The role of syllables in word recognition among beginning Finnish readers: Evidence from eye movements during reading

Tuomo Häikiö, Jukka Hyönä, & Raymond Bertram

University of Turku

Finland

Acknowledgements

This study was supported by Suomen Kultuurirahasto [Finnish Culture Foundation]; Suomen Kulttuurirahaston Varsinais-Suomen Rahasto [Finnish Culture Foundation, Varsinais-Suomi Regional Fund]; Oskar Öflunds Stiftelse [Oskar Öflund Foundation] to the first author.

Disclosure Statement: There is no conflict of interest of any kind for any of the authors.

Corresponding author: Tuomo Häikiö, Department of Psychology, University of Turku, FIN-20014 Turku, Finland; Tel: +358-2-3338501, E-mail: tuilha@utu.fi

Note: This is the final, accepted version of the article published in Journal of Cognitive Psychology, 27(5), 562-577. https://doi.org/10.1080/20445911.2014.982126

Abstract

The eye movements of Finnish 1st and 2nd graders were monitored as they read sentences where polysyllabic words were either hyphenated at syllable boundaries, alternatingly colored (every second syllable black, every second red) or had no explicit syllable boundary cues (e.g., ta-lo vs. talo vs. talo = 'house'). The results showed that hyphenation at syllable boundaries slows down reading of 1st and 2nd graders even though syllabification by hyphens is very common in Finnish reading instruction, as all 1st grade textbooks include hyphens at syllable boundaries. When hyphens were positioned within a syllable (t-alo vs. ta-lo), beginning readers were even more disrupted. Alternate coloring did not affect reading speed, no matter whether colors signaled syllable structure or not. The results show that beginning Finnish readers prefer to process polysyllabic words via syllables rather than letter by letter. At the same time they imply that hyphenation encourages sequential syllable processing, which slows down the reading of children, who are already capable of parallel syllable processing or recognizing words directly via the whole-word route.

Keywords: beginning reading, reading development, eye movements, syllables, word recognition

To be able to read, readers need to understand how graphemes and phonemes correspond to each other. This is a central principle in developmental theories of reading (e.g., Ehri, 1995; Ehri & McCormick, 1998). For instance, according to Ehri's (1995) model, after completing the pre-alphabetic phase readers start to utilize some letter-sound pairings in the partial-alphabetic phase before moving to the full-alphabetic phase, in which they can fully utilize these correspondences to read words letter by letter. When reading skill improves, readers start to utilize larger reoccurring units such as syllables in a phase that is called the consolidated-alphabetic phase.

The multiple-route model of orthographic processing (Grainger & Ziegler, 2011; see also Grainger, Lété, Bertand, Dufau, & Ziegler, 2012) incorporates a similar view of reading development. According to this model, increasing reading skill offers more possibilities for word recognition. Beginning readers usually read words letter by letter, but as they become more skilled they do not have to rely solely on serial letter identification and can also process letters in a word in parallel. At this point, they can use the fine-grained route which utilizes letter chunks with sensitivity to the precise letter order. In other words, this route presumes that letter chunks like syllables can be used as functional processing units. Parallel letter processing also allows readers to make use of the so-called coarse-grained route, by which they can recognize words on the basis of letter clusters that are not necessarily contingent. Thus, via this route more direct access to whole-word representations can be accomplished.

Several studies have demonstrated that syllables are used by skilled readers in lexical access. For instance, Macizo and Van Petten (2007) showed that syllable frequency affects word reading times in English. Furthermore, Ashby (2010) showed in an ERP study that phonological representations of syllables become activated early on during visual word recognition. Similarly, Ashby and Martin (2008) found a syllable congruency effect in a modified lexical decision task where words were preceded by parafoveal previews that either were or were not congruent with the initial syllable of the target word. Furthermore, Ferrand, Segui and Humphreys (1997) found syllable priming effects in English. In addition, studies in French and English show reliable number-of-syllable effects in lexical decision (Chetail, 2014; Yap & Balota, 2009). Finally, the syllable structure may – at least for relatively short words – be already extracted from the parafovea, as demonstrated by a larger skipping rate for 5-letter monosyllabic than 5-letter disyllabic words (Fitzsimmons & Drieghe, 2011).

However, the evidence for the role of syllables in word processing is not unequivocal. For instance, Schiller (2000) failed to replicate the syllable priming effect of Ferrand et al. (1997). Furthermore, even though there is evidence that increasing the number of syllables increases word recognition times, this effect either disappears (e.g., Ferrand & New, 2003) or reverses (Yap & Balota, 2009) for high frequency words. There is further evidence showing that word frequency modifies syllable effects. Colé, Magnan and Grainger (1999) found a syllable compatibility effect for low-frequency (e.g., format) but not for high-frequency words (e.g. balcon) in an experiment where participants responded whether a carrier word (e.g., balcon) started with a previously presented target syllable (e.g., bal) or not. Ashby (2006) found that reading times for low frequency (e.g. detest) but not for high frequency words were faster when the parafoveally presented preview was identical to the initial syllable of the target word (de) than when it contained an extra letter (det). These findings imply that for adults the syllable is a functional processing unit, but more in low than in high frequency words.

Several studies have found syllable effects for children as well (e.g., Colé et al., 1999; González & Valle, 2000; Hautala, Aro, Eklund, Lerkkanen, & Lyytinen, 2012; Maïonchi-Pino, Magnan, & Écalle, 2009). For instance, syllable compatibility effects with high-frequency syllables were found for French 1st grade readers; for 3rd and 5th grade readers the same effects were found independent of syllable frequency (Colé et al., 1999; Maïonchi-Pino et al., 2009). Similarly, González and Valle (2000) demonstrated that the presence of frequent initial syllables facilitated beginning readers' word recognition. For Finnish, Hautala et al. (2012) demonstrated a number-of-syllables effect for dysfluent 2nd grade readers reading easy words, but found no reliable number-of-syllables effect for more proficient 2nd graders; that is, they responded equally fast to 6-letter bisyllabic words as to length-matched trisyllabic words. All in all, the findings suggest that syllables are used in word processing but that with increasing proficiency the role of the syllable may change. It is possible though that the role of syllables in polysyllabic word processing is language-dependent, as syllable structure is more complex and syllables boundaries more obscure in some languages than others (for a comparison between 16 languages, see Seymour, Aro & Erskine, 2003).

The present study was conducted in Finnish. Finnish is a syllable-stressed language with straight-forward syllabification rules[1]. In fact, Finnish is the least complex European language with regard to the number of existing syllables (ca. 3000 syllables), syllable structure (no heavy consonant clusters for instance), and the unambiguity of syllable boundaries (Seymour et al., 2003). Furthermore, in Finnish the primary stress is always on the first syllable of a word. Finally, Finnish has a transparent orthography with an almost perfect grapheme-phoneme correspondence (Karlsson, 1999). Due to these characteristics, it is not unlikely that Finnish readers would make more frequent use of smaller reading units in word recognition than is the case in other languages. This would be in line with the psycholinguistic grain size theory (Ziegler & Goswami, 2005), which posits that readers of transparent orthographies make more frequent use of smaller units in word recognition than readers of more opaque orthographies. One could hypothesize that for Finnish syllables are likely to be involved in word recognition due to words being typically polysyllabic and having simple syllable structure and clear syllable boundaries.

The importance of syllables in word recognition is emphasized in Finnish reading instruction. One example of this is that in all Finnish 1st grade ABC books (be it mother tongue, religion, or natural sciences) the syllable structure of words is signaled by hyphens at syllable boundaries. From the 2nd grade onwards, hyphens do not appear in bisyllabic words anymore, but they are preserved in difficult and long words until the end of the 2nd grade. The use of hyphenation reflects the idea that making children aware of syllables and syllable structure aids the development of reading and writing.

However, despite the use of hyphenation in early reading instruction, recent evidence shows that hyphens within multi-unit words may disrupt word recognition. For instance, Häikiö, Bertram, and Hyönä (2011) had 2nd grade readers read sentences containing compound words which were either hyphenated at the morpheme boundary (as prescribed by Finnish spelling regulations, e.g., ulko-ovi = ‘front door’) or concatenated (sivuovi = ‘side door’). Hyphenation was beneficial for the less proficient readers, but the more proficient 2nd graders actually read concatenated compounds faster than hyphenated compounds.[2] Häikiö et al. argue that the hyphen directs initial attention mainly to the first morpheme, and that consequently morphemes are processed more sequentially than when they are presented without a morpheme boundary hyphen. They further argue that sequential morpheme-based processing is disruptive for proficient 2nd graders who might be able to recognize these words as holistic units or may gather information from both morphemes simultaneously. In contrast, the hyphen would be beneficial for less proficient readers who process polymorphemic words in a morpheme-by-morpheme fashion anyway. In addition, it has been shown that hyphenation at syllable boundaries slows down 2nd graders despite its prominent use in reading instruction (Häikiö, Bertram, & Hyönä, submitted). In this study, 2nd grade children read sentences with target words which were either concatenated or hyphenated at the syllable level. Häikiö et al. also manipulated word length and showed that hyphenation was even more disruptive for longer words with more than two syllables. They argue that hyphenation directs initial attention to the first syllable and thus interferes with parallel access to multiple syllables or with direct whole-word access. In contrast, 1st graders in the early stages of reading instruction are likely to process syllables sequentially during polysyllabic word recognition. If so, hyphens at syllable boundaries would be facilitative, as the syllables are easier to detect in hyphenated words. The current study tested this hypothesis.

As mentioned above, hyphenation at syllable level disrupts reading – at least for 2nd graders. Yet, there are other ways to visually signal syllable boundaries. One such cue is alternate coloring (every second syllable black, every second red). In fact, it has been established that both adults and children can use alternate coloring as a syllable boundary cue (e.g., Carreiras, Vergara, & Barber, 2005; Chetail & Mathey, 2009; Prinzmetal, Hoffman, & Vest, 1991; Rouibah & Taft, 2001). Prinzmetal et al. (1991) demonstrated that when participants were required to report the briefly presented color of a letter, they made more errors when the color was incongruent with the syllable structure than when it was congruent. Doignon and Zagar (2006) replicated the findings of Prinzmetal et al. (1991) for 1st to 5th grade children. Carreiras et al. (2005) showed in an ERP study a P200 effect for incongruent syllable coloring. Finally, Chetail and Mathey (2009) showed that incongruent syllable coloring slowed down weaker 2nd grade children in a lexical decision task. Interestingly, for more proficient 2nd graders this effect was reversed. Chetail and Mathey interpreted these findings as a demonstration of syllable coloring activating syllable information. When readers do not have access to many lexical candidates, as is the case for weaker 2nd graders, syllabic activation helps them to identify the word swiftly. For proficient 2nd graders syllabic activation leads to lexical competition since the colored syllables activate more words than for weaker 2nd graders, leading to slower word recognition. It may also be the case that the more proficient readers were slowed down due to being encouraged to access words via syllables, whereas normal recognition would go by via the whole-word orthographic route. In sum, the lexical decision and ERP studies indicate that alternate coloring may facilitate word recognition. The current study investigated whether alternate coloring of syllables would facilitate word recognition during reading.

In sum, the current study investigates the role of syllables and syllable cues (hyphenation and alternate syllable coloring) in polysyllabic word recognition. In Experiment 1 we investigated these issues in early reading development by testing 1st graders who had received only 2 months of formal reading instruction. In Experiment 2 we investigated the development in polysyllabic word recognition by comparing the reading performance of 1st and 2nd graders.

Experiment 1

The evidence reviewed above suggests that hyphenation is disruptive for proficient Finnish 2nd grade readers but may be beneficial for less proficient readers. In Experiment 1, we investigated whether 1st grade readers benefit from hyphenation. In addition, we tested whether a visually less salient cue, alternate coloring of syllables (using red and black, e.g., talo) would speed up reading among 1st graders. Participants read normal non-cued sentences and sentences where words contained explicit syllable boundary cues (hyphenation or alternate colors). We hypothesized that if the 1st graders have reached the consolidated-alphabetic phase (i.e., they can make use of letter chunks like syllables during polysyllabic word recognition; see Ehri, 1995), they should benefit from explicit syllable boundary cues.

However, it is also possible that 1st graders are still in the full-alphabetic phase (Ehri, 1995) and process polysyllabic words on a letter-by-letter basis. In order to investigate this issue, we introduced conditions where hyphens or alternate coloring were inserted within a syllable (t-alo, talo); we called these conditions ‘illegal’, since the hyphens do not inform readers about syllable boundaries as they usually do in 1st grade textbooks. If syllables are utilized in polysyllabic word processing, inserting a hyphen or alternating colors at a non-syllable boundary should slow down reading in comparison to syllable boundary hyphenation. On the other hand, if 1st graders read words on a letter-by-letter basis, ‘illegal’ hyphenation/alternate coloring should not lead to a different pattern of results when compared to ‘legal’ hyphenation/alternate coloring.

Finally, it is possible that 1st graders already use larger units than syllables or access multiple syllables in parallel, in which case syllable boundary cues would be disruptive. If so, hyphenation – being a more salient visual cue – is expected to be more disruptive than alternate coloring.

Method

Participants

Seventeen monolingual 1st graders (on average 7:5 years, range 7:0-7:10) were recruited from a Finnish elementary school. At the time of testing (October/November) they had received approx. 2 months of formal reading instruction but had learned to read to some extent already before this. All participants had normal or corrected-to-normal vision. None of them had previously participated in an eye movement experiment. Permission from children's parents was acquired prior to the experiment. The participants received candy or stickers as reward for participation.

Apparatus

Eye movements were recorded monocularly with a table-mounted model of Eyelink 1000 (SR Research, Canada). A sampling rate of 1000 Hz was used. The eyetracker is an infrared video-based tracking system with hyperacuity image processing and spatial resolution of 0.5 degrees. A chin rest was used to minimize head movements. The texts were presented on a 20-inch ViewSonic G225f computer screen (refresh rate of 100 Hz, resolution 1024*768).

Materials

The target word selection started with a list of words for which the age of acquisition (AoA) and familiarity ratings had been acquired earlier. The list was then complemented with additional words. For these words, the AoA and familiarity ratings were acquired from 12 members of the Turku University community. For all the words, AoA was rated on a 7-point scale (1 = 0-2 years, 2 = 3-4 years, 3 = 5-6 years, 4 = 7-8 years, 5 = 9-10 years, 6 = 11-12 years, 7 > 12 years). Familiarity was rated on a 5-point scale ("How often do you encounter or use the word?"; 1 = very seldom, 2 = seldom, 3 = sometimes, 4 = often, 5 = very often). On the basis of these ratings, we selected 90 words for the experiment. With respect to AoA, all words were rated by at least 80% of the raters with 4 or lower and had an average rating below 4 (i.e., acquired before 7-8 years).

The 90 words were divided into three groups since there were three experimental conditions (hyphenated, colored and control). The groups were matched with the other groups on word length, word frequency, average bigram frequency, initial trigram frequency, family size (i.e., the number of derivations and compounds beginning with the stem of the target word), AoA, and familiarity, each $p > .3$. All values, apart from AoA and familiarity, were extracted from a newspaper corpus containing 22.7 million word forms (Laine & Virtanen, 1999).[3] The lexical statistics of the target words are presented in Table 1.

(Insert Table 1 about here.)

The target words were embedded in sentences. Since there were three matched groups of 30 target words, 30 sentence triplets were created. A sentence frame identical up to word N+1 was constructed for each word triplet. The target word never appeared in the beginning or end of the sentence. The sentences of each triplet were rated by 9 Turku University employees to be equally natural. In addition to the control condition sentences, two versions of the sentences were created, one with a cue at the syllable boundaries in the target words, so-called legal syllable boundary cues, and one with cues within a syllable, so-called illegal syllable boundary cues. The other words in the sentences with cued target words were 'legally' cued in both conditions, that is, all cues were at syllable boundaries. In the illegal cue condition, for half of the target words the cue was moved one character toward to the beginning of the word, and for the other half one character toward the end of the word. The experimental sentences were preceded by six practice sentences, followed by a block of 45 experimental sentences, containing 15 control condition sentences, 15 sentences with legal target word cues (7 or 8 colored sentences, and 8 or 7 hyphenated sentences), and 15 sentences with illegal target word cues (8 or 7 colored sentences, and 7 or 8 hyphenated sentences). The order of sentences within each block was randomized for each participant. Each participant saw each target word only once, presented either within the hyphenated, colored or control sentences, and either in the legal or illegal condition. An example with the five conditions is presented in Table 2.

(Insert Table 2 about here.)

The sentences were presented in Courier font so that each character position was of equal width. With a viewing distance of 60 cm, one character space subtended approximately 0.4 degrees of visual angle. Sentences were always presented halfway between the top and the middle of the screen. Sentences were all single-line sentences with a maximum of 67 characters per line.

After the eye movement experiment, the children's reading skill was assessed with a classroom subtest (word recognition) of ALLU (Lindeman, 1998), a standardized Finnish reading test for elementary-school children. In this task, children see a picture (e.g., of a pencil = *kynä*) and have to choose which of the four words, all phonologically similar to the correct word (e.g., *kylä* = village, *kyllä* = yes, *kynä* = pencil, and *kylmä* = cold), corresponds to the picture. The subtest comprises 80 picture-word pairings. There is a time limit of five minutes, and one point is awarded for each correctly chosen word. The descriptive statistics of the ALLU test included in the analyses are presented in Table 3.

(Insert Table 3 about here.)

Procedure

The participants were instructed to read sentences for comprehension at their own pace and were encouraged to read silently. They were further told that after varying intervals they would get an oral question (after approx. every 6 sentences) about a sentence to which they had to give a yes/no answer. The participants answered the questions with a minimum of 75% accuracy. Furthermore, the participants were told that text presentation could be somewhat unusual but that they should nevertheless read the sentences to their best ability. The eye-tracker was calibrated using a three-point calibration grid extending horizontally over the computer screen. Before presenting a target sentence, the participant fixated on a calibration point at the left side of the screen, after which the sentence appeared.

Dependent variables and predictors

We used two standard, word-based eye movement measures as the dependent variables, gaze duration (i.e., summed fixation durations on the target word before exiting it for the first time), and go-past time (i.e., summed durations of fixations made before exiting the word to the right for the first time). Furthermore, we analyzed one global measure, namely sentence reading time. We analyzed the data twice. For the first analyses, we included the critical predictor Syllable Boundary Cue with three levels, *Control*, *Color* and *Hyphen*. Here legality of the cue was not considered. In the second analyses, we excluded the control condition and entered next to Syllable Boundary Cue (which now had two levels, *Color* and *Hyphen*) the factor Legality with two levels (*Legal* and *Illegal*).

Furthermore, for the two target word measures, we considered a number of variables that have been established as reliable predictors of word processing: word length (*Len*), word frequency (*Freq*), *AoA*, word familiarity (*Fmlrty*), and bigram trough ratio[4] (*Trough*). We were also interested in how the number of syllables affected the target word reading. As expected, there was a high correlation between number of syllables and word length, so we computed the residuals of number of syllables over word length and included the resulting variable as a predictor (*Syl.res*). For both the word and sentence measures, we included one participant measure, reading level as assessed by the reading skill test (*Allu*). Despite the fact that there were many predictors, we had no convergence problems within the models. The kappa value measuring collinearity of the continuous predictors was below 8.5 in each model.

Statistical considerations

The measures were log-transformed to normalize the data. Furthermore, values 2.5 SDs smaller or larger than the grand mean were excluded from the duration measures. Finally, all the continuous predictor variables were centered.

We used multiple regression mixed-effects modelling with participants and items as crossed random effects. Other variables did not improve the random effect structure. We will only report models with the effects retaining statistical significance in the stepwise backward elimination procedure. In this procedure, we first included all the predictors (including all relevant interactions) in the model. We then removed the least predictive predictor in each round until we ended up with a model in which all the predictors were significant, $|t| > 1.96$. We also made sure by model comparison that each predictor significantly improved the explanatory value of the model. The analyses were conducted using the languageR library of R statistical software (R Development Core Team, 2007). The models reported in Appendix 1 present the output of the pvals.fnc() function of the library.

Results

Trials in which the target word was initially skipped were excluded from the analyses (3.2% of trials). Furthermore, the analyses were restricted to first-pass sentence reading, which ended when the reader reached the last word and when possible regressions were positioned before the second to last word in the sentence. No further data trimming was done. The non-transformed means for each measure are presented in Table 4. The full models are presented in Appendix 1. One participant was excluded due to not completing the whole experiment, yielding 16 participants to be included in the analyses.

(Insert Table 4 about here.)

Gaze duration. There was a significant difference between *Control* and *Hyphen*, $t > 2$; hyphenation disrupted target word reading. The difference between *Control* and *Color* was not significant, $t < 1$, whereas the contrast between *Color* and *Hyphen* was significant, $t > 2$ (not shown in the model). There were also significant main effects of *Len*, *Fmlrty*, *Trough*, and *Allu*, all *t*s > 2. Shorter and more familiar words shortened gaze duration, as did bigram trough at the syllable boundary. Good readers had shorter gaze durations than weaker readers. None of these variables interacted with *Hyphen* or *Color*.

Additional analyses did not reveal a main effect of *Legality*, but there was a tendency for a *Syllable Boundary Cue* x *Legality* interaction, $t = 1.63$. This tendency reflected that words including hyphens within syllables elicited longer gaze durations than words with hyphens at syllables boundaries ($t = 2.30$, 293 ms), whereas the difference between colors alternating within syllables or at syllable boundaries was negligible (t < 1, 20 ms).

Go-past time. The differences between *Control* and *Hyphen* as well as *Color* and *Hyphen* were significant, *t*s > 2. Hyphenation slowed down go-past times in comparison to the other two conditions, which did not differ from each other, $t < 1$. Moreover, the main effects of *Syl.res*, *AoA*, *Trough*, and *Allu* were significant as well as the interaction between *Len* and *Hyphen*, *t*s > 2. Go-past times were longer when the words had more syllables, were acquired later, or did not have a bigram trough at the syllable boundary. Weak readers had longer go-past times than good readers. In addition, the disruptive effect for hyphenation was enlarged in long words in comparison to short words.

Additional analyses showed a significant *Hyphen* x *Legality* interaction. Illegality disrupted reading of hyphenated words, $t = 3.57$, but not of colored words, $t = -.21$. For hyphenated words the within-syllable manipulation elicited go-past times that were on average 484 ms longer than the syllable-boundary manipulation, whereas the difference between within-syllable and syllable-boundary color alternation was only 22 ms.

Sentence reading time. For these analyses, only sentences with legally cued target words were included. There was a significant difference between *Control* and *Hyphen*, $t > 2$, as well as between *Color* and *Hyphen*, $t > 2$, indicating that syllable-boundary hyphenation lengthened sentence reading times in comparison to the other two conditions. There was no significant difference between *Control* and *Color*, $t < 1.96$. Furthermore, there was a significant main effect of *Allu*, $t > 2$: Good readers read sentences faster than weak readers. None of the variables interacted with *Hyphen* or *Color*.

Discussion

In Experiment 1, it was shown that hyphenation disrupted reading both at word and sentence level. Moreover, coloring did elicit shorter reading times than hyphenation, but no disruption or facilitation appeared in comparison to the control condition. With regard to the legality of the syllable boundary cues, we found a Hyphenation x Legality interaction in go-past time as well as a trend for an interaction in gaze duration. Illegal hyphenation disrupted reading more than legal hyphenation.

The results imply that already during the 1st grade proficient beginning readers have moved past the stage of accessing words on a letter-by-letter basis. If they still did so, the hyphen should have affected word processing in the same way independent of the position where it is inserted. Since the hyphen at the syllable boundary clearly facilitates reading in comparison to hyphens inserted at non-boundary position, we can conclude that 1st graders can make use of syllables during polysyllabic word processing. Other effects point to the same direction. Go-past time became longer with increasing number of syllables, independent of word length. In addition, another cue indicating syllable structure, namely the bigram trough at the first syllable boundary, speeded up word reading. Thus, in line with earlier research in other languages (e.g., Ashby, 2010; Colé et al., 1999; González & Valle, 2000; Maïonchi-Pino et al., 2009), it seems that syllables are indeed functional processing units during lexical access among beginning Finnish readers.

However, the results also show that even though hyphenation at the syllable boundary might be beneficial in comparison to hyphenation within the syllable, it is certainly not beneficial in comparison to the non-cued, ‘normal’ presentation of words. This is best seen in the sentence reading times, where syllable boundary hyphenation elicited a one-second delay in comparison to normal presentation. How can this disruptive hyphenation effect be explained?

One possibility is that visual acuity plays a role; the further the letters are away from the fixation point, the more the final letters of the word are visually degraded. Since hyphens extend the length of a word, it brings the final letters further away from the fixation point. Visual acuity constraints may indeed contribute to the 29 ms effect in gaze duration, even though it is unlikely that one or two hyphens more would generate an effect of this size. However, it can only account a small part of the 183 ms effect in go-past time, as this measure includes regressions unlikely to be significantly influenced by visual acuity.

Häikiö et al. (2011) argued that in compound word processing a hyphen at the constituent boundary of two-constituent compound words (e.g., *ulko-ovi* = front door) directs attention to the first constituent resulting in the processing of the second constituent being delayed. As a result, hyphenated compound words are processed constituent-by-constituent, whereas short compound words without hyphens can be processed via whole-word forms or via parallel processing of constituents. Following this line of reasoning, it may be argued that hyphenated polysyllabic words are processed syllable-by-syllable, whereas their normal counterparts can be processed holistically or via simultaneous syllable processing.

Given the bigram trough and the number of syllable effects, one may be tempted to interpret the results to support the latter option. However, the results also show whole-word related effects. Familiar words elicited shorter gaze durations than less familiar words, and early-acquired words elicited shorter go-past times than late-acquired words (see reviews of Blythe & Joseph, 2011; Rayner, 1998, 2009, for similar evidence). These effects indicate that words may also be recognized as holistic units without the involvement of syllables. The most likely scenario is that both syllable-based and holistic processing make a contribution to the recognition of polysyllabic words, in line with the parallel route architecture of the Grainger and Ziegler (2011) model.

Alternate coloring of syllables did not disturb reading in comparison to normal presentation and generated faster reading times than syllable boundary hyphenation. This implies that alternate coloring does not direct attention to the first syllable in a similar way as hyphenation does. Chetail and Mathey (2009) showed that at least 2nd graders were affected by syllable coloring, be it positive (non-proficient 2nd graders) or negative (proficient 2nd graders). Our data show trends in both directions. First, there was a non-significant numerical trend of coloring slowing down reading. Second, there was a trend for a Color x Trial Number interaction ($t = 1.85$) in total fixation duration (a measure not reported here since it yielded very similar results to go-past time). The trend suggests that while alternate coloring disrupted word reading in the beginning of the experiment, it started to facilitate word reading towards the end of the experiment. To explore whether these trends could turn into significant effects, we increased the statistical power in Experiment 2.

Experiment 2

A possible reason for beginning readers not relying on syllable boundary cues in Experiment 1 may be that these cues were partly unreliable, as illegal cues were also included. Therefore, in Experiment 2 we only included polysyllabic words with legal syllable-boundary cues. To increase statistical power, we doubled the number of trials per participant. As a result, we expected to obtain larger effects of both hyphenation and alternate coloring. Furthermore, since Chetail and Mathey (2009) demonstrated that the effect of alternate coloring was facilitatory for weaker 2nd graders and inhibitory for more proficient 2nd graders, we expected to find an interaction with reading skill. To this end, and to directly assess the development of polysyllabic word processing from 1st to 2nd grade, children of both grades were included in Experiment 2.

Since early reading development proceeds at a rapid pace, we tested the same children twice at different stages. In this way we could examine possible developmental changes within the same participant groups. We also examined whether hyphenation is more beneficial in oral than silent reading. Since syllables are emphasized in Finnish reading and writing instruction, signaling syllable structure may facilitate oral reading more than silent reading where words are not mouthed. Oral reading was included in the study also for the reason that beginning readers are more used to oral than silent reading in early reading instruction.

Method

Participants

Twenty monolingual 1st graders (on average 7:3 years during the first testing, range 6:10-7:9) and twenty-one monolingual 2nd graders (on average 8:6 years during the first testing, range 7:10-8:9) were recruited from a Finnish elementary school. At the time of the fall testing (October/November) the 1st graders had received approx. 2 months and the 2nd graders approx. 1 year and 2 months of formal reading instruction. Practically all 1st graders had learned to read before entering school. All participants had normal or corrected-to-normal vision and participated in an eye movement experiment for the first time. Permission from the children's parents was acquired prior to the experiment. All the participants were also tested during spring (February/March). The participants received candy or stickers as reward for participation.

Apparatus

The eye-tracker was identical to the one used in Experiment 1.

Materials

For the fall testing, the sentences from Experiment 1 were used with the difference that the participants read all 90 sentences instead of 45 sentences. For the spring testing, half of the words were embedded in new sentences, while half of the old sentences were used to make sure that the observed effects were not context-dependent. The target words were matched between the old and new sentences on the same variables as in Experiment 1, each $p > .2$, apart from familiarity, which was slightly higher for the old than the new material (3.30 vs. 2.93, respectively). The new sentences were constructed following the same criteria as in Experiment 1. Each new sentence was rated to be equally natural with the other sentences of the triplet by eight Turku University employees.

The experiment proper was preceded by six practice sentences and included 90 sentences for 1st graders and 120 sentences for 2nd graders, divided into two blocks of equal size. First-graders read only the experimental sentences, whereas the 2nd graders also read 30 filler sentences. The sentences were assigned to three matched lists counterbalanced across participants. Each participant saw 30 hyphenated, 30 colored and 30 non-cued sentences. The block order was counterbalanced across participants. The sentence order was randomized for each participant within each block. For the fall testing, each participant saw each target word only once, either in the hyphenated, colored or control condition. In the spring testing, each participant saw the target words in the same format as in the fall testing, with the only difference being that half of the sentences were new.

After the fall experiment, the children's reading skill was assessed with a classroom subtest of ALLU (Lindeman, 1998). For 1st graders, the word recognition task was used (see Experiment 1). Second graders completed a sentence comprehension task where they saw a picture (e.g., people sailing) and had to choose which of the four sentences (e.g., "They swim", "They dive", "They sail", and "They dance") corresponds to the picture. The subtest comprises 20 picture-sentence pairings. There is a time limit of 120 seconds, and one point is awarded for each correctly chosen sentence. The descriptive statistics of the ALLU test are presented in Table 5.

(Insert Table 5 about here.)

Procedure

The eye-tracker calibration and sentence presentation were performed identically to Experiment 1. The instructions about reading and comprehension questions were almost the same as in Experiment 1 the main difference being that children got instructions about the oral reading session as well. The participants answered the questions with a minimum of 75% accuracy. For the fall experiment, half of the participants read the first block orally and the second block silently, while the other half did the opposite. For the spring experiment, each participant did the opposite; those who had previously first read orally now read the first block silently and the second block orally, and vice versa. For both testing sessions, there was a short optional break between blocks.

Design and data analyses

There were a few changes in the analyses of Experiment 2 in comparison to those of Experiment 1. First, the legality of the syllable boundary cue was not manipulated and therefore not included as a factor. Second, a number of new predictors were included in the analyses: *Grade (1 vs. 2)*, Reading Mode (*Mode*: oral vs. silent), Sentence novelty (*Novelty*: new vs old) and Experiment (*Xp:*

Experiment 1, fall vs. Experiment 2, spring) as predictors. Third, because 1st and 2nd graders completed different *Allu* subtests, standardized scores were used in the analyses. The kappa value measuring collinearity of the continuous predictors was below 5 in each model. The models were constructed in a similar way as in Experiment 1.

Results

Trials in which the target word was initially skipped were excluded from the analyses (4.9% of trials). Furthermore, the analyses were restricted to the first-pass sentence reading as in Experiment 1. No further data trimming was done. The non-transformed means for each measure are presented in Table 6. The full models and *t* values of the predictors can be found in Appendix 2. Seven 1st graders and four 2nd graders were excluded due to excessive track loss or not completing the whole experiment after getting too tired. In the end, thirteen 1st graders and seventeen 2nd graders were included in the analyses. In the following, we will not discuss results for *Mode*, *Novelty,* and *Xp* since, they did not interact significantly with the Syllable Boundary Cue, even though the effects of *Mode* and *Xp* were significant in each measure, all *t*s < -5.5. (For *Novelty*, all *t*s < 1.36). Silent reading was overall faster than oral reading and reading was generally faster at the 2nd than at the 1st testing point for each measure. Because of the lack of interactions involving *Xp, Mode and Novelty*, we collapsed all of the data in the same analyses to increase statistical power. In case Syllable Boundary Cue interacted with *Grade*, we followed up the interaction by testing the effect of Syllable Boundary Cue individually for both grades.

(Insert Table 6 about here.)

Gaze duration. The effect of *Hyphen* was qualified by significant interactions with *Allu* and *Grade, t*s > 2. The interaction with *Allu* indicated that the more proficient readers were more disrupted by hyphenation than the less proficient ones. With regard to *Grade,* hyphenation disrupted 2nd graders while the effect was non-significant for 1st graders (Control vs. Hyphen, t = 1.72 for 1st graders, and t = 3.03 for 2nd graders; Color vs. Hyphen, t = .64 for 1st graders, and t = 3.85 for 2nd graders). There was neither a difference between *Control* and *Color* nor an interaction involving *Color, t*s < 1. The main effects of *Len*, *Syl.res*, *AoA* and *Freq* were significant. Words were read faster when they were frequent or early acquired. Furthermore, gaze durations increased as a function of word length and number of syllables.

Go-past time. *Hyphen* interacted significantly with *Len*, *Allu*, and *Grade, t*s > 2. The disruptive effect of hyphenation was larger for long than for short words. With regard to reading level (*Allu*), hyphenation disrupted more proficient readers more than less proficient ones. Finally, hyphenation disrupted 2nd graders to a larger degree than 1st graders (Control vs. Hyphen, t = 3.57, Cohen's d = .110 for 1st graders, and t = 7.91 for 2nd graders, Cohen's d = .220; Color vs. Hyphen, t = 2.24, Cohen's d = .169 for 1st graders, and t = 7.27, Cohen's d = .257 for 2nd graders). For *Color*, there was neither a significant difference with the Control condition nor interactions, *t*s < 1. Finally, there were significant main effects of *Syl.res*, *Fmlrty*, and *AoA*. Go-past time was longer when the word was late acquired, less familiar, or contained several syllables.

Sentence reading time. The effect of *Hyphen* was qualified by significant interactions with *Allu* and *Grade, t*s > 2. Hyphenation disrupted more proficient readers to a larger extent than less proficient ones and 2nd graders more than 1st graders (Control vs. Hyphen t = 6.66, Cohen's d = .201 for 1st graders, and t = 11.08, Cohen's d = .288 for 2nd graders; Color vs. Hyphen t = 5.58, Cohen's d = .167 for 1st graders and t = 9.83, Cohen's d = .257 for 2nd graders). The difference between *Color* and *Control* was not significant, t < 1.

Discussion

For the most part, the main results of Experiment 1 were replicated in Experiment 2. The color alternation condition yielded similar reading times as normal presentation in every measure, and interactions involving color did not come close to significance. We conclude that at least in normal sentence reading color alternation does not provide a sufficiently strong cue to affect polysyllabic word recognition.

In contrast, and similarly to Experiment 1, hyphenation disrupted polysyllabic word processing in all measures. In addition, we found that 2nd graders were disrupted by hyphenation to a larger degree than 1st graders. In gaze duration the hyphenation effect was present only for 2nd graders; in go-past and sentence reading time it was larger for 2nd than 1st graders. If habituation to hyphenation would have been an issue, 2nd graders should have been less disturbed by hyphens than 1st graders, as they have been exposed to hyphenated textbooks at school for a longer time. A similar pattern as for grade was found for reading proficiency, with stronger disruptive effects of hyphenation obtained for more proficient readers.

Regarding the other predictor variables, increase in word length was associated with longer gaze durations, the number of syllables with both increased gaze durations and go-past times, late acquired words with longer gaze durations and go-past times than early acquired words, and familiar words with shorter go-past times than less familiar words. These effects are similar to the ones obtained in Experiment 1 and in line with earlier research (see reviews of Blythe & Joseph, 2011; Rayner, 1998, 2009). Unlike in Experiment 1, there was no effect of bigram trough. Finally, as expected, readers were overall faster during the 2nd testing which took place approx. 4 months later than the 1st testing. The improvement in reading skill was considerable: First graders read on average 60 words per minute (wpm) in the fall and 85 wpm in the spring term; 2nd graders went up from 85 wpm in the fall to 107 wpm in the spring.

With regard to oral vs. silent reading, we witnessed the usual finding (for a review, see Rayner, 1998) that oral reading is slower than silent reading. However, reading mode did not modify the impact of syllable boundary cues. Against our expectations, syllable boundary cues did not facilitate oral reading, even though it can be hypothesized that syllables – by virtue of having strong phonological representations – play a more prominent role in oral than in silent reading. However, we noted that among beginning readers oral and silent reading may not qualitatively differ from each other. Even when asked to read silently, they often mouthed words and were even whispering what they were reading. This was quite pronounced for 1st graders but also apparent for 2nd graders. Thus, even though syllables are used as word recognition units, reading mode does not modify their role and the hyphen is equally disruptive in both oral and silent reading.

General discussion

The present study demonstrated that hyphenation disrupts 1st graders more when hyphens appear within a syllable (t-alo) than when they appear at a syllable boundary (ta-lo). We took this as evidence that Finnish 1st graders do not recognize words solely on a letter-by-letter basis anymore, but that they make use of syllabic units in written word recognition and hence have moved past the alphabetic phase (see Ehri, 1995). However, the results indicate that this is only part of the story, as hyphens at syllable boundaries clearly disrupt polysyllabic word processing of beginning Finnish readers. Most strikingly, the disruptive hyphenation effect is already found in the very early stages of reading instruction, in the beginning of the 1st grade.

As we argued in Häikiö et al. (2011), the hyphen is a visually very salient cue that clearly divides the word into separate parts. It is likely that in hyphenated polysyllabic words visual attention is initially mainly allocated to the first syllable demarcated from the rest of the word by a salient visual cue and attentional allocation to the subsequent syllables is therefore delayed. In addition, as the visual acuity decreases the further the letters are away from the fixation point, the more the final letters of the word are visually degraded. This problem is magnified the longer the word is and the more hyphens the word includes, as is the case with the trisyllabic and quadrosyllabic words used in this study. The interactions between hyphen and word length in go-past time in Experiment 1 and 2 of the present study and in Häikiö et al. (submitted) indeed reflect that hyphenation becomes more problematic with increased word length. However, the visual acuity account does not explain the sizeable effect in go-past times. Moreover, it does not explain why the disruptive effect of hyphenation is larger for 2nd graders than 1st graders, even though the perceptual span gradually develops during the elementary school years (Häikiö et al., 2009). It thus seems that hyphens encourage children to process syllables within words serially even when they are capable of processing more than one syllable at a time.

As argued in the Introduction, recognition of polysyllabic words in Finnish may take different shapes. First, it is possible that children recognize words via syllables by simultaneously accessing more than one syllable. The number-of-syllable effect obtained in Experiment 1 and 2 supports this interpretation. However, we also found an AoA and familiarity effect, which point to holistic processing of polysyllabic words.

In Figure 1 we adapted the model of Grainger and Ziegler (2011) to describe how the development of reading a polysyllabic word like *talo* (= house) may take place. It is clear that at the beginning of reading development children start out using Route 1 and read words via serial letter identification and phonological recoding (see Ziegler et al., 2012, for experimental evidence), a stage called the alphabetic stage by Ehri (1995). As reading skill develops, words may start to be recognized via the so-called fine-grained route (Route 2) whereby readers make use of frequently co-occurring letter clusters like syllables. We suspect that initially this route may be phonologically mediated, so the syllables *ta*- and *–lo* may address the phonological representation /talo/, before the orthographic representation and the meaning of the word are retrieved. However, eventually words may be processed entirely via the orthographic routes. Also, as argued above, syllables may be addressed sequentially or simultaneously, with the latter option related to a more advanced stage in reading development. In this more advanced stage, hyphens at constituent boundaries will become disruptive.

At the most advanced stage of reading development, orthographic representations may be accessed via the holistic or coarse-grained route without phonological involvement. At this stage letters would be recognized in parallel and directly mapped onto an orthographic representation (Route 3b) or mediated – as Grainger and Ziegler (2011) suggest – via open bigrams (combinations of letters that are not necessarily contingent).[5] Routes may operate in parallel and the balance between the routes may depend on factors such as word frequency. We argue that hyphenation is disruptive for either route, as it narrows down the initial focus on the first syllable, and thereby eliminates a number of open bigrams and/or the possible direct mapping of the whole stimulus onto the whole-word orthographic representation.

(Insert Figure 1 about here.)

The present results suggest that Finnish 1st and 2nd graders recognize words mainly via the fine-grained route and the more direct routes (Routes 3a and 3b). Less proficient readers make more use of the fine-grained route than more proficient readers, and – given that they are less disturbed by hyphenation – they may sometimes access syllables serially. It is nevertheless slightly surprising that the balance is already so much shifted towards the more direct routes for our early 1st graders, as they have been exposed to formal reading instruction for a few months only. However, it should be noted that the majority of the children were able to read to some extent already when entering the elementary school. Moreover, the shallow orthography and the clear syllabic system in Finnish allows for quick reading development (see Seymour et al., 2003). Prolonged use of hyphens at syllable boundaries in

Finnish ABC books may therefore well disturb or delay reading development rather than facilitate it.

Footnotes

1. In Finnish, a syllable boundary is always 1) before the last letter of a consonant cluster, or 2) between two different vowels that do not constitute a diphthong.
2. It can be argued that faster reading does not necessarily correlate with more proficient reading. However, in the Häikiö et al. (2011) study as well as – to foreshadow the present results – in the present study, reading skill (measured by a standardized reading test) was connected to faster reading speed.
3. We acknowledge that a newspaper corpus is not ideal for indexing word frequencies for children. However, as we had no access to a children’s corpus, we used it as an approximation. Moreover, we collected AoA-ratings to make sure that all the target words were familiar to the participants.
4. A bigram trough (Seidenberg, 1987; Seidenberg & McClelland, 1989) is a letter pair of lower frequency than the preceding and following bigram. We assessed the effect of bigram trough and its interaction with explicit syllable boundary cues. To this end, we constructed a trough ratio using the formula { [(preceding bigram frequency + following bigram frequency) / 2] / bigram frequency at first syllable boundary }. Thus, words without a trough had a ratio below 1 and words with a trough a ratio above 1.
5. Open bigram coding is not the only possible mechanism for extracting the relative position of letters in words (for an alternative coding scheme, see Gomez, Ratcliff, & Perea, 2008).

References

Ashby, J. (2006). Prosody in skilled silent reading: Evidence from eye movements. *Journal of Research in Reading*, 29, 318-333. doi: 10.1111/j.1467-9817.2006.00311.x

Ashby, J. (2010). Phonology is fundamental in skilled reading: Evidence from ERPs. *Psychonomic Bulletin & Review*, 17, 95-100. doi: 10.3758/PBR.17.1.95

Ashby, J., & Martin, A.E. (2008). Prosodic phonological representations early in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 224–236 doi: 10.1037/0096-1523.34.1.224

Blythe, H.L., & Joseph, H.S.S.L. (2011). Children's eye movements during reading. In S.P. Liversedge, I.D. Gilchrist & S. Everling (Eds.), *Oxford handbook on eye movements* (pp. 643-662). Oxford: Oxford University Press.

Carreiras, M., Vergara, M., & Barber, H. (2005). Early event-related potential effects of syllabic processing during visual word recognition. *Journal of Cognitive Neuroscience*, 17, 1803-1817. doi: 10.1162/089892905774589217

Chetail, F. (2014). Effect of number of syllables in visual word recognition: new insights from the lexical decision task. *Language, Cognition and Neuroscience*. doi: 10.1080/23273798.2013.876504

Chetail, F. & Mathey, S. (2009). The syllable frequency effect in visual recognition of French words: a study in skilled and beginning readers. *Reading and Writing*, 22, 955-973. doi: 10.1007/s11145-008-9135-9

Colé, P., Magnan, A., & Grainger, J. (1999). Syllable-sized units in visual word recognition: Evidence from skilled and beginning readers of French. *Applied Psycholinguistics*, 20, 507–532. doi: 10.1017/S0142716499004038

Ehri, L.C. (1995). Phases of development in learning to read by sight. *Journal of Research in Reading*, 18, 116-125. doi: 10.1111/j.1467-9817.1995.tb00077.x

Ehri, L.C., & McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers. *Reading and Writing Quarterly*, 14, 135-163. doi: 10.1080/1057356980140202

Ferrand, L., & New, B. (2003). Syllabic length effects in visual word recognition and naming. *Acta Psychologica*, 113, 167-183. 10.1016/S0001-6918(03)00031-3

Ferrand, L., Segui, J., & Humphreys, G.W. (1997). The syllable's role in word naming. *Memory & Cognition*, 25, 458-470. doi: 10.3758/BF03201122

Fitzsimmons, G., & Drieghe, D. (2011). The influence of number of syllables on word skipping during reading. *Psychonomic Bulletin & Review*, 18, 736-741. doi: 10.3758/s13423-011-0105-x

Gomez, P., Ratcliff, R., & Perea M. (2008). The Overlap Model: A Model of Letter Position Coding. *Psychological Review*, 115, 577-600. doi: 10.1037/a0012667

González, J., & Valle, I. (2000). Word identification and reading disorders in the Spanish language, *Journal of Learning Disabilities*, 33, 44–60. doi: 10.1177/002221940003300108

Grainger, J., Lété, B, Bertand, D., Sufau, S., & Ziegler, J.C. (2012). Evidence for multiple routes in learning to read. *Cognition, 123*, 280-292. doi: 10.1016/j.cognition.2012.01.003

Grainger, J., & Ziegler, J.C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, 2:54. doi: 10.3389/fpsyg.2011.00054

Häikiö, T., Bertram, R., & Hyönä, J. (2011). The development of whole-word representations in compound word processing: Evidence from eye fixation patterns of elementary school children. *Applied Psycholinguistics*, 32, 533-551. doi: 10.1017/S0142716411000208

Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology*, *102*, 167-181.

Häikiö, T., Bertram, R., & Hyönä, J. (submitted). The role of syllables in word recognition among 2nd grade children: An eye movement study.

Hautala, J., Aro, M., Eklund, K., Lerkkanen, M.-K., & Lyytinen, H. (2012). The role of letters and syllables in typical and dysfluent reading in a transparent orthography. *Reading and Writing: An Interdisciplinary Journal.* doi: 10.1007/s11145-012-9394-3

Karlsson, F. (1999). *Finnish: An essential grammar*. London and New York: Routledge.

Laine, M., & Virtanen, P. (1999). *WordMill lexical search program*. Turku, Finland: University of Turku, Center for Cognitive Neuroscience.

Lindeman, J. (1998). *ALLU: Ala-asteen Lukutesti [ALLU: Reading test for elementary school]*. Turku, Finland: University of Turku, Center for Learning Research.

Macizo, P., & Van Petten, C. (2007). Syllable frequency in lexical decision and naming of English words. *Reading and Writing*, 20, 295-331. 10.1007/s11145-006-9032-z

Maïonchi-Pino, N., Magnan, A., & Écalle, J. (2010). Syllable frequency effects in visual word recognition: Developmental approach in French children. *Journal of Applied Developmental Psychology*, 31, 70-82. doi: 10.1016/j.appdev.2009.08.003

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. doi: 10.1037/0096-1523.17.4.902

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422. doi: 10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62, 1457-1506. doi: 10.1080/17470210902816461

R Development Core Team (2007). *R: A language and environment for statistical computing*. Vienna, Austria; http://www.R-project.org

Rouibah, A. & Taft, M. (2001). The role of syllabic structure in French visual word recognition. *Memory & Cognition*, 29, 373-381. doi: 10.3758/BF03194932

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* (pp. 245-263). Hillsdale, NJ: Erlbaum.

Seidenberg, M., & McClelland, J. L. (1989). A distributed, developmental model of visual word recognition and naming. *Psychological Review*, 96, 523-568. doi: 10.1037/0033-295X.96.4.523

Seymour, P.H.K., Aro, M., & Erskine, J.M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143-174. doi: 10.1348/000712603321661859

Schiller, N.O. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 512–528. doi: 10.1037/0278-7393.26.2.512

Yap. M.J., & Balota, D.A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, 60, 502-529. doi: 10.1016/j.jml.2009.02.001

Ziegler, J.C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3-29. doi: 0.1037/0033-2909.131.1.3

Table 1. Lexical Statistics of the Target Words and Sentences.

[a] Per million; [b] Per thousand; [c] On a scale of 1-7; [d] On a scale of 1-5

Measure	Mean	SD	Min.	Max.
Frequency[a]	83	148	0	929
Length in characters	6.31	1.47	4	10
Length in syllables	2.60	0.61	2	4
Bigram frequency[b]	7.81	3.18	0.84	14.00
Initial trigram frequency[b]	0.69	0.74	0.01	6.15
Family size	105	124	3	649
Age of acquisition[c]	2.47	0.57	1.30	3.95
Familiarity[d]	3.11	0.77	1.76	4.76
Sentence length in characters (excluding hyphens)	38.98	5.73	29	54
Sentence length in words	5.74	0.70	5	7

Table 2. An example sentence triplet. Target words are presented here in bold for illustrative purposes.

Condition	Finnish Sentence	Translation
Control	Äiti kertoi, että **ikkuna** oli likainen.	Mother said that the **window** was dirty.
Color-legal	Äiti kertoi, että **aurinko** oli kirkas.	Mother said that the **sun** was bright.
Color-illegal	Äiti kertoi, että **aurinko** oli kirkas.	Mother said that the **sun** was bright.
Hyphenated-legal	Äi-ti ker-toi, et-tä **naa-pu-ri** o-li vi-hai-nen.	Mother said that the **neighbor** was angry.
Hyphenated-illegal	Äi-ti ker-toi, et-tä **na-apu-ri** o-li vi-hai-nen.	Mother said that the **neighbor** was angry.

Table 3. Descriptive statistics of ALLU reading performance in Experiment 1.

	Mean (SD)	Min	Max
Raw score[a]	35.6 (12.3)	19	58
Reading level[b]	4.8 (1.4)	3	7

[a] Maximum of 80; [b] On scale 1-9.

Table 4. Means for Each Eye Movement Measure (in ms) in Experiment 1, as a Function of Word Presentation Style and Syllable Cue Legality. Standard Deviations Are in Parentheses.

	Control	Color		Hyphenated	
		Legal	Illegal	Legal	Illegal
Gaze Duration	1284 (1005)	1248 (883)	1268 (943)	1313 (1087)	1606 (1213)
Go Past Time	1433 (1036)	1442 (1032)	1464 (1099)	1616 (1262)	2100 (1581)
Sentence Reading Time	6630 (3570)	6948 (3768)	-	7637 (3961)	-

Note. For sentence reading time only legally cued sentences were included.

Table 5. Descriptive statistics of ALLU reading performance in Experiment 2.

	1st grade			2nd grade		
	Mean (SD)	Min	Max	Mean (SD)	Min	Max
Raw score[a]	36.2 (13.0)	19	66	7.8 (2.5)	4	13
Reading level[b]	4.7 (1.6)	3	8	2.8 (1.6)	1	6

[a] Maximum of 20 for 1st grade, 80 for 2nd grade; [b] On scale 1-9.

Table 6. Means for Each Eye Movement Measure (in ms) in Experiment 2, as a Function of Word Presentation Style and Grade. Standard Deviations Are in Parentheses.

	Control		Color		Hyphenated	
	1st grade	2nd grade	1st grade	2nd grade	1st grade	2nd grade
Gaze Duration	1305 (1229)	939 (769)	1326 (1200)	931 (742)	1337 (1170)	1012 (786)
Go Past Time	1461 (1324)	1033 (880)	1496 (1335)	1043 (860)	1579 (1388)	1177 (929)
Sentence Reading Time	6647 (4086)	4698 (2551)	6778 (4238)	4732 (2479)	7440 (4439)	5394 (2763)

Figure caption

Figure 1. Model of Grainger and Ziegler (2011) adapted to elucidate the reading development of Finnish children when reading polysyllabic words like talo (= house).

Appendix 1

Final models for each measure in Experiment 1. *T* values below -1.96 and above 1.96 correspond to *p* values below 0.05. If an interaction was significant, its main effects are reported as well. Please note that these effects are not independently interpretable in the lmer() output. We separately present the model with *Legal* in go-past time since it was the only measure where we got a significant interaction.

(Insert Tables A1-A4 about here.)

Appendix 2

Final models for each measure in Experiment 2. *T* values below -1.96 and above 1.96 correspond to *p* values below 0.05. If an interaction was significant, its main effects are reported as well. Please note that these effects are not independently interpretable in the lmer() output.

(Insert Tables A5-A7 about here.)

Table A1. Experiment 1, Gaze Duration: Fixed Effects

	Estimate	SE	*t* value
(Intercept)	7.805974	0.290164	26.902
Color	0.001357	0.053361	0.025
Hyphen	0.110735	0.05384	2.057
Trough	-0.023036	0.011	-2.094
Fmlrty	-0.073988	0.033092	-2.236
Len	0.121075	0.017961	6.741
Allu	-0.025279	0.007571	-3.339

Note: Gaze duration values have been log-transformed.

Table A2. Experiment 1, Go-past Time: Fixed Effects

	Estimate	SE	*t* value
(Intercept)	7.874534	0.289333	27.216
Color	0.027308	0.042896	0.637
Hyphen	0.248257	0.043273	5.737
Len	0.080213	0.023696	3.385
Allu	-0.023364	0.007501	-3.115
AoA	0.134561	0.041154	3.270
Syl.res	0.192788	0.074844	2.576
Trough	-0.023552	0.010120	-2.327
Color:Len	0.031835	0.029758	1.070
Hyphen:Len	0.077072	0.029951	2.573

Note: Go-past time values have been log-transformed.

Table A3. Experiment 1, Go-past Time with Legality Included: Fixed Effects

	Estimate	SE	*t* value
(Intercept)	7.862555	0.298296	26.358
Hyphen	-0.123957	0.135828	-0.913
Legal	-0.004445	0.060218	-0.074
Syl.res	0.179259	0.088654	2.022
Len	0.134976	0.019319	6.987
AoA	0.117217	0.048151	2.434
Trough	-0.025479	0.011954	-2.131
Allu	-0.021953	0.007382	-2.974
Hyphen:Legal	0.231410	0.085697	2.700

Note: Go-past time values have been log-transformed.

Table A4. Experiment 1, Sentence Reading Time: Fixed Effects

	Estimate	SE	*t* value
(Intercept)	9.666621	0.285504	33.86
Color	0.049952	0.026349	1.90
Hyphen	0.167758	0.026552	6.32
Allu	-0.028639	0.007588	-3.77

Note: Sentence reading time values have been log-transformed.

Table A5. Experiment 2, Gaze Duration: Fixed Effects

	Estimate	SE	*t* value
(Intercept)	8.464647	0.218885	38.67
Color	-0.036877	0.062597	-0.59
Hyphen	-0.159727	0.062633	-2.55
Allu	-0.285896	0.039770	-7.19
Grade	-0.769127	0.139736	-5.50
Syl.res	0.120882	0.050117	2.41
Mode	-0.132324	0.013760	-9.62
Freq	-0.025361	0.010505	-2.41
Xp	-0.282228	0.013763	-20.51
Len	0.106148	0.010851	9.78
AoA	0.139794	0.027706	5.05
Color:Allu	0.012766	0.011374	1.12
Hyphen:Allu	0.041675	0.011395	3.66
Color:Grade	-0.003973	0.039809	-0.10
Hyphen:Grade	0.126237	0.039814	3.17

Note: Gaze duration values have been log-transformed.

Table A6. Experiment 2, Go-past Time: Fixed Effects

	Estimate	SE	*t* value
(Intercept)	8.709415	0.232893	37.40
Color	-0.006386	0.055403	-0.12
Hyphen	-0.061819	0.055481	-1.11
Len	0.116766	0.012954	9.01
Allu	-0.305500	0.042324	-7.22
Grade	-0.854613	0.148718	-5.75
Mode	-0.162171	0.012188	-13.31
Syl.res	0.186051	0.052251	3.56
AoA	0.112354	0.030413	3.69
Xp	-0.293758	0.012190	-24.10
Fmlrty	-0.045520	0.022392	-2.03
Color:Len	0.005947	0.010305	0.58
Hyphen:Len	0.027136	0.010327	2.63
Color:Allu	0.007378	0.010070	0.73
Hyphen:Allu	0.030909	0.010095	3.06
Color:Grade	-0.000362	0.035252	-0.01
Hyphen:Grade	0.130117	0.035266	3.69

Note: Go-past time values have been log-transformed.

Table A7. Experiment 2, Sentence Reading Time: Fixed Effects

	Estimate	SE	*t* value
(Intercept)	8.46032	0.22276	37.98
Color	-0.0645	0.06585	-0.98
Hyphen	-0.15204	0.06596	-2.3
Allu	-0.28519	0.04027	-7.08
Grade	-0.76567	0.14147	-5.41
Xp	-0.28653	0.01444	-19.85
Mode	-0.1285	0.01443	-8.9
Color:Allu	0.01482	0.01198	1.24
Hyphen:Allu	0.03992	0.012	3.33
Color:Grade	0.02344	0.04182	0.56
Hyphen:Grade	0.12345	0.04183	2.95

Note: Sentence reading time values have been log-transformed.