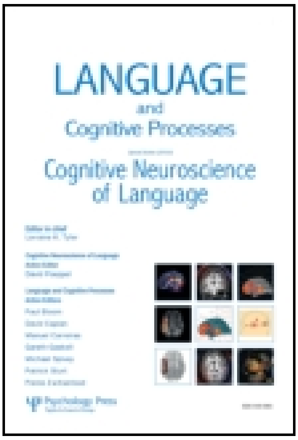

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

Graphemic cohesion effect in reading and writing complex graphemes

Elsa Spinelli [a b e], Sonia Kandel [a b d], Helena Guerassimovitch [a] & Ludovic Ferrand [c]

[a] Laboratoire de Psychologie et NeuroCognition, Université Pierre Mendès France, Grenoble, France

[b] Institut Universitaire de France

[c] CNRS and Université Blaise Pascal, Clermont-Ferrand, France

[d] GIPSA-Lab, Grenoble, France

[e] Department of Linguistics, University of California, Berkeley, CA, USA

Available online: 25 Oct 2011

To cite this article: Elsa Spinelli, Sonia Kandel, Helena Guerassimovitch & Ludovic Ferrand (2012): Graphemic cohesion effect in reading and writing complex graphemes, Language and Cognitive Processes, 27:5, 770-791

To link to this article: http://dx.doi.org/10.1080/01690965.2011.586534

LANGUAGE AND COGNITIVE PROCESSES
2012, 27 (5), 770–791

Psychology Press
Taylor & Francis Group

Graphemic cohesion effect in reading and writing complex graphemes

Elsa Spinelli[1,2,5], Sonia Kandel[1,2,4], Helena Guerassimovitch[1], and Ludovic Ferrand[3]

[1]Laboratoire de Psychologie et NeuroCognition, Université Pierre Mendès France, Grenoble, France
[2]Institut Universitaire de France
[3]CNRS and Université Blaise Pascal, Clermont-Ferrand, France
[4]GIPSA-Lab, Grenoble, France
[5]Department of Linguistics, University of California, Berkeley, CA, USA

AU /o/ and *AN* /ɑ̃/ in French are both complex graphemes, but they vary in their strength of association to their respective sounds. The letter sequence *AU* is systematically associated to the phoneme /o/, and as such is always parsed as a complex grapheme. However, *AN* can be associated with either one phoneme (/ɑ̃/ in e.g., *CRAN* /kRɑ̃/ "notch") and be parsed as a complex grapheme; or with two phonemes (/an/ in e.g., *CANE* /kan/ "duck"), thus being parsed as two simple graphemes. As a consequence, *AU* would be a more cohesive grapheme than *AN*, for which there is a parsing ambiguity. We examined whether the reading and writing systems take into account this potential parsing ambiguity due to the graphemes' degree of cohesion when processing complex graphemes. Experiment 1 consisted of a letter detection task. The participants had to detect, for example *A* in strongly cohesive complex graphemes (e.g., *AU* /o/) or weakly cohesive complex graphemes (e.g., *AN* /ɑ̃/). *A* was detected faster in weakly cohesive complex graphemes than in strongly cohesive ones. In a handwriting task (Experiment 2) we found that weakly cohesive complex

---

Correspondence should be addressed to Elsa Spinelli, Laboratoire de Psychologie et NeuroCognition, Université Pierre Mendès France, Bâtiment Sciences de l'Homme & Mathématiques, BP 47 38040 Grenoble Cedex 9, France. E-mail: elsa.spinelli@upmf-grenoble.fr

We thank Piennie Zwitserlood and two anonymous reviewers for their helpful comments on earlier versions of this manuscript. We are also grateful to Melinda Woodley for her valuable comments.

http://www.psypress.com/lcp http://dx.doi.org/10.1080/01690965.2011.586534

graphemes (e.g., *ON*) yielded longer programming times than strongly cohesive ones (e.g., *OU*), suggesting that the handwriting system also takes into account the potential decomposability of the complex graphemes into either one (ON /ɔ̃/) or two (O + N /on/) units. Overall, our results show an effect of parsing ambiguity due to graphemic cohesion of complex graphemes; these results should be accounted for by current models of written word processing and spelling.

Keywords: Graphemes; Letter-detection task; Handwriting; Parsing.

Graphemes are commonly defined as the written representation of phonemes (Coltheart, 1978; Henderson, 1985). There can be a direct match between letters and graphemes, in the French word *tuba* /tyba/ "snorkel" (T = /t/, U = /y/, B = /b/, A = /a/). One letter can either correspond to a simple grapheme (one letter grapheme) such as *A* in *tuba* or be a part of a complex grapheme—i.e., embedded in a multi-letter grapheme—such as *A* in the French word *pause* /poz/ 'pause'. The aim of the present research was to gain insight into the way complex graphemes are processed. In French, complex graphemes may vary in the strength of association to a given sound. The complex grapheme *AU* is systematically associated to the phoneme /o/ whereas letter sequences like *AN* can be associated with either one phoneme /ɑ̃/ (in *CRAN* /kRɑ̃/ "notch") and constitute a complex grapheme or two phonemes (/an/ in *CANE* /kan/ "duck") and constitute a sequence of two simple graphemes (*A* and *N*). Thus, *AU* should be unambiguously parsed as a complex grapheme, giving rise to a cohesive unit because it is phonologically consistent (always associated with the same sound). On the contrary, for *AN* there could be an ambiguity in parsing. Because *AN* is not phonologically consistent (not always associated with the same sound), *AN* is not a cohesive unit in the sense that it could be parsed as a complex grapheme AN = /ɑ̃/ or as two simple graphemes A + N = /an/. This study investigated whether the orthographic representations that are activated in reading and writing processes take into account the graphemic cohesion of complex graphemes.

An important debate in the field of written word recognition concerns the processes underlying the mapping of the sensory information from the visual input to the stored entries in the lexicon. More precisely, a central question is to determine which abstract components such as features, letters, graphemes, or syllables are involved in the process of accessing the lexicon (see Carreiras & Grainger, 2004 for a review). Unlike letter units, grapheme units allow a more direct correspondence between orthographic and phonological word forms. Hence, graphemes have been proposed as functional units to mediate access to the lexicon (Rastle & Coltheart, 1998). A number of studies have shown that letter/grapheme mismatches in words and nonwords require more processing time than when the stimuli are composed of simple graphemes. Rastle and Coltheart (1998) compared naming latencies to five letter long nonwords differing in number of graphemes (e.g., *FOOFH* /fuf/: 3 graphemes and *FROLP* /frolp/: 5 graphemes). Naming latencies were shorter for the stimuli showing no mismatch between letters and graphemes, i.e., when the nonwords contained the same number of graphemes and letters (*FROLP*). In a similar vein, Rey, Jacobs, Schmidt-Weigand, and Ziegler (1998) found in a perceptual identification task that response times were shorter for words containing as many graphemes as letters (*BLAST*: 5 graphemes) compared to words with fewer graphemes than letters (*TEETH*: 3 graphemes). These two studies therefore indicate that mismatches between the number of graphemes with respect to the number of letters affect performance in reading. They suggest that during the reading process there is a stage of letter clustering into bigger grapheme units.

Further research with a letter-detection task conducted by Rey, Ziegler, and Jacobs (2000) showed that in French and English the reading system processes graphemes as perceptual units. In their experiments, the target letters could either be single-letter graphemes (e.g., detect *A* in *place,* /plas/ "place") or part of a multi-letter grapheme (e.g., detect *A* in *pause,* /poz/ "pause"). Rey et al. (2000) found shorter response times for single-letter graphemes than for multi-letter graphemes, suggesting that graphemes are functional units in reading (see also Royer, Spinelli, & Ferrand, 2005). Hence, there are several studies supporting the idea that letters are combined into grapheme units that could mediate lexical access. This tendency to group small units into bigger chunks is a well-known phenomenon that is particularly efficient for the memorisation of strings with several elements (cf. Jenkins & Russel, 1952). It is therefore likely that orthographic representations encode letter strings by chunking them into their grapheme constituents.

Spelling processes in writing also seem to require the grouping of letters into chunks. A first line of evidence comes from neuropsychological data. Tainturier and Rapp (2004) analysed the spelling performance of two English patients with acquired dysgraphia. Their spelling errors revealed that orthographic representations store information on two-letter graphemes, like *ph* = /f/ in *phone*. This information is different from grapheme sequences that correspond to single phonemes like *pl* = /pl/ in *place*. According to the authors, graphemes "are represented as units with an internal structure that specifies the identity and order of its constituents" (Tainturier & Rapp, 2004, p. 130). The patients' performance indicates that complex graphemes have a unitary representation and are "unpacked" at the moment of serial production, to specify letter identity and order.

This is in line with Houghton and Zorzi's (2003) connectionist model of spelling processes that posits two distinct representational levels, one for grapheme-units and another for letter-units. The system associates the phonemes to their graphemic counterparts before activating letter strings. For example, to spell the word *seat*, the system activates $S + EA + T$ at the grapheme level and then $s + e + a + t$ at the letter level. The model processes the grapheme *ea* as a unit at the grapheme level, providing a more straightforward mapping from phonology to orthography than if there was a direct mapping from sounds to letters. The authors showed that simulations of the spelling process are more accurate when considering both grapheme and letter levels than when excluding the grapheme level.

Recent data on non-brain-damaged participants also reveal that graphemic complexity affects the timing of handwriting programming (Kandel & Spinelli, 2010). Adult French participants wrote words on a digitiser that recorded the spatial and kinematic parameters of the handwriting movement. The words contained simple graphemes (e.g., A in *clavier*, /klavje/ "keyboard") or complex graphemes (e.g., AI /ɛ/ in *prairie* /pRɛRi/ "meadow" and AIN /ɛ̃/ in *plainte* /plɛ̃t/ "complaint"). The results revealed that graphemic complexity affected movement duration of the letter preceding the target grapheme. Most of the effects found in handwriting production are anticipatory such that writers program the unit *n* while writing unit *n-1*. For example, the duration of L in *cl<u>a</u>vier* was shorter than the first R in *pr<u>ai</u>rie* (the target grapheme is underlined). The authors observed that the degree of grapheme complexity also modulated the timing of movement processing before starting to write the target grapheme. The duration of R in *pr<u>ai</u>rie* was shorter than L in *pl<u>ain</u>te*. This indicates that the writing system is sensitive to grapheme complexity. Similar results were observed in a developmental perspective (Kandel, Soler, Valdois, & Gros, 2006).

In sum, the studies described in the previous paragraphs indicate that grapheme-like units mediate reading and spelling processes. They all compared the perceptual and spelling processes involving simple graphemes with respect to complex graphemes. One concern is that they considered complex graphemes as if they were all the same. In French there are at least 34 graphemes of more than one letter (Catach, 1995). Complex graphemes can differ in the number of letters they contain—like AI and AIN in Kandel and Spinelli (2010)—and also on other aspects in relation to their correspondence to sounds.

As mentioned earlier in the introduction, complex graphemes vary in their strength of association to a given phoneme. For example, the complex grapheme *AI* is systematically associated to the phoneme /ɛ/ and cannot be associated to any other phoneme. The univocal association of this two-letter sequence to a particular phoneme makes the grapheme unit cohesive and therefore unambiguously parsed: *AI* must be parsed as a single two-letter unit *AI*. The status of the complex grapheme *AN* is different. The letter sequence *AN* can represent one phoneme (e.g., /ɑ̃/ in *CRAN* /kRɑ̃/ "notch") but also the two phonemes /a/ and /n/ (e.g., in *CANE,* /kan/ "duck"). As a consequence *AN* could be parsed either as a single two-letter unit *AN* or as two single one-letter units *A* and *N.* Indeed, *AI* (always /ɛ/) cannot be split whereas *AN* may also represent a sequence of two simple graphemes and can therefore be decomposed into *A* = /a/ and *N* = /n/. In other words, the probability of having *AI* = /ɛ/ is much higher than having *AN* = /ɑ̃/ because *AN* can sometimes be realised /an/ (e.g., in *CANE,* /kan/). One might thus expect that although *AI* and *AN* are both complex graphemes, *AI* would be a more cohesive unit than *AN* for which there is a parsing ambiguity. The present study focuses on the processing of two letter complex graphemes and aims at evaluating whether their graphemic cohesion may regulate their processing, in both perception (letter-detection task) and production (handwriting task). Therefore we compared the processing of strongly cohesive (*AU* /o/, *OU* /u/) to that of weakly cohesive (*AN* /ɑ̃/, *ON* /ɔ̃/) complex graphemes.

EXPERIMENT 1 (LETTER-DETECTION TASK)

Method

As in Rey et al. (2000) in this experiment, participants had to monitor for a target letter in a subsequently briefly presented carrier word.

Participants

Twenty students of the University Pierre Mendès France, Grenoble, participated in the experiment for course credit. All participants were native speakers of French and had normal or corrected vision.

Stimuli

Target-present trials. Sixteen pairs of words were selected from the French database Lexique (New, Pallier, Ferrand, & Matos, 2001). In a first condition, the target vowel (e.g., *O*) was part of a complex grapheme that is cohesive (e.g., *O* in *bijou*-/biju/ "jewel"- OU is always realised /u/ in French, hence always parsed as a complex grapheme OU). In a second condition, the target vowel (e.g., *O*) was part of a complex grapheme that is less cohesive (e.g., *O* in *gazon*-/gazɔ̃/ "lawn", -ON can be realised /ɔ̃/ or /on/ depending on the words, hence either parsed as a complex grapheme ON or as two simple graphemes O and N). The 16 pairs were matched for Frequency (5.51 occurrences per million for the strongly cohesive condition and 8.24 occurrences per million for the weakly cohesive condition), number of letters

(5.3, 5.5, respectively), number of phonemes (4.3, 4.2, respectively) and number of syllables (2.1, 2.0, respectively). The pairs were also matched for the position of the target letter in the word (always the penultimate one). Two types of targets were chosen: *A* and *O* consisting of eight pairs for each target (see Appendix 1).

Target-absent trials. Sixteen pairs of words that did not contain the targets *A* or *O* were matched to the experimental pairs in frequency (7.04, 7.06 occurrences per million), number of letters (5.5, 5.6), number of phonemes (4.4, 4.1), and in number of syllables (2.1, 2.0).

Fillers. Because the target letters were always the penultimate letter of the words, we also included 16 target-present trials in which the target letters (8 *A*, and 8 *O*) appeared in various positions in the word (e.g., *A* in *audio* "audio", *O* in *souci* "worry"). We needed to prevent the participants from focusing on the ends of words in monitoring for targets. Sixteen target-absent trials were also added in which the target letters (8 *A*, and 8 *O*) were not present in the word (e.g., *A* in *devin* "soothsayer"). We also added 52 target-present trials for which the target letters were consonants instead of vowels and could appear in various positions in the word (e.g., *S* in *satin* "satin", *T* in *ultime* "ultimate", *P* in *pont* "bridge") and 52 target-absent trials in which the consonant target letters were not present in the word (e.g., S in *objet* "object").

Procedure

Participants were tested individually in a quiet room. As in Rey et al.'s (2000) procedure, each trial began with a 700 ms presentation of a target letter in the centre of a computer screen (e.g., *O*). Then a fixation mark (:) was presented for 1,000 ms and was replaced by a stimulus word (e.g., *gazon*), which remained on the screen for 50 ms. The stimulus word was followed by a blank interval of 70 ms. The target letter was presented in uppercase and the stimulus word in lowercase in order to prevent detection from being based on low level perceptual analysis. Then a mask (######) appeared and remained on the screen until the participant responded. The participant had to decide as accurately and as quickly as possible whether the target letter was in the stimulus word or not by using one of two response buttons (the « yes » button or the « no » button). They had to press the « yes » button with the forefinger of their dominant hand. The experiment was controlled by E-prime. The RT was measured from the presentation of the mask on the screen to the participants' response. Response latencies and errors were collected. The session began with 10 practice trials. The 200 trials were then presented in a randomised order for each participant. The session lasted approximately 20 minutes.

Results and discussion

Incorrect responses (5.3%), RTs longer than 1,500 ms, shorter than 200 ms as well as RTs above and below 2*SD* from the participants' means per condition (4.7%) were excluded from the analysis. The results were evaluated using a two-way repeated measures analyses of variance (ANOVAs) by participants (*F*1) and by items (*F*2), with Target Letter (A vs O) and Graphemic Cohesion (strongly cohesive, weakly cohesive) as main factors (see Table 1).

The effect of *Graphemic Cohesion* was significant with longer durations for the strongly cohesive condition (507 ms) than for the weakly cohesive condition (454 ms), $F_1(1, 19) = 17.25$, $p < .001$, $F_2(1, 14) = 14.85$, $p < .005$. There was a main effect of *Target Letter* significant by subject only, $F_1(1, 19) = 6.03$, $p < .05$, $F_2(1, 14) = 2.81$, $p = .11$, *ns*, but no interaction between *Target Letter* and *Graphemic Cohesion,* $F_1(1, 19) = 2.69$, $p = .11$, *ns*, $F_2(1, 14) = 1.77$, $p = .20$, *ns*. The analyses conducted on errors showed no effect of *Graphemic Cohesion* (5.3% for both conditions, both Fs < 1), no effect of *Target Letter,* $F_1(1, 19) < 1$, $F_2(1, 14) = 2.03$, $p = .17$, *ns,* and no interaction between these two factors (both Fs < 1).

This experiment showed that single-letter detection times were influenced by the graphemic cohesion of the two-letter sequences. Single letters (e.g., *O*) were easier to detect in complex graphemes that were weakly cohesive (e.g., *ON*, /ɔ̃/ parsed as a complex grapheme ON in *gazon* but that can be realised /on/ and parsed as two simple graphemes in other words) than in complex graphemes that were strongly cohesive (e.g., *OU*, always realised /u/, hence always parsed as a complex grapheme OU). The degree of graphemic cohesion of two-letter sequences thus influences their processing times in reading. It seems that weakly cohesive graphemes give rise to both parses, i.e., complex grapheme ON and simple graphemes O + N, this latter parse

TABLE 1
Mean reaction times in milliseconds (ms), standard deviations for correct targets detection (A and O) for the strongly cohesive and weakly cohesive complex graphemes

	Strongly cohesive	*Weakly cohesive*	*Effect*
Target A			
RT (ms)	534	459	+75 ms
SD	148	95	
Errors (%)	6.3	6.3	
Target O			
RT (ms)	479	449	+30 ms
SD	137	126	
Errors (%)	4.4	4.4	

favouring detection of the letter O in this condition. Conversely, the univocal parsing OU—complex grapheme—for the strongly cohesion condition hindered the detection of the letter O.

EXPERIMENT 2 (HANDWRITING TASK)

Method

In this experiment participants had to handwrite a word that was visually displayed.

Participants

Thirty-three students of the University Pierre Mendès France, Grenoble, participated in the experiment for course credit. All participants were native speakers of French and had normal or corrected vision. None of them had participated in the previous experiment.

Stimuli

Forty pairs of words were selected from the French data base Lexique (New et al., 2001) corresponding to two conditions. In the first condition, the words contained target complex graphemes that are strongly cohesive (e.g., *OU* in *mouler*- /mule/ "to mould", *OU* is always realised /u/ in French hence always parsed as a complex grapheme, see Appendix 2). In the second condition, the words contained target complex graphemes that are weakly cohesive (e.g., *ON* is realised /ɔ̃/ in *monter*- /mɔ̃te/ "to go up" thus parsed as a complex grapheme *ON*; but can be realised /on/ in other words like *monnaie* /monɛ/ "currency" hence parsed as two simple graphemes *O* and *N*). Three pairs of target graphemes were chosen: *OU* /u/ vs *ON* /ɔ̃/, *AU* /o/ vs *AN* /ɑ̃/, and *OU* /u/ vs *OM* /ɔ̃/ for the strongly cohesive versus weakly cohesive condition respectively (see Appendix 2). The 40 pairs were matched for grapheme position (after the first letter or the first cluster), frequency (11.28 occurrences per million for the strongly cohesive condition vs 10.77 occurrences per million for the weakly cohesive condition), number of letters (7.3 vs 7.3, respectively), number of syllables (2.1 vs 2.1, respectively), and bigram frequency (6,301.7 vs 7,149.7, respectively).

Procedure

The experiment was conducted with Ductus, which is a new handwriting software package developed in our laboratory for the study of handwriting production (Guinet & Kandel, 2010). Participants were presented with a word on a computer screen and asked to write it on a digitiser (Wacom

Intuos 2, sampling frequency 200 Hz, accuracy 0.02 mm). They were asked to write with a special pen (Intuos Inking Pen) on lined paper covering the digitiser. The words appeared on the centre of the screen of a laptop written in upper-case Times New Roman size 18. Before the presentation of the word there was auditory signal that indicated the beginning of the trial and a fixation point for 200 ms. Participants wrote the words in upper-case letters and lifted the pen between each letter in a small wrist upward–downward movement. Participants were asked to start writing as soon as they saw the word and to write at a normal speed. The 80 words were presented in random order. The experiment lasted approximately 50 minutes. Two practice trials preceded the experiment.

Data processing

The data were smoothed with a Finite Impulse Response filter (Rabiner & Gold, 1975) with a 12 Hz cut-off frequency. Then, the analysis was conducted with Ductus' semi-automatic handwriting analysis module (Guinet & Kandel, 2010). We segmented the words into their letter constituents by hand so we could obtain data on the timing of the movement that produced each letter. We focused on letter duration because it is a very sensitive measure that is widely used in studies investigating the linguistic components of handwriting production (Bogaerts, Meulenbroek, & Thomassen, 1996; Van Galen, 1991). Letter duration refers to the time the participants took to write each letter.

Since Kandel and Spinelli (2010) showed that most of the processing of graphemic complexity was achieved while the participants were writing the letter that preceded the target grapheme, in this experiment we focused on the duration of the letter preceding the strongly cohesive complex graphemes (e.g., *M* in *MOULER* /mule/) and the weakly cohesive complex graphemes (*M* in *MONTER* /mɔ̃te/). We also measured the duration of the first letter of the complex graphemes (e.g., *O* in *MOULER* /mule/, strongly cohesive condition, and *O* in *MONTER* /mɔ̃te/, weakly cohesive condition).

In some cases, we needed to compare the durations of letters of different sizes. The N in "nausée", for instance, should be longer than the L in "lancée" because the former has three lines whereas the latter has two lines. We therefore normalised the absolute duration following a procedure that has already been used in previous handwriting studies (see for example, Bogaerts et al., 1996; Kandel, Herault, Grosjacques, Lambert, & Fayol, 2009; Kandel & Spinelli, 2010). The absolute duration is divided by the number of strokes per letter. Unlike cursive lower-case letters, there is no standard definition for stroke segmentation for handwritten upper-case letters (see Meulenbroek & Van Galen, 1990). We therefore used Kandel and Spinelli's (2010) letter segmentation procedure. For the letters that are composed of straight lines (e.g., *N*, *L*, *T*, *F*, *E*), we considered each line to be one stroke. The absolute duration was thus divided by the number of strokes composing the letter. Thus given durations such as $N = 240$ ms and $L = 160$ ms, we did 240/3 = 80 and 160/2 = 80. The resulting stroke durations were 80 ms for both letters. We concluded that there was no difference between the durations of the two letters. For the letters with curved traces, such as *C*, *S*, *R*, *O* we used the classical stroke segmentation procedure that consists of segmenting the continuous trace according to tangential velocity minimal values (see, Orliaguet, Kandel, & Boë, 1997 for an example). Ten productions of each letter were recorded and segmented on the basis of the velocity minima in the tangential velocity profile. In these productions, *S* for instance, always presented three strokes. So if in our experiment the participant produced *S* in 300 ms, the stroke duration was 300/3 = 100 ms. All the analyses presented in the Results section refer to stroke durations (see Appendix 3 for more details on the normalisation procedure).

RESULTS

Figure 1 shows the normalised durations of the letter preceding the grapheme (e.g., *M* in *MONTER, MOULER*) and of the first letter of the complex grapheme (e.g., *O* in *MONTER, MOULER*) in words containing strongly cohesive and weakly cohesive complex graphemes. The results were evaluated using a two-way repeated measures analyses of variance (ANOVAs) by participants (*F*1) and by items (*F*2), with Letter Position (preceding letter, first letter) and Graphemic Cohesion (strongly cohesive, weakly cohesive) as main factors.

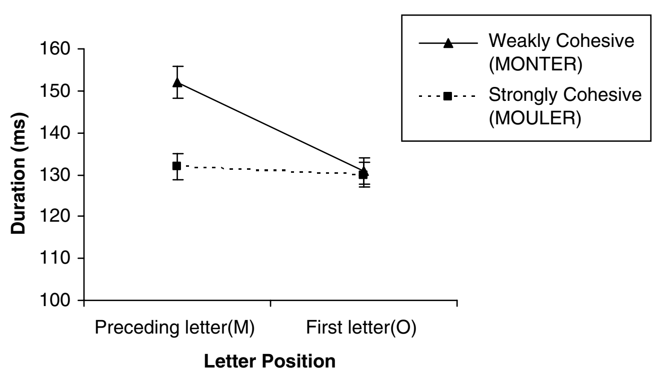

Figure 1. Normalized durations of the letter preceding the grapheme (e.g., *M*) and of the first letter of the complex grapheme (e.g., *O*) in words containing strongly cohesive (e.g., *MOULER*) and weakly cohesive complex graphemes (e.g., *MONTER*). Error bars show the standard errors of the means.

The main effect of *Letter Position* was significant, with shorter durations for the first letter of the complex grapheme (130.5 ms) than for the letter preceding the complex grapheme (142 ms), $F_1(1, 32) = 39.65$, $p < .001$, $F_2(1, 39) = 5.99$, $p < .01$. The effect of *Graphemic Cohesion* was also significant with shorter durations for the strongly cohesive condition (131 ms) than for the weakly cohesive condition (141.5 ms), $F_1(1, 32) = 113.10$, $p < .001$, $F_2(1, 39) = 5.71$, $p < .05$. Moreover, the interaction between *Letter Position* and *Graphemic Cohesion* was significant, $F_1(1, 32) = 254.76$, $p < .001$, $F_2(1, 39) = 5.50$, $p < .05$. Planned comparisons showed that the duration of the letter preceding the complex grapheme was significantly longer in the weakly cohesive condition (*M* in *MONTER,* 152 ms) than in the strongly cohesive condition (*M* in *MOULER,* 132 ms); $F_1(1, 32) = 203.23$, $p < .001$, $F_2(1, 39) = 5.66$, $p < .05$. There was no difference in the duration of the first letter of the complex grapheme between the weakly cohesive condition (*O* in *MONTER,* 131 ms) and the strongly cohesive condition (*O* in *MOULER,* 130 ms); both *Fs* < 1.

This pattern of results shows an effect of parsing ambiguity. *OU* like graphemes are strongly cohesive and are unambiguously parsed as one unit whereas *ON* like graphemes are less cohesive and can potentially be parsed in two ways (*ON* or *O* + *N*). Our results show that the writing times reflect the degree of cohesion of complex grapheme and that this processing is done before we start writing the complex grapheme. Once we start writing the first letter of the complex grapheme, there is no more influence of its degree of cohesion.

GENERAL DISCUSSION

This research investigated the way complex graphemes are processed by the reading/writing system. We examined whether the graphemic cohesion in complex graphemes could regulate their processing in reading and writing tasks. Our results in the perception task showed that single-letter detection times in two-letter graphemes were influenced by the graphemic cohesion of the grapheme unit. Single letters (e.g., *O*) were faster to detect in complex graphemes that were weakly cohesive (e.g., *ON*), than in complex graphemes that were strongly cohesive (e.g., *OU*). The results of the handwriting task also showed that spelling processes take into account the graphemic cohesion of complex graphemes. The letter preceding weakly cohesive complex graphemes (*ON*) took longer to write than the letter preceding strongly cohesive complex graphemes (*OU*).

The main result of this study is that complex graphemes are to a certain extent breakable units, some of them (the weakly cohesive units) being more breakable than others (the strongly cohesive units). In this sense, our results are compatible with studies showing activation of complex grapheme constituents. For example, in a letter detection task, Peereman, Brand, and Rey (2006) found that single letters are converted into their respective sounds when they form multi-letter graphemes. In their study, participants had to detect phonemes in visually presented nonwords. Target phonemes (e.g., /o/, /y/, or /i/) were never present in the carrier nonwords (e.g., /Rud/) but corresponded either to the first letter of the multi-letter grapheme of the nonword (e.g., *O* in *roude* /Rud/), the second (e.g., *U* in *roude* /Rud/), or none of them (e.g., *I* in *roude* /Rud/). Peereman et al. (2006) observed an interference in phoneme decision latencies to target phonemes corresponding to both letters of the complex grapheme, compared to the absent letter condition, suggesting that both letters of a complex grapheme are converted into their respective sounds. Our letter detection results are also compatible with the view that both elements of the multi-letter grapheme are activated, at least for weakly cohesive units (like ON). Our results showed that the weakly cohesive units can be broken into their constituents. The shorter detection times for O in ON compared to OU suggests that both simple graphemes O and N are computed out of ON in addition to the complex grapheme ON. The two possible parses for the weakly cohesive graphemes seem to be considered by the reading system.

A second interesting aspect of the letter-detection results is that we found a difference in detection time for the first letter of two-letter graphemes (the letter *O* was detected faster in *ON* than *OU*). However, Brand, Giroux, Puijalon, and Rey (2007) found an effect of grapheme complexity (slower response times to detect a letter in a multi-letter grapheme than in a simple grapheme) only when the target letter was in second position (*U* in *boule* /bul/ "ball" vs *brune* /bRyn/ "brown") and not when it was in the first position (*A* in *gland* /glɑ̃/ "acorn" vs *glace* /glas/ "mirror"). They argued that chunking (i.e., grouping the two letters of the complex grapheme into one single unit) is position dependent. Under the assumption that letter strings are processed serially (Rastle & Coltheart, 1998), the first letter of a multi-letter grapheme is processed as a single grapheme unit and the second letter is combined with the preceding letter to form the complex graphemic unit.

The fact that we found a difference in detection time for the first letter of two-letter graphemes could thus suggest that chunking had already occurred while processing the first letter of the complex grapheme. Alternatively, it could be that there is a quick deactivation of the grapheme *O* once the *U* letter is processed in *OU* whereas such deactivation of the grapheme *O* when encountering *N* is not possible because parsing is here dependent on the following letter (the letter following *N*) and the system maintains the two parses.

The results of the handwriting task are in agreement with the ones observed in the letter detection task since there was a clear difference in the processing of the weakly and strongly cohesive complex graphemes. Weakly cohesive complex graphemes (*ON*) required more processing time than strongly cohesive complex graphemes (*OU*). It is also interesting to point out that these differences were only observed at the letter preceding the complex grapheme, indicating that graphemic parsing is done before the participants started to write it. Since the differences did not reach significance at the first letter of the complex graphemes, it is likely that graphemic parsing was finished when the participants started to write it. This pattern of results suggests that the handwriting system takes into account the potential decomposability of complex grapheme *ON* into either one or two units. For strongly cohesive complex graphemes (*OU*) the handwriting system only has one unit to program without ambiguity. Because graphemic cohesion relates to the strength of association between graphemes and their respective sounds, our results also indicate that the processing of graphemes is tightly linked to multiple print-to-sound associations, and are in line with more general grapho-phonological consistency effects in handwriting production. For example, Delattre, Bonin, and Barry (2006) found in a task of writing spoken words to dictation that irregular French words (i.e., containing low probability phoneme to grapheme mappings) elicited longer latencies than regular ones. This effect was larger for low than for high frequency words. Jones, Folk, and Rapp (2009) induced orthographic working memory disruption through the use of a distractor shadowing task while participants performed a task of writing under dictation. Their study showed that segments with low versus high levels of sound-letter convergence, a measure of the frequency of sublexical mappings, were more vulnerable to disruption. For example, the E in BRIEF was less susceptible to error than the *I* in *BRIEF* because it has a stronger sound-letter convergence. Globally this latter study showed that individual letters of digraphs are affected by the frequency of the sublexical phoneme-grapheme mappings.

In our case, the graphemic cohesion effect results from a parsing ambiguity that is resolved on the basis of the analysis of the following context. The parsing procedure in handwriting production is necessarily left to right, hence sometimes giving rise to multiple potential segmentations. In the case of the word MONTER /mɔ̃te/, activation of the simple grapheme units *O* and *N* will compete with activation of the complex grapheme unit *ON*. The complex grapheme unit *ON* will prevail only when the following context *T* will be taken into account. If the following context is a vowel (as in *monument /monymɑ̃/* "monument") or the geminate *N* (as in *monnaie /monɛ/* "currency"), *O* and *N* as separate simple graphemes should be produced. The system then has to maintain two different parses associated to units of different complexity until context disambiguates. In the case of MOULER /mule/ however, the grapheme unit *OU* will take precedence over the simple grapheme units *O* and *U* because, irrespective of the following context, chunking will always favour *OU* as a complex grapheme as its individual letters would never be phonologically realised separately. Note that such ambiguity in parsing would also lead to predict a processing cost for simple graphemes such as *A* and *N* in "*cane*" /kan/ "duck" compared to *A* and *L* in "*cale*" /kal/ "wedge" because *A* and *N* could correspond to complex graphemes in other words (as *AN* in "*cran*" /krɑ̃/) whereas *A* and *L* are always realised as two separate graphemes.

It could be argued that the duration differences in the handwriting task could be due to reading processes, i.e., the visual processing of the letter string on the screen. However, there are recent handwriting data on the role of syllable-sized chunking showing a syllabic effect both with a task involving visual input (Kandel, Álvarez, & Vallée, 2006) and with written picture naming and dictation tasks (Álvarez, Cottrell, & Afonso-Hernández, 2009) suggesting that the effects found in handwriting may not be accounted for in terms of reading processes.

We thus found converging results from the letter detection and the handwriting task raising questions concerning how reading/writing proceeds when the mapping from letters to phonemes is not straightforward. Single letters are grouped into grapheme-sized units but we have shown that the strength of the chunking depends on the potential realisation of the single letter units alone, thus making some complex graphemes more cohesive than others. Another final interesting aspect of our results concerns the processing of grapheme units with regards to their position within the word. In French *AN* and *ON* are, as we defined them, weakly cohesive graphemes when the serial position in which these letters occur is not taken into account. In word-final position however, they are always pronounced as nasal vowels, hence always parsed as a complex grapheme, and thus strongly cohesive. In the letter detection task, the complex graphemes were always located in word-final position, thus reducing the potential decomposition of the weakly cohesive units. Despite this reduction of ambiguity for weakly cohesive units, we found a difference between strongly cohesive and weakly cohesive units that suggests that the potential decomposability of graphemes is processed independently of their position within the words.

Let us now examine the consequences of our findings for models of visual word recognition and spelling. We will now consider how our results can be integrated in two influential models of visual word recognition using graphemes, namely the Dual Route Model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) and the CDP+ model (Perry, Ziegler, & Zorzi, 2007).

Our results impose new constraints for modelling. For example, within the framework of the DRC model; (Coltheart et al., 2001; Rastle & Coltheart, 1998), the graphemic cohesion effect has to be located within the nonlexical route (the processing route that converts any letter sequence into a sequence of phonemes by applying a set of grapheme-to-phoneme conversion rules). According to this model, the graphemic cohesion effect would be a consequence of the serial letter-by-letter processing performed within the nonlexical route. This explanation holding mainly for nonwords can also be extended to words, provided that these are low-frequency words. In the DRC model, words are usually processed within the lexical route where the orthographic representation of words is stored and accessed directly from the letter level. However, the two routes (lexical and nonlexical) are working in parallel and processing of low-frequency words can be influenced by processing conflicts within the nonlexical route. In our experiments, we used words of low frequency (words in Experiment 1 had a mean frequency of 6.8 occurrences per million, while words of Experiment 2 had a mean frequency of 11 occurrences per million). Given that low-frequency words are supposed to be accessed less rapidly in the lexical route, their processing might therefore be slowed down by the parsing difficulty generated in the nonlexical route with strongly cohesive complex graphemes (compared to weakly cohesive complex graphemes). Computer simulations need to be done in order to check whether the processing dynamics of the DRC model can indeed handle this graphemic cohesion effect on words (although the letter detection task is not implemented in the DRC model).

Within the connectionist dual-route model of spelling developed by Houghton and Zorzi (2003), which contains a level of grapheme representations, this latter level is explicitly linked to the notion of a graphemic buffer, which is endorsed by cognitive models of the spelling system (e.g., Caramazza, Miceli, Villa, & Romani, 1987; Ellis, 1988; Shallice, 1988). Specifically, Houghton and Zorzi (2003) assume that "graphemic representations are syllabically structured, and complex graphemes (e.g., SH, TH, EE) are locally represented" (p. 112). Perry et al. (2007) implemented the graphemic buffer of Houghton and Zorzi (2003) in their CDP+ model of reading, as the input level of the sublexical orthography-to-phonology. In their paper, they discussed the problem of how graphemic parsing is achieved. Letters must be parsed and segmented into graphemes to obtain a level of representation compatible with the graphemic buffer (i.e., the input of the sublexical orthography-to-phonology network). They propose a serial parsing of the string of letters from left to right, each individual grapheme being submitted to the sublexical network. This has the effect of serialising the sublexical route, which resembles the serial assembly of the nonlexical route of the DRC model (Coltheart et al., 2001). In their extended version adapted to disyllabic words (CDP++), Perry, Ziegler, and Zorzi (2010) propose that "whenever a letter string is presented to the model, graphemes are first identified by a graphemic parser (with complex graphemes being preferred over simpler ones whenever there is potential ambiguity)" (p. 114). The graphemic parser computes grapheme representations from the letters available and activation spreads to the phoneme units in the sublexical network, generating a plausible sublexical phonological representation. However, nothing is said in the model about graphemic cohesion resulting in parsing difficulty of some complex graphemes as the one we found in the present experiments.

The fact that our results found different processing speed between strongly cohesive complex graphemes and weakly cohesive complex graphemes has to be taken into account in the CDP+ model. As the authors put it themselves, they "did not explicitly simulate grapheme identification" (p. 282) but it would be interesting to see if the current versions of the model (CDP+ and CDP++) can simulate our empirical results (i.e., the graphemic cohesion effect) obtained in both perception and production.

Overall, our results show an effect of parsing ambiguity for complex graphemes and these results should be accounted for by current models of written word processing.

Manuscript received 27 May 2010
Revised manuscript received 3 May 2011
First published online 18 October 2011

REFERENCES

Álvarez, C. J., Cottrell, D. Y, & Afonso-Hernández, O. (2009). Writing dictated words and picture names: Syllabic boundaries affect execution in Spanish. *Applied Psycholinguistics*, *30*, 205–223.

Bogaerts, H., Meulenbroek, R. G. J., & Thomassen, A. J. W. M. (1996). The possible role of the syllable as a processing unit in handwriting. In M. L. Simner, C. G. Leedham, & A. J. W. M. Thomassen (Eds.), *Handwriting and drawing research: Basic and applied issues* (pp. 115–126). Amsterdam: IOS Press.

Brand, M., Giroux, I., Puijalon, C., & Rey, A. (2007). Syllable onsets are perceptual reading units. *Memory and Cognition*, *35*, 966–973.

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: Evidence from a case of acquired dysgraphia. *Cognition*, *26*, 59–85.

Carreiras, M., & Grainger, J. (2004). Sublexical units and the "front end" of visual word recognition. *Language and Cognitive Processes*, *19*, 321–331.

Catach, N. (1995). *L'orthographe Française* [French orthography]. Paris: Nathan Université.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.) *Strategies of information processing* (pp. 151–216). London: Academic Press.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A computational model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.

Delattre, M., Bonin, P., & Barry, C. (2006). Written spelling to dictation: Sound-to-spelling regularity affects both writing latencies and durations. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*, 1330–1340.

Ellis, A. W. (1988). Normal writing processes and peripheral acquired dysgraphias. *Language and Cognitive Processes*, *3*, 99–127.

Guinet, E., & Kandel, S. (2010). Ductus: A software package for the study of handwriting production. *Behavior Research Methods*, *42*, 326–332.

Henderson, L. (1985). On the use of the term "grapheme". *Language and Cognitive Processes*, *2*, 135–148.

Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology*, *20*, 115–162.

Jenkins, J. J., & Russel, W. A. (1952). Associative clustering during recall. *Journal of Abnormal and Social Psychology*, *47*, 818–821.

Jones, A. C., Folk, J. R., & Rapp, B. (2009). All letters are not equal: Subgraphemic texture in orthographic working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *35*, 1389–1402.

Kandel, S., Álvarez, C., & Vallée, N. (2006). Syllables as processing units in handwriting production. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 18–31.

Kandel, S., Herault, L., Grosjacques, G., Lambert, E., & Fayol, M. (2009). Orthographic vs. phonologic syllables in handwriting production. *Cognition*, *110*(3), 440–444.

Kandel, S., Soler, O., Valdois, S., & Gros, C. (2006). Graphemes as motor units in the acquisition of writing skills. *Reading & Writing: An Interdisciplinary Journal*, *19*, 313–337.

Kandel, S., & Spinelli, E. (2010). Processing complex graphemes in handwriting production. *Memory & Cognition*, *38*, 762–770.

Meulenbroek, R. G. J., & Van Galen, G. P. (1990). Perceptual–motor complexity of printed and cursive letters. *Journal of Experimental Education*, *58*, 95–110.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Lexique: Une base de données lexicales du français contemporain sur internet [Lexique: An online lexical database for modern French]. *L'Année Psychologique*, *101*, 447–462.

Orliaguet, J.-P., Kandel, S., & Boë, L.-J. (1997). Visual perception of cursive handwriting: Influence of spatial and kinematic information on the anticipation of forthcoming letters. *Perception*, *26*, 905–912.

Peereman, R., Brand, M., & Rey, A. (2006). Letter-by-letter processing in the phonological conversion of multi-letter graphemes: Searching sounds in printed pseudowords. *Psychonomic Bulletin & Review*, *13*, 38–44.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested modeling and strong inference testing in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *27*, 301–333.

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.

Rabiner, L. R., & Gold, B. (1975). *Theory and application of digital signal processing*. Englewood Cliffs, NJ: Prentice-Hall.

Rastle, K., & Coltheart, M. (1998). Whammy and double whammy: Length effects in nonword naming. *Psychonomic Bulletin and Review*, *5*, 277–282.

Rey, A., Jacobs, A. M., Schimdt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition*, *68*, 41–50.

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, *75*, B1–B12.

Royer, C., Spinelli, E., & Ferrand, L. (2005). On the status of mute letters in French: Simple graphemes or part of complex graphemes? *Current Psychology Letters*, *16*. Retrieved from http://cpl.revues.org/index465.html

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, UK: Cambridge University Press.

Tainturier, M.-J., & Rapp, B. (2004). Complex graphemes as functional spelling units: Evidence from acquired dysgraphia. *Neurocase*, *10*, 122–131.

Van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science*, *10*, 165–191.

APPENDIX 1

Material used for Experiment 1

	Strongly cohesive						Weakly cohesive					
Target letter	*Wd*	*Syll.*	*Phon.*	*Let.*	*Freq.*	*Mean Rts*	*Wd*	*Syll.*	*Phon.*	*Let.*	*Freq.*	*Mean Rts*
a	boyau 'gut'	2	5	5	4.6	518	toucan 'toucan'	2	4	6	0.1	441
a	noyau 'kernel'	2	5	5	24.5	491	bilan 'results'	2	4	5	11.4	452
a	tuyau 'pipe'	2	5	5	7.2	494	ruban 'tape'	2	4	5	11.7	416
a	fléau 'plague'	2	4	5	2.9	581	divan 'couch'	2	4	5	11.8	472
a	esquimau 'eskimo'	3	6	8	0.6	625	courtisan 'courtier'	3	7	9	1	470
a	étau 'vice'	2	3	4	2.5	553	flan 'flan'	1	3	4	1.8	445
a	joyau 'jewel'	2	5	5	1.2	449	brelan 'brelan'	2	5	6	0.4	510
a	préau 'yard'	2	4	5	2	544	boucan 'din'	2	4	6	1.3	468
o	flou 'fuzzy'	1	3	4	6.3	437	thon 'tuna'	1	2	4	2.4	447
o	acajou 'mahogany'	3	5	6	5.6	536	écusson 'escutcheon'	3	5	7	1.4	479
o	bambou 'bamboo'	2	4	6	2.6	492	melon 'melon'	2	4	5	3.6	405
o	genou 'knee'	2	4	5	14.6	425	talon 'heel'	2	4	5	8.6	467
o	bisou 'kiss'	2	4	5	0.3	530	jupon 'petticoat'	2	4	5	2.3	525
o	écrou 'nut'	2	4	5	1.9	465	salon 'lounge'	2	4	5	49.4	449
o	verrou 'lock'	2	4	6	5	493	carton 'cardboard'	2	5	6	20.7	411
o	bijou 'jewel'	2	4	5	6.8	516	gazon 'lawn'	2	4	5	4.3	428
	Mean	2.1	4.3	5.3	5.5	509	Mean	2	4.2	5.5	8.2	455

Wd, carrier word; Syll., number of syllables; Phon., number of phonemes; Let., number of letters; Freq., number of occurrences per million.

APPENDIX 2

Material used for Experiment 2

Strongly cohesive							*Weakly cohesive*						
Word	*Syll.*	*Let.*	*Freq.*	*Bigr. Freq.*	*Prec. Let.*	*1st let.*	*Word*	*Syll.*	*Let.*	*Freq.*	*Bigr. Freq.*	*Prec. Let.*	*1st let.*
mouler 'to mold'	2	6	0.47	3,386	131	137	monter 'go up'	2	6	96.89	3,386	141	148
coudre 'to sew'	1	6	8.65	9,708	214	149	contre 'against'	1	6	5.74	9,708	225	147
foudre 'lightning'	1	6	12.64	1,995	112	139	fondre 'to melt'	1	6	17.7	1,995	113	144
poudre 'powder	1	6	27.57	4,235	131	138	pondre 'to lay'	1	6	1.69	4,235	106	134
mourir 'to die'	2	6	130.61	3,386	127	140	bondir 'to pounce'	2	6	9.05	2,934	113	136
coupage 'cutting'	2	7	0.14	9,708	220	140	pontage 'bypass'	2	7	0.07	4,235	109	133
couvert 'covered'	2	8	17.09	9,708	220	143	concert 'concert'	2	7	24.86	9,708	214	140
moutarde 'mustard'	2	8	4.93	3,386	123	138	montagne 'mountain'	2	8	49.8	3,386	122	135
coureuse 'runner'	2	8	6.28	4,346	218	143	conteste 'contests'	2	8	1.01	9,708	213	139
touriste 'tourist'	2	8	1.08	9,708	149	123	tondeuse 'mower'	2	8	1.49	4,346	134	136
boulimie 'bulimia'	3	8	0.47	2,934	113	139	mongolie 'mongolia'	3	8	0.07	3,386	127	137
lourdeur 'heaviness'	2	8	4.86	4,884	117	140	jongleur 'juggler'	2	8	0.74	676	122	140
souligner 'underline'	3	9	1.65	3,703	108	137	consigner 'record'	3	9	1.55	9,708	211	141
gourmande 'greedy'	2	9	1.22	1,693	130	150	constance 'constancy'	2	9	2.84	9,708	214	141
couvrante 'covering'	2	9	1.49	9,708	212	140	constante 'constant'	2	9	0.41	9,708	225	143
jaunir 'to yellow'	2	6	0.68	454	129	115	bandit 'gangster'	2	6	4.59	3,006	119	115
sauter 'to jump'	2	6	43.31	6,940	108	110	vanter 'to praise'	2	6	7.5	2,562	134	113
trauma 'trauma'	2	6	0.47	19,342	114	121	planté 'planted'	2	6	3.04	8,528	105	113
clause 'clause	1	6	1.01	8,528	112	107	transe 'trance'	1	6	3.51	19,342	100	116
saumon 'salmon'	2	6	3.65	6,940	113	113	pantin 'dummy'	2	6	4.19	4,657	140	117
nausée 'nauséa'	2	6	7.91	5,838	118	113	lancée 'launched'	2	6	3.78	8,528	118	113
sauver 'to save'	2	6	36.89	6,940	110	111	danser 'to dance'	2	6	35.41	2,272	138	115
dauphin 'dolphin'	2	7	1.22	2,272	142	112	manchon 'muff'	2	7	1.35	5,581	125	116

Appendix 2 (*Continued*)

Strongly cohesive							*Weakly cohesive*						
Word	*Syll.*	*Let.*	*Freq.*	*Bigr. Freq.*	*Prec. Let.*	*1st let.*	*Word*	*Syll.*	*Let.*	*Freq.*	*Bigr. Freq.*	*Prec. Let.*	*1st let.*
laurier 'laurel'	2	7	3.58	8,528	123	112	janvier 'january'	2	7	30.61	454	124	113
exaucer 'to fulfill'	3	7	0.81	278	128	111	oranger 'orange tree'	3	7	2.09	19,342	114	114
nautisme 'boating'	2	8	0.06	5,838	119	111	fantasme 'fantasy'	2	8	1.49	1,697	107	114
sauteuse 'jumper'	2	8	0.95	6,940	106	109	mangeuse 'eater'	2	8	0.27	5,581	120	118
fraudeur 'fraudster'	2	8	0.07	19,342	120	115	grandeur 'greatness'	2	8	26.49	19,342	117	111
fournil 'bakehouse'	2	7	3.92	1,995	124	141	nombril 'navel'	2	7	5.2	2,881	128	139
boucler 'to curl'	2	7	6.55	2,934	118	143	combler 'to fill in'	2	7	7.97	9,708	219	140
crouler 'to collapse'	2	7	2.91	9,834	116	144	plomber 'to seal'	2	7	0.54	4,884	109	144
grouper 'to group'	2	7	2.03	9,834	121	146	tromper 'to deceive'	2	7	22.16	9,834	116	138
soulever 'to lift'	3	8	17.16	3,703	106	138	composer 'to compose'	3	8	7.3	9,708	225	140
mouliner 'to mill'	3	8	0.47	3,386	120	137	combiner 'to combine'	3	8	1.55	9,708	228	144
soulager 'to relieve'	3	8	6.49	3,703	110	141	comparer 'to compare'	3	8	8.11	9,708	222	144
couvreur 'slater'	2	8	38.65	3,703	219	143	combatif 'combative'	3	8	0.54	9,708	228	142
souvenir 'memory'	2	8	0.74	9,708	113	144	nombreux 'numerous'	2	8	37.03	2,881	118	139
groupage 'grouping'	2	8	0.14	9,834	117	138	trombone 'trombone'	2	8	0.54	9,834	115	139
rouspéter 'to grouch'	3	9	0.27	9,834	125	137	comprimer 'to squeeze'	3	9	0.61	9,708	215	140
boulevard 'avenue'	2	9	52.03	2,934	113	136	compétent 'competent'	3	9	1.08	9,708	226	141
Mean	2.08	7.33	11.28	6,301.75	134	131		2.13	7.3	10.77	7,149.73	152	132

Syll., number of syllables; Let., number of letters; Freq., number of occurrences per million; Bigr. Freq., bigram frequency; Prec. Let, normalised durations of the letter preceding the grapheme; 1st Let., normalised durations of the first letter of the grapheme.

APPENDIX 3

Procedure for normalisation used in Experiment 2. Writing durations presented in Figure 1 refer to durations of movements while the pen is on the surface of the digitiser. The duration of the "in air" movements to produce the second stroke of T, for example, is not considered in the calculation of duration. This means that the duration for L refers to the duration of two strokes (vertical line + horizontal line in the bottom) and the duration for T refers to the duration of two strokes (vertical line + horizontal line in the top). For the letters that are composed of straight lines (e.g., *N*, *L*, *T*, *F*, *E*), we considered each line to be one stroke. The absolute duration was thus divided by the number of strokes composing the letter. Thus given durations such as $N = 240$ ms and $L = 160$ ms, we did 240/3 = 80 and 160/2 = 80. The resulting stroke durations were 80 ms for both letters. For the letters with curved traces, such as *C*, *S*, *R*, *O* we used a stroke segmentation procedure that consists of segmenting the continuous trace according to tangential velocity minimal values. Ten productions of each letter were recorded and segmented on the basis of the velocity minima in the tangential velocity profile. In these productions, *S* for instance, always presented three strokes. So if in our experiment the participant produced *S* in 300 ms, the stroke duration was 300/3 = 100 ms. Note that the direction of the movement—i.e., clockwise (e.g. D) or counter-clockwise (e.g. C)—as well as the anticipation of forthcoming strokes while producing a curved line (e.g. the diagonal line of the R) are factors that produce a decrease in velocity. Strokes are defined by the presence of two consecutive velocity minima. This is the reason why C for example has one stroke and S has three.

Number of strokes for each letter of the alphabet that we used for the normalisation.

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Number of strokes	3	5	1	3	4	3	4	3	1	3	3	2	4	3	2	2	3	5	3	2	2	2	4	2	2	3