Grapheme coding during sublexical processing in French third and fifth graders

Eva Commissaire^{a,†}, Anne-Sophie Besse^a, Elisabeth Demont^a,
Séverine Casalis^b

^a Laboratoire de Psychologie des Cognitions (LPC-EA4440), Université de Strasbourg, 67000 Strasbourg, France

^b SCALab (UMR CNRS 9193), Université de Lille, 59653 Villeneuve d'Ascq, France

article info

Article history:

Received 21 July 2017

Revised 12 March 2018

Available online 22 April 2018

Keywords:

Grapheme coding

Sublexical orthographic units

Visual word recognition

Developing readers

Letter detection task

Grapheme cohesion

abstract

This study aimed to investigate grapheme coding during silent word reading in French developing readers from Grades 3 and 5. Children performed a letter detection task in which three conditions were used; the letter to detect was (a) presented as a single-letter grapheme (simple condition; A in phare), (b) embedded within a multi-letter grapheme that is considered as a unit or not depending on context (weakly cohesive complex condition; A in chant where “an” is a unit but not in other words such as cane), or (c) embedded within a multi-letter grapheme that is systematically considered as a unit (highly cohesive complex condition; A in chaud). Results showed a grapheme condition effect in Grade 5 children only. In this group, both complex grapheme conditions were processed more slowly than the simple condition, but this complexity effect was much stronger for the highly cohesive condition. We suggest that graphemes can be coded as sublexical orthographic units from Grade 5 and that such orthographic fine-grained coding is affected by the degree of grapheme cohesion.

[†] Corresponding author. Fax: +33 (0) 3 68 85 15 46.

E-mail address: commissaire@unistra.fr (E. Commissaire).

Introduction

Learning to read is a long and laborious process that requires children to connect multiple types of linguistic information (i.e., phonological, orthographic, and semantics) and to eventually reach quick, fluent, and automatic reading. Before entering into print, children have already acquired large oral vocabularies and semantic information is already linked to phonological forms. Thus, learning to read involves relating this information with an orthographic code that is more or less transparent according to the language (Seymour, Aro, & Erskine, 2003). Once children come to master letter names and sounds, they start by applying a print-to-sound conversion mechanism, a procedure that is referred to as phonological decoding or the sublexical reading route and is considered as the “sine qua non” of reading acquisition (Share, 1999). Despite great interest in the phonological component of this sublexical reading route (Ehri, 1992; Share, 1999), how the sublexical orthographic code is previously extracted and processed during reading acquisition remains unclear.

This issue is of particular relevance when considering languages such as English and French in which the orthography-to-phonology mapping is not straightforward and involves sublexical orthographic representations of different grain sizes, including single-letter and multi-letter graphemes, also referred to as complex graphemes (e.g., “oa”, “sh”). How these complex graphemes are represented and processed during reading has received great attention by reading modelers, especially within the connectionist dual process (CDP+/++) theoretical framework developed by Perry, Ziegler, and Zorzi (2007, 2013, 2014), see also Grainger and Ziegler (2011). The model proposes a stage dedicated to grapheme parsing within the sublexical route to reading where letters from the visual input are extracted and combined into these larger significant units (following a subsyllabic onset–vowel–code scheme) before they are converted into phonemes. For example, when reading the word CHECK, the grapheme CH is allocated to the onset position, E to the vowel position, and CK to the coda position, and this preliminary stage of grapheme parsing is then followed by a conversion into their corresponding phonemes. Although the existence of this level of representation has been supported by several studies in expert readers (e.g., Rey, Ziegler, & Jacobs, 2000), how these grapheme units develop during reading acquisition remains little addressed.

To our knowledge, only a few studies by Marinus and de Jong addressed this issue with Dutch children. They showed that words and pseudowords including a complex grapheme were named faster than words without one (matched on letter length) and that this effect was larger in Grade 2 children and dyslexic readers compared with Grade 4 children (Marinus & de Jong, 2010). Whether graphemes are processed as sublexical orthographic units was more specifically assessed in Grade 4 children (mean age = 9;9 [years;months]) in two other studies. Using a visual segmentation paradigm along with naming and lexical decision tasks, reading speed was slower when a symbol was inserted within a unit (o#u) than between units (ou#), a finding indicating that complex graphemes were processed as units (Marinus & de Jong, 2008). Similar conclusions were reached by Marinus and de Jong (2011), who found that detecting a letter in a word took longer when it was embedded within a complex grapheme (A in *bea_ch*) compared with when it was presented as a simple grapheme (A in *black*), an effect referred to as the grapheme complexity effect. Based on previous adult studies (Rey et al., 2000), these grapheme effects were interpreted as reflecting the existence of sublexical orthographic representations at the grapheme level beyond letters. Detecting a letter embedded in a complex grapheme would lead to competition between the grapheme and letter levels and, thus, a processing cost.

Young readers seem to code complex graphemes as orthographic units during reading (when using the sublexical reading procedure; see Commissaire & Casalis, 2017) after a short experience with reading. How the mapping from letters to graphemes is learned was also more recently investigated and implemented by Perry et al. (2013). According to the authors, grapheme units that people learn to use are phonologically constrained. That is, although graphemes are represented at an orthographic level that is independent of phonology, whether adjacent letters are eventually converted into grapheme units may depend on phonological properties. Thus, grapheme coding could differ according to specificities of the grapheme under study such as grapheme cohesion. As pointed out by Spinelli, Kandel, Guerassimovitch, and Ferrand (2012), some complex graphemes, especially in French, are ambiguous as to whether they should be chunked into a grapheme unit or not. Some

are systematically processed as a unit (e.g., “au” corresponding to the phoneme /o/) and can be considered as highly cohesive, whereas others can be processed as units or not depending on the context (e.g., “an” is a unit corresponding to the phoneme /ã/ in *flan* /flã/ ‘custard’, but it should be processed as two separate single-letter graphemes in other words such as *cane* /kan/ ‘duck’) and, thus, are considered as weakly cohesive. [Commissaire and Casalis \(2017\)](#) recently reported in French adult readers that only highly cohesive complex graphemes were processed as units. A processing cost was found when detecting a letter (A) embedded in a highly cohesive complex grapheme (*chaud*) compared with when it was presented as a single letter (*phare*), but this complexity effect did not emerge with a weakly cohesive complex grapheme (*chant*).

To date, we do not know how sensitive children are to grapheme complexity and cohesion and whether these effects are modulated by reading experience. In the current study, two groups of developing readers attending Grades 3 and 5 needed to perform a letter detection task in French where three conditions were presented: (a) highly cohesive complex grapheme (e.g., detect A in *chaud*), (b) weakly cohesive complex grapheme (e.g., detect A in *chant*), and (c) simple grapheme condition (e.g., detect A in *phare*). Our goal was twofold. First, we aimed to assess grapheme coding in French-speaking developing readers and, thus, compare it with previous studies conducted with Dutch children. Although both languages are considered as quite transparent in the print-to-sound direction ([Seymour et al., 2003](#)) and contain a large proportion of complex graphemes, the existence of ambiguous graphemes seems to be more prominent in French. How developing readers deal with more or less cohesive complex graphemes is informative on the mechanisms underlying such a sub-lexical orthographic representational level ([Perry, Ziegler, & Zorzi, 2010](#); [Perry et al., 2013](#)). Based on [Commissaire and Casalis \(2017\)](#), our hypothesis was that highly cohesive complex graphemes would be more developed as grapheme units compared with weakly cohesive ones; grapheme complexity effects could then emerge only for the former condition. Second, whether graphemes are coded as sub-lexical units in developing readers was examined in two groups of children with different reading experience; thus, we extended the investigation of these mechanisms to a younger group of children than previously tested.

Method

Participants

A total of 88 French-speaking children participated in the study: 39 from Grade 5 (mean age = 10;8) and 49 from Grade 3 (mean age = 8;7). The standardized test *L’Alouette* ([Lefavrais, 1965](#)) was used to assess French reading. Children were asked to read a nonsense text composed of 265 words as accurately and quickly as possible. Reading accuracy was calculated by dividing the number of correct words read by the total number of words read within 3 min (the time limit to read the text), and reading speed corresponded to the number of correctly read words multiplied by 180, divided by the time spent to read the text. A total of 3 participants were excluded due to low reading performances (lower than 2.5 standard deviations from the mean of the group): 1 from Grade 3 and 2 from Grade 5. When examining reading performances across grade with this restricted sample, data revealed, as expected, higher reading accuracy in Grade 5 compared with Grade 3 (mean accuracy index = 96.16, SD = 1.96 for Grade 5 children and mean = 94.18, SD = 2.63 for Grade 3 children), $t(83) = 3.82$, $p < .001$, as well as higher reading speed in Grade 5 compared with Grade 3 (mean speed index = 347.42, SD = 115 for Grade 5 children and mean = 236.38, SD = 83.43 for Grade 3 children), $t(83) = 5.15$, $p < .0001$.

Materials

The stimuli were taken from [Commissaire and Casalis \(2017\)](#); see Appendix) and included letter-present and letter-absent trials. The 54 letter-present trials were divided into three conditions: (a) simple grapheme, where the letter to detect (e.g., A) was presented as a single letter in the word (e.g., *phare* ‘lighthouse’); (b) highly cohesive complex grapheme, where the target letter was embedded in a complex grapheme that is systematically processed as a unit in French (e.g., *chaud* ‘warm’);

and (c) weakly cohesive complex grapheme, where the letter was embedded in a complex grapheme of weaker cohesion; that is, the letter cluster in which the target letter appeared was considered as a grapheme unit in the experimental word (e.g., chant /Să/ ‘sing’) but could have corresponded to two single-letter graphemes, not a unit, if it had appeared in another word (e.g., ‘an’ in cane /kan/ ‘duck’). All words were four or five letters long and monosyllabic. The nature and position of the letter to be detected was perfectly controlled; there was the same number of A, E, and O target letters at exactly the same position in the word (second or third position) across the three conditions. The target always appeared as the first letter of the complex grapheme unit (e.g., A in chaud and in chant). Stimuli were matched on consonant–vowel (CV) structure, number of letters ($M = 4.39$, $SD = 0.50$), $F < 1$, ns, and lexical frequency (simple grapheme condition: $M = 155$, $SD = 168$; highly cohesive complex grapheme condition: $M = 160$, $SD = 126$; weakly cohesive complex grapheme condition: $M = 154$, $SD = 177$), $F < 1$, ns, which was estimated based on the Manulex database (Lété, Sprenger-Charolles, & Colé, 2004). The two complex grapheme conditions were also matched on phoneme length ($M = 2.44$, $SD = 0.51$), $F < 1$, ns. A total of 54 target-absent trials were also constructed in the same way as the letter-present trials.

Procedure

A target detection task was performed following Rey et al.’s (2000) procedure. The target letter was first presented for 700 ms in uppercase letters in the middle of the screen followed by a fixation point for 1000 ms. The target word then appeared in lowercase letters for 57 ms and was replaced by a blank screen presented for 70 ms followed by a 50-ms mask consisting of hashes. Participants needed to press ‘yes’ with their dominant hand if they detected the target letter in the word or to press ‘no’ with their nondominant hand. The experiment was preceded by a 10-trial training phase. The whole testing procedure lasted about 25 min.

Results

A total of 4 participants, 2 from Grade 5 and 2 from Grade 3, were excluded due to high error rates (>2.5 standard deviations above the mean of the group), which led to 35 remaining participants in Grade 5 and 46 in Grade 3. An analysis of variance was conducted on both reaction times and errors by participants (F_1) and items (F_2). In the analysis by participants, grade was entered as a between-participants variable and grapheme condition as a within-participant variable. In the analysis by items, it was the reverse pattern; grade was entered as within-participant variable and grapheme condition as a between-participants variable. Data cleaning on reaction times was conducted by removing all data points above 5000 ms and then discarding those below or above 2.5 standard deviations of participants’ individual mean ($<3\%$ of data for both Grade 5 and 3 participants). Reaction times were inverse transformed ($1/RT$, where RT is reaction time). Raw reaction times can be found in Table 1.

Table 1
Reaction times in microseconds and percentages of errors (and standard deviations) in Grade 3 and Grade 5 children according to the grapheme condition.

	Simple grapheme	Weakly cohesive complex grapheme	Highly cohesive complex grapheme
Grade 3			
Reaction time (ms)	1275 (310)	1251 (299)	1283 (318)
Errors (%)	8.7 (8.5)	9.4 (9.1)	8.9 (10.8)
Grade 5			
Reaction time (ms)	1008 (268)	1031 (237)	1056 (258)
Errors (%)	9.5 (7.9)	8.9 (6.8)	10 (10.4)
Mean reaction time (ms)	1160 (319)	1156 (294)	1185 (313)

Note. Standard deviations are in parentheses.

Reaction times

The effect of grade was significant by participants and items, $F_1(1, 79) = 11.77$, $p < .001$, $\eta^2 = .13$, $F_2(1, 51) = 375.14$, $p < .001$, $\eta^2 = .88$, and reflected faster reaction times in Grade 5 children (1032 ms) compared with Grade 3 children (1270 ms). The effect of grapheme condition was also significant, $F_1(2, 158) = 5.80$, $p < .01$, $\eta^2 = .07$, $F_2(2, 51) = 9.09$, $p < .001$, $\eta^2 = .26$. Importantly, the interaction between grade and grapheme condition was significant in the analysis by items, $F_2(2, 51) = 3.44$, $p < .04$, $\eta^2 = .12$, and as a trend by participants, $F_1(2, 158) = 2.72$, $p = .07$. When examining grapheme effects by grade, we found that the grapheme effect was significant in the Grade 5 group, $F_1(2, 68) = 4.80$, $p < .02$, $\eta^2 = .12$, $F_2(2, 51) = 10.28$, $p < .001$, $\eta^2 = .29$. Letters were detected more slowly in complex highly cohesive graphemes compared with the simple condition (48 ms), $F_1(1, 34) = 8.27$, $p < .01$, $F_2(1, 51) = 20.51$, $p < .001$, and compared with the complex weakly cohesive grapheme condition, albeit by items only, $F_2(1, 51) = 5.99$, $p < .02$, and not by participants (25 ms), $F_1(1, 34) = 2.78$, $p = .10$, ns. Detecting a letter in the weakly cohesive grapheme condition was also slightly longer compared with that in the simple condition (23 ms) by items only, $F_1(1, 51) = 4.33$, $p < .05$, but not by participants, $F_2(1, 34) = 2.57$, $p = .11$, ns. In contrast, in the Grade 3 group, the grapheme effect was clearly not reliable, $F_1(2, 90) = 2.34$, $p = .10$, ns, $F_2(2, 51) = 2.87$, $p = .07$, $\eta^2 = .10$.

Errors

There was no effect of grade or grapheme condition, all $F_s < 1$, ns. The interaction between grade and grapheme condition was not significant, all $F_s < 1$, ns.

Discussion

The goal of the study was to examine whether grapheme complexity effects could be observed in developing readers from Grades 3 and 5 and whether these effects would be modulated by grade and by the specific type of grapheme used, either highly or weakly cohesive. The grapheme condition effect proved to be significant in Grade 5 children only. In this group, a grapheme complexity effect emerged for both highly and weakly cohesive complex graphemes, although it was much more robust for the former. Our finding of a grapheme complexity effect in children aged 10 and 11 years confirms the unique previous letter detection study in Dutch-speaking children by [Marinus and de Jong \(2011\)](#). This effect emerged even if the letter to detect was at the first position of the complex grapheme, a result in line with [Marinus and de Jong \(2011\)](#), who found no difference according to the position of the target letter within the complex grapheme (but see [Brand, Giroux, Puijalon, & Rey, 2007](#)). Thus, based on previous interpretations of this effect ([Rey et al., 2000](#)), our data showed that grapheme sublexical orthographic representations are already developed at Grade 5; they also support parallel processing of the component letters of these grapheme representations, as postulated by the CDP theoretical framework ([Perry et al., 2010](#)). It is noteworthy that the procedure we used slightly differed from that of [Marinus and de Jong \(2011\)](#) in that the target word was presented for a 57-ms duration (as with adults in [Rey et al., 2000](#)), whereas it lasted until participants' response in their study. The use of a short duration of presentation allowed us to better compare our findings with adult data and to reduce potential phonological activation during the task in order to more specifically focus on sublexical orthographic representations per se.

Importantly, this effect was much stronger for highly cohesive complex graphemes and less reliable for weakly cohesive ones. Detecting a letter embedded in a highly cohesive complex grapheme (e.g., A in chau_d) took longer than when it was presented as a single-letter grapheme (e.g., A in phare), whereas only a small and less reliable detection cost emerged when the letter to detect was embedded in a weakly cohesive complex grapheme (e.g., A in chan_t). This finding, which is very close to that of adult expert readers from [Commissaire and Casalis \(2017\)](#), indicates that the sublexical grapheme level is much more activated, and thus triggers more competition with the letter level, for highly cohesive graphemes. When grapheme parsing is more ambiguous, as is the case for weakly cohesive complex graphemes (e.g., “an”), graphemes would be less activated during sublexical orthographic

processing. Thus, although grapheme units can probably be activated as a pure orthographic code, in the absence of phonological activation during the task (see [Commissaire & Casalis, 2017](#), Experiment 1c), letter clusters that developing readers learn to use as units probably initially rely on how a sound is mapped onto print, in line with proposals by [Perry et al. \(2013\)](#); see [Brand et al., 2007](#), and [Schmalz, Porshnev, & Marinus, 2017](#), for sublexical representations other than graphemes). It is noteworthy that the less confident pattern for weakly cohesive complex graphemes could also reflect individual variability in the way that participants' responses vary with grapheme condition. Mixed-level modeling (with a larger sample) and the inclusion of additional reading-related measures would help to explore and further understand this individual variability issue.

In Grade 3, the grapheme condition effect did not reach significance. This was also the case when excluding 25% of the slowest or least accurate readers from the analysis. Thus, considering that grapheme complexity effects reflect fast and automatized activation of grapheme orthographic representations, beyond the letter level, our data suggest that Grade 3 children did not automatically process complex graphemes as units in this task, where duration of presentation of the word is very short with no overt response required and rather limited phonological activation. Of course, this does not preclude the use of complex graphemes during reading aloud given that these must be somehow activated to correctly read words with complex graphemes. [Marinus and de Jong \(2010\)](#) found that Dutch children from Grade 2 (mean age = 7;8), but not from Grade 4, named words with complex graphemes faster than words without them when these were matched on their letter length. Interestingly, the pattern was reversed with slower reading rates for words with complex graphemes compared with words without them when these were matched on their phoneme length in children from both Grades 2 and 4. Thus, what could seemingly be interpreted as grapheme coding from Grade 2 might be due to a phoneme length effect and a letter-by-letter reading strategy. Despite their task probably tapping different mechanisms, these results fit well with our observation of grapheme coding as units at Grade 5 only. Interestingly, a follow-up analysis did not show any significant correlation between the robust grapheme complexity effect obtained for highly cohesive graphemes in Grade 5 participants and either their reading accuracy or speed. Although it is beyond the scope of this study, this would support the idea that developing readers need subsequent print exposure for automatic activation of graphemes to occur and that the sensitivity of these sublexical orthographic units that is needed to develop automatized coding is rather independent of participants' actual reading skills (at least measured via text reading). Learning (complex) graphemes relies on explicit teaching throughout the first years of primary school but also exposure to common sequences of letter clusters that co-occur with phonemes via unsupervised learning ([Perry et al., 2013](#)). Although computational models such as CDP++ parser ([Perry et al., 2013](#)) are able to quickly and very efficiently learn the mapping from letters to graphemes, our study reveals that it takes some time for developing readers to automatically process complex graphemes as units and that such coding also varies according to grapheme cohesion, a finding that could be integrated into the French version of the model ([Perry et al., 2014](#)). Future behavioral and computational work should now shed more light on the specific mechanisms allowing development of such sublexical orthographic units and extend our understanding of how these are then processed during more ecological reading situations.

Acknowledgments

This work was supported by the University of Strasbourg and Mission Investissement d'Avenir (IDEX Grant No. W14RAT34). We thank the teachers and pupils of the primary schools in Strasbourg, France, who participated in the project.

Appendix

Simple grapheme condition: A: place, phare, plate, page, rare; E: clerc, vert, sept, cerf; O: chose, score, sport, fort, port, rose, hors, mode, mort.

Highly cohesive complex grapheme condition: A: frais, chaud, trait, faux, saut; E: creux, peur, feux, ceux; O: froid, croix, droit, noir, bout, toit, noix, doux, mois.

Weakly cohesive complex grapheme condition: A: blanc, chant, franc, banc, rang; E: prend, sens, dent, cent; O: blond, tronc, front, donc, long, fond, mont, pont, rond.

References

- Brand, M., Giroux, I., Puijalon, C., & Rey, A. (2007). Syllable onsets are perceptual reading units. *Memory & Cognition*, 35, 966–973.
- Commissaire, E., & Casalis, S. (2017). The use and nature of grapheme coding during sub-lexical processing and lexical access. *Quarterly Journal of Experimental Psychology*. <https://doi.org/10.1080/17470218.2017.1318294>. Advance online publication.
- Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–143). Hillsdale, NJ: Lawrence Erlbaum.
- Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00054>.
- Lefavrais, P. (1965). *Test de l'Alouette* (revised in 2005). Paris: ECPA.
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). Manulex: A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers*, 36, 156–166.
- Marinus, E., & de Jong, P. F. (2008). The use of sublexical clusters in normal and dyslexic readers. *Scientific Studies of Reading*, 12, 253–280.
- Marinus, E., & de Jong, P. F. (2010). Variability in the word reading performance of dyslexics: Effects of word length and digraph presence. *Cortex*, 46, 1259–1271.
- Marinus, E., & de Jong, P. F. (2011). Dyslexic and typical-reading children use vowel digraphs as perceptual units in reading. *Quarterly Journal of Experimental Psychology*, 6, 504–516.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, 114, 273–315.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology*, 61, 106–151.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2013). A computational and empirical investigation of graphemes in reading. *Cognitive Science*, 37, 800–828.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2014). When silent letters say more than a thousand words: An implementation and evaluation of CDP++ in French. *Journal of Memory and Language*, 72, 98–115.
- Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, 75, 1–12.
- Schmalz, X., Porshnev, A., & Marinus, E. (2017). Two distinct parsing stages in nonword reading aloud: Evidence from Russian. *Quarterly Journal of Experimental Psychology*, 70, 2548–2559.
- Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174.
- Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, 72, 95–129.
- Spinelli, E., Kandel, S., Guerassimovitch, H., & Ferrand, L. (2012). Graphemic cohesion effect in reading and writing complex graphemes. *Language and Cognitive Processes*, 27, 770–791.