The effect of syllable-level hyphenation on reading comprehension: Evidence from eye movements

Tuomo Häikiö[1], Timo T. Heikkilä[1], & Johanna K. Kaakinen[1,2]

[1] Department of Psychology, University of Turku, Finland

[2] Turku Institute for Advanced Studies, University of Turku, Finland

Author note

Portions of the data have been presented at Scandinavian Workshop on Applied Eye Tracking, Turku, Finland, June 19-21, 2016; Annual Conference of Society for Text and Discourse, Kassel, Germany, July 18-20, 2016; and Developmental Symposium, Southampton, UK, August, 17-19, 2016. The dataset can be downloaded at https://osf.io/6px2d. 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; E-mail: tuilha@utu.fi

Abstract

Syllabification by hyphens (e.g., hy-phen-a-tion) is a standard procedure in early Finnish reading instruction. However, recent findings indicate that hyphenation slows down children's reading already during the first grade (Häikiö et al., 2015, 2016). In the present study, it was examined whether this slowdown is indicative of deeper processing and/or more strategic reading. To this end, 36 second-grade children read short expository and narrative stories while their eye movements were registered. The presence of Syllable Boundary Cue (SBC) was manipulated; for half of the stories, each word was hyphenated at syllable boundaries whereas the other half included no hyphenation. After each story, story comprehension (SC) was measured by three types of oral questions, namely Free Recall, Cued Recall, and True/False Questions. With regard to reading behavior, SBC interacted with independently measured reading comprehension scores for both Forward and Regressive Fixation Times during first pass sentence reading. Hyphenation slowed down reading of good comprehenders to a larger extent than weaker comprehenders in comparison to non-hyphenated condition, especially for Regressive Fixation Times. With respect to SC, Cued Recall scores were lower in the hyphenated than non-hyphenated condition. There was no effect of SBC in Free Recall or True/False Questions. Hyphenation seems to promote phonological encoding even when readers might want to access words via orthographic codes, which are obscured by hyphenation, especially at the whole-word level. This more piecemeal reading style then makes it harder to integrate the pieces into a bigger whole, affecting not only reading speed but also reading comprehension.

Keywords: Reading, Reading comprehension, Syllables, Hyphenation, Eye movements

Educational Impact And Implications Statement

Syllable structure is routinely marked with hyphens (e.g., ‘syl-la-ble’) in early Finnish reading instruction. The present study suggests that hyphens slow down 2nd grade children’s reading and hinder their reading comprehension. These findings imply that the use of hyphenation in Finnish reading instruction needs to be reconsidered, especially as a one-style-fits-all approach.

Introduction

Several theories of reading development suggest that early readers utilize small, letter-size units in word encoding before moving to larger chunks as their reading skills develop (e.g., Ehri, 1995; 2005; Ehri & McCormick, 1998; Frith, 1985). For instance, according to the phase theory developed by Ehri (1995, 2005), once children have moved past the pre-alphabetic phase (during which children engage in a type of sight reading based on company logos etc.), they start to form connections between written letters (graphemes) and spoken sounds (phonemes) in the alphabetic phase.[1] Thanks to these grapheme/phoneme connections, children are then able to read even words they have not encountered before by decomposing letter strings into individual graphemes and sounding out the phonemes. When developing readers are able to utilize most or all of the grapheme/phoneme connections, they can start to use recurring letter patterns such as syllables and morphemes in the consolidated-alphabetic phase. During this phase, children do not need to divide every word into smallest possible units anymore; instead, a word such as *remember* can be decoded via three syllables instead of eight letters. It needs to be noted that this does not mean that more proficient readers do not use grapheme-level decoding anymore. The possibility of utilizing both individual letters and letter chunks such as syllables makes reading more flexible as one can rely on the most suitable mode of processing. In the final phase, proficient readers are even able to process some words as wholes (see also Grainger, Lété, Bertand, Dufau, & Ziegler, 2012; Grainger & Ziegler, 2011).

As mentioned above, syllables can be used as processing units in written word recognition. This has been shown for both adults and children across many languages (e.g., Ashby, 2010; Ashby & Martin, 2008; Chetail, 2014; Chetail & Mathey, 2008; Chetail & Mathey, 2012; Colé, Magnan, & Grainger, 1999; González & Valle, 2000; Häikiö, Hyönä, & Bertram, 2015; Hasenäcker & Schroeder, 2017; Hautala, Aro, Eklund, Lerkkanen, & Lyytinen, 2012; Macizo & Van Petten, 2007, Maïonchi-Pino, Magnan, & Écalle, 2010). This is also true for Finnish (Häikiö et al., 2015; Hautala et al., 2012), a highly transparent language with simple syllable structure. In principle, there are only two rules concerning the syllabification in Finnish; a syllable boundary is always located a) before the last consonant of a consonant cluster (or the only consonant), and b) between two different vowels that do not form a diphthong. Furthermore, the number of syllables is relatively low (ca. 3,000 syllables), and there are no heavy consonant clusters (Seymour, Aro, & Erskine, 2003). Because of these characteristics, Finnish children learn to read fluently earlier than in most other European languages (Seymour et al. 2003). Already 1st graders utilize syllable-level information in reading Finnish (Häikiö et al., 2015).

Because syllables are such prominent processing units in Finnish language, it comes as no surprise that syllabification is also heavily emphasized in Finnish reading instruction, especially in the most widely used method, KÄTS (Karppi, 1983). For instance, in the beginning of reading instruction, all words with two or more syllables are hyphenated at syllable boundaries (e.g., kis-sa; cat). The hyphens are then gradually removed but they are preserved in longer and more difficult, new words until the end of the 2nd grade. This way, early reading instruction is based on decomposition of words at syllable level, activation of phonological codes and assembling the whole-word phonology from these pieces. The word meaning is then retrieved via its phonological representation. Because children are exposed to hyphenation to a large extent in their early reading development, one might assume that they are good at dealing with hyphenated words. However, it has been shown that hyphenation in fact slows down reading in comparison to a non-hyphenated condition (Häikiö et al., 2015; Häikiö, Bertram, & Hyönä, 2016).

The effect of hyphenation on reading speed has been demonstrated in a series of eye tracking studies conducted by Häikiö et al. (2015, 2016). In the experiments, 1st and 2nd grade children read single sentences for comprehension while their eye movements were recorded.

In each of these experiments, there was a hyphenated condition, in which all of the words in the sentence were hyphenated at the syllable boundaries, as well as control condition, in which there were no hyphens. It was shown that hyphenation slowed down both word and sentence reading. For 1st graders, this effect was not readily observed during the first encounter of the word apart from longer 3-4-syllable words (Häikiö et al., 2016). However, there was a delayed effect for all words in go-past time (Häikiö et al., 2015), a measure that includes all of the fixations starting from the very first fixation on the word and ending when the word is exited to the right for the first time. In other words, this measure also takes into account regressive fixations and first-pass re-reading. For 2nd graders, the slowdown was observed already during the first encounter of the word (Häikiö et al., 2015, 2016). Furthermore, the effect was more pronounced when subsequent re-reading was taken into account (i.e., go-past time). Finally, hyphenation slowed down both 1st and 2nd graders on sentence level (Häikiö et al., 2015). Taken together, the authors took these findings as evidence for hyphenation directing attention to smaller processing units than the readers would prefer. Instead of being able to access the word meanings via whole word representations, the readers were forced to utilize more fine-grained processing units (i.e., syllables), which are processed serially instead of faster, parallel activation of more coarse-grained units, such as whole words (see also Grainger et al., 2012; Grainger & Ziegler, 2011). The more pronounced effects in go-past time were interpreted as hyphenation making integration of the words into the sentence context more difficult, thus causing readers to make more regressions.

However, there is also an alternative interpretation for these latter findings. It has been established that 10-15% of fixations adults make go back in text (see Rayner, 1998, 2009, for a review). For children, this can be up to 30% of fixations (Blythe & Joseph, 2011). Furthermore, it has been shown that regressions are partly used to correct erroneous interpretations (Bicknell & Levy, 2011). Therefore, they may in fact lead to better reading comprehension. This was nicely demonstrated by Schotter, Tran, & Rayner (2014) who conducted a trailing-mask experiment, in which their adult participants had to read sentences for comprehension while their eye movements were recorded. In the experimental condition, once the participant had exited the word, it was masked with a string of X's, making rereading of any word impossible. It was shown that when rereading was denied, the participants answered the comprehension questions with a lower accuracy than when rereading was possible. This was true for both ambiguous and unambiguous sentences. Therefore, an alternative interpretation to the results of the Häikiö et al. studies (2015, 2016) is that while hyphenation does lead to slowdown, the extra time spent on text due to the increased number of regressions increases comprehension. This would go well in line with the finding that readers with good comprehension skills do reread specific parts of the text (van den Broek, White, Kendeou, & Carlson, 2009). This may also be connected to the level of strategic reading.

It has been established that slow linear adult readers that do not go back to earlier parts of the text have lower comprehension scores than topic structure processors, who engage in strategic rereading (Hyönä, Lorch, & Kaakinen, 2002). The amount of time spent looking back in the text is also positively correlated with the level of recalling the main points of the text (Hyönä & Nurminen, 2006). The finding that the more strategic adult readers have better comprehension scores (Hyönä et al., 2002) is in line with the finding that good 9-10-year-old comprehenders were better at detecting inconsistencies than poor comprehenders (Oakhill, Hartt, & Samols, 2005). Interestingly, poor comprehenders did not have particular problems with the task when the inconsistencies were between two adjacent sentences. Oakhill et al. hypothesized that this may have been influenced by less stable mental model of the text as a whole. Similarly, van der Schoot, Reijntjes, and van Lieshout (2011) showed that poor 10-12-year-old readers did not generally notice global inconsistency in a story whereas good readers did. On sentence level, it has been shown that 5th grade children react to a more difficult sentence with heightened rereading times, indicative of active comprehension monitoring (Vorstius, Radach, Mayer, & Lonigan, 2013). However, this does not necessarily lead to a correct answer on whether the sentence in question seems generally fine (i.e., plausible).

To our knowledge, the effect of hyphenation on reading comprehension has not been examined before. Even though in the studies of Häikiö et al. (2015, 2016), participants had to read sentences for comprehension, said sentences were presented one at a time, and they did not form a coherent text. When assessing reading comprehension, longer texts alongside sentence-level measures (as opposed to word-level measures in single sentence reading) should be used (see Hyönä, Lorch, & Rinck, 2003). To this end, in the present study, 2nd grade children were given longer narrative and expository texts while their eye movements were registered. The stories were presented either in hyphenated or non-hyphenated (i.e., control) format. After each story, the story comprehension was assessed by different types of comprehension questions. With regard to eye movement data, we hypothesized that the participants will be slowed down by the hyphenation, and that this effect is more pronounced in regressive fixations. With regard to story comprehension data, there were two possible outcomes. If, on the one hand, longer re-reading times (as indexed by regressive fixations) facilitate comprehension and/or are an index of a more strategic reading, we should witness higher comprehension scores in the hyphenated condition when compared with the control condition. If, on the other hand, the regressive fixations are indicative of integration problems, we should see lower comprehension scores in the hyphenated condition.

Method

Participants

Thirty-six monolingual 2nd graders (13 male, 23 female, on average 8:4 years, range 7:10-8:10) were recruited from two class rooms in a Finnish elementary school. At the time of testing (October/November) they had received approx. 1 year 2 months of formal reading instruction. All participants had normal or corrected-to-normal vision. The school authorities granted permission for the study to take place. Children's parents signed a written informed consent form prior to the experiment. The participants received candy or stickers as a 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. A chin-and-forehead rest was used to minimize head movements. The texts were presented on a 24-inch BenQ XL2420Z computer screen (refresh rate of 100 Hz, resolution 1920*1080).

Materials

The experimental material consisted of eight animal stories written in Finnish. They were taken from two standardized reading/listening comprehension tests, Diagnostiset testit 2 (Vauras, Mäki, Dufva, & Hämäläinen, 1995) for 1st grade, and YTTE (Kajamies, Poskiparta, Annevirta, Dufva, & Vauras, 2003) for 2nd grade. Both of these tests have been designed to be used by teachers, special education teachers, and school psychologists to assess the level of reading and listening comprehension of early readers. For both of them, data from 210-240 children (depending on the text) has been acquired to provide comprehension level norms. Two of the stories were short enough to fit on one screen, and six of them were presented on two consecutive text screens. The title page was presented separately before each story but was not used in the analyses. Each text screen consisted of a maximum of eight lines that extended horizontally up to 131 characters in length. The sentences were presented in a proportional Calibri font with font size 20 and line spacing 3.5. With a viewing distance of 60 cm, one character space subtended from .16 to .56 degrees of visual angle. On the top and bottom of the screen there was a 100 pixel margin, and on the left and right side a 300 pixel margin. The stories consisted of 93 to 140 words. Four of the stories were aimed at the first-grade level, and another four at the second-grade level. Furthermore, four of the stories were simple stories with regard to their structure (theme – aim – means – goal), and another four were complex (theme – aim – means – goal – change of theme – aim – means – goal – cause). Finally, four of the stories were expository texts, and another four were narrative. Since there were three 2-level dimensions for these stories (i.e., grade level, complexity, and genre), they were balanced according to a 2x2x2 latin square. The stories were divided into two blocks so that both blocks included a first-grade expository text, a first-grade narrative text, a second-grade expository text, and a second-grade narrative text. The participants read the blocks in two separate sessions on different days.

Two versions of each story were created, one with hyphens at every syllable boundary, and one without explicit syllable boundary cues. In each block, two stories were always hyphenated and two were not. Half of the participants started with two hyphenated stories, followed by two non-hyphenated stories, and vice versa. Participants who first read hyphenated stories, started the second session with non-hyphenated stories, and vice versa. The order of the texts was fixed into four possible lists, the order of which was counterbalanced between the participants.

After each story, there were three types of questions that were presented orally; free recall, cued questions, and true/false questions, given in this order. For free recall, the children were asked to recall everything they remembered about the previous story. For cued questions, the children were asked open questions (5 or 6 questions per story) such as “How does the cat climb down the tree?” Finally, true/false questions (6 or 7 questions per story) were detailed claims such as “The cat climbs down the tree using its sharp claws”. Participants’ responses to the questions were recorded, and answers were later scored according to the test manual criteria (Kajamies et al., 2003; Vauras et al., 1995). Two independent raters scored the answers for 8 participants and 8 texts. The interrater agreement was 91.88% for the free recall, 88.13% for the cued recall, and 99.03% for the true/false question scores, respectively, and the rest of the answers were scored by only one rater.

After the eye movement experiment, the children's reading skill was assessed with two classroom subtests (decoding/technical reading skill, TL4; reading comprehension, LY) of ALLU (Lindeman, 1998), a standardized Finnish reading test for elementary-school children. Furthermore, the children completed the Word Chain test (Nevala & Lyytinen, 2000).[2] As for reliability, Kuder-Richardson's coefficient of internal consistency (KR20) was KR20 = .82 for TL4, and KR20 = .88 for LY (Lindeman, 1998). For the Word Chain test, the test-retest reliability was r = .72 (Nevala & Lyytinen, 2000). During the first classroom session, children completed Word Chain test and the first text of the reading comprehension test, and during another session the technical reading skill test and another text from the reading comprehension test.

In the test of technical reading skill, children saw 20 pictures of scenes (e.g., people sailing) and had to choose from four sentence alternatives which sentence corresponded to the picture (e.g., "They swim", "They dive", "They sail", and "They dance"). Children responded to as many picture-sentence combinations as possible within 120 seconds. The descriptive statistics are presented in Table 1.

In the reading comprehension test, children read a short expository text (first session) and a short narrative (second session) and had to answer 12 multiple-choice questions on the basis of both texts. They were allowed to have the texts next to the questions throughout the testing. They had approximately 25 minutes to complete each reading task. For the analyses, we combined the scores from both texts. The descriptive statistics are presented in Table 1.

In the Word Chain test, one must recognize and separate words in a string of letters (e.g., catcomputerprincessstoneheavy) by drawing a vertical line at word boundaries (cat|computer|princess|stone|heavy) as quickly as possible. There is a time limit of 85 seconds, and the test is scored by giving 1 point for each correctly placed vertical line. The descriptive statistics are presented in Table 1.

[Insert Table 1 about here.]

Procedure

In both eye-tracking sessions, the participants were instructed to read stories for comprehension at their own pace. They were encouraged to read silently. They were further told that after each story they would get oral questions about the story they just read. Furthermore, the participants were told that for some of the stories the words would be hyphenated at the syllable level, but that the other stories would be presented without hyphenation. The eye-tracker was calibrated using a nine-point calibration grid extending over the computer screen before each story. Calibration was deemed successful if the average error across the calibration points was <0.75 degrees. However, for 16 stories spread across 8 participants we used a less stringent calibration criteria (M=0.93 degrees, SD=0.14, range 0.78-1.28); however, as the visual inspection of the data indicated that the accuracy was satisfactory in order to differentiate eye fixations on different sentences, we kept these data in the analyses. Before presenting each text screen, the participant fixated on a calibration point at the left side of the screen, after which the text appeared. Before the experimental stories, the participants read one practice story and answered the questions. After the practice, the four experimental stories were presented, each followed by story comprehension questions.

Dependent variables and predictors

The eye fixation data was analyzed on the sentence level (i.e., each sentence was a separate unit of analysis). We used three dependent variables to assess the reading behavior; Forward Fixation Time (i.e., summed durations of progressive fixations before exiting the sentence for the first time) and Regressive Fixation Time (i.e., summed durations of regressive fixations before exiting the sentence for the first time) for first-pass reading, and Probability Of A Look-From (i.e., the probability of making a regression out of the sentence) for re-reading (see Hyönä et al., 2003).

For the story comprehension data, the dependent variables were the accuracy scores for Free Recall, Cued Questions, and True/False Questions. Since different texts had different maximum scores, all of the scores were transformed to percentages relative to the maximum score (see Table 2).

The critical predictor was Syllable Boundary Cue (SBC) with two levels, *Control* and *Hyphen*. Furthermore, we used two continuous, independent reading skill scores as predictors, namely *Comprehension* and *Decoding* (reading comprehension and technical reading skill tasks from ALLU, respectively). Finally, since the children completed the task in two sessions, *Session* number was used as a predictor.[3]

Statistical considerations

Even though the participants were able to read stories at their own speed, testing took place during the school days. Because of this, there was a natural time limit of 40 minutes (i.e., length of a lesson) per session. This lead to six participants not reading all of the stories due to slow reading pace. However, the data from the completed stories is included in the analyses. Data from one participant was excluded due to the task being too demanding and two participants were excluded as they did not understand the task (all female). Screens for which there was significant data loss either due to technical problems (shift in calibration) or disturbance during the course of reading were excluded; this lead to the exclusion of data from twelve screens. Finally, reading data from 113 sentences was excluded due to track loss. In total, there were 3064 sentences included in the reading analyses. For story comprehension analyses, we included data from all of the stories for which we had at least partial eye movement data. Not all of the questions could be scored due to noise in recordings but this was taken into consideration by transforming the scores to percentages from the theoretical maximum of the scored questions for each story. For Cued Recall, we had to exclude 4 data points for one participant due to technical problems (excessive noise in recordings). For True/False Questions we had to exclude one data point due to increased stress which led to the termination of the testing. In total, there were 245, 241, and 244 texts in story comprehension analyses for Free Recall, Cued Recall, and True/False Questions, respectively.

The duration measures were log-transformed to normalize the data. For Regressive Fixation Duration, all the zero values were excluded (6.4% of trials). Furthermore, values 2.5 SDs smaller or larger than the participant mean were excluded from the duration measures separately for both conditions. This lead to the exclusion of 1.6% of the remaining data for Forward Fixation Time, and 1.0% of the remaining data for Regressive Fixation Duration. Finally, the continuous predictor variables were centered.

We used multiple regression mixed-effects modelling with participants and items as crossed random effects for the duration and accuracy measures. For the probability measure, we used binomial generalized multiple regression mixed-effects modelling with participants and items as crossed random effects. 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, $|t$ and $z| > 1.96$. We are aware that this cutoff point may be considered anti-conservative (see Luke, 2017). To reduce the possibility of Type I error, we conducted the analyses for the duration and accuracy measures using Kenward-Roger approximation for degrees of freedom. We also made sure by model comparison that each predictor significantly improved the explanatory value of the model. The analyses were conducted using the lme4 package (version lme4_1.1.10; Bates, Maechler, Bolker, & Walker, 2015) for R statistical software (version 3.3.2; R Core Team, 2016). The final models are reported in Appendix 1. The 95% confidence intervals for the model estimates have been computed using Wald estimation.

Results

Reading data

The eye movement measures as a function of SBC are presented in Table 2. For Forward Fixation Time, SBC interacted with Comprehension, $B = .0068$, CI 95% [.0030; .0105], and Decoding, $B = -.011$, [-.0198; -.0029]. The nature of the significant interactions can be seen in the Figures. In each Figure, the weaker readers are on the left side and the better readers on the right side. The envelopes denote 95% confidence intervals. As can be seen in Figure 1, for the participants with weakest comprehension skills, there was no difference between hyphenated and non-hyphenated conditions. While better reading comprehension as such made Forward Fixation Times faster in both conditions, the disruptive effect of hyphenation became larger as a function of reading comprehension. On the other hand, with regard to Decoding, the weaker decoders were disrupted by hyphenation to a larger degree than the better decoders (see Figure 2). Finally, there was a main effect of Session, $B = -.0424$, [-.0565; -.0282]; Forward Fixation Times were shorter during the second session. The descriptive statistics by session are reported in Appendix 2.

For Regressive Fixation Time, there was an interaction between SBC and Comprehension, *B* = .0243, [.0099; .0388]. For better comprehenders, hyphenation was disruptive in comparison to the non-hyphenated condition, whereas for the weaker comprehenders, there was no such distinction (see Figure 3). Furthermore, there were main effects of Decoding, *B* = -.0165, [-.2505; -.0801] and Session, *B* = -.1372, [-.1926; -.0819]. Regressive Fixation Times were faster for better decoders, and during the second session.

[Insert Tables 2, and 3 about here.]

[Insert Figures 1, 2, and 3 about here.]

For the Probability Of A Look-From, the only predictor that survived in the model was Session, *B* = -.2580, [-.4751; -.0409]; the probability was lower in the second session.

Story comprehension data

The data from the story comprehension question accuracy as a function of SBC is presented in Table 3. There were no interactions for any of the story comprehension question types (Free Recall, Cued Questions, True/False Questions). However, there was a main effect of SBC in Cued Recall, *B* = -7.65, [-13.15; -2.16]; story comprehension was lower in the hyphenated condition. Finally, there was a main effect of Comprehension in every type of story comprehension question, *B* = 1.11, [.45; 1.77], *B* = 2.41, [1.32; 3.50], and *B* = 1.16, [.48; 1.84], for Free Recall, Cued Questions, and True/False Questions, respectively. Better comprehenders (according to an independent comprehension test) performed better in every type of story comprehension question than weaker comprehenders.

Discussion

In the present study, the effect of hyphenation at syllable level on reading comprehension (dubbed story comprehension in order to differentiate it from independently measured reading comprehension scores) of Finnish 2nd graders was examined. It was shown that after reading hyphenated stories, the accuracy of Cued Recall was lower than when the stories did not have hyphens. Furthermore, hyphenation slowed down first-pass reading at the sentence level. This disruption was modulated by the level of reading comprehension. Better comprehenders were more disrupted by hyphenation than less skilled comprehenders, and this effect was more pronounced in first-pass regressive fixations times. Finally, hyphenation did not have an effect on second-pass rereading; in fact, participants rarely engaged in second-pass rereading at all.

Previously, it has been established that hyphenation slows down reading at word level for 1st and 2nd graders (Häikiö et al., 2015, 2016). Häikiö et al. (2015) showed that this is also true at the sentence level when reading single sentences but here we extend these findings to sentences in longer texts. Furthermore, the slowdown is more pronounced for children with better reading comprehension. This is analogous to the earlier findings that the better decoders as well as older children are disrupted to a larger degree by hyphenation at word-level (Häikiö et al., 2015, 2016). The same goes for the more pronounced effect of hyphenation in regressive fixation times for better comprehenders; the more skilled young readers had relatively longer go-past times (a measure that includes both first-pass reading and regressive fixations) in the hyphenated condition in the Häikiö et al. (2015) study. Häikiö et al. (2015, 2016) interpreted these findings as reflecting problems in integrating the words into the sentence context as hyphenation directs attention to smaller processing units than preferred.[4]

In the present study, we examined an alternative hypothesis with regard these regressive fixations. Since re-reading and regressive fixations have been shown to reflect better comprehension (e.g., Hyönä et al., 2002; Hyönä & Nurminen, 2006; Schotter et al., 2014), the larger regressive fixation times in the hyphenated condition may therefore reflect deeper processing and/or more strategic reading. Nevertheless, our story comprehension data goes against this account; in Cued Recall there was a reliable negative effect of hyphenation. The recall accuracy was lower after reading hyphenated stories than non-hyphenated stories. Therefore, it seems that signaling the syllable structure by hyphenation is indeed disruptive to early readers, not only with regard to reading speed but story comprehension as well, even though this disruption was restricted to one type of story comprehension measure. In her textbook, Lerkkanen (2006) voiced a possible concern regarding this. Lerkkanen noted that while the KÄTS reading instruction method (Karppi, 1983), which focuses on the use of phonology (as enforced by hyphenation), may be suitable for children at risk in reading development, for normally developing children it may even delay the development of reading comprehension skills. Our findings show that this concern is valid. Hyphenation seems to promote phonological encoding even when readers might want to access words via orthographic (i.e., visual) codes, which are obscured by hyphenation, especially at the whole-word level. This more piecemeal reading style then makes it harder to integrate the pieces into a bigger whole, not only on sentence level but on a whole text level as well.

Interestingly, better decoders were disrupted less by hyphenation in forward fixation times. It has been established that children with better technical reading skills also read more (for a meta-analysis, see Mol & Bus, 2011). Therefore, they also have more experience in finding the phonological codes in the text. Our finding points to hyphenation facilitating this process. However, this facilitation lies on a quite superficial level. While it is beneficial to crack the phonological code quickly to free more resources for the integration, the actual integration is then disrupted by the continued exposure to the small phonological chunks (i.e., syllables separated by hyphenation).

As already discussed, hyphenation hampered story comprehension, as indexed by a lower accuracy in Cued Recall. We did not find such effects in Free Recall and True/False Questions. However, one needs to notice that on average, the performance was close to floor in Free Recall and close to ceiling in True/False Questions. Naturally, to detect possible effects of text presentation on comprehension, one needs to use the most sensitive measures possible.[5] The fact that we found the effect of hyphenation only in one out of three story comprehension measures demonstrates this point nicely. For instance, to recall every detail one remembers after reading a story seems pretty difficult to young readers, no matter whether the stories are hyphenated or not. This led to a floor effect in Free Recall. On the other hand, 2nd grade children who are able to read several stories at one go (as was the case with our participants) are relatively good decoders already. Therefore, they have relatively little problems answering true/false questions even when hyphenation makes the overall comprehension of the stories more difficult. Hyphenation does not seem to disrupt extracting details from the text, at least when children are required to identify whether a particular detail was present or not, as manifested in the ceiling effect in True/False Questions. This is in line with the finding of Vorstius et al. (2013) who showed that while 5th grade children are able to detect inconsistencies in sentences, this does not necessarily lead to a better accuracy in judging whether the sentence is plausible or not. Taken together, these findings nicely demonstrate that a mere detection of a detail does not lead to heightened comprehension. One also needs to be able to integrate these smaller details into a bigger whole, which in turn is more difficult when the attention is directed to smaller, more superficial parts of the text due to hyphenation.

We did not find an effect of hyphenation on the probability of making a regression to previous sentences. In fact, the regression probabilities to previous sentences were very low (around 13%), reflecting that our 2nd graders did not generally regress to earlier parts of the text. Instead, they tackled each sentence one at a time. Once the sentence was processed, they usually moved forward in the text to read the next sentence. This demonstrates a lack of strategic rereading for the 2nd graders, which goes well in line with the findings of Kaakinen, Lehtola, and Paattilammi (2015). They tested Finnish children and adults whose task was to read scientific texts taken from children's textbooks in order to answer "why" questions. Kaakinen et al. showed that 2nd graders spent more time inspecting sentences during first-pass reading in the question condition when compared to a "read for comprehension" condition whereas for the older children and adults, the question condition caused more second-pass re-inspections (i.e., regressions back to previous sentences). This suggests a shift towards more strategic reading somewhere between the 2nd and 4th grade. It is likely that 2nd grade Finnish children are still in the process of perfecting their decoding skills (however, see Seymour et al., 2003), and only after this has happened, they start to practice more demanding skillsets such as higher order comprehension strategies (see Perfetti, 2007). The mere presence of regressions during first-pass reading of a sentence without later re-inspections cannot be postulated to be an index of such strategies. It has been shown that children with poor text comprehension skills generally have no difficulties in detecting inconsistencies when they are within what one can consider a local area but this is not true when the inconsistencies span a larger distance in text (Oakhill et al., 2005; van der Schoot et al., 2011). For better comprehenders the pattern is different; they are able to detect such anomalies even when there is intervening text. Taken together, these findings suggest that a hallmark of strategic reading would be the ability to integrate information across a longer text, in which selective rereading of earlier parts of the text plays a big role.

In conclusion, hyphenation at syllable level disrupts both reading speed and reading comprehension. There may be a short phase when hyphenation facilitates reading, but this is likely to be at different times for different individuals, depending on when they shift from very fine-grained (phonological) decoding to a more coarse-grained processing (see Grainger &

Ziegler, 2011). Alternatively, hyphenation may facilitate reading of children at risk, as suggested by Lerkkanen (2006). Both of these possibilities remain to be examined. At any rate, we think that the use of hyphenation in Finnish reading instruction needs to be reconsidered, especially as a one-style-fits-all approach. One viable option would be to use electronic reading devices where hyphenation could be turned on or off by tapping the screen, both on the whole text level but also for individual words. This would give children the opportunity to use the help of hyphenation when needed while not forcing the others to use a less effective approach.

Footnotes

1. This phase can be divided into partial-alphabetic phase and full-alphabetic phase on basis of whether a reader does or does not master all of the grapheme/phoneme connections.
2. Since both technical reading skill and Word Chain test tap into decoding, we will only present models including technical reading skill. The models with Word Chain were very similar.
3. Since half of the stories were simple and half of them complex, we also run analyses that included *Complexity* as a predictor. Because it did not interact with SBC in any of the measures, we left it out of the final models.
4. It would have been interesting to look into the regressive behavior in the word-level data in the present study as well. Unfortunately, there was too much noise in the data for the word-level data to be reliable in this respect.
5. A more thorough discussion of the nature of reading comprehension tests and what cognitive skills they tap is beyond the scope of this paper. (However, see e.g., Cutting & Scarborough, 2006; Keenan, Betjemann, & Olson, 2008).

References

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67,* 1–48. doi:10.18637/jss.v067.i01

Bicknell, K., & Levy, R. (2011). Why readers regress to previous words: A statistical analysis. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 931–936). Austin, TX: Cognitive Science Society.

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.

Chetail, F. (2014). Effect of number of syllables in visual word recognition: New insights from the lexical decision task. *Language, Cognition and Neuroscience, 29*, 1249–1256. doi:10.1080/23273798.2013.876504

Chetail, F., & Mathey, S. (2008). 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

Chetail, F., & Mathey, S. (2012). Effect of syllable congruency in sixth graders in the lexical decision task with masked priming. *Scientific Studies of Reading, 16*, 537–549. doi:10.1080/10888438.2011.607480

Colé, P., Magnan, A., & Grainger, J. (1999). Syllable-sized units in visual word recognition: Evidence from skilled and beginning readers. *Applied Psycholinguistics*, *20*, 507–532. doi:10.1017/s0142716499004038

Connor, C. M., Day, S. L., Phillips, B., Sparapani, N., Ingebrand, S. W., McLean, L., …, Kaschak, M. P. (2016). Reciprocal effects of self-regulation, semantic knowledge, and reading comprehension in early elementary school. *Child Development, 87*, 1813–1824. doi:10.1111/cdev.12570

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277–299, doi:10.1207/s1532799xssr1003_5

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. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*, 167–188. doi:10.1207/s1532799xssr0902_4

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

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. Marshall & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 301–330). London, UK: Erlbaum.

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. (2016). The hyphen as a syllabification cue in reading bisyllabic and multisyllabic words among Finnish 1st and 2nd graders. *Reading and Writing, 29*, 159–182. doi:10.1007/s11145-015-9584-x

Häikiö, T., Hyönä, J., & Bertram, R. (2015). The role of syllables in word recognition among beginning Finnish readers: Evidence from eye movements during reading. *Journal of Cognitive Psychology*, *27,* 562–577. doi:10.1080/20445911.2014.982126

Hasenäcker, J., & Schroeder, S. (2017). Syllables and morphemes in German reading development: Evidence from second graders, fourth graders, and adults. *Applied Psycholinguistics*, *38,* 733–753. doi:10.1017/S0142716416000412

Hautala, J., Aro, M., Eklund, J., Lerkkanen, M.-K., & Lyytinen, H. (2012). The role of letters and syllables in typical and dysfluent reading in a transparent orthography. *Reading and Writing, 26*, 845–864. doi:10.1007/s11145-012-9394-3

Hyönä, J., Lorch, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology, 94*, 44–55. doi:10.1037//0022-0663.94.1.44

Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.) *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier Science.

Hyönä, J., & Nurminen, A. M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology, 97*, 31–50. doi:10.1348/000712605X53678

Kaakinen, J. K., Lehtola, A., & Paattilammi, S. (2015). The influence of a reading task on children's eye movements during reading. *Journal of Cognitive Psychology, 27*, 640–656. doi:10.1080/20445911.2015.1005623

Kajamies, A., Poskiparta, E., Annevirta, T., Dufva, M., & Vauras, M. (2003). *YTTE. Luetun ja kuullun ymmärtämisen ja lukemisen sujuvuuden arviointi* [YTTE. Assessment of reading and listening comprehension and reading fluency] Turku, Finland: University of Turku, Center for Learning Research.

Karppi, S. (1983). *Lukutaidon ABC. Johdatus lukemisen ja kirjoittamisen perustekniikan opetukseen* [The ABC of reading skill. Introduction to teaching the basic techniques of reading and writing.] Espoo, Finland: Weilin + Göös.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*, 281–300. doi:10.1080/10888430802132279

Lerkkanen, M.-K. (2006). *Lukemaan oppiminen ja opettaminen esi- ja alkuopetuksessa* [Learning and teaching reading in preschool and early education.] Helsinki, Finland: WSOY Oppimateriaalit.

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

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*, 1494–1502. doi:10.3758/s13428-016-0809-y

Macizo, P., & Van Petten, C. (2007). Syllable frequency in lexical decision and naming of English words. *Reading and Writing, 20*, 295–331. doi: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

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*, 267. doi:10.1037/a0021890

Nevala, J., & Lyytinen, H. (2000). *Sanaketjutesti* [Word chain test]. Jyväskylä, Finland: Niilo Mäki Institute and Child Research Center of University of Jyväskylä.

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing, 18*, 657–686. doi:10.1007/s11145-005-3355-z

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading, 11*, 357–383. doi:10.1080/10888430701530730

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 Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once) comprehension is supported by regressions during reading. *Psychological science, 25,* 1218–1226. doi:10.1177/0956797614531148

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

van den Broek, P., White, M. J., Kendeou, P., & Carlson, S. (2009). Reading between the lines. Developmental and individual differences in cognitive processes in reading comprehension. In. K. Wagner, C. Schatschneider, & C. Plythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension* (pp. 107–123). New York, NY: Guilford Press.van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2011). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing, 25*, 1665–1690. doi:10.1007/s11145-011-9337-4

Vauras, M., Mäki, H., Dufva, M., & Hämäläinen, S. (1995). *Diagnostiset testit 2: Kuullun ja luetun ymmärtäminen* [Diagnostic tests 2: Listening and reading comprehension] Turku, Finland: University of Turku, Center for Learning Research.

Vorstius, C., Radach, R., Mayer, M. B., & Lonigan, C. J. (2013). Monitoring local comprehension monitoring in sentence reading. *School Psychology Review, 42*, 191–206.

Table 1. The descriptive statistics of the independent reading skill measures.

	Raw scores			Reading level			*z* scores		
	M	SD	Range	M	SD	Range	M	SD	Range
Technical reading	9.73[1]	1.79	7-14[1]	4.18[2,3]	1.21	2-7[2,3]	-0.49[3]	0.61	-1.40-1.00[3]
Reading comprehension	17.21[4]	3.96	8-23[4]	5.33[2,3]	1.98	1-8[2,3]	0.26[3]	0.97	-2.00-1.70[3]
Word chain	9.27[5]	4.67	2-19[5]	4.21[6]	1.36	1-7[6]	-0.21[7]	0.81	-1.65-1.48[7]

1 = Theoretical maximum 20

2 = On scale of 1-9

3 = Norms gathered in March

4 = Theoretical maximum 40

5 = Theoretical maximum 34

6 = On scale of 1-8

7 = Norms gathered in September/October

Table 2. Means and standard deviations of eye movement data in each condition.

	Hyphenated		Concatenated	
	M	SD	M	SD
Forward Fixation Time (ms)	6975	2671	5993	2625
Regressive Fixation Time (ms)	3098	1384	2642	1345
Probability Of A Look-From	.139	.095	.122	.070

Table 3. Means and standard deviations of story comprehension scores overall and in each condition in percentages.

	Overall				Hyphenated		Concatenated	
	M	SD	Min.	Max.	M	SD	M	SD
Free Recall	13.8	8.9	0.0	48. 9	13.5	9.6	14.2	8.9
Cued Questions	38.3	16.1	0.0	91. 7	35.0	17.4	41.5	16.9
True/False Questions	86.3	8.8	33.3	100.0	87.0	11.2	85.7	9.3

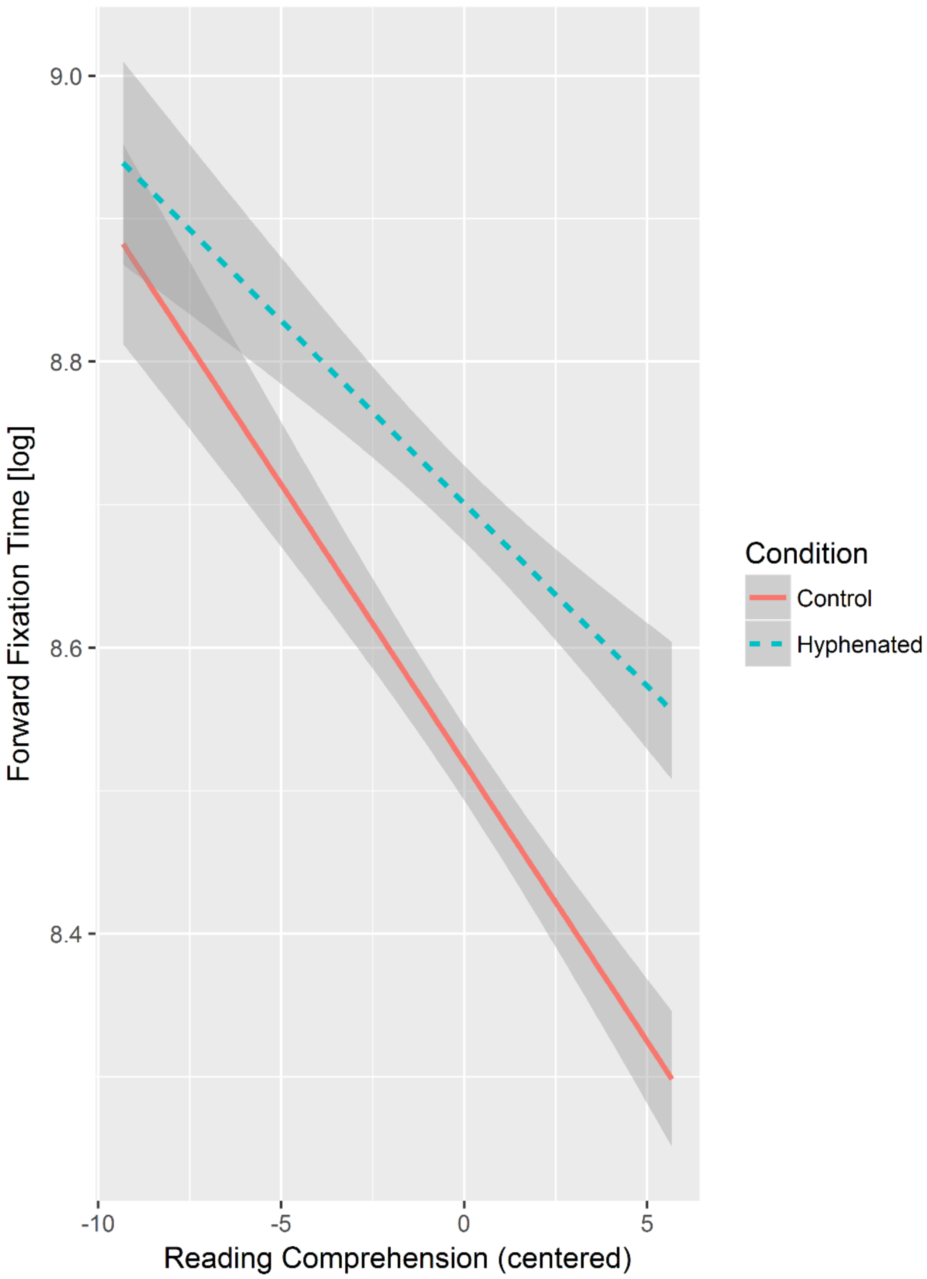

Figure 1. The effect of Syllable Boundary Cue on Forward Fixation Time as a function of Reading Comprehension. The envelopes denote the 95% confidence intervals.

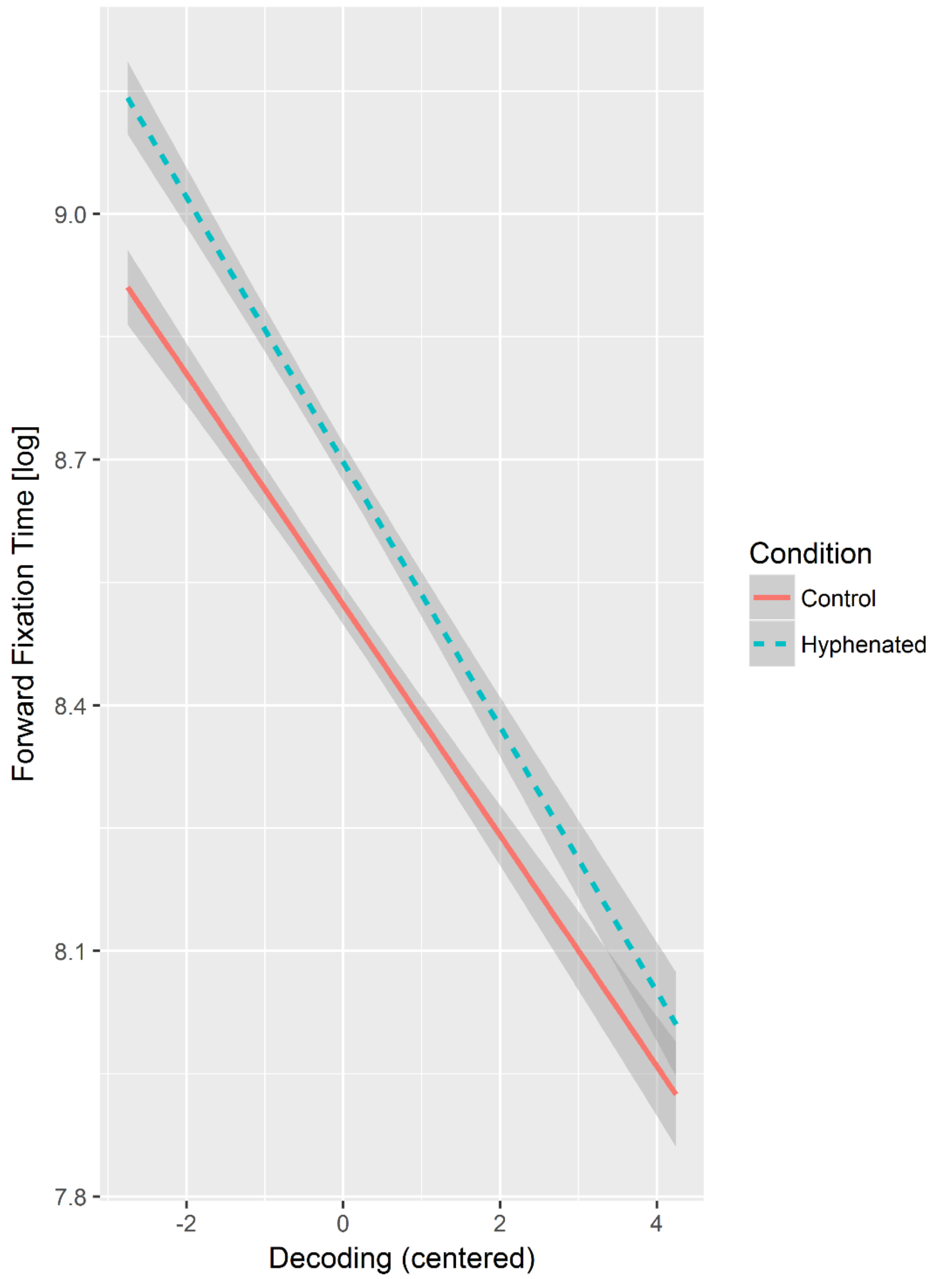

Figure 2. The effect of Syllable Boundary Cue on Forward Fixation Time as a function of Decoding skill. The envelopes denote the 95% confidence intervals.

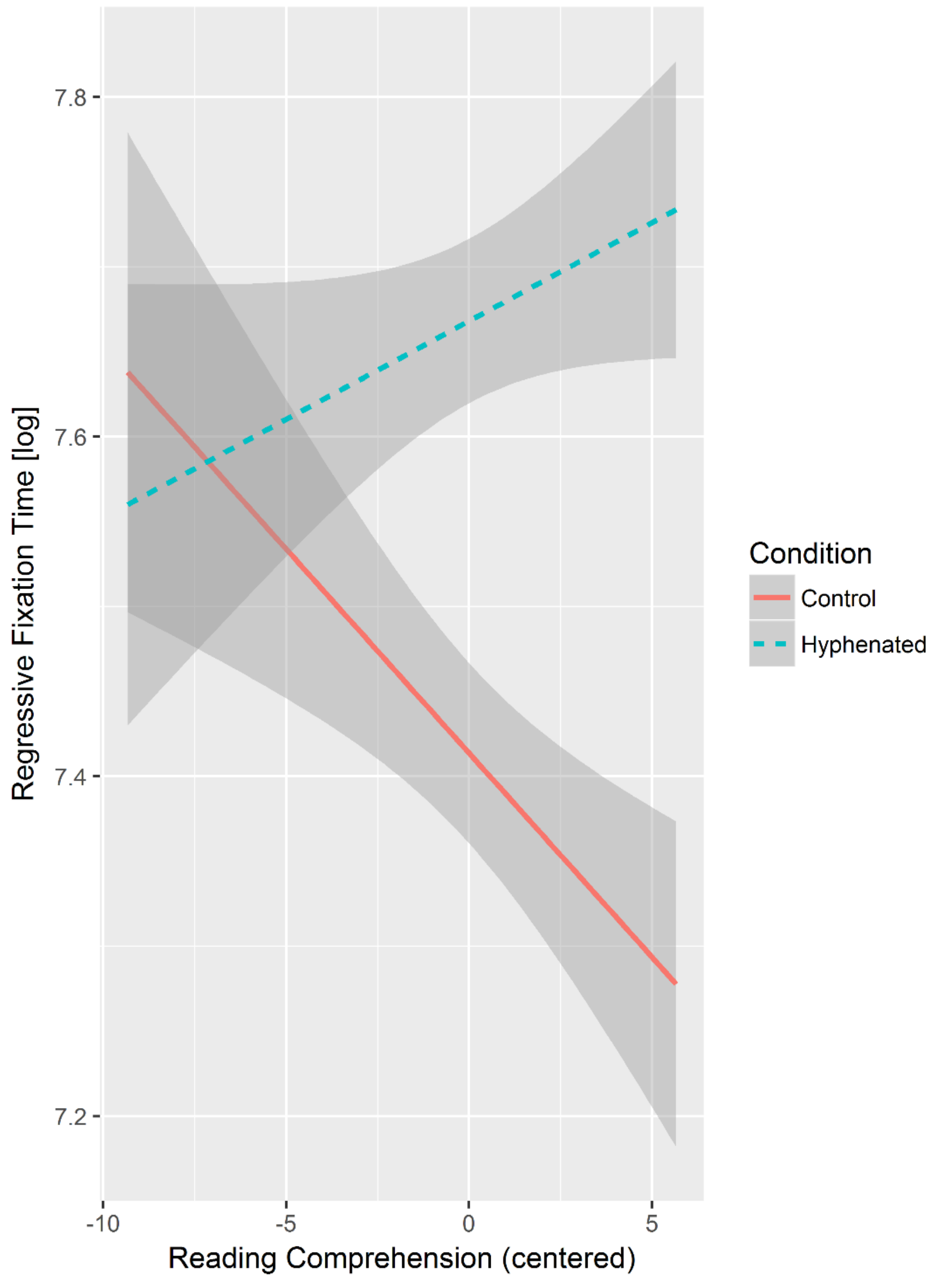

Figure 3. The effect of Syllable Boundary Cue on Regressive Fixation Time as a function of Reading Comprehension. The envelopes denote the 95% confidence intervals.

Appendix 1

Final models for each measure. *T* and *z* values below -1.96 and above 1.96 correspond to *p* values below 0.05. The analyses for the duration and accuracy measures have been conducted using Kenward-Roger approximation for degrees of freedom. If an interaction is significant, its main effects are reported as well.

[Insert Tables A1-A3 about here.]

Appendix 2

[Insert Table A4 about here.]

Table A1. Duration Measures: Fixed Effects

	B	SE	*t* value
Forward fixation time			
(Intercept)	8.612648	0.060060	138.68
SBCHyphen	0.167379	0.043161	3.87
Comprehension	-0.029568	0.013134	-2.15
Decoding	-0.134075	0.029116	-4.40
Session	-0.042375	0.007229	-5.86
SBCHyphen:Comprehension	0.006759	0.001908	3.54
SBCHyphen:Decoding	-0.011326	0.004317	-2.62
Regressive fixation time			
(Intercept)	7.614169	0.096176	76.84
SBCHyphen	0.246595	0.062285	3.95
Comprehension	-0.009752	0.019982	-0.47
Decoding	-0.165334	0.043475	-3.63
Session	-0.137215	0.028239	-4.86
SBCHyphen:Comprehension	0.024345	0.007393	3.29

Note: Durations have been log-transformed. SBC = Syllable Boundary Cue.

Table A2. Probability Of A Look-From: Fixed Effects

	B	SE	*z* value
(Intercept)	-1.7399	0.1983	-8.776
Session	-0.2580	0.1108	-2.329

Table A3. Story Comprehension measures: Fixed Effects

	B	SE	*t* value
Free recall			
(Intercept)	13.9271	1.3787	9.80
Comprehension	1.1117	0.3383	3.19
Cued questions			
(Intercept)	42.6108	2.7953	14.60
SBCHyphen	-7.6522	2.8036	-2.60
Comprehension	2.4072	0.5558	4.20
True/false questions			
(Intercept)	86.3008	1.5395	54.13
Comprehension	1.1585	0.3454	3.26

Note. SBC = Syllable Boundary Cue.

Table A4. Means and standard deviations in each condition as a function of test session.

	Session 1				Session 2			
	Hyphenated		Concatenated		Hyphenated		Concatenated	
	M	SD	M	SD	M	SD	M	SD
Forward Fixation Time (ms)	6938	2807	5973	2663	6672	2526	5525	2295
Regressive Fixation Time (ms)	3144	1234	2634	1268	2830	1445	2343	1307
Probability Of A Look-From	.140	.115	.127	.084	.128	.090	.103	.085
Free Recall (%)	14.9	10.1	16.3	10.4	13.1	9.4	13.2	9.2
Cued Questions (%)	34.9	17.1	42.7	18.3	35.5	18.5	42.0	15.6
True/False Questions (%)	86.7	11.9	83.6	11.7	87.9	13.5	88.7	9.2