Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 51 (2004) 325–345
Journal of Memory and Language
www.elsevier.com/locate/jml

Morphological parsing and the use of segmentation cues in reading Finnish compounds☆

Raymond Bertram,[a,*] Alexander Pollatsek,[b] and Jukka Hyönä[c]

[a] Department of Psychology, University of Turku, Turku, Finland
[b] University of Massachusetts at Amherst, USA
[c] Department of Psychology, University of Turku, Turku, Finland

Received 9 April 2004; revision received 9 June 2004
Available online 7 July 2004

Abstract

This eye movement study investigated the use of two types of segmentation cues in processing long Finnish compounds. The cues were related to the vowel quality properties of the constituents and properties of the consonant starting the second constituent. In Finnish, front vowels never appear with back vowels in a lexeme, but different quality vowels can appear in different constituents in compounds. Experiments 1 and 2 showed that compounds with different vowel quality constituents are processed faster than those with same vowel quality constituents, but only if the first constituent is long. This indicates that the use of segmentation cues in processing long compounds depends on the ease of encoding the first constituent. Experiment 3 established that (a) the effect does not depend on the crucial vowels being adjacent and (b) processing is affected by the type of consonant beginning the second constituent (i.e., whether or not it could end a first constituent).

Keywords: Vowel quality; Decomposition; Segmentation; Eye movements; Finnish; Compounds; Ambiguous consonants

For many languages, the lexicon includes many words that are compound words: words constructed out of two or more independent words such as day/light. (Here, and in the remainder of the article, forward slashes indicate constituent boundaries.) In theory, compound words could either be encoded by directly matching the entire letter string to a full form representation in a mental lexicon, or by a compositional process in which the letters in the constituents are matched to entries for those constituents in the mental lexicon followed by some sort of constructive process that combines the constituents. Of course, some combination of the two processes may also occur. If one believes that the actual process of compound word recognition involves, to some extent, recognition of the components, a major question that needs to be solved is where the components are and where the constituent boundary or boundaries are. (We will focus on two-constituent compounds.)

One possible solution to the problem of locating the constituent boundary is that the reader processes the beginning of the compound word (somehow) until a constituent is found, and then the rest of the letters are treated as the second constituent. We will return later to the issue of whether this kind of processing strategy would plausibly work in "real time," but for now, we

☆ This study was financially supported by the Academy of Finland (grant to the third author). We thank Annu Haapakangas, Tuomo Häikiö, Anna-Mari Nurminen, Anna Piponius, and Juha Toivonen for their help in collecting and analyzing the data. We are also grateful to Harald Baayen, Lee Wurm, and two anonymous reviewers for their comments on previous versions of the article.

* Corresponding author. Fax: +358-2-3335060.
E-mail address: rayber@utu.fi (R. Bertram).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2004.06.005

want to consider whether it is logically sufficient or whether something more is needed. It should be clear that this strategy may succeed in "parsing" the compound word in many cases (e.g., *straw/berry*), but it may often run into problems. For example, in English, in a word like *hailstorm*, it is not clear whether the initial constituent is *hail* or *hails*. Thus, if the reader initially took *hails* to be the initial constituent, there would be processing difficulty; however, the reader could presumably eventually parse the word by realizing that *torm* is not a constituent and try a second parse. Another problem is that there are truly ambiguous compounds such as *clamprod* (= >*clamp/rod* or *clam/prod*) for which there are multiple parses and readers do consider these alternatives (Libben, 1994).

The above discussion suggests that parsing compound words may be far from trivial and that, although the reader may parse a compound word by going through the "dirty work" of recognizing the constituents, it would help such a parsing process if it could use "short-hand" cues to determine where the constituent boundary is. For example, parsing *strawberry* may be easier than parsing *hailstorm* not only because *strawb* or *stra* are not words, but because *wb* is a bigram that indicates a syllable boundary. In Finnish, there is a potential parsing cue that is similar to the bigram *wb* in English because it is not only a reliable indicator of a syllable boundary, but also a reliable indicator of a morpheme boundary. Our goal in this paper is to see whether this morphemic cue, and another morphemic cue that we will discuss later, are actually used in parsing Finnish compound words, and if so, how they enter into the process of word recognition when these words are read in sentences. However, we first have to give some facts about Finnish to explain this first cue.

This cue is related to a principle in Finnish called *vowel harmony*, which applies to monomorphemic, inflected, and derived words. In these types of words, the front vowels /æ/, /y/, and /ϕ/ can never appear together with the back vowels /ɑ/, /u/, and /o/ in the same word. The other "neutral" vowels, /e/ and /i/, can appear in the same simplex word with either front or back vowels (e.g., *väline* 'instrument, tool' or *urheilu* 'sport'). However, in compound words, the vowel quality of the constituents is the same as when these constituents are in isolation, so that vowels of different quality can appear together in the same word, but they can never appear in the same constituent. Thus, two adjacent characters of different vowel quality as *äo* in *selkäongelma* ('back problem') would therefore logically imply that the constituent boundary is between the it ä and the *o*. Therefore, as indicated above, this cue unambiguously indicates both a syllable boundary and a morpheme boundary. As a result, we hypothesized that if any cue could work as a segmentation cue, this 'vowel disharmony' cue is a very likely candidate and a good place to start. In contrast, a word such as *opinto/uudistus* ('study reform'), in which the letters at the boundary (*ou*) do not necessarily define either a morpheme boundary or a syllable boundary, would be devoid of any such cue. This vowel harmony principle is not unique to Finnish; there are other vowel harmony languages, such as the related Finno-Ugric language Hungarian or Ural-Altaic languages like Turkish (see Suomi, McQueen, & Cutler, 1997; Vroomen, Tuomainen, & De Gelder, 1998, for vowel harmony effects in spoken word recognition). It also shares properties with potential parsing cues that are not unique to vowel harmony languages such as cues mentioned above that denote a syllable boundary.

A second property that vowel harmony has that is likely related to being a syllable boundary is that (in the example above) the bigram at the boundary (*äo*) is a low-frequency bigram. Of course, there is no requirement in Finnish that the letters in the bigram at the constituent boundary of a compound need to be vowels, so that one can have a word such as *nuoriso/säätiö* 'youth foundation' where the vowel disharmony information is not carried by a bigram. In this case, only the approximate location of the boundary is indicated by the closest disharmonious vowels. As a result, we felt that looking at vowel disharmony would be an interesting arena to get insights into the way the morphological parser functions. In particular, we thought it was a good way to understand whether this particular vowel disharmony cue that unambiguously signals a morphemic boundary serves as a cue above and beyond its low bigram frequency.

Earlier studies

Over the past 5 years, we have conducted a number of eye movement studies on two-noun Finnish compound words (Bertram & Hyönä, 2003; Hyönä & Bertram, 2004; Hyönä, Bertram, & Pollatsek, 2004a; Hyönä & Pollatsek, 1998; Pollatsek, Hyönä, & Bertram, 2000a). In all the experiments, the compound words were embedded in sentences and readers' eye fixations on them were recorded while they read the sentences for comprehension. The results of Hyönä and Pollatsek (1998) and Pollatsek et al. (2000a, 2000b) suggested that, for long compounds, both decompositional and direct look-up processes are involved in the identification of the word. Specifically, the gaze duration (the summed duration of all fixations on a word before exiting to another word) of long compounds was affected by both constituent frequency and whole-word frequency manipulations. First, the gaze duration on long compounds was much shorter when the whole word was high frequency than when it was low frequency. In addition, long compounds with either a high frequency first or second constituent elicited much shorter gaze durations than those with either a low frequency first or second constituent. However, when looking at an early measure of lexical processing, first fixation duration, the picture was different. For this measure, the first constituent frequency manipulation produced a significant effect, whereas the second constituent frequency manipulation had no effect and there was only a hint of an effect for the whole-word frequency. This suggests that readers access the constituents of a long compound serially, starting with the first constituent before they turn to the second. At the same time, there may be some early involvement of the whole word form.

Hyönä et al. (2004a) employed an eye-movement–contingent display change paradigm in which a preview of the second constituent of the compound was either present or partly replaced by orthographically similar letters before the second constituent was fixated. There was a large preview benefit when the second constituent was present, implying that even though the second constituent is not lexically accessed during an early stage of processing, some orthographic features of this constituent are picked up before readers leave the first constituent. However, it should be noted that this effect almost exclusively appeared after the boundary was crossed and the second constituent was fixated, whereas the frequency of the first constituent had most of its effect before the eyes moved on to the second constituent. Together these results imply that in processing long compounds lexical access of the first constituent is attempted before orthographic features of the second constituent are assessed.

Bertram and Hyönä (2003), however, showed that not all compounds are initially accessed via their first constituent. In their study, first-constituent frequency and whole-word frequency were manipulated for both long and short compounds. (Long compounds were, on average, 13 characters long and almost always elicited two or more fixations, whereas short compounds were, on average, about 7.5 characters and typically elicited only one fixation.) Bertram and Hyönä replicated the results of the earlier studies with long compounds; however, the processing of short compounds had a much more holistic flavor. The first fixation duration on short compounds was affected by the whole-word frequency manipulation, but not by the first-constituent frequency manipulation. In addition, gaze durations on short compounds were strongly affected by whole-word frequency, but were only mildly affected by first-constituent frequency. Bertram and Hyönä (2003) concluded that visual acuity constraints modulate morphological processing. That is, when compound words are long, initial fixations fall around the fourth character, which implies that the latter part of the word falls outside foveal vision. As a result, if a reader simply analyzes what is in or near the fovea, the second part of long compounds is not in the region of high visual acuity, and thus the first constituent is the first entity to be accessed. For short compounds, where all characters usually fall within foveal vision on the initial fixation, the whole-word form dominates the access procedure. These claims are supported by studies in English and French, in which the impact of morphological structure on lexical processing is greater the longer the complex word (Beauvillain, 1996; Colé, Beauvillain, & Segui, 1989; Niswander, Pollatsek, & Rayner, 2000).

In sum, the results suggest a model of compound processing in which: (a) compositional and holistic access procedures operate simultaneously; (b) the length of the compound determines which procedure dominates; (c) for long compounds, access to the constituents is in some sense serial, starting with the first constituent and followed by the second; and (d) the holistic access procedure is more than accessing the whole word form via the constituents (i.e., whole word access representations do exist). To elaborate on points (a), (b), and (c), it seems that processing long compounds starts with at least an attempt to access the first constituent, after which orthographic features and then lexical features of the latter part of the word are encoded, leading to access of both the whole word form and the second constituent. Our assertion in (d) is supported by the findings (a) that for long compounds a whole word frequency effect surfaces somewhat earlier than a second constituent frequency effect and (b) that for short compounds the whole word frequency effect is more pronounced and earlier than the first constituent frequency effect. That is, if whole word frequency effects are only due to the speed of combining the first and second constituent together—quickly for high frequency compounds and slowly for low frequency compounds—the effects would not appear as early as they do because the first and second constituent would have to be accessed before the gluing procedure can start. Thus, the eye movement results are at odds with models that claim that whole word representations are only accessed after morphological components are accessed (e.g., Taft, 2004).

It should be noted that the frequency of the whole compound word can modify the effect of word length on the relative roles of whole-word and componential processing. More specifically, for very low frequency or novel compounds it is unlikely that the whole word form plays an important role in processing (see e.g., Van Jaarsveld & Rattink, 1988). Thus even though all the letters of a novel short compound can be identified during one single fixation, the lack of a whole-word representation would mean that it has to be processed by means of the constituents anyway. In general one could assume that the higher the frequency of a compound, the more pronounced the role of holistic processing will be (see e.g., Alegre & Gordon, 1999, for empirical evidence on this position with inflected words).

The time course of the segmentation process

An important stage that is neglected in most models including the above-sketched model is the stage of segmentation (but see Schreuder & Baayen, 1995, for an exception)—the process of segmenting morphologically complex words into meaningful sublexical components such as morphemes. That is, in order to identify the separate constituents in compounds, one needs to know where the first constituent ends and where the second begins. Although it is possible that the segmentation stage precedes the identification of the first constituent (e.g., on the basis of salient bigrams), it seems unlikely that this occurs in a majority of cases. First, the first constituent frequency effect during the first fixation indicates that at least some identification of the first constituent is occurring quite early in processing (Bertram & Hyönä, 2003; Hyönä & Pollatsek, 1998). Second, low probability transitions between letters do not necessarily indicate morphemic boundaries, but could also indicate (non-morphemic) syllabic boundaries (see Seidenberg, 1987). It seems therefore more plausible that, in the early stages of compound processing, there is some kind of interaction between first constituent access and localization of the constituent boundary. If access to the first constituent is unfinished when access to the second constituent is initiated, the localization of the constituent boundary becomes particularly important because knowing where the second constituent begins seems like a crucial step in identifying it. That is, if one initially misidentifies the beginning letters of the second constituent, it would likely lead to costly reanalyses and a lot of intraword regressions. Thus, even though it has been assumed that segmentation mostly has an impact on the very early stages of processing (e.g., Schreuder & Baayen, 1995), it might well be that much of its impact is felt at a later stage. One goal of the current experiments is to monitor the time course of segmentation by using the eye movement record to see when the orthographic segmentation cues we manipulate exert their effects.

The segmentation process: Empirical evidence

As noted above, for long compounds, constituents have been shown to play an active role in the access process, not only in Finnish, but also in other languages, most notably in English (e.g., Andrews, Miller, & Rayner, 2004; Juhasz, Starr, Inhoff, & Placke, 2003; Lima & Pollatsek, 1983). This implies successful segmentation of the whole word into its constituents.

The question that then arises is whether the segmentation process is some kind of trial-and-error procedure, or whether the reader searches more actively for the constituent boundary to delimit the first constituent. Even though most psychomorphological models assign a role to decomposition in lexical processing, they do not specify the exact nature of the prelexical segmentation procedure. However, several linguistic computation models of parsing have been more specific about the segmentation procedure. Hankamer (1989) and Koskenniemi (1983, 1984) proposed models in which parsing is thought to take place by scanning an input string in a left-to-right fashion, while the system attempts to match the input with one or more lexical representations, taking into account any formally determined differences between the input and underlying representations. Even though these two-level morphology approaches have been successful in parsing huge language corpora (e.g., the FINTWOL parser is based on Koskenniemi's theory), they have been criticized for lacking psychological validity (see, e.g., Baayen & Schreuder, 1996; Forster, 1989; Sproat, 1992). That is, a strict left-to-right scanning procedure and the attempt to match each possibility with lexical representations may be very time-consuming and may lead to many misparses.

It seems that a psychologically more valid segmentation procedure would include a parser that makes use of cues that are derivable from the lexical statistics of a given language. As noted earlier, in Finnish the 'vowel disharmony cue' would be a good cue to determine the morpheme boundary. In English, there are several studies that indicate that bigram troughs facilitate the identification of morpheme boundaries (a word contains a bigram trough when one of its bigrams is much less frequent than the adjacent bigrams). In particular, Seidenberg (1987), Hay (2003), and Hay and Baayen (2003) all found evidence for the position that low transitional probabilities within bigrams are effectively employed in discovering and employing sublexical structure. In addition, Newport and Aslin (2000, 2004) showed that, in support of their statistical learning theory, language acquirers are sensitive to transitional probabilities of syllables and, to some extent, even to non-adjacent regularities.

It should be noted that the results reported above come from very different paradigms and language modalities. Seidenberg addressed written word comprehension using the Prinzmetal paradigm (Prinzmetal, Hoffman, & Vest, 1991; Prinzmetal, Treiman, & Rho, 1986), Hay addressed speech perception using typical speech perception paradigms, and Newport and Aslin typically employed nonsense language using a two alternative forced-choice test, in which participants had to decide (after some exposure to the nonsense language) which of the two presented letter strings was more word-like. The only study on compound word processing during normal reading we know off did not find any effect for the role of constituent boundary bigram frequency (Inhoff, Radach, & Heller, 2000). In three reading experiments, they found no processing differences between compounds with common, intermediate, and uncommon bigrams around the constituent boundary in several eye movement measures. However, they also manipulated other possible segmentation cues in compounds in the same experiments, such as putting (illegal) spacing between the constituents and (illegally) capitalizing initial constituent letters. Inclusion of these other possibly stronger cues might have obscured the effect of bigram frequency. In addition, they did not check whether the uncommon bigrams also constituted a bigram trough (i.e., were between substantially higher frequency bigrams). Hence, we do not know whether segmentation cues like bigram troughs are used in accessing morphologically complex words during normal reading.

In three experiments, we investigated the role of two cues that can unambiguously signal where the constituent boundary is. The primary cue we studied is *vowel quality*: that is, whether there is a vowel quality difference across constituents. The second cue we studied is the *ambiguity* of the consonant that begins the second constituent. We contrasted consonants that could both begin the second constituent and end the first constituent with consonants that could not be the end of the first constituent. (We will explain this manipulation in more detail in Experiment 3.) In Experiment 1, we investigated whether long compounds whose first and second constituent differed in vowel quality were processed faster than compounds with the same vowel quality throughout the word. In this experiment, the final letter of the first constituent and the initial letter of the second constituent were always both vowels. The length of the first constituent varied from 4 to 10 characters. In Experiment 2, we investigated whether first constituent length affects the use of different vowel quality in the two constituents as a segmentation cue by examining compounds with short and long first constituents separately. In Experiment 3, we examined (a) whether vowel quality affects parsing even when the disharmonious letters are nonadjacent in the word and (b) whether the ambiguity of the consonant beginning the second constituent also affects processing of the compound word. We also ran analyses to examine the extent to which the effects produced by these constituent boundary cues were explainable by differences in bigram frequency at the constituent boundary. We used gaze duration as the primary measure of compound word processing. It was supplemented with more detailed analyses of the durations, probabilities, and locations of individual fixations that were designed to examine the time course of the effects.

Experiment 1

Method

Participants. Thirty students of the University of Turku participated in the experiment. All were native speakers of Finnish, and had normal or corrected-to-normal vision.

Apparatus. Eye movements were monitored by the EYELINK II eyetracker manufactured by SR Research (Canada). The eyetracker is a second-generation 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 total weight of the headband is 450 g. The cameras sample pupil location and pupil size at the rate of 500 Hz. Registration is monocular and is performed for the selected eye by placing the camera and the two-infrared light sources at a distance of 4–6 cm from the eye. The spatial accuracy is better than 0.5° and the spatial resolution (i.e., the differential accuracy) of the system is 15 min of arc. Head position with respect to the computer screen is tracked with the help of a head-tracking camera mounted on the center of the headband at the level of the forehead. Four LEDs are attached to the corners of the computer screen, and are viewed by the head-tracking camera when the subject 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.

Materials. Twenty-three two-constituent compounds in which the vowels in the first and second constituent were of different quality (the *different vowel quality* condition, hereafter the dvq-condition) and 23 two-constituent compounds in which the vowels were of same quality (the *same vowel quality* condition, hereafter the svq-condition) were selected from an unpublished computerized newspaper corpus of 22.7 million word forms with the help of the WordMill database program of Laine and Virtanen (1999). For all the words in both conditions, the last letter of the first constituent and the first letter of the second constituent were both vowels. An example, for the dvq-condition would be SÄHKÖ/ASENTAJA 'electricity expert' (the constituent boundary bigram containing front vowel Ö and back vowel A); an example, for the svq-condition would be SATU/OLENTO 'fairytale creature' (the constituent boundary bigram containing two back vowels, U and O). The two conditions were matched on several indices, including word length, first constituent length, whole word frequency, first and second constituent frequency, average bigram frequency, initial trigram frequency, and final trigram frequency.[1] Naturally, the two conditions differed on the frequency of the bigram around the constituent boundary. The lexical–statistical properties of both conditions are listed in Table 1.

[1] Here and in the remainder of the paper, no differences between the conditions on the matched variables are even remotely close to significant.

Table 1
Lexical–statistical properties of the two vowel quality conditions in Experiment 1

Lexical–statistical property	DVQ[a]	SVQ[a]
Whole word frequency per 1×10^6	0.74	0.70
First constituent frequency per 1×10^6	93	108
Second constituent frequency per 1×10^6	241	230
Word length in characters	12.8	12.2
First constituent length in characters	6.0	5.5
Average bigram frequency per 1000	6.08	6.77
Initial trigram frequency per 1000	0.69	0.60
Final trigram frequency per 1000	0.83	1.06
Frequency of the bigram around the constituent boundary per 1000	0.09	4.06

[a] Here and in all other tables, DVQ stands for compounds of which the vowels in the first constituent are of different quality than the vowels of the second constituent. SVQ stands for compounds of which the vowels in the first constituent are of the same quality as the vowels of the second constituent.

The target words were embedded in sentences with each target word appearing in a separate sentence. Each of the compound words in the dvq-condition was paired with a compound word in the svq-condition, and a sentence frame was constructed that was identical up to the word following the target word; the remainder of the sentences differed.

dvq-condition:

Ystäväni kertoi, että **selkä/ongelma** oli nyt taaksejäänyttä elämää.

’My friend said, that the **back problem** was now behind him.

svq-condition:

Ystäväni kertoi, että **ryöstö/yritys** oli jättänyt hänelle pysyviä traumoja.

My friend said, that the **robbery attempt** had left him with permanent traumas.

The sentences were presented in Courier font one at a time, starting from the center left position on the computer screen.[2] Each sentence took up a maximum of three lines of text, and the critical word never appeared as the initial or final word of a line. With a viewing distance of about 65 cm, one character space subtended approximately 0.5° of visual angle. The 46 experimental sentences were mixed with 34 filler sentences. The sentences were presented in two blocks; the two sentences of a pair never appeared in the same block. The order of the blocks was counterbalanced across participants and the order of the sentences was randomized within a block. There was a short break between the two experimental blocks. A practice block of eight sentences, whose structures were similar to those in the experimental sentences, preceded the two experimental blocks.

Procedure. Prior to the experiment, the eyetracker was calibrated using a 9-point calibration grid that extended over the entire computer screen. In addition, prior to each sentence, the calibration was checked by presenting a fixation point in the center left position of the screen; if needed, the calibration was automatically corrected, after which a sentence was presented to the right of the fixation point.

Participants were instructed to read the sentences for comprehension at their own pace. They were further told that they would periodically be asked to paraphrase the last sentence they had read to make sure that they were attending to what they read. It was emphasized that the task was to comprehend, not to memorize the sentences. (Participants were asked to paraphrase a sentence after approximately every seven sentences.) The experimental session lasted a maximum of 50 min.

Results and discussion

For the target word, the primary measure of the time to encode the word that we will employ is the gaze duration. This is the summed duration of all fixations on the word in the first pass through the text before exiting to another word. It does not include regressions back to the word. In addition, skips of the target word (which were extremely rare) are excluded rather than counted as zero fixation time. In addition, the duration, location, and probability of individual fixations were assessed to get a more detailed picture of the time course of processing. Individual fixation durations shorter than 50 ms were excluded from the analyses as other analyses have shown that they are not likely to be real fixations, but instead, momentary artifacts of the eye movement system. The averages of the measures for the two conditions are presented in the second and third column of Table 2. For both the participant and item analyses, vowel quality type was a within factor. (Here, and in Experiments 2 and 3, the values in the tables are means over the items.)

Duration measures. The gaze duration on the target word is probably the best single measure of how long it

[2] Courier font was employed so that all characters subtended equal widths. As a result, words matched on number of characters were also matched on width.

Table 2
Averages of eye movement measures for compounds with same vowel quality and different vowel quality across constituents (Experiment 1)

Eye movement measure[a]	DVQ	SVQ
Gaze duration	521	564
First fixation duration	193	203
Second fixation duration	196	198
Third fixation duration	186	186
Probability of second fixation	0.91	0.92
Probability of third fixation	0.56	0.57
Probability of fourth fixation	0.18	0.23
First fixation location	4.01	3.98
Second fixation location	8.05	7.73
Third fixation location	9.51	8.18

[a] Here and in Tables 4 and 6, the duration values are given in ms and the location values in character spaces.

takes to recognize the target word. As can be seen in Table 2, gazes on the different vowel quality (dvq) compounds were 43 ms shorter than those on the same vowel quality (svq) compounds, $t_1(29) = 3.11$, $p < .01$, $t_2(45) = 1.99$, $p = .05$, indicating that the vowel quality manipulation had a substantial impact on processing time for the compound words.[3] Some of this effect occurred early, as the first fixation duration for dvq compounds was 10 ms less than for svq compounds, $t_1(29) = 4.32$, $p < .01$, $t_2(45) = 2.77$, $p < .01$. In contrast, there was virtually no vowel-quality effect on either second or third fixation duration, (second: $t_{1,2} < 1$; third: $t_1(28) = 1.61$, $p > .10$; $t_2 < 1$).

Probability of fixation measures.[4] There was virtually no vowel quality effect for the probability of a second or third fixation, a11 Fs < 1, but the svq-compounds were 5% more likely to be fixated at least four times than the dvq-compounds, $t_1(29) = 2.05$, $p < .05$, $t_2(45) = 1.70$, $p < .10$.

Location measures. There was no vowel quality effect for the location of the first fixation, $t_{1,2} < 1$, but there was for the location of the second and third fixations. Second fixations were 0.32 letters further in the word for dvq-compounds than for svq-compounds, $t_1(29) = 2.76$, $p < .01$, $t_2(45) = 2.00$, $p = .05$, and for third fixations, the difference was 1.33 letters, $t_1(29) = 4.32$, $p < .001$, $t_2(45) = 5.18$, $p < .001$.

Interim Summary. The data of Experiment 1 show that the vowel quality manipulation affected the speed of processing of the target word as indicated by the gaze duration data. However, it is interesting that, although there was a 10 ms effect on first fixation duration, most of the 43 ms difference in gaze duration surfaced in later measures, mainly in the probability of a fourth fixation. This later effect appeared to be mediated by the positioning of the second and third fixations, which was closer to the constituent boundary for same vowel quality compounds.

Follow-up analysis. An interesting question is whether the length of the first constituent modulated the effect of vowel quality. That is, when the first constituent is short, the characters of the first constituent would usually fall in the area of high visual acuity, so that the constituent boundary might easily be determined by full identification of the first constituent and not need any additional orthographic cues. In contrast, for long first constituent compounds, it is unlikely that all characters of the initial constituent would be in sharp foveal vision, and thus there is room for orthographic cues to help indicate where the first constituent ends and thus aid in processing of the constituents and the word. This suggests that short first constituent compounds may be less affected by the vowel quality difference than long first constituent compounds. To test this hypothesis, we divided our experimental materials into relatively short and relatively long first constituent compounds.

In the experiment, the compounds were matched so that the difference between the first constituent lengths in the two conditions was at most one letter. The first constituents ranged from 4 to 10 letters, and many were four or five characters long. We selected the items whose first constituents were four and five letters for the subset of short first constituent compounds ($n = 11$) and the items whose first constituents were six or more letters for the subset of long first constituent compounds ($n = 8$). The item pairs for which the first constituent length differed by one character around the boundary (i.e., with five letters in one condition and six in the other) were excluded ($n = 4$). This division produced no difference in the first fixation duration effect: the effect was 12 ms in favor of the dvq-compounds for long first constituents, $t_1(29) = 2.30$, $p < .05$, and 10 ms for short first constituent compounds, $t_1(29) = 2.30$, $p < .05$. However, the division produced a difference in the gaze duration effect consistent with our hypothesis: for the short first constituents there was only a 23 ms difference between the dvq-compounds and the svq-compounds, $t_1(29) = 1.24$, $p = .22$, but there was a 49 ms difference for the long first constituents, $t_1(29) = 2.18$, $p < .05$. It should be noted that, due to reduced power, none of the item analyses came out significantly (all ps $> .08$).

[3] In case of a 'blind' parser, it would be possible that all the morphemes and pseudomorphemes in a compound word are parsed out. However, the effect size was almost exactly the same (42 ms) for the 7 svq-compounds that contained a pseudo-constituent spanning the constituent boundary. This is evidence against a view that all possible morphemes included in a word would get automatically activated.

[4] The probability of the nth fixation reflects the probability of at least n fixations. For example, the probability of a third fixation, also includes trials on which a word was fixated four or more times.

It thus appears that the ability to localize the constituent boundary is important for smooth processing of a long compound. The results of this experiment showed that vowel quality is a factor that helps to parse the compound into its constituents. However, the post hoc analyses indicated that this factor might mainly be needed when the first constituent is of considerable length. These latter analyses, though suggestive, were post hoc and not significant. Moreover, in such post hoc analyses, matching of individual item pairs on other aspects is more difficult. A better and more straightforward way to test the hypothesis that the vowel quality effect is modulated by first constituent length is to factorially manipulate the length of the first constituent. Thus, in Experiment 2, we independently varied first constituent length and vowel quality to test whether the vowel quality effect crucially depends on the first constituent being relatively long.

Experiment 2

Method

Participants. Thirty-five students of the University of Turku participated in the experiment. All were native speakers of Finnish, and had normal or corrected-to-normal vision. None of them had participated in Experiment 1.

Apparatus. The apparatus was identical to that used in Experiment 1.

Materials. Eighty two-constituent target compounds were included in this experiment. These compounds were selected so that there were 40 whose first constituent was relatively long (7–9 characters) and 40 whose first constituent was relatively short (3–5 characters). Each of these two sets was further divided into 20 compounds in which the vowels in the first and second constituent were of different quality (dvq-condition) and 20 in which the vowels in the first and second constituent were of the same quality (svq-condition). As in Experiment 1, all compounds were selected from the unpublished computerized newspaper corpus of 22.7 million word forms with the help of the WordMill database program of Laine and Virtanen (1999). The vowel quality conditions within the long and short first constituent compounds were matched on several factors, including word length, first constituent length, whole word frequency, first and second constituent frequency, average bigram frequency, initial and final trigram frequency. Naturally, the same vowel quality and different vowel quality conditions differed on the frequency of the bigram around the constituent boundary. However, the two short first constituent conditions were well matched with the long first constituent conditions on all relevant indices. The lexical–statistical properties of all four conditions are listed in Table 3.

The target words were embedded in sentences with each target word appearing in a separate sentence. Each of the long first constituent compound words in the dvq-condition was paired with a long first constituent compound word in the svq-condition, and a sentence frame was constructed that was identical up through the word following the target word; the rest of the sentence was different. Similarly, each of the short first constituent compound words in the dvq-condition was paired with a short first constituent compound word in the svq-condition and embedded in a sentence frame that was identical up through the word following the target word.

The 80 target sentences were mixed with 66 filler sentences. The sentences were presented in two blocks, so that the paired sentences never appeared in the same block. The order of the blocks was counterbalanced across participants, and within a block the order of sentences was randomized. There was a short break between the two experimental blocks. A practice block that included eight sentences with a similar kind of structure as those in the experiment proper preceded the experimental blocks.

Table 3
Lexical–statistical properties of the four vowel quality conditions in Experiment 2

Lexical–statistical property	Long first constituent		Short first constituent	
	DVQ	SVQ	DVQ	SVQ
Whole word frequency per 1×10^6	0.84	0.65	0.59	0.63
First constituent frequency per 1×10^6	69	61	98	98
Second constituent frequency per 1×10^6	172	156	140	136
Word length in characters	12.3	12.4	12.2	11.8
First constituent length in characters	7.5	7.8	4.5	4.4
Bigram frequency per 1000	5.98	7.33	5.99	6.39
Initial trigram frequency per 1000	0.81	0.80	0.48	0.67
Final trigram frequency per 1000	0.71	1.16	1.12	1.23
Frequency of the bigram around the constituent boundary per 1000	0.05	2.05	0.06	3.82

Procedure. The procedure was identical to that of Experiment 1.

Results and discussion

Again, individual fixation durations shorter than 50 ms were excluded from the analyses. For the target word, the same dependent measures as in Experiment 1 were analyzed. For all target word measures, analyses of variance were conducted with either participants or items as the random variable. For the participant analyses, both vowel quality type and constituent length were within factors, whereas for the item analyses, vowel quality type was a within factor and constituent length was a between factor. The means of the various eye movement measures for the four conditions are listed in Table 4.

Duration measures. As can be seen in Table 4, the predicted pattern of results came out very clearly in gaze duration: gaze durations were 114 ms shorter for the long first constituent dvq-compounds than for the long first constituent svq-compounds, but there was only a 2 ms gaze duration difference due to vowel quality for the short first constituent compounds. Both the main effect of vowel quality and the interaction between first constituent length and vowel quality were significant, $F_1(1,34) = 24.88$, $p < .001$, $F_2(1,38) = 5.67$, $p = .02$, $F_1(1,34) = 47.66$, $p < .001$, $F_2(1,38) = 5.34$, $p = .03$, respectively. Subsequent t tests indicated that the vowel quality effect was significant for the long first constituents, $t_1(34) = 6.56$, $p < .001$; $t_2(39) = 2.61$, $p = .01$, but not for the short first constituents, $t_{1,2} < 1$.

For first fixation duration, there was only a 3 ms difference between dvq-compounds and svq-compounds, $F_1(1,34) = 2.54$, $p = .12$; $F_2(1,38) = 3.54$, $p = .07$, and there was no interaction between constituent length and vowel quality, $F_{1,2} < 1$. For second fixation duration, there was only a 1 ms difference, and no interaction, Fs < 1. For third fixation duration, there was a 10 ms main effect of vowel quality, $F_1(1,32) = 5.12$, $p = .03$; $F_2(1,38) = 3.55$, $p = .07$, and no interaction, $F_{1,2} < 1$. Thus, it seems that little of the gaze duration difference for the long first constituent words was due to differences in individual fixation durations.

Probability of fixation measures. It might come as no surprise that the probability effects are pretty much in line with the gaze duration effect (a sizeable vowel quality effect for long first constituent compounds and little vowel quality effect for short first constituent compounds), as there were only small vowel quality differences for individual fixation durations. The main effect of vowel quality was significant for the probability of making a second fixation, $F_1(1,34) = 7.54$, $p = .01$; $F_2(1,38) = 3.97$, $p = .05$, as well as the interaction, $F_1(1,34) = 7.38$, $p = .01$; $F_2(1,38) = 5.47$, $p = .03$. The interaction reflected that there was no effect for the long first constituent compounds, both ts < 1, but somewhat unexpectedly, a reversed effect of vowel quality for the short first constituent compounds, $t_1(34) = 3.66$, $p < .01$; $t_2(39) = 2.63$, $p = .01$. For the probability of making at least three fixations, the main effect of vowel quality was not reliable, $F_1(1,34) = 2.09$, $p = .15$; $F_2 < 1$, but the interaction was, $F_1(1,34) = 10.47$, $p < .01$; $F_2(1,38) = 4.54$, $p < .05$. The interaction reflected that for long first constituents, the dvq-compounds were 8% less likely to be refixated, $t_1(34) = 2.91$, $p < .01$; $t_2(39) = 1.93$, $p = .06$, but for short first constituents, they were 4% more likely to be refixated, $t_1(34) = 1.46$, $p = .15$; $t_2 < 1$. For the probability of making four or more fixations, the main effect of vowel quality and the interaction were significant, $F_1(1,34) = 19.42$, $p < .001$; $F_2(1,38) = 6.19$, $p < .02$; $F_1(1,34) = 12.22$, $p = .001$; and $F_2(1,38) = 3.96$, $p = .05$, respectively: for long first constituents, dvq-compounds were 15% less likely to be refixated, $t_1(34) = 5.74$, $p < .001$; $t_2(39) = 2.70$, $p = .01$, but for short first constituents, they were only 2% less likely to be refixated, ts < 1. Thus, a lot of the vowel quality effect on gaze duration for long compounds was due to

Table 4
Averages of eye movement measures for long and short first constituent compounds with same vowel quality and different vowel quality across constituents (Experiment 2)

Eye movement measure	Long first constituent		Short first constituent	
	DVQ	SVQ	DVQ	SVQ
Gaze duration	561	675	544	546
First fixation duration	206	209	215	218
Second fixation duration	216	219	213	212
Third fixation duration	205	217	191	198
Probability of second fixation	0.92	0.92	0.91	0.86
Probability of third fixation	0.49	0.57	0.46	0.42
Probability of fourth fixation	0.18	0.33	0.17	0.19
First fixation location	4.31	4.37	4.34	4.28
Second fixation location	7.95	8.08	7.54	7.17
Third fixation location	8.51	8.15	7.78	7.34

differing probability of refixating them a third and fourth time.

Since there was a difference between the long first constituent conditions in the third fixation probability and 4th fixation probability, we further analyzed the pattern of fixations after the second fixation. We reasoned that, as both conditions diverged from the second fixation onwards, it is of interest to know whether that is accompanied by a difference in the probability of making a fixation back toward the word beginning. In fact, for long first constituent compounds, different vowel quality compounds elicited significantly fewer intraword regressions than same vowel quality compounds (0.28 vs. 0.37), $t_1(28) = 3.50$, $p < .01$; $t_2(39) = 2.25$, $p < .03$, but not for the short first constituent compounds (0.33 vs. 0.35), $t_{1,2} < 1$.

Location measures. For first fixation location, there was neither an effect of vowel quality nor an interaction, $F\text{s} < 1$. For second fixation location, there was also no main effect of vowel quality, $F_1(1, 28) = 2.07$, $p = .16$, $F_2(1, 38) = 1.30$, $p = .26$, but there was a significant interaction between vowel quality and constituent length, $F_1(1, 28) = 8.89$, $p < .01$, $F_2(1, 38) = 5.58$, $p = .02$, as the second fixation location was 0.37 characters further into the word for short dvq-compounds, $t_1(34) = 3.02$, $p < .01$; $t_2(39) = 3.28$, $p < .01$, whereas for long first constituents, it was 0.13 characters less far, $t_1(34) = 1.58$, $p = .12$, $t_2(39) = 1.20$, $p = .24$. The third fixation location was 0.44 characters further into the word for dvq-compounds, $F_1(1, 26) = 3.66$, $p = .06$; $F_2(1, 38) = 3.10$, $p = .09$, but there was no hint of an interaction, $F_{1,2} < 1$.

To summarize, it is quite clear from the gaze duration data that the processing of long first constituent compounds was strongly influenced by whether the vowels that straddled the constituent boundary were of the same or different vowel quality. In contrast, for short first constituent compounds, there was no consistent effect of vowel quality on gaze duration. It thus seems that for short first constituent compounds, "low level" orthographic cues in the word that logically should help to determine the constituent boundary do not seem to matter, presumably because the boundary falls practically always in the area of high foveal vision.

Post hoc analyses. Above, we tacitly assumed that the vowel quality difference across constituents was the cause of the effects that we found in Experiments 1 and 2. However, it is also possible that the processing benefits for the different vowel quality compounds found in these experiments have nothing or little to do with the quality of the vowels in the first and second constituent as such, but are merely due to the frequency of the bigram that straddles the constituent boundary. As can be seen in Tables 1 and 3, the bigram frequencies of the two characters around the constituent boundary in the svq-condition were at least 40 times as frequent as those in the dvq-condition. (Of course, when manipulating vowel quality across constituents it is impossible to match these frequencies.) Thus, it is possible that our effects are a result of the difference in bigram frequency at the constituent boundary between dvq-compounds and svq-compounds.

We conducted two post hoc analyses for Experiment 2 to assess this problem. In Experiment 2, there was a natural division in the svq-group between compounds with very high frequency bigrams and those with relatively low frequency bigrams around the constituent boundary: there were nine high frequency bigrams with a mean frequency of 3.89 (per thousand) and a range of 2.04–6.03, and 11 low frequency bigrams with a mean frequency of 0.54 and a range of 0.14–1.10. This division corresponds with bigram cues for syllable structure, so that for the compounds with very high frequency bigrams, the first character of the second constituent can be part of the final syllable of the first constituent. For example, in *va.luut.ta.|u.ni.o.ni* 'currency union' (with dots indicating syllable boundaries), the 'u' after the constituent boundary could have been part of the frequently used syllable 'tau' (e.g., *tau.ko* 'break' and *tau.ti* 'disease'). Thus, it was possible to select a subgroup of 11 svq-compounds (the other set) that was more or less matched on the bigram frequency of the constituent boundary with the corresponding dvq-compounds. However, the vowel quality effect (svq-condition minus dvq-condition) in gaze duration was exactly the same for this subgroup as it was for the subgroup with a greater mismatch in bigram frequency at the constituent boundary (114 ms) in both cases. Another way to look at this issue is to see whether the difference in bigram frequency of the vowels around the constituent boundary predicts the difference in gaze duration between the svq-condition and the dvq-condition, using a regression analysis. This regression analysis, using items as cases, showed a nearly flat slope ($t < 1$) and an intercept of 106 ms that was close to significant, $t = 1.81$, $p < .09$. Thus, the difference in gaze duration could not be predicted by the difference in bigram frequency, and the intercept was nearly of the same size as the gaze duration effect in the original analyses.

Our post hoc analyses thus indicate that the vowel quality cue at the constituent boundary is not merely a bigram frequency effect, but instead has a different locus. We can think of two reasons why it may be qualitatively different from bigram frequency effects that are reported in the literature. First, this different vowel quality feature may be a unique defining feature for morphemic boundaries not shared by other low transition probability bigrams. That is, the cues related to bigram frequency also define non-morphemic syllable boundaries as well as morphemic syllable boundaries. In addition, it may be unique because of its phonological properties, as several studies have shown that phonological coding appears to enter early and importantly into printed word identification (Perfetti, Zhang, & Berent, 1992; Pollatsek, Lesch, Morris, & Rayner, 1992; Pollatsek, Tan, & Rayner, 2000b; Spinks, Liu, Perfetti, & Tan, 2000).

The above speculations raise the question of whether the vowel quality cue is dependent on the two vowels being adjacent to each other (i.e., as a bigram). On the one hand, it seems that this would be important, as the exact location of the constituent boundary would only be uniquely defined by the presence of the different quality vowels when they are adjacent. However, it is possible that this cue, in addition to other constraints, might still help to define the constituent boundary even if the "mismatching" vowels were not adjacent (e.g., if the second constituent started with a consonant followed by the different quality vowel). In Experiment 3, we tested whether adjacency of the matching and mismatching vowels was necessary to produce the vowel quality effect observed in Experiments 1 and 2 by separating them by a consonant. This also gave us the opportunity to manipulate the cueing properties of the consonant at the constituent border, which also allowed us to probe further the distinction between cues for syllable boundaries and cues for constituent boundaries.

Experiment 3

The post hoc analyses of Experiment 2 indicated that the vowel quality effect did not crucially depend on the bigram frequency of the constituent boundary letters. However, it is not clear whether the vowel quality effect observed in the first two experiments is crucially dependent on the mismatching vowels forming a bigram, and thus unambiguously defining the constituent boundary, or whether there is a more general phenomenon involving a global perception of two constituents with different vowel qualities. Nonetheless, even if there is a perception of two constituents when the vowels are non-adjacent, that still leaves open the question of where the constituent boundary is. In Finnish, however, there are other cues that can help to locate the constituent boundary. The one we chose to focus on is the following: there are certain consonants that cannot end a constituent. Thus, if the reader (for other reasons) believes that a consonant is at the constituent border, then for these consonants, it has to be the beginning of the second constituent rather than the end of the first constituent. (In contrast, there are other consonants that can either be the end of the first constituent or the beginning of the second constituent.) This cueing property of the consonant, however, is different in one important aspect from the vowel harmony cue of the prior two experiments. That is, when two disharmonious vowels are adjacent, that bigram unambiguously indicates that the constituent boundary is between those letters, whereas when there is a consonant that cannot end a constituent, neither its presence, nor either of the bigrams containing it, necessitate that there is constituent boundary there. Instead, the consonant is a conditional cue in the following sense: if one assumes the constituent boundary either follows or precedes the consonant, then the consonant unambiguously determines which of the two alternatives is true.

In Experiment 3, we had two primary cueing manipulations. The first was vowel quality. However, it was different from that in the first two experiments in that the disharmonious vowels were not adjacent. For all of the target words, the first constituent ended in a vowel, but the second constituent began with a consonant. The second was the above constituent cueing property of the initial consonant of the second constituent. For some of the items, this consonant could not end a constituent but could begin a constituent, and for the rest of the items, this consonant could either begin or end a constituent. We expected that the latter type of consonants, the so-called ambiguous consonants, would be of less help in constituent boundary determination (resulting in longer gaze durations) than the former type of consonants, the unambiguous ones. At a second level of analysis, there was another property of the consonant that was manipulated. For the consonants that could not end a constituent, there was a further distinction: the consonants in one subset also could not end a syllable, whereas the consonants in the other subset could end a syllable even though they could not be the last letter in a constituent. We expected that the consonants that could not end a syllable would give the strongest cue for a sublexical boundary and therefore generate shortest gaze durations. Both this latter distinction and post hoc tests involving bigram frequency serve as tests of the extent to which such a consonantal constituent boundary cue is specific to indicating the location of constituent boundaries instead of being a cue indicating some more generic statistical property of orthography or phonology.

Method

Participants. Thirty-four students of the University of Turku participated in the experiment. All were native speakers of Finnish, and had normal or corrected-to-normal vision. None of them had participated in Experiment 1 or 2.

Apparatus. The apparatus was identical to that used in Experiments 1 and 2.

Materials. Eighty-four two-constituent target compounds were included in this experiment in a 2 by 2 design with vowel quality and constituent boundary consonant ambiguity as independent variables. Forty-two compounds were selected in which the vowels in the first and second constituent were of different quality (dvq-condition) and 42 were selected in which the vowels in the first and second constituent were of the same quality (svq-condition). It should be noted that this design allowed us to match the vowel quality conditions on constituent boundary bigram frequency as well. The compounds in both groups were chosen so that: (a) the final letter of the first constituent was a front or back vowel; (b) the first letter of the second constituent was a consonant; and (c) the second letter of the second constituent was a front or back vowel. Furthermore, for 28 of the 42 items in each vowel quality condition, the initial consonant of the second constituent was *unambiguous*, in the sense that it could not be the last letter of an initial constituent. For the other 14 items in each vowel quality condition, the initial consonant of the second constituent (*s*) was *ambiguous*, in the sense that it could either be the last letter of the first constituent or the first letter of the second constituent. (We will generally refer to this manipulation as *consonant ambiguity* below.) Furthermore, as indicated above, there was a secondary manipulation of interest. That is, of the 28 items in each vowel quality condition that were *unambiguous*, 14 began the second constituent with a consonant (*j* or *v*) that cannot end a syllable, whereas the other 14 began the second constituent with a consonant (*h*, *p*, or *m*) that can end a syllable. However, the frequency of the bigram spanning the two constituents was similar in all four sub-conditions (svq-jv, 1.76; dvq-jv 1.49; svq-hpm, 1.74; and dvq-hpm, 1.83).

The lexical characteristics of the four main conditions are listed in Table 5. As in Experiments 1 and 2, all compounds were selected from the unpublished computerized newspaper corpus of 22.7 million word forms with the help of the WordMill database program of Laine and Virtanen (1999) and the target words were embedded in sentences with each target word appearing in a separate sentence.

Procedure. The procedure was identical to that of Experiments 1 and 2.

Results and discussion

Again, individual fixation durations shorter than 50 ms were excluded from the analyses. For the target word, the same dependent measures as in Experiments 1 and 2 were analyzed. For all target word measures, analyses of variance (ANOVA) were conducted with both participants and items as the random variable. For the participant analyses, both vowel quality type and the ambiguity of the consonant with respect to the location of the constituent boundary were treated as within factors, whereas for the item analyses vowel quality type was treated as a within factor and consonant ambiguity as a between factor. The data for the various eye movement measures are listed in Table 6.

Duration measures. For gaze duration, there was a 60 ms main effect of vowel quality, $F_1(1,33) = 44.05$, $p < .001$, $F_2(1,40) = 7.83$, $p < .01$, and a 52 ms main effect of consonant ambiguity, $F_1(1,33) = 66.25$, $p < .001$, $F_2(1,40) = 4.31$, $p < .05$. Dvq-compounds were read faster than svq-compounds and compounds with unambiguous consonants at the constituent boundary were read faster than those with ambiguous constituents. The vowel quality effect was slightly bigger when the consonants were unambiguous than when they were ambiguous (65 ms vs. 51 ms), but the interaction was far from significant, $F_{1,2} < 1$. Thus, the difference in the quality of the constituents' vowels affected processing even though there was no bigram containing different

Table 5
Lexical–statistical properties of the four vowel quality conditions in Experiment 3

Lexical–statistical property	DVQ, unambiguous consonant	SVQ, unambiguous consonant	DVQ, ambiguous consonant	SVQ, ambiguous consonant
Number of items	28	28	14	14
Whole word frequency per 1×10^6	0.60	0.59	0.50	0.47
First constituent frequency per 1×10^6	122	125	116	118
Second constituent frequency per 1×10^6	85	83	86	89
Word length in characters	13.3	13.3	13.1	13.4
First constituent length in characters	6.8	7.0	6.7	6.9
Average bigram frequency per 1000	6.12	7.19	6.76	6.93
Initial trigram frequency per 1000	0.82	0.76	0.74	0.96
Final trigram frequency per 1000	1.04	1.06	0.59	0.99
Frequency of the bigram around the constituent boundary per 1000	1.66	1.75	4.68	5.81
Frequency of the bigram preceding the constituent boundary per 1000	6.28	8.44	6.79	5.68
Frequency of the bigram following the constituent boundary per 1000	4.09	6.37	3.97	6.65

Table 6
Averages of eye movement measures for compounds with same vowel quality and different vowel quality across constituents and with ambiguous consonants possibly ending or beginning a constituent and unambiguous consonants possibly only beginning a constituent (Experiment 3)

Eye movement measure	DVQ, unambiguous consonant	SVQ, unambiguous consonant	DVQ, ambiguous consonant	SVQ, ambiguous consonant
Gaze duration	492	557	551	602
First fixation duration	206	205	204	204
Second fixation duration	200	206	202	205
Third fixation duration	183	193	195	191
Probability of second fixation	0.87	0.90	0.88	0.90
Probability of third fixation	0.44	0.50	0.52	0.54
Probability of fourth fixation	0.12	0.22	0.25	0.29
Probability of regression after second fixation	0.35	0.39	0.37	0.46
First fixation location	4.77	4.66	4.69	4.64
Second fixation location	8.59	8.49	8.45	8.68
Third fixation location	9.04	8.60	8.64	8.16

quality or same quality vowels that straddled the constituent boundary. In addition, the cueing property of the consonant beginning the second constituent had an equally big effect on processing time, and this effect appeared to be about the same regardless of whether the vowels in the two constituents were the same or different quality. It is also worth noting that the sum of the two main effects here appeared to be roughly equal to the vowel quality effect observed in Experiment 2.

For first fixation duration, both main effects and the interaction between them were less than 2 ms, Fs < 1.1. The second fixation duration for dvq-compounds was 5 ms less than for svq-compounds, $F_1(1,33) = 3.56$, $p = .08$, $F_2(1,40) = 2.33$, $p = .14$, but there was neither a main effect of consonant ambiguity nor an interaction, Fs < 1. For third fixation duration, the main effect of vowel quality was 5 ms, $F_1(1,33) = 1.12$, $p = .30$; $F_2 < 1$ and the effect of consonant ambiguity was also 5 ms, $F_1(1,33) = 11.52$, $p < .01$; $F_2(1,40) = 1.14$, $p = .29$. The vowel quality effect was somewhat bigger for unambiguous consonant words, but the statistical support for this was minimal, $F_1 < 1$, $F_2(1,40) = 1.58$, $p = .22$. Thus, only a small part of either main effect in gaze duration was reflected in the individual fixation durations.

Probability measures. The effect of vowel quality on the probability of making at least two fixations was significant in the participant analysis and almost significant in the item analysis, $F_1(1,33) = 4.64$, $p < .05$; $F_2(1,40) = 3.36$, $p = .07$, but the effect of consonant ambiguity and the interaction were not close to significant, all Fs < 1. For the probability of making at least three fixations, there were indications of main effects for both vowel quality, $F_1(1,33) = 6.36$, $p = .02$; $F_2(1,40) = 2.47$, $p = .12$, and consonant ambiguity, $F_1(1,33) = 13.61$, $p = .01$; $F_2(1,40) = 1.85$, $p = .18$, but little evidence for an interaction, both ps $> .20$. For the probability of making four or more fixations, the main effects of vowel quality, $F_1(1,33) = 27.57$, $p < .001$; $F_2(1,40) = 4.63$, $p < .05$ and consonant ambiguity, $F_1(1,33) = 38.66$, $p < .001$; $F_2(1,40) = 9.56$, $p < .01$, were significant. The interaction here mirrors the small interaction found in the gaze duration, although it was far from significant in the item analysis, $F_1(1,33) = 3.51$, $p = .07$, $F_2 < 1$. In all the above cases, the probability of making a refixation is (somewhat) higher in case of svq-compounds than in case of dvq-compounds, but the consonant ambiguity at the constituent boundary seems to exert its effect only beginning with the probability of making a third or fourth fixation.

As in Experiment 2, we further analyzed the pattern of fixations after the second fixation because the effect of vowel quality and/or consonant ambiguity in the probability measures and third fixation location (see below) could be related to changes in the probability of making a fixation back toward the word beginning. In fact, there was a greater tendency to regress back to an earlier position on the word for svq-compounds than for dvq-compounds, $F_1(1,32) = 8.57$, $p < .01$; $F_2(1,40) = 3.33$, $p < .08$, and for ambiguous consonant compounds than for unambiguous consonant compounds, $F_1(1,33) = 6.34$, $p = .02$; $F_2(1,40) = 1.99$, $p = .17$. There was no interaction between the two variables, $F_{1,2} < 1.2$.

Location measures. The location of the first fixation was clearly not affected by vowel quality or consonant ambiguity, all ps $> .15$. There was also no main effect of either variable on the location of the second fixation, Fs < 1, but there was a suggestion of an interaction between the two, $F_1(1,33) = 4.91$, $p < .05$, $F_2 < 1$. The third fixation location was 0.45 characters farther into the word for dvq-compounds than for svq-compounds,

$F_1(1, 32) = 6.94$, $p = .01$; $F_2(1, 40) = 3.23$, $p = .08$, and it was 0.46 characters farther into the word for the unambiguous consonant compounds than for the ambiguous consonant compounds, $F_1(1, 32) = 9.70$, $p < .01$, $F_2 < 1$. However, there was no interaction, $F_{1,2} < 1.1$. In sum, only the location of the third fixation (which is, of course, 'decided upon' during the second fixation) appeared to be at all affected by these two variables.

Summary. It seems quite clear that vowel quality not only exerts an effect on gaze duration when the constituent boundary is formed by two vowels, but also when one of the constituent boundary letters is a consonant. In addition, Experiment 3 established an independent effect of the ambiguity of the consonant beginning the second constituent, with compounds with consonants at the beginning of the second constituent that could not logically end the first constituent being processed faster than compounds with consonants at the boundary that could either be the last letter of the first constituent or the first letter of the second. Finally, this consonant ambiguity effect was largely independent of the vowel quality effect. This is a bit surprising because if the consonant is unambiguous, the VCV trigram containing the two vowels unambiguously defines the boundary when there are different quality vowels (and thus the unambiguous consonant is truly unambiguous), whereas when the vowels in the two constituents have the same quality, the VCV trigram does not unambiguously define the constituent boundary. In the following, we will consider the other measures more thoroughly, in order to get a better grip on the time course of the observed effects.

Subsidiary analyses of stimulus characteristics. As noted above, there were two types of unambiguous consonants at the boundary. Although both were unambiguous in their signaling properties of a morphemic boundary, they differed in their characteristics of signaling the end of a syllable: the second constituent of half the unambiguous compounds began with a consonant (*j* or *v*) that cannot end a syllable, whereas for the other half, the second constituent began with a consonant (*h*, *p*, or *m*) that can end a syllable. We will refer to these conditions as *unambiguous syllable* and *ambiguous syllable*. We thus ran a subsidiary analysis to see whether this syllabic property had any effect on gaze durations. We expected that unambiguous syllables would form a better sublexical segmentation cue than ambiguous syllables. In fact, the syllabic property had a slight effect in the opposite direction, that is, compound words with ambiguous syllables were read slightly faster than those with unambiguous syllables, although the difference was far from significant. For different vowel quality compounds, the gaze durations for ambiguous syllable and unambiguous syllable consonants were 486 and 498 ms, respectively, and for same vowel quality compounds, they were 551 and 564 (vowel quality effect: $F_1(1, 33) = 53.03$, $p < .001$, $F_2(1, 26) = 12.62$, $p < .01$; syllabic ambiguity effect: $F_1(1, 33) = 2.66$, $p = .11$, $F_2 < 1$; and interaction: $F_{1,2} < 1$).[5]

Our post hoc analyses of Experiment 2 indicated that bigram properties at the constituent boundary had little, if any effect on the gaze duration on the compound word. We also explored the data of Experiment 3 to see whether either the vowel quality effect or the (constituent) ambiguity of the consonant beginning the second constituent were plausibly due to possible confounding with the frequency of the bigram spanning the boundary. In fact, when one examines the mean frequency of the bigram spanning the constituent border in each of the four conditions (see Table 5), they were appreciably higher in the ambiguous conditions than in the unambiguous conditions; however, there were only small differences between the same vowel quality and different vowel quality conditions. As a next step, we computed the correlations within each of the four cells between this bigram frequency and the item means for the gaze duration. For the unambiguous conditions, the correlations were small and inconsistent: +.094 for the same vowel quality condition and −.114 for the different vowel quality condition. These correlations might have been small due to the small variability of the bigram measure within the cells. However, for the ambiguous conditions, there was appreciable variability (*SD*s > 2) and the correlations were larger. However, they were still inconsistent: −.461 for the same vowel quality condition and +.291 for the different vowel quality condition. This analysis suggested again that the bigram frequency at the boundary had little effect on gaze duration. To check this out, we also performed an analysis of covariance on the gaze duration data (using the frequency of the boundary bigram as the covariate in the analysis of item means) and found that this analysis produced the same result—significant main effects of vowel quality and consonant ambiguity (*p*s $< .01$)—as our primary analysis of variance reported earlier (bigram frequency exerted a negligible effect as the covariate, $F < 1$).

It is possible, of course, that there may be other measures that are more revealing than this bigram frequency. The one that we thought was most plausible is the frequency of the trigram that includes the consonant at the boundary and the two vowels that flanked it. Needless to say, this frequency is much lower when the vowels are of different quality, but we did not find that of much interest because of the subsidiary analysis in Experiment 2, where the frequency of the bigram

[5] We did a similar analysis on all the eye movement measures, and only one measure, third fixation duration, showed a significant syllabic ambiguity effect, $F1(1, 31) = 7.36$, $p = .01$; $F2 = 4.39$, $p < .05$. However, like the gaze effect, it was in the opposite direction to the one that would be expected.

involving the two adjacent vowels had virtually no effect. There was some difference between the consonant conditions, but not in a way that was plausibly related to the gaze duration differences. That is, this trigram frequency was actually higher for the (constituent) unambiguous consonants than for the ambiguous consonants, but only for the syllabically ambiguous consonants. For the same vowel conditions, the mean trigram frequencies were 0.29, 1.13, and 0.27 (per thousand trigrams) for the "unambiguous constituent ambiguous syllable," "unambiguous constituent unambiguous syllable," and ambiguous conditions, respectively, whereas the gaze durations were 551, 564, and 602. These differences in trigram frequency (as you might expect) were quite small for the different vowel conditions (but had the same pattern as above), and thus also had no explanatory power in explaining the effect of the (constituent) ambiguity effect. Thus, as in Experiment 2, it appeared that little, if any of the effects that were observed were due to uncontrolled low level measures such as bigram or trigram frequency.

Time course analyses

To get a better insight into the nature of the vowel quality and consonant ambiguity effects, it seemed advisable to try to get a reasonable picture of the time courses of these effects. From the analyses above, it seems reasonably clear that the effects are not immediate. That is, it does not appear to be the case that these cues aid a parse of the compound word that is prior to any serious processing of the compound word, as the effects are clearly not localized to the first and second fixations. However, it is hard to assess the time course using only fixation duration measures or only using refixation probabilities, as there was no significant effect for the duration of individual fixations on either the first, second or third fixation, and only a significant probability effect for the probability of at least four fixations. However, there were large gaze duration effects for both vowel quality and consonant ambiguity that had to be reflecting processing at some time during the first pass on the target words. Furthermore, gaze duration is logically just a cumulation of the durations of individual fixations and the probability of various refixations. Thus, it seems advisable to construct more powerful measures that give a better idea of the time course of processing that led to these large effects. We also wanted to do a similar analysis of the vowel quality effect in Experiment 2 to see whether the effects in the two experiments followed a similar time course.

As a result, we constructed two measures of *effect size on fixation*$_n$ that were successive attempts to reflect how much of the gaze duration difference could be attributed to the processing that occurred on fixation$_n$. Two design features of our measures were (a) that the sum of the effect sizes should equal the gaze duration effect and (b) that the measure should incorporate both individual fixation durations and the probability of subsequent fixations. The second measure that we came up with is better, but unfortunately, it is not expressible as a simple algebraic formula. As a result, we will try to explain it by starting with a first measure that does have a simple expression, and then indicate why and how the second measure is adapted from it. To simplify exposition, we will explain the method using the vowel quality effect, although the logic is the same for the consonant ambiguity effect.

The question, then, is how much of the processing difference in gaze duration can be attributed to processing efforts on each of the fixations. Our discussion below assumes immediacy of effect, which seems most parsimonious and is consistent with many current models of reading including the E-Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998). As a result, we assumed that the difference between the fixation durations on a given fixation is a result of processing on that fixation. However, one then has to weight that difference by the probability of making that fixation. For example, if there is a 60 ms fixation time difference on the 5th fixation, but only a .02 probability of making a 5th fixation, the impact on the gaze duration is going to be .02 times 60 ms. Similarly, the decision to make fixation$_{n+1}$ is likely decided on during fixation$_n$. The impact of this decision affects gaze duration by adding more fixations "downstream" and the difference in this refixation probability between conditions will affect the difference in gaze durations. This impact should also be weighted by the probability that one is still fixating the word when this decision is made. The initial formula we developed was the following:

$$
\begin{aligned}
\text{Effect size at fix}_n &= \text{ProbFixDVQ}_n * (\text{FixDurSVQ}_n - \text{FixDurDVQ}_n) \\
&\quad + \text{FixDurSVQ}_{n+1} * (\text{ProbFixSVQ}_{n+1} \\
&\quad - \text{ProbFixDVQ}_{n+1}).
\end{aligned}
$$

Thus, as indicated above, the first part of the formula captures the impact of the differing fixation durations on fixation$_n$ and the second part captures the impact of the differing probabilities of making the next fixation. To give a simple example, suppose: (a) there are at least one and at most two fixations in both the SVQ and DVQ conditions; (b) the mean first and second fixation durations in the SVQ condition are 230 and 250 ms and those in the DVQ condition are 210 and 200 ms, and (c) the probabilities of a second fixation in the two conditions are 0.6 and 0.4. Then the gaze duration for the SVQ condition is $230 + 0.6 \times 250 = 380$ ms, the gaze duration in the DVQ condition is $210 + 0.4 \times 200 = 290$ ms, and the difference between the two conditions is 90 ms. How much of this 90 ms effect can be attributed to processing on fixation 1?

Presumably, both the fixation duration difference on fixation 1 is due to processing on fixation 1 as is the difference in refixation probability. The formula above attributes both the 20 ms fixation duration difference on fixation 1 and the difference in refixation probability (0.2) times 250 ms to the effect on fixation 1 (for a total of 70 ms). It then attributes the duration difference on fixation 2 (50 ms) times 0.4 (or 20 ms) as the effect size on fixation 2. This is obviously a special case as there are no further fixations to add to the effect size for fixation 2, but it indicates that the terms in the above formulas do indeed add up to the overall gaze duration effect. (This is true in general; the terms always add up to the gaze duration.)

One problem with the above formula is that choosing which conditions to put outside the parentheses in the two terms is relatively arbitrary. (They do have to be from different conditions to make the sum equal to the gaze duration difference—that is, if one chose Prob-FixSVQ$_n$ in the first term, one would have to choose FixDurDVQ$_n$ for the second term.) However, it only made a slight difference which pairing one chooses, and to keep things reasonably simple, we present only the pairing in the above formula. It tacitly assumes that the SVQ condition is the baseline in the second term and that the impact of increasing the refixation probability will be on the subsequent fixation duration in this "baseline" condition. A more serious problem is that the above formula assumes that the probability of making a third fixation is completely dependent on what happens during the second fixation. However, that is clearly not the case, as for example, if the probability of making a second fixation is zero, the probability of each subsequent fixation is also zero. More generally, all subsequent probabilities of a refixation after fixation$_n$ are potentially affected by the decision to refixate on fixation$_n$, and not just the probability of the subsequent fixation, which is the tacit assumption of the above formula. As a result, the formula underpredicts the effects of early fixations to some extent. We devised a second, more complex, computation that adjusts for this problem. For a description of the method and a detailed description and presentation of the results, see Hyönä, Bertram, and Pollatsek (2004b).

The results of our analyses for the data of both the long first constituent compounds of Experiment 2 and all compounds of Experiment 3 are presented in Fig. 1. The analysis indicates that the bulk of the effect appears to occur in the second through fourth fixations. (The modified computation made the contribution from the fifth fixation substantially smaller.) This corroborates one's first impression that the effect of making the division of the word more or less apparent is not a particularly early effect. The effect of consonant ambiguity appears to come in a bit later than the vowel quality effect, as it is completely non-existent on the first fixation. This makes sense, as the boundary was fairly far into the word and unlikely to be processed in detail on the first fixation, whereas the disharmony of the vowels plausibly would be perceptible, at least some of the time, on the initial fixation, because it is a global cue not depending on one specific character.

One interesting question is whether the vowel quality effect in Experiment 2 could be viewed as the sum of the two effects in Experiment 3. That is, in Experiment 3, the different vowel quality manipulation contained *approximate* information about where the constituent boundary was (i.e., within a trigram) and the consonant ambiguity condition contained *exact* information about where the boundary was conditional on locating the boundary to that trigram. In Experiment 2, the different vowel quality condition, by itself, allowed one to find the exact location of the boundary. Is the local effect of Experiment 2 the sum of a local and a more global effect of Experiment 3? If one looks at the gaze durations, the

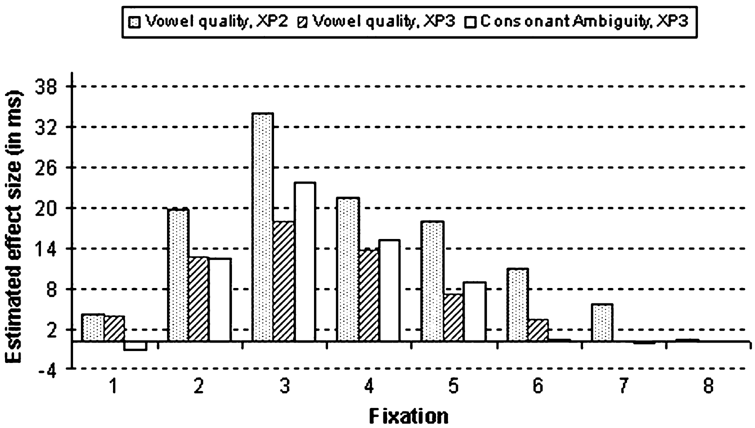

Fig. 1. The estimated size of the observed effects in Experiments 2 and 3 on each fixation during the course of processing (from Fixation 1 onwards).

answer is "yes," as the two gaze duration effects in Experiment 3 add almost perfectly to the vowel quality effect in Experiment 2. However, the time courses do not exactly line up—the biggest difference is that the effect in Experiment 2 seems longer lasting than either of the effects in Experiment 3. However, it is possible that our computation, even in the second method, does not fully account for the impact of early decisions on later processing. (For example, differences in fixation location may have effects "downstream" that are not accounted for in the computation.)

Since segmentation or parsing is generally seen as a very early stage of word processing, it might seem odd that these "parsing aids" would have such late effects. For instance, in Schreuder's and Baayen's model (1995) the segmentation stage is the earliest possible stage in word processing, and in Taft's theorizing the parser is generally seen as prelexical (e.g., Taft, 1979, 1994, 2004). However, as we were speculating in the introduction, our results show that identifying the first constituent or any other morphological unit in a complex word and localizing morphological boundaries is likely to be a much more interactive type of process. We will discuss this issue in more detail in the following section.

General discussion

The question that we addressed in this paper is whether parsing constituents in two-constituent compounds can be aided by morphemic segmentation cues. Specifically, in Finnish, different quality vowels can never appear in a simplex word and their co-occurrence in a bigram is thus a completely unambiguous cue for the presence of a constituent boundary. In Experiment 1, we found that this cue facilitated lexical processing: compounds with vowels of different quality at the constituent boundary elicited shorter first fixation durations and gaze durations than compounds with vowels of same quality at the constituent boundary. This effect, however, seemed largely restricted to compound words with longer initial constituents even though the overall length of the compound words was controlled. In Experiment 2, we divided the stimulus set into compounds with long and short first constituents and obtained a large and robust effect of the vowel quality manipulation on the gaze durations for long first constituent compounds (114 ms) but not for short first constituent compounds (2 ms). In Experiment 3, we separated the same and different quality vowels by a consonant to determine whether the vowel quality effect for long first constituent compounds observed in the prior two experiments depended on the different quality vowels being adjacent. The effect was still highly reliable, but it decreased to about half the size of the effect observed in Experiment 2. In Experiment 3, we also manipulated the consonant that began the second constituent: in some cases, it logically could not end a first constituent and in other cases, it could. This cue also had a large and reliable effect on the gaze duration on the target word. In addition, post hoc analyses indicated that both these morphemic boundary marking cues were not in any simple way related to orthographic cues such as bigram or trigram frequency.

Another important aspect of our data was that both boundary marking cues surfaced relatively late in processing, with the effects peaking at around the third fixation. This indicates that these lower-level segmentation cues do not operate as a preliminary stage prior to processing the constituents, but instead work together with attempted encoding of the constituents. This conclusion is also consistent with the findings in Experiments 1 and 2 that the vowel harmony cue had little or no impact when the initial constituent was short. That is, when the initial constituent is short, it is likely that it will be successfully decoded early, and this successful decoding of the initial constituent is sufficient to parse the word, and hence the vowel harmony cue is of little value. On the other hand, as we argue below, the vowel harmony cue that clearly surfaces by the third fixation (with small effects earlier) is serving more than a late "mop-up" function (i.e., if all other ways of parsing the word fail, try using vowel harmony). Thus, we will argue that the more plausible model is that these cues for "parsing" are used in the process of determining the morphemic structure of the compound word and in encoding the compound word constituents.

The work of the parser

Although we cannot make a definitive statement about the exact mechanism by which the two segmentation cues—vowel disharmony and consonant unambiguity—facilitated processing, we think that our findings raise some interesting general questions about how compound words are parsed and how this relates to the encoding of compound words. First, it is important to point out that the compounds we studied were fairly long—most were 12 or more characters. In our Finnish database, almost all words of this length have more than one morpheme: of the 732,823 noun forms of 12 characters or more, only 4,114 (0.56%) are morphologically simplex (in fact fewer, for the automatic parser does not succeed in identifying all derivations). Of all the 948,429 word forms of 12 characters or more, 672,902 (70.95%) are compound words. For nouns, the likelihood that a long word is a compound is even higher (611,499 out of 734,823, 83.44%). Thus, logically, if one encounters a long word, it is highly probable that it is a compound word. It is a well-known fact in the eye movement literature that readers pick up certain aspects of an upcoming word from the parafovea before they fixate it.

In particular, they certainly acquire information of the length of an upcoming word, as the longer a word is, the longer the incoming saccade into the word is (see e.g., Bertram & Hyönä, 2003; Inhoff, 1989; McConkie, Kerr, Reddix, & Zola, 1988; O'Regan, 1979; Rayner, 1979). Of course, it does not logically follow from the fact that the length of the upcoming word affects the planning of the length of the saccade into it that this information is also used to plan how to process the word. Nonetheless, it does not seem totally implausible that, given a long word, the reader is, in some way, prepared to process a compound word.[6]

It is therefore not unlikely that the search for the first constituent (and the constituent boundary) starts on the first fixation and that the parser uses all the possible information to delimit the first constituent as soon as possible. If this is true, a strict left-to-right scanning procedure, as suggested by computational linguistic models (e.g., Koskenniemi, 1984), seems unlikely, for it would not be able to explain that some orthographic cue in the middle of the word can enhance segmentation and therefore facilitate subsequent processing. In case of our compounds, Koskenniemi's model would predict that left-to-right scanning would take place until the last letter of the first constituent is reached. At that moment, the scanned letters can be matched with a unit in the root lexeme lexicon, and a pointer to a continuation lexicon triggers the search for the second constituent. This 'scan-and-match' operation is independent of the type of constituent boundary and therefore it would not predict any processing difference between the same and different vowel quality compounds and between the short and long first constituent compounds. Our results suggest, instead, that there is a much more dynamic interaction going on between gaining access to the first constituent and the search for a constituent boundary.

More specifically, we would argue that the encoding of the first constituent and the identification of the constituent boundary are processes running, to some extent, in parallel. When the first constituent is short, it appears that the more bottom-up routine for identifying the first constituent proceeds rapidly enough for the orthographic boundary information cue to be of little benefit because all letters of the short first constituent are in foveal vision. In contrast, when the first constituent is long, this bottom-up routine appears to need help from these orthographic routines. If there is no segmentation cue, the access of a long first constituent may remain incomplete when a second fixation is made to the other part of the word, or it may be misidentified. In either case, gathering information about the second constituent during the subsequent refixation helps in disambiguating (or reanalysing) the first constituent and in locating the morpheme boundary. This is why the parsing problem when there is no helpful boundary cue is reflected in later eye movement measures, such as the third fixation duration and location, and especially the probability of making a regression after the second fixation. Overall, it seems that lengthening first constituents per se does not necessarily delay processing if there is a segmentation cue such as a different vowel-quality bigram at the constituent boundary. For example, in Experiment 2, the gaze duration was only 17 ms longer for long first constituent compounds with a vowel quality segmentation cue than for short first constituent compounds with a vowel quality segmentation cue (see Table 4).

We summarize the main results of the present study in a "processing tree" depicted in Fig. 2. The process starts off by fixating the stimulus (i.e., the compound word). If the compound word is short (shorter than 9 characters or so), most (if not all) letters fall on the foveal vision and the word can usually be processed holistically (see Bertram & Hyönä, 2003). If it is longer

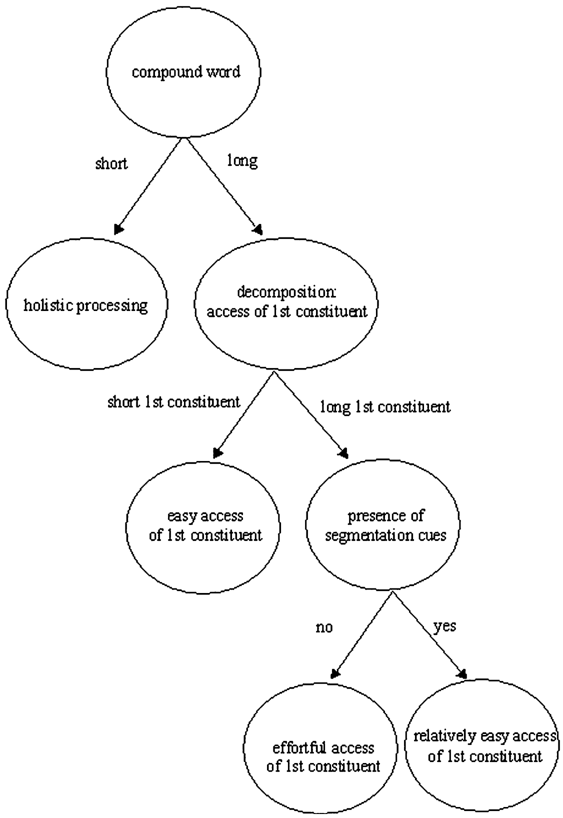

Fig. 2. A "processing tree" for identifying compound words.

[6] We should point out that there is no indication that the readers extract any information about the actual morphemic structure of the word from the parafovea. For example, Hyönä and Pollatsek (1998) found that the location of the first fixation on a compound word was not affected by the length of the first constituent.

than that the word needs to be morphologically decomposed, and its identification starts by first attempting to access the first constituent. If the first constituent is short (less than 5 characters) its access is readily achieved, and the process smoothly proceeds to the next step where the second constituent and the whole-word form are accessed. On the other hand, if the first constituent is long (7 or more characters), the relative ease of its access is determined by the presence of segmentation cues to separate out the first constituent out of the rest of word. If no segmentation cues exist, effortful processing will ensue.

Nature of the segmentation cues

Our analyses indicated that the segmentation cues studied here are not reducible to frequency of the bigram and trigram that span the constituent boundary. It is logically possible that some other, similar, type of frequency measure would do a better job of explaining the phenomenon, but that seems implausible, as these two measures are the most focused in terms of capturing the orthographic properties of the boundary. However, we wish to clarify what we mean by a similar type of measure. That is, we view the essence of these two measures is that they are capturing a sequential property of letters that is independent of their location in the word. It is possible, however, that some location-specific measure might do a better job, such as the frequency of the bigram conditional on the bigram being in location 4–9 in the word. However, casting the cue in these terms seems less parsimonious than simply describing the cue as "a bigram containing different quality vowels," "finding two different quality vowels separated by a single consonant," or "finding a consonant that cannot end a constituent."

Our results thus partly converge on and partly diverge from earlier findings. They converge on earlier findings of Seidenberg (1987), Hay (2003), and Hay and Baayen (2003) in showing that subtle properties at sublexical boundaries come to aid in morphological processing. However, contrary to these studies, but in line with Inhoff et al. (2000), our results indicate that the frequency of the bigram at a morphological boundary is not directly affecting morphological processing. It is possible that these differences derive from a paradigmatic distinction (reading vs. single word paradigms) or from the type of complex words investigated (compounds, inflections, and derivations). It is also likely that many segmentation cues are language specific. Whereas a cue like vowel quality difference across constituents in Finnish might make bigram troughs a secondary cue, readers of a language like English or Dutch might use such cues to a greater extent. Thus, in English, it may be that bigrams that never can be in the same syllable (e.g., *km* in *bookmark*) may have similar effects on compound processing as the different vowel quality bigrams in Finnish. It should also be noted that the feature of vowel harmony is not unique to Finnish and therefore a similar segmentation cue may be employed in other vowel harmony languages, such as the related Finno-Ugric language Hungarian or a Ural-Altaic language like Turkish.

Our results indicate that parsing difficulties for compound words only start to become obvious for long compound words and words with long first constituents. As a result, a place to start testing the generalizability of our results to other orthographic boundary cues appears to be in languages with productive compounding like German or Dutch. If the phenomenon does generalize to other, more universal, types of segmentation cues in those languages, one might then try seeing whether they generalize to languages such as English, where the number of long compound words is limited. More generally, it appears that processing long compounds relies heavily on compositional processing whose success depends on successfully parsing the compound word. This appears to depend on both (a) orthographic segmentation cues and (b) the length of the first constituent, the latter probably modulating the ability of the system to successfully encode the initial constituent in a bottom-up manner.

References

Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language, 40*, 41–61.

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

Baayen, H., & Schreuder, R. (1996). Modelling the processing of morphologically complex words. In T. Dijkstra & K. de Smedt (Eds.), *Computational psycholinguistics* (pp. 166–191). London: Taylor & Francis.

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

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from reading short and long Finnish compounds. *Journal of Memory and Language, 48*, 615–634.

Colé, 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.

Forster, K. I. (1989). Basic issues in lexical processing. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 75–107). Cambridge, MA: MIT Press.

Hankamer, J. (1989). Morphological parsing and the lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 392–408). Cambridge, MA: MIT Press.

Hay, J. (2003). *Causes and consequences of word structure*. New York: Routledge.

Hay, J., & Baayen, R. H. (2003). Phonotactics, parsing and productivity. *Italian Journal of Linguistics, 15*.

Hyönä, J., & Bertram, R. (2004). Do frequency characteristics of nonfixated words influence the processing of fixated words during reading? *European Journal of Cognitive Psychology, 16*, 104–127.

Hyönä, J., Bertram, R., Pollatsek, A. (in press). Are long compound words identified serially via their constituents? Evidence from an eye-movement contingent display change study. *Memory & Cognition*.

Hyönä, J., Bertram, R., Pollatsek, A. (in press). Reading Finnish compound words: Eye fixations are affected by component morphemes. In G. Underwood (Ed.), *Eye guidance and cognitive processes*. Oxford: Oxford University Press.

Hyönä, J., & Pollatsek, A. (1998). *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1612–1627.

Inhoff (1989). Parafoveal processing of words and saccade computation during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 544–555.

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

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology, 94*, 223–244.

Koskenniemi, K. (1983). *Two-level morphology: A general computational model for word form recognition and production*. University of Helsinki.

Koskenniemi, K. (1984). A general computational model for word form recognition and production. In *Proceedings of the 10th international conference on computational linguistics (COLING '84)*.

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

Libben, G. (1994). How is morphological decomposition achieved? *Language and Cognitive Processes, 9*, 369–391.

Lima, S. D., & Pollatsek, A. (1983). Lexical access via an orthographic code. The Basic Orthographic Syllable Structure (BOSS) reconsidered. *Journal of Verbal Learning and Verbal Behavior, 22*, 310–322.

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading. I. The location of initial eye fixation on words. *Vision Research, 28*, 1107–1118.

Newport, E.L., Aslin, R.N. (2000). Innately constrained learning: Blending old and new approaches to language acquisition. In S.C. Howell, S.A Fish & T. Keith-Lucas (Eds.), *Proceedings of the 24th annual Boston University conference on language development.* Somerville, MA: Cascadilla Press.

Newport, E. L., & Aslin, R. N. (2004). Learning ar a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127–162.

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

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

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a universal phonological principle. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning advances in psychology* (pp. 227–248). Oxford, UK: North-Holland.

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

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 148–162.

Pollatsek, A., Tan, L. H., & Rayner, K. (2000b). The role of phonological codes in integrating information across saccadic eye movements in Chinese character identification. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 607–633.

Prinzmetal, W., Hoffman, H., & Vest, K. (1991). Automatic processes in word perception: An analysis from illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 902–923.

Prinzmetal, W., Treiman, R., & Rho, S. H. (1986). How to see a reading unit. *Journal of Memory and Language, 25*, 461–475.

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

Rayner, K. (1979). Eye-guidance in reading: Fixation location within words. *Perception, 8*, 21–30.

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

Seidenberg, M. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*. Hillsdale, NJ: Erlbaum.

Spinks, J. A., Liu, Y., Perfetti, C. A., & Tan, L. (2000). Reading Chinese characters for meaning: The role of phonological information. *Cognition, 76*, B1–B11.

Sproat, R. (1992). *Morphology and computation*. Cambridge, MA: MIT Press.

Suomi, K., McQueen, J., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36*, 422–444.

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

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

Taft, M. (2004). Morphological processing and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*.

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

Vroomen, J., Tuomainen, J., & De Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*, 133–149.