’ on it was 86 ms longer \((F_1(1, 23) = 37.36, p < 0.001, F_2(1, 27) = 88.65, p < 0.001)\), and it was skipped during first-pass reading 19% less often \((F_1(1, 23) = 105.29, p < 0.001)\).

---

5 The average length of the initial constituent was 6.0 characters for the frequent constituent and 7.6 characters for the infrequent constituent. In Experiment 2, the average length of the frequent second constituent was 7.4 and that of the infrequent second constituent 6.6. It is highly unlikely that these minor differences in constituent lengths could account for the observed effects.
Table 1

First fixation duration (in ms), gaze duration (in ms), and probability of fixation (first-pass and second-pass reading) for compound words with a frequent and an infrequent first morpheme, separately for the first and second constituent (Experiment 1)

<table>
<thead>
<tr>
<th></th>
<th>Fixation measures on 1st constituent</th>
<th>Fixation measures on 2nd constituent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequent 1st constituent</td>
<td>Infrequent 1st constituent</td>
</tr>
<tr>
<td>First fixation duration</td>
<td>183</td>
<td>197</td>
</tr>
<tr>
<td>Gaze duration</td>
<td>229</td>
<td>315</td>
</tr>
<tr>
<td>Probability of fixation</td>
<td>First pass</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>Second pass</td>
<td>0.06</td>
</tr>
</tbody>
</table>

$F_2(1, 27) = 15.80, p < 0.001$. The probability of coming back to it after having fixated the second constituent (i.e., what we are calling ‘second-pass reading’) was also greater when the initial constituent was low-frequency ($F_1(1, 23) = 51.52, p < 0.001, F_2(1, 27) = 10.23, p < 0.01$).

There were also some significant effects of the first constituent frequency on fixation times on the second constituent; however, some were quite different from ‘spillover’ effects commonly seen when reading separate words. Most notably, when the initial constituent was low-frequency, the gaze duration on the second constituent was actually 66 ms shorter ($F_1(1, 23) = 30.49, p < 0.001, F_2(1, 27) = 22.54, p < 0.001$). (The initial fixation on the second constituent was also 5 ms shorter, but $p$-values were >0.20.) In addition, the probability of skipping the second constituent during the first-pass reading was actually greater when the initial constituent was low-frequency ($F_1(1, 23) = 46.83, p < 0.001, F_2(1, 27) = 22.89, p < 0.001$). However, the probability of fixating the second constituent during the second-pass reading was more ‘normal’: it was greater when the initial constituent was low-frequency ($F_1(1, 23) = 15.83, p = 0.001, F_2(1, 27) = 4.02, p = 0.055$).

The latter result, together with an increased number of second-pass fixations on the first constituent, indicate that compounds with a low-frequency first constituent were generally refixed more often. However, the other fixation time results on the second constituent are quite counterintuitive if one views fixation times on a constituent as a simple reflection of processing that constituent or simple spillover effects from processing the prior constituent. One possible explanation for those apparently counterintuitive results is that the mean frequencies of the second constituents were not perfectly equated between the two conditions: the high-frequency first constituents had on average somewhat lower-frequency second
Reading of Finnish compound words

Table 2

First fixation duration (in ms), gaze duration (in ms), and probability of fixation (first-pass and second-pass reading) for compound words with a frequent and an infrequent second morpheme, separately for the first and second constituent (Experiment 2)

<table>
<thead>
<tr>
<th></th>
<th>Fixation measures on 1st constituent</th>
<th>Fixation measures on 2nd constituent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequent 2nd constituent</td>
<td>Infrequent 2nd constituent</td>
</tr>
<tr>
<td>First fixation duration</td>
<td>214</td>
<td>214</td>
</tr>
<tr>
<td>Gaze duration</td>
<td>354</td>
<td>311</td>
</tr>
<tr>
<td>Probability of fixation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First pass</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>Second pass</td>
<td>0.14</td>
<td>0.26</td>
</tr>
</tbody>
</table>

constituents (188 vs. 288 per million). However, this relatively small frequency difference is not particularly likely to produce the fairly large differences in gaze duration or skipping. Instead, we suspect that processing of the two constituents did not neatly relate to where people were fixating. First, note that the probability of initially skipping the first constituent was quite different between the low- and high-frequency first constituents. Thus, one possible reason for the increased gaze duration on the second constituent for high-frequency first constituents is that the constituent, though skipped, was not fully processed, necessitating ‘catch up’ processing when the second constituent was processed. In addition, some of the increased gaze duration on the first constituent in the infrequent first constituent condition could have reflected increased processing on the second constituent.

In sum, the data of Table 1 suggest that though the constituents may be acting somewhat like separate words, there are ways in which the processing of them is more interrelated than for separate words. This could be true either for linguistic reasons (e.g., that even transparent compounds are not linguistically like separate words) or because the lack of a space between them alters where the eye lands. We will return to this issue in the next section.

In Experiment 2, the frequency of the initial constituent was matched, but the frequency of the second constituent differed considerably between the two compound word types. As expected, when the second constituent was low-frequency, the duration of first fixation on it was 20 ms longer \((F_1(1, 23) = 22.75, p < 0.001, F_2(1, 31) = 5.63, p = 0.02)\), and the gaze duration on it was 46 ms longer \((F_1(1, 23) = 37.19, p < 0.001, F_2(1, 31) = 25.20, p < 0.001)\). Moreover the second constituent was skipped 18% less often when it was low-frequency \((F_1(1, 23) = 69.43, p < 0.001, F_2(1, 31) = 32.77, p < 0.001)\); see Table 2). In
addition, there was one reliable ‘first-pass’ effect on the first constituent: the gaze duration on the initial constituent was 43 ms less when the second constituent was low-frequency ($F_1 (1, 23) = 27.02, p < 0.001, F_2 (1, 31) = 7.51, p = 0.01$). Note, however, that there was no effect of second constituent frequency on the first fixation duration of the first constituent, nor much of an effect on the probability of skipping the first constituent ($F_1 (1, 23) = 4.00, p = 0.06, F_2 (1, 31) = 2.00, p > 0.1$). Hence, the evidence for the effects of processing the second constituent during the first pass on the first constituent was fairly modest.

Interestingly, the probability of making a second-pass fixation on both the first and second constituents was significantly greater for low-frequency second constituents ($F_1 (1, 23) = 36.21, p < 0.001, F_2 (1, 31) = 15.80, p < 0.001$, and $F_1 (1, 23) = 24.74, p < 0.001, F_2 (1, 31) = 16.32, p < 0.001$, respectively). This suggests that difficulty in processing the second constituent led to reprocessing of the entire word.

To sum up the follow-up analyses of Experiment 2, most of the results are reasonably consistent with the two constituents being processed sequentially. Almost all the measures indicated that the second constituent frequency affected only first-pass measures on the second constituent. However, the finding that gaze duration on the first constituent was longer with a high-frequency second constituent calls for an explanation. As the probability of fixation data indicate, high-frequency second constituents were skipped more frequently than low-frequency constituents. What this probably indicates is that there are occasions when the last fixation on the initial constituent (either the first or second fixation) is near enough to the boundary between the two constituents to allow processing of the second constituent, and that, on some occasions, this fixation is lengthened because the second constituent is processed and then skipped. This would be analogous to the finding that fixation duration on word $n$ is lengthened when word $n + 1$ is skipped (Hogaboam, 1983; Pollatsek, Rayner and Balota, 1986).

More generally, however, the patterns of data from Experiments 1 and 2 suggest that even transparent compound words are not processed the way they would be if they were two separate words. Although we indicated above that some part of the difference may be linguistic, there is recent evidence from German (Inhoff, Radach and Heller, in press) that at least part of the difference is due to the lack of spaces between the constituents. Most notably, German readers were actually faster to process German compounds when spaces were inserted between the constituents even though the resulting separated compound words were both unfamiliar and grammatically illegal.

Above we have presented evidence showing that the frequency of compound word constituents influences fixation durations pretty much the same way that word frequency does. This evidence is consistent with our working hypothesis, which assumes that compound words are identified via their constituents. However,
the finding from Experiment 3 that the effect of whole-word frequency shows up relatively early during the encoding process suggests that a completely serial model is not viable. Thus, the complete pattern of results suggests that there may be two identification mechanisms operating in parallel: (a) a mechanism that decomposes the compound word into its constituents and tries to identify the word via its constituents (the decomposition route), and (b) another mechanism that tries to access the mental lexicon by using the word’s full form as an entry (the full-form route). If one assumes that the two mechanisms operate in parallel, one can explain both the observation that whole-word frequency exerts a relatively early effect and the findings that constituent frequencies influence compound word processing when word frequency is controlled for. (Search models like the one by Taft and Forster, 1976, cannot account for our results; see Hyönä and Pollatsek, 1998; Pollatsek et al., 2000.)

Before closing this section, we would like to discuss the model of O’Regan (1992), which posits that fixation location is the primary determinant of fixation durations in words and that fixation durations are influenced by cognitive factors only under special circumstances (i.e., when a fixation is optimally positioned around the word’s center). We conducted analyses of the frequency effects observed in Experiments 1–3 as a function of initial landing position (we compared fixations landing on character positions 0–4 to those landing on character positions 5–9). In Experiment 1, the frequency of initial constituent affected the duration of the first fixation when it was positioned around the word’s center, but did not do so for fixations landing in the word’s beginning (i.e., positions 0–4). A similar trend was observable in Experiment 2 (in which the frequency of the second constituent was varied), but the effects did not prove statistically significant. In Experiment 3 (in which whole-word frequency was varied), there was not even a hint that the frequency effect was modulated by the initial fixation location. In all, there is some support (but not consistent) for O’Regan’s hypothesis that the location of the initial fixation crucially determines whether frequency affects the duration of individual fixations (see also Rayner, Sereno and Raney, 1996; for effects of fixation position on fixation duration, see also Chapter 7). In contrast, there is ample evidence for the view that cognitive factors related to word identification determine fixation durations in words. Moreover, the follow-up analyses of Experiments 1 and 2 that we presented above indicate that fixation durations are modulated by the frequency of the constituent of a compound word that is being fixated.

**Where decisions**

In this section, we summarize data on individual fixation locations and saccade lengths. The issue here is to what extent constituent properties of compound words
have the potential to influence where fixations in a word land. In Experiment 4, the
length of the compound word constituents was manipulated, and in Experiments
1–3, the frequency of the constituents or the whole word was varied.

The effect of constituent length

With respect to the effect of constituent length, we reasoned that if constituents are
treated similarly to separate words (note that the constituents were not separated
by a space or hyphen), the incoming saccade into the compound word and the
first forward saccade in the word should go farther into the word when the initial
constituent is longer. The prediction concerning the incoming saccade is based
on an assumption that the compound word is decomposed parafoveally into its
constituents while still fixating on the previous word, and this information is then
utilized in saccadic programming. (Most probably, the target would be the middle
of the first constituent.) The prediction about the location of the second fixation
merely assumes that the decomposition into constituents is accomplished some time
during the first fixation. If it is, then fixations targeted either to the center of the
first constituent or to the center of the second constituent should be further to the
right when the first constituent is long. (Because there is no previous evidence for
morphological decomposition of words before they are fixated — see Inhoff, 1989,
Lima and Inhoff, 1985 — the former prediction seemed less likely to be true.)

As it turned out, the location of initial fixation did not differ between the long
and short initial constituents. However, reliable effects of constituent length were
observed for the first within-word saccade. First, when the first constituent was
long, the first forward within-word saccade was launched further toward the end
of the word (i.e., closer to the center of the second constituent) than when the
first constituent was short. This is analogous to the finding that an exit saccade
from a long word is longer than from a short word. Second, the probability of
making a regression back to the beginning of the word was significantly greater for
compounds with a short initial constituent. This makes sense as the mean location
of initial fixation on the word was on the 5th letter, which is a less optimal landing
site for identifying a short (3–5 characters) initial constituent. In other words, when
the initial constituent was skipped over (the saccade presumably being guided by
the length of the entire word), a regressive saccade was often launched toward the
beginning of the word.

The finding that the length of the initial constituent affected the saccadic
trajectory of the forward refixations suggests that saccadic programming is affected
by compound word constituents in a manner similar to that when two words
are separated by a space. However, there was evidence indicating that saccadic
programming in the absence or presence of spaces between morphemes is not
identical. First, as indicated above, the initial fixation was on the 5th letter of
the compound word, which means that the first constituent, when short, was typically skipped over resulting in an increased tendency to program a regressive saccade toward the word beginning. Second, the first forward refixation did not land optimally (i.e., near the center of the second constituent), but was further into the word than would be expected for the short initial constituent and not as far into the word as would be expected for the words with long initial constituents. Thus, fixation locations in long compound words appear to be a compromise between a pure morphological guidance mechanism and a pure visual (i.e., space-guided) mechanism. In morphemic guidance, the reader needs to recognize the boundary between the constituents and use that information in targeting the subsequent refixation. However, Hyönä and Pollatsek (1998) showed that the frequency of the morpheme-spanning bigram was not responsible for the morphemic effects. Thus it appears that morphological structure of compound words did affect fixation locations, and that the effect was not due to a low bigram frequency at the constituent boundary popping out as a visually salient feature. (Other potential low-level artifacts were also ruled out by Hyönä and Pollatsek, 1998.)

The effect of constituent frequency

The predictions concerning fixation locations in Experiments 1–3, where constituent frequency or whole-word frequency was manipulated, were derived from the processing difficulty hypothesis of saccadic computation proposed by Hyönä (1995; see also Henderson and Ferreira, 1990). According to the hypothesis, the perceptual span around the fixation from which useful information is picked up is narrowed down with increasing difficulty in parafoveal and foveal processing. Thus, when a word in foveal or parafoveal vision is low-frequency, less parafoveal processing will be done, which should then lead to a shorter forward saccade.

There is some previous evidence in support of the processing difficulty hypothesis. Beauvillain et al. (1996) and Hyönä (1995) observed that the initial fixation tends to land closer to the beginning of a word having an irregular letter cluster in the beginning part of the word. By assuming that an irregular letter cluster in the word’s beginning is more difficult to process parafoveally, these findings can be interpreted as being consistent with the processing difficulty hypothesis. Second, Kennison and Clifton (1995) manipulated foveal and parafoveal load by varying the frequency of two adjacent words. When both words \( N - 1 \) and \( N \) were of low frequency (i.e., when both foveal and parafoveal load was high when fixating on \( N - 1 \)), the probability of skipping over the word \( N \) was decreased for good readers (i.e. readers with a high working memory span). For readers with a low working memory span, on the other hand, the skipping rate was highest when the foveal and parafoveal load was low (i.e., when both words were high-frequency). Thus, the Kennison and Clifton study suggests that foveal and parafoveal processing difficulty
can affect saccadic trajectory during reading, at least in the sense of programming saccades to skip over entire words.

In Experiments 1–3, we tested the validity of the processing difficulty hypothesis by examining the effects of constituent frequency on saccadic programming. In Experiment 1, where the frequency of the initial constituent was varied, we observed that the incoming saccade into the compound word landed somewhat closer to the word beginning and the second fixation (preceded by a forward saccade) clearly landed closer to the word beginning when the initial constituent was of low frequency. These findings are consistent with the processing difficulty hypothesis, as they suggest that both the difficulty in parafoveal preprocessing (as indexed by the effect in the incoming saccade) and in foveal processing (as indexed by the effect in the within-word saccade) are capable of influencing the saccade trajectory during reading.

The processing difficulty hypothesis was further tested against the data of Experiments 2 and 3. The hypothesis predicts that in Experiment 2, where the frequency of the second constituent was varied, the location of the second fixation in the word will be closer to the word beginning for compounds with a low-frequency than a high-frequency second constituent. This is because during the initial fixation on the compound word parafoveal processing of the second constituent is more difficult when it is of lower frequency. On the other hand, the landing site of the incoming saccade into the word is not predicted to differ between the two conditions, because the frequency of the initial constituent was equated and it would be highly unlikely that properties of the second constituent would influence saccadic programming from a long distance.

The data were consistent with the above predictions. The landing site of the initial fixation was very similar for the two word types, and the location of the second fixation was closer to the word beginning with a low-frequency second constituent (although the effect was reliable only in the subject analysis). The third prediction concerning the data of Experiment 2 is that the saccade leaving the word would be shorter when the second constituent is low in frequency, because increased difficulty in processing the second constituent would hamper with the parafoveal processing of the subsequent word thus leading to a shorter exit saccade.

6 Subsequent regression analyses indicated that the frequency effect on the initial landing position could not be accounted for by differences in the initial trigram frequency.

7 The effect cannot be explained by the predictability of the second constituent based on the first constituent (i.e., a longer saccade associated with an easily predictable second constituent). The predictability of the second constituent was actually much higher for low-frequency first constituents, as there were only one or two compound words in the dictionary starting with a particular low-frequency constituent (for high-frequency first constituents, there was a minimum of 45 compounds starting with a particular first constituent).
We conducted an analysis of the length of the exit saccade separately for the trials when there were exactly one or two fixations on the compound word. For the one-fixation trials, the length of the exit saccade was in the predicted direction (the means in character spaces were 11.51 and 11.87, for the low- and high-frequency second constituents, respectively), but the effect was far from significant ($F_1 < 1$; it should be noted that there were a lot of missing data for the one-fixation trials). There was also no effect on saccade length in the analysis of two-fixations trials ($F_1(1, 23) = 2.85, p > 0.1, F_2 < 1$). If anything, the trend is in the opposite direction to what was predicted (the means were 12.47 and 12.23, for the low- and high-frequency second constituents, respectively).

With regard to the low- and high-frequency compound words employed in Experiment 3, the processing difficulty hypothesis does not predict any difference for the length of the incoming saccade, because the frequency of the initial constituent was carefully matched between low- and high-frequency compounds. On the other hand, the predictions concerning the first within-word saccade and the exit saccade (i.e., the saccade leaving the word) depend on when the effect of compound word frequency is assumed to ‘kick in’. If the frequency effect shows up relatively early in processing, the first within-word saccade should be affected, but if it appears only very late, as predicted by our three-stage model, only the saccade leaving the word would be influenced by the compound word frequency.

The data did not reveal any effects of word frequency on landing positions; the locations of the initial and second fixations on the compound words were not affected by word frequency. Moreover, the length of the exit saccade remained unaffected by word frequency. The length of the exit saccade was analyzed separately for trials with a single fixation (the data for 6 subjects had to be discarded due to missing data points) or with two fixations on the target word. For compound words, word frequency did not exert any significant effects on exit saccade either for single fixation trials ($t_1(17) = 1.85, p = 0.08, t_2 < 1$), or for two fixation trials ($t_1(23) < 1, t_2 < 1$). If anything, the trend was in the opposite direction to that predicted. That is, there was a longer exit saccade for low-frequency compounds: for the single fixation trials, the means were 9.74 vs. 10.85 character spaces, and for the two fixation trials, 9.62 vs. 9.25 character spaces, for low- and high-frequency compounds, respectively. For the monomorphemic words, the effect of word frequency was also non-significant, both for the single fixation trials ($t_1(21) < 1, t_2 < 1$) and for the two fixation trials ($t_1(23) < 1, t_2 < 1$).

In sum, the following effects were predicted by the processing difficulty hypothesis and observed in the data: (a) the frequency of the initial constituent influenced the location of both first and second fixations on the word; (b) the frequency of the second constituent influenced the location of the second fixation on the word. On the other hand, the processing difficulty hypothesis predicted two effects that did not occur: (1) an effect of second constituent frequency on the length of the
exit saccade; and (2) an effect of whole-word frequency on the location of second fixation and on the length of the exit saccade.

One way of reconciling the pattern of results is to suggest that intra-word saccades are affected by processing difficulty, while inter-word saccades are not. However, the finding that the incoming saccade was affected by the frequency of the initial constituent, poses a problem for this interpretation (as the effect is small in size, it should probably be replicated before basing any really strong conclusions on it, however). We offer another view which is based on the notion that frequency effects associated with the fixated word would not spill over to the parafoveal processing of the next word, at least in so far as saccadic programming is concerned. In other words, a compound word is identified to a sufficient degree during its fixation before attention is shifted and a saccade is programmed to a succeeding word. This notion would explain why the frequency effect does not carry over to the exit saccade. On the other hand, intra-word saccades are affected as they are intimately linked with the ongoing identification of the currently fixated word. Finally, entry saccades are influenced, because attention shifts to the to-be-fixated word prior to executing a saccade to it, and consequently the frequency of initial constituent is capable of affecting the saccadic trajectory.

**Pupil size**

As noted above, pupil size has been shown to reflect differences in the attentional effort needed to carry out various information processing tasks (see Beatty, 1982; Kahneman, 1973). However, we do not know of any study where cognitive effort associated with lexical access during reading has been examined using the pupillometric method. The EYELINK tracker used in collecting the data also measured the size of the pupil, so we were in a position to employ this measure to determine whether pupil diameter would reflect differences in the relative ease in identifying long compound words (i.e., larger pupil size reflecting more effort). In the following, we report a detailed analysis of the pupil size for the data of Experiment 2 accompanied with an additional analysis of the data from Experiment 3. We employed pupil diameter as the measure to comply with the tradition in the field (the other available option would have been pupil area). 8

Because of the relatively small number of observations on the third fixation, we concentrated our analysis of the mean pupil size on the first two fixations

---

8 The pupil diameter was recorded in pixels, but in the absence of any baseline measure we were unable to convert the pixels into any absolute measure, such as millimeters. Thus, we measured only relative changes in pupil diameter.
Table 3

Pupil diameter (in pixels) as a function of second constituent frequency and whether the initial fixation landed on letters 0-4 or letters 5-8 of the target word in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Region 1: Fix1 on 0-4</th>
<th>Region 2: Fix1 on 5-8</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequent 2nd constituent</td>
<td>Infrequent 2nd constituent</td>
<td></td>
</tr>
<tr>
<td>First fixation</td>
<td>53.9</td>
<td>53.0</td>
<td>52.1</td>
</tr>
<tr>
<td>Second fixation</td>
<td>52.3</td>
<td>51.5</td>
<td>51.1</td>
</tr>
</tbody>
</table>

made on the target words in Experiment 2 (the one in which the frequency of the second constituent was varied). As seen in Table 3, there was a tendency for the pupil size to be larger in the frequent constituent condition (apparently contrary to the effort hypothesis), but the frequency effect was not significant, even averaged over the two fixations: $F_1(1, 23) = 2.22, p > 0.1, F_2(1, 31) = 2.31, p > 0.1$. However, the difference in pupil size between the two fixations was highly reliable: $F_1(1, 23) = 61.0, p < 0.001, F_2(1, 31) = 21.9, p < 0.001$. This effect is potentially interesting as it suggests that encoding a word takes more effort in its initial stages and less effort as the candidates are narrowed down. Obviously, there are many potential stimulus artifacts that might also explain the effect. One is that the fixation locations are different on the two fixations and thus there could be differential luminance levels. However, because the first fixation is nearer the space between the words (which is a source of increased luminance when the background is light), one might expect the opposite effect to the one we observed (i.e., smaller pupil size near the beginning of the word).

One way to assess potential artifacts is to analyze pupil size conditional on the initial landing position. There does appear to be a location effect of initial fixation location, as pupil size on the first fixation was larger when the initial fixation location was closer to the beginning of the word ($F_1(1, 23) = 7.96, p = 0.01, F_2(1, 31) = 5.51, p = 0.025$; see Table 3). The decrease in pupil size from first to second fixation, however, can not be completely due to the fixation location. In the above analysis, the mean pupil size was still significantly smaller on the second fixation than on the first, even when the initial fixation location was around the word center (i.e., on characters 5-9; $F_1(1, 21) = 40.7, p < 0.001, F_2(1, 30) = 60.8, p < 0.001$). Unfortunately, this comparison is not perfect as the mean first fixation location in these cases was 6.07, as contrasted with 8.05 for the mean second fixation location. As a result, we did a second analysis in which we tried to equate fixation locations even more closely. In this analysis, we only examined trials on which the first fixation was to the left of the 6th letter of the word. This restriction led to the
Table 4

Pupil diameter (in pixels) as a function of word frequency and word type (compound vs. monomorphemic words) in Experiment 3

<table>
<thead>
<tr>
<th></th>
<th>Compound words</th>
<th></th>
<th>Monomorphemic words</th>
<th></th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low frequency</td>
<td>High frequency</td>
<td>Low frequency</td>
<td>High frequency</td>
<td></td>
</tr>
<tr>
<td>First fixation</td>
<td>57.0</td>
<td>57.0</td>
<td>56.5</td>
<td>56.3</td>
<td>56.7</td>
</tr>
<tr>
<td>Second fixation</td>
<td>55.0</td>
<td>55.5</td>
<td>54.5</td>
<td>54.6</td>
<td>54.9</td>
</tr>
</tbody>
</table>

mean first fixation location actually being slightly further into the word than the mean second fixation location (7.9 vs. 7.5). Unfortunately, this restriction led to a smaller set of data, but there was still a pupil size difference between first and second fixation (53.5 vs. 53.0), which was significant ($F_1(1, 13) = 7.41, p < 0.017$). (Because pupil size differed widely between participants, item analyses based on widely differing numbers of data points from different participants seemed pointless.)

Another possible artifact that could have produced the pupil size difference between the first and second fixation is the differing size of the saccade preceding the fixation (although we have not found any reports of such an artifact). That is, saccades from word to word are longer than intraword saccades. To try to control for this, we conducted an analysis in which only trials were examined in which the fixation prior to landing on the target word was within four characters of the space preceding the target word and the first fixation on the target word was within the first four characters of the target word. In addition, all trials on which there were regressive saccades were eliminated. This did not exactly equate for mean saccade length, as the mean saccade length prior to the first fixation was 5.7 characters and the mean saccade length prior to the second fixation was 4.8 characters. Again, these restrictions led to elimination of a lot of data; however, the difference in pupil diameter between the first and second fixations (51.3 vs. 50.2 pixels) was still significant, $F_1(1, 13) = 77.1$ with $p < 0.001$, and because the size of the effect was just about the same size as in the main analysis (where there was a large difference in mean saccade size), we doubt whether the small remaining difference in saccade size could be the cause of the difference in pupil size.

A highly reliable fixation order effect was also found in the pupil size data of Experiment 3, where high- and low-frequency compounds and length-matched and frequency-matched monomorphemic words were employed as stimuli ($F_1(1, 18) = 203.6, p < 0.001$; see Table 4). Here we wanted to examine a possible artifact that was not considered in the previous analysis: that the duration of the first fixation tends to be longer than that of the second fixation. Thus, we did a regression analysis by items, in which the pupil diameter was predicted from both the fixation duration
Reading of Finnish compound words

and the fixation number (either first or second). If the fixation order effect in pupil size is merely an artifact of fixation duration, fixation duration should predict pupil size and fixation number should exert no effect. However, the regression analysis showed that fixation duration had virtually no predictive power and fixation order essentially predicted the entire effect.

In Experiment 3, word frequency did not exert an effect on pupil size, $F_1 < 1$, but there was a marginally significant effect of word type, $F_1(1, 18) = 3.79$, $p = 0.067$, suggesting that the average pupil size for compound words was slightly larger than for monomorphemic words. Although the finding is only suggestive, it is consistent with the view that compound words would be more effortful to identify than monomorphemic words. The gaze duration data are in line with this view by showing a 60-ms decrement for the monomorphemic words over the compound words.

Across the two experiments, the data on pupil size strongly suggest that encoding a long word takes more effort in its initial stages and less effort as the potential lexical candidates are narrowed down. Thus, it may be worthwhile in future eye movement research on word identification to include the pupil size as an additional measure to assess relative processing effort.

One may argue against our interpretation that the fixation order effect is due to mental effort by pointing out that the pupil is known to dilate fairly slowly in response to a triggering event. The peak amplitude is typically observed only approximately 1000 ms after the beginning of an experimental trial (Beatty, 1982), and the dilation is shown to begin with a delay of about 300 ms (Hoeks and Levelt, 1993). As a typical fixation during reading lasts about 200–250 ms, a skeptic would claim that the duration of initial fixation does not last long enough to reflect changes in the relative mental effort related to the identification of the fixated word. We do not think this is a valid argument, because it is now well established that the processing of a given word is already initiated while fixating on the previous word. Because of this parafoveal preprocessing, the triggering event for the pupil dilation may appear well before making the first fixation on the target word, and thus it is not at all implausible that an effect can be detected during the initial fixation. Perhaps a part of the effect could be explained as a reaction to the length of the word (long words being more effortful to identify). However, assuming that the difference between the monomorphemic and compound words is real, word length cannot be the only explanation for the pupil dilation effect.

Summary

The above data indicate that Finnish compound words are decomposed into their primary constituents on-line (at least some of the time), and that this decomposition influences both when and where decisions. Most importantly, the frequencies of
the first and second constituents each affect both how long the constituent is
fixed and how long the whole compound is fixed. In addition, the time course
of processing is consistent with a serial processing model, in which the first
constituent is processed prior to the second. However, a pure decomposition model
is not tenable, as the frequency of the whole word has an effect which surfaces in
the eye movement record at least as early as the effects of the frequency of the
second constituent.

The pattern of fixation duration data are compatible with a parallel dual-route
model (see also Pollatsek et al., 2000), which assumes that two processes go on
in parallel when accessing morphologically complex words — one that tries the
access by looking up an entry in the mental lexicon that directly corresponds
to the processed word as a whole, and another that decomposes the word into its
morphological constituents and accesses the word meaning via the constituents. The
finding that constituent frequencies reliably influence individual fixation durations
is evidence supporting the existence of a decomposition route, and the finding
that the whole-word frequency exerts an early effect in processing is evidence
for a direct look-up route. Moreover, by assuming that these two processes occur
in parallel, we are in the position to explain both that the 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.

Our analyses of the constituents as if they were separate words also indicated that
significant processing of the second constituent occurred when the first constituent
was fixated, resulting in the frequency of the first constituent having significant
effects on the probability that the second constituent was fixated, and the gaze
duration on the second constituent. In contrast, there was little effect of the
frequency of the second constituent on initial processing of the first constituent.
The only effect was on the gaze duration on the first constituent, and this effect could
have been due to refixations on the first constituent that were intended as initial
fixations on the second constituent.

These data suggest that although compound word constituents to some extent
may behave like separate words, there is considerably more overlap in their
processing than is the case with separate words. First, the effect of word frequency
does not typically spill over to the processing of next word, a finding that is at
odds with our observation that the processing of first constituent continues when
fixating the second constituent. Second, with words it is typically not the case that
the frequency of the parafoveal word to the right of the fixated word exerts an effect
prior to its fixation (but see Kennedy, 1998; Chapter 8), again a finding we observed
for compound word constituents. Thus, compound word identification does not
reduce to the identification of the constituents as if they were separate words.
Properties of the constituents also affected where on the word a saccade landed. Most notably, the length of the initial morpheme affected where the refixations on the compound word were directed, even when word length and various frequency measures were controlled. This indicated that the size of constituents can guide saccadic targeting decisions even in the absence of spaces between them (and in the absence of other lower-level cues such as low bigram frequencies at constituent boundaries). The frequency of the constituents also influenced where a word is fixated (both the initial fixation and refixations). The latter effects are most parsimoniously explained by a processing difficulty hypothesis, which assumes that saccadic amplitude can be affected by foveal and parafoveal processing difficulty. Note that the constituent length effect can not be plausibly explained in this way, as fixations go further into a word when the first constituent is long (and hence, all other things being equal, would be harder to process).

The processing difficulty hypothesis can take two forms. According to one version, processing difficulty influences the probability of fixating or skipping a word (or a compound word constituent) and the probability of making a refixation on a word (or constituent). According to a stronger version, saccade length can also be affected by processing difficulty in a graded fashion (see Radach and McConkie, 1998, for a similar distinction). The former version translates to a when decision in the sense that the decision of where to saccade next is governed by how hard a foveal or parafoveal word or constituent is to process. In other words, it assumes that processing difficulty only influences which word or constituent is selected as the target for a saccade. The latter version of the hypothesis adds a mechanism which is capable of fine-tuning the exact saccadic amplitude as a function of foveal or parafoveal processing difficulty. In other words, a mechanism is postulated that directly modulates the saccade amplitude computation. It is this latter version that has created some controversy (Radach and McConkie, 1998; see also Chapter 7).

The weaker version of the processing difficulty hypothesis that is generally accepted by the eye movement community was supported by the finding that high-frequency constituents were selected as the saccadic targets less often than low-frequency constituents. On the other hand, we also found evidence for the more controversial version of the hypothesis: fine-tuning of the saccadic amplitude was carried out by the readers on the basis of the frequency of compound word constituents. This leads us to conclude that processing difficulty influences both the target selection as well as the amplitude of saccades in reading.

Finally, we observed some evidence suggesting that the size of the pupil may reflect the difficulty of accessing long words in reading with the pupil being larger during the initial processing stages and relatively smaller during the later stages. As this is apparently the first time a relationship is observed between pupil size and lexical access during reading, it awaits confirmation by corroborative evidence from future studies.
References


