Applied Psycholinguistics (2025), **46**, e21, 1–20
doi:10.1017/S0142716425100106

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

French is not so easy to decode: a pilot study

Anne-Françoise de Chambrier[1], Myrto Atzemian[1], Michel Fayol[2], Pascal Zesiger[3] and Catherine Martinet[1]

[1]Special Needs Education Unit, University of Teacher Education of State of Vaud, Lausanne, Switzerland, [2]LAPSCO CNRS, University of Clermont Auvergne, Clermont, France and [3]Psycholinguistic and Speech Therapy, University of Geneva, Geneva, Switzerland
Corresponding author: Anne-Françoise de Chambrier; Email: anne-francoise.de-chambrier@hepl.ch

(Received 16 November 2023; revised 30 January 2025; accepted 20 June 2025)

Abstract
Although easier to read than English, French has several inconsistent grapheme-to-phoneme correspondences (GPCs) whose impact on decoding performance has been little studied. In the current pilot study, 27 adult participants were asked to read aloud 60 pseudowords containing the ambiguous adjacent letters "an," "on," and "in"; the contextual graphemes "g," "s," and "e"; and the final consonants "d," "p," "s," and "t"; as well as 60 matched control pseudowords without these characteristics. Results indicated that the grapheme "e" corresponding to /ə/; the final consonants meant to be silent; the grapheme "s" corresponding to /z/; the graphemes "an," "on," and "in" corresponding to /ɑ̃/, /ɔ̃/, and /ɛ̃/; and the grapheme "g" corresponding to /ʒ/ gave rise to more unexpected answers than their respective control pseudowords. The unexpected answers seem to be explained by dominant rules partly moderated by the position of the GPC in the pseudowords. These findings highlight that the difficulty of decoding French should not be underestimated and suggest that such GPCs might be the subject of particular educational attention.

Keywords: Decoding; expert readers; French orthographic system; inconsistent graphemes; sublexical phonological procedure

French is not so easy to decode: a pilot study

Learning to decode printed words is an important part of learning to read (e.g., Castles et al., 2018), and it is more difficult in some writing systems than in others. With reference to the dual route model (Coltheart et al., 2001; Ziegler et al., 2014), words can be read either by the sublexical phonological procedure or by the lexical procedure. Within the sublexical phonological procedure, small written units (letters, graphemes, syllables) are converted into their phonological correspondents, whereas in the lexical procedure, written words are recognized as a whole and linked to their oral form and meaning stored in the internal lexicon. The sublexical

The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

phonological procedure is used very predominantly at the beginning of learning to read, and its proper development is crucial. Indeed, its efficiency strongly supports the acquisition of reading and spelling since orthographic representations of words are gradually acquired as they are successfully decoded (Martinet et al., 2004; Share, 1995). The ease with which children identify words is also the most important predictor of written comprehension up to the middle of primary school, the identification of written words being the *sine qua non* of reading comprehension (Share, 1995; Vellutino et al., 2007). Moreover, the main source of difficulties in learning to read is problems in the sublexical procedure (Gentaz et al., 2015; Spencer et al., 2014), while difficulties in decoding are more frequent in less consistent orthographic systems (Seymour et al., 2003; Ziegler & Goswami, 2005). Taken together, these observations suggest that the grapheme-to-phoneme conversion rules that are most likely to cause challenges in a particular orthographic system should be precisely identified.

The French orthographic system has many particularities, which have led it to be described as "the least Latinate of all Romance languages" (Posner, 1996, p. 245, our translation). With 130 graphemes for about 36 phonemes, it is an asymmetric orthographic system, being easier in the grapheme-to-phoneme direction than in the phoneme-to-grapheme one (Jaffré & Fayol, 2005). Even so, it is one of the most difficult writing systems to read after English. In a cross-language study, Seymour et al. (2003) recorded the percentage of words correctly read by pupils after 1 year of reading instruction in 13 orthographies and ranked French as a language of intermediate difficulty (alongside Danish and Portuguese), even though the items were frequent words that did not comprise the various inconsistencies of this orthographic system.

One of the main specificities of the French orthographic system is its high number of multiletter graphemes (i.e., groups of letters corresponding to one phoneme). Most of them are composed of two letters ("ou," "ch," "oi," "ai," "au," "an," "in," "on," etc.) and some of three letters ("eau," "ain," or "ein"). Some of them are highly cohesive, meaning that they always correspond to a grapheme unit in print (Chetail, 2020). For example, "au" systematically corresponds to a single grapheme and thus to one phoneme (/ɔ/). However, a considerable amount of French multiletter graphemes are ambiguous or low cohesive (Commissaire et al., 2018). For example, "an" can correspond either to one grapheme (i.e., /ɑ̃/ such as in "rang," /rɑ̃/—"rank"; "blanc," /blɑ̃/—"white") or to two simple graphemes when followed by a vowel or the duplicated "n" (i.e., /a/-/n/ such as in "canard," /kanar/—"duck"; "banir," /banir/—"to ban"; "année," /ane/—"year"). A few words are composed of two or even three of such ambiguous sequences of letters ("banane," /banan/—"banana"; "cinema," /sinema/—"cinema"; "inanimé," /inanime/—"inanimate"). Such multiletter graphemes thus represent difficulties of orthographic segmentation into graphemes beyond those of converting graphemes into phonemes.

French also has a considerable number of contextual graphemes, i.e., graphemes that correspond to different phonemes according to the surrounding graphemes. Among the most frequent are "c" and "g," which are also contextual in other languages such as English, Italian, or Spanish. In French, they respectively correspond to /k/ and /g/ when followed by "a," "o," "u," "r," or "l" (such as in

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

"cave," /kav/—"cellar"; clou, /klu/—"nail"; "goût," /gu/—"taste"; "glace," /glas/—"ice cream") and to /s/ and /ʒ/ when followed by "e," "i," or "y" (such as in "cil," /sil/—"lash"; "cerf," /sɛr/—"deer"; "gilet," /ʒilɛ/—"vest"; "geler," /ʒəle/—"freeze"). The letter "s" is another contextual grapheme in French, corresponding to /z/ between two vowels (such as in "visage," /vizaʒ/—"face") but to /s/ at the beginning of words or between a vowel and a consonant (such as in "soupe," /sup/—"soup"; "veste," /vɛst/—"jacket"). Less often presented as such, the grapheme "e" inside words is also contextual (Afonso et al., 2015; Léon et al., 2009). It makes /ɛ/ in almost all cases where it is followed by two consonants ("veste," /vɛst/—"jacket"; "ferme," /fɛrm/—"a farm"; "persil," /pɛrsil/—"parsley"). On the opposite side, it makes /ə/ when followed by a single consonant ("devoir," /dəvwar/—"homework"; "melon," /məlɔ̃/—"melon"; "cerise," /səriz/—"cherry").

An additional challenge with French is its final consonants, which are largely unstable (Perry et al., 2014). Some final consonants are mostly sounding ("l," 99.2%, "cheval," /ʃəval/—"horse"; "c," 92%, "lac," /lak/—"lake"; Peereman et al., 2007), while others are almost always silent ("t," 99.5%, "bout," /bu/—"piece"; "p," 98.8%, "trop," /tro/—"too much"; "d," 98.7%, "tard," /tar/—"late"; "s," 96.1%, "souris," /suri/—"mouse"). Many silent final consonants have a function at the derivational level ("plat," "plateau"—"flat," "tray"), and there can also be two final silent letters at this word level ("temps"—"time"; "instinct"—"instinct"). One to three silent final letters also often appear at the inflectional level ("il fait"—"he does"; "des fourmis"—"ants"; "tu prends"—"you take," "ils trouvent"—"they find"). This is due to the fact that French inflectional morphology is predominantly silent. Indeed, while written French resembles other Romance languages, spoken French has undergone its own evolution, characterized in particular by an influence from Germanic languages (Barra-Jover, 2009), making it a particularly complex written language to acquire morphosyntactically (Ågren, 2016).

The consequences of these specificities on decoding performance are not sufficiently documented. First, for multiletter graphemes, results are not consistent as regards the automaticity with which they are processed. On the one end, studies showed that English and French adults were slower to detect a target letter in a word when the target letter was embedded in a multiletter grapheme (i.e., "U" in "LOUPE") than when it corresponded to a single-letter grapheme (i.e., "U" in "CHUTE"; Rastle & Coltheart, 1998; Rey et al., 2000), showing that such graphemes are processed as perceptual units. Among French children from Grades 1 to 4, Sprenger-Charolles et al. (2005) showed that two-letter graphemes (ch, ou, on, etc.) did not generate more mistakes or longer latencies than those with one-letter graphemes only, leading the authors to conclude that multiletter graphemes are processed as perceptual units from early on. On the other hand, using five experimental tasks among French adults, Chetail (2020) found no reliable grapheme effect, supporting the claim that graphemes are not perceptual units in skilled visual word recognition. For their part, Spinelli et al. (2012) highlighted that graphemic effects depend on the grapheme cohesion. In a letter decision task, they found that A was detected faster in weakly cohesive complex graphemes (e.g., AN) than in strongly cohesive ones (e.g., AU) (see also Commissaire and Casalis, 2018). Furthermore, studies highlighted that expert readers are significantly slower in reading pseudowords containing multiletter graphemes than only simple ones, an

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

effect known as the graphemic complexity effect (Joubert & Roch Lecours, 2000; Rastle & Coltheart, 1998; Rey et al., 1998; Rey & Schiller, 2005). Results are therefore inconsistent on the ease with which multiletter graphemes are processed, and data are above all lacking regarding the possible decoding challenge that low-cohesive multiletter graphemes, which are prominent in French, might represent.

Regarding contextual graphemes, most studies have been conducted in English on the reading of adjacent vowels since their spelling-sound relations are particularly complex in that orthography (Kessler & Treiman, 2001; Treiman et al., 2006). A few English studies still investigated the context use for the consonants "c" and "g" (Treiman & Kessler, 2019; Treiman et al., 2007). Treiman and Kessler (2019) found that readers from early elementary school to university were not as influenced by the context as would be expected given the contextual effects in the English vocabulary and concluded that the use of context developed slowly. For example, for "c" preceded by "e" or "i," the percentage of front pronunciations (i.e., /s/) increased from 17% for students from first- to third-grade level to 79% for students from a post-high school level. The contextual consonants "c" and "g" are, however, not as consistent in English as they are in French, even though the alternation of pronunciation of these graphemes came from French, borrowed in the medieval period (Emerson, 1997). Indeed, in English, the letter "g" before "e" or "i" is usually pronounced /g/ at the beginning of non-Latinate words (e.g., get, give) and /ʒ/ in Latinate vocabulary (e.g., genetic, gingivitis) (Treiman & Kessler, 2019), while it is always pronounced /ʒ/ before these letters in French. Despite this higher consistency, Sprenger-Charolles et al. (2005) found among French children that words containing "c" and "g" were read less accurately than words without any contextual grapheme up to Grade 4. Similarly, Alegria and Mousty (1994) showed that in the direction of writing, the contextual letters "g" and "s" caused many difficulties for second graders with and without reading difficulties. In Italian, in which the spelling-sound relations for these graphemes are similar to French, it has been shown that 3rd and 5th graders read low-frequency words containing "c" or "g" more slowly and less accurately than words containing simple rules (see also Burani et al., 2006). Together, these observations show that the decoding challenges generated by French contextual graphemes need further study, especially those for which we found no data, such as "s" or "e."

Finally, few data are available for final consonants. Royer et al. (2005) showed that expert French readers detected a silent final consonant more quickly ("t" in "chat") than letters embedded in multiletter graphemes ("i" in "quai"), suggesting that silent letters are processed as single-letter graphemes rather than being bonded to their preceding vowel. In their connectionist model of reading developed for French, Perry et al. (2014) reported that inconsistencies with silent letters were hard to process, that the activation produced by words with these properties was higher, and that the model produced slower-than-average reaction times in these cases. They also found that the adult participants mainly pronounced the silent consonants, a tendency that was faithfully reproduced by their model. Since this propensity was found on monosyllabic pseudowords and since monosyllabic words are particularly inconsistent in French (Ziegler et al., 1996), the decoding challenges that final consonants might represent should be further studied on other types of items.

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

The current study

Therefore, to find out whether these inconsistent and insufficiently studied French graphemes cause decoding challenges to typical readers, adults were required to read aloud pseudowords containing the ambiguous adjacent letters "an," "on," and "in"; the contextual graphemes "g," "s," and "e"; and the final consonants "d," "p," "s," and "t," as well as matched control pseudowords without them. Comparing the performance of adults reading these two kinds of pseudowords allowed us to test to what extent the conversion rules of the inconsistent graphemes are automatized. Indeed, automatized GPC rules for the inconsistent scenarios should result in targeted pseudowords read as accurately and as quickly as control pseudowords without inconsistent graphemes. Our hypothesis is, however, that when adults have to convert graphemes into phonemes, pseudowords containing inconsistent correspondences will give rise to more unexpected answers and to longer latencies than control pseudowords.

Method

Participants

This pilot study was planned to be conducted among students from our university, but this was not possible because of the COVID pandemic. Therefore, adults were recruited in the authors' neighborhood, having strictly followed the hygiene rules in force. Twenty-seven participants (20 women and 7 men) took part in the study. They were 35 years old on average ($SD = 12.7$; min = 16 years old; max = 57 years old). They were all native French speakers from the French-speaking part of Switzerland and all learned to read in French, reported no history of learning disorders, and all achieved or were about to achieve a tertiary level of education. They received a 15-franc (14 GBP) voucher to thank them for their participation.

Material

Sixty pseudowords containing the graphemes of interest were created. The targeted pseudowords were matched to sixty control pseudowords that were similar in every way, except that the graphemes under study were replaced by stable graphemes corresponding to a phoneme of similar pronunciation length. Beyond that constraint, sequences of letters surrounding the graphemes of interest were replaced by sequences of letters of as comparable frequency as possible according to the Manulex infra database (Peereman et al., 2007; see Appendix 1). The frequency of the sequences of letters in the target and control pseudowords was controlled because sequences of letters make the occurrence of a given letter (or grapheme) more or less predictable (Pacton et al., 2005). All pseudowords were composed of three syllables. The position of the graphemes of interest varied between the beginning, middle, and end of the pseudoword. The positions were the same in each pair of target and control pseudowords, and in all categories of graphemes, the number of each possible position was the same.

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

More precisely:

- To test the possible difficulty generated by the graphemes "in," "on," or "an" corresponding to /ɛ̃/, /ɔ̃/, or /ɑ̃/ (when followed by a consonant), six pseudowords were created. In the control pseudowords, the graphemes "in," "on," or "an" were replaced by a stable multiletter grapheme corresponding to a vocalic phoneme ("au" or "ou") by choosing sequences of three letters whose frequency was as similar as possible. For example, the pseudoword "bonfivule" was matched to "baufivule," and "létonvir" to "létauvir."
- The criteria were the same for the sequences of letters "in," "on," or "an" corresponding to /i/-/n/, /ɔ/-/n/, or /a/-/n/ (when followed by a vowel). In the control pseudowords, the sequences of letters "in," "on," or "an" were replaced by a stable sequence of letters corresponding to one vocalic and one long phoneme ("ir," "aj," "il," "of," "or," "ar") by choosing sequences of three letters whose frequency was as similar as possible. For example, the pseudoword "binavule" was matched to "biravule," and "boutanore" to "boutajore."
- To test the possible difficulty generated by the subcase of two following sequences of letters, such as "in," "on," "an," "im," "om," or "am," meant to be respectively read /i/-/n/, /ɔ/-/n/, /a/-/n/, /i/-/m/, /ɔ/-/m/, and /a/-/m/ (when followed by a vowel), three pseudowords were created. In the control pseudowords, the sequences of letters were replaced by two stable sequences of letters corresponding to one vocalic and one long phoneme ("or," "av," "il," "ur," "az," "ir") by choosing sequences of letters that were as frequent as possible. For example, the pseudoword "binumate" was matched to "bilurate," and "lomanube" to "loravube."
- To test the possible difficulty generated by the grapheme "s" inside a word corresponding to /s/ (between one vowel and one consonant), six pseudowords were created. In the control pseudowords, the grapheme "s" was replaced by a stable grapheme corresponding to a long phoneme ("r," "l") by choosing sequences of three letters around them whose frequency was as similar as possible (see Appendix 1). For example, the pseudoword "misbudole" was matched to "mirbudole," and "nafuspé" to "nafulpé."
- The criteria were the same to test the possible difficulty generated by the grapheme "s" corresponding to /z/ (between two vowels). For example, the pseudoword "posido" was matched to "porido," and "bilusore" to "bilufore."
- To test the possible difficulty generated by the grapheme "e" inside a word corresponding to /ɛ/ (when in the middle of a syllable, generally followed by two consonants), six pseudowords were created. The grapheme "e" was in the first syllable in half of the items and in the second syllable in the other half of the items. In the control pseudowords, the grapheme "e" was replaced by a stable grapheme corresponding to a vocalic phoneme ("a," "i," or "o") by choosing sequences of three letters around them whose frequency was as similar as possible. For example, the pseudoword "neltari" was matched to "noltari," and "chudelpa" to "chudilpa."
- The criteria were the same to test the possible difficulty generated by the grapheme "e" corresponding to /ə/ (at the end of an initial or median syllable, generally followed by a single consonant). For example, the pseudoword "refapile" was matched to "rofapile," and "chadelu" to "chadilu." For some of

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

the targeted items (buvedire, chadelu, vateril), the "e" could also be silent (schwa), in which cases such answers were also considered correct.

- To test the possible difficulty generated by the grapheme "g" corresponding to /g/ (when followed by "a," "o," "u," "r," or "l"), six pseudowords were created. In the control pseudowords, the grapheme "g" was replaced by a stable grapheme corresponding to an occlusive phoneme ("t," "b," "p," or "d") by choosing sequences of three letters around them whose frequency was as similar as possible. For example, the pseudoword "mogalou" was matched to "mobalou," and "tolugar" to "tolupar."
- The criteria were the same to test the possible difficulty generated by the grapheme "g" corresponding to /ʒ/ (followed by "e" or "i"). In the control pseudowords, the grapheme "g" was replaced by a stable grapheme corresponding to a long phoneme ("f," "r," "v"). For example, the pseudoword "bégilor" was matched to "béfilor," and "tapogé" to "tapofé."
- To test the possible difficulty generated by the final consonants "d," "p," "s," and "t," which are most often silent in French, nine pseudowords were created. In the control pseudowords, the final consonants were replaced by the silent final vowel "e" (always silent in French) by choosing final sequences of letters that were as frequent as possible. For example, the pseudoword "firédoup" was matched to "firédoue," and "nujélard" to "nujélare."

Procedure

The task was created using the E-prime 3 software. The items were presented in a pseudorandomized order. More precisely, eight lists were created with the pseudowords in a different order, having ensured that the two pseudowords of the same pair were separated by at least five items to minimize the priming effect (Stark & McClelland, 2000). The pseudowords were individually displayed for 500 ms in the middle of a 15-inch computer screen in Arial font, lowercase letters, and size 40. They were preceded by a "+" fixation point displayed for 500 ms in the center of the screen and followed by a mask composed of eight capital "X," covering the place occupied by the pseudowords, displayed for 1,500 ms. The experiment began with five practice items. Participants were individually tested in a quiet room, comfortably seated on a chair approximately 40 cm from the screen. They were told that meaningless words would be presented to them and were required to read them aloud, as quickly and as accurately as possible, as if they were French words. The experiment lasted approximately six minutes per participant. Answers were considered correct if the pseudowords were read as expected according to the absolute or largely grapheme-to-phoneme correspondences (GPC) rules. An interjudge agreement was completed on a third of the data (nine participants) and yielded 96.3% agreement. Latencies (Perfetti, 1985) were registered in ms through a voice key (Cedrus SV-1) on the basis of the pseudowords that were read as expected.

Results

Since most of the distributions were not Gaussian, Wilcoxon tests (for paired samples) were computed on the accuracy and reaction time (RT) scores for each scenario of the

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

Table 1. Mean accuracy and RT for each grapheme-to-phoneme correspondences (GPC) scenario and its control pseudowords and differences between them

Pseudowords with inconsistent graphemes			Control pseudowords		Differences					
	Accuracy (M, SD)	RT (M, SD)	Accuracy (M, SD)	RT (M, SD)	Accuracy			RT		
s = /s/	0.80 (0.15)	0.90 (0.17)	0.88 (0.14)	0.88 (0.11)	W = 3	z = –2.051	p = 0.040	t = –1.038		p = 0.154
s = /z/	0.64 (0.21)	0.79 (0.14)	0.88 (0.15)	0.82 (0.13)	W = 0	z = –4.065	$p < 0.001$**	t = –2.186		p = 0.019
e = /ɛ/	0.91 (0.12)	0.86 (0.13)	0.91 (0.13)	0.87 (0.17)	W = 6	z = –0.182	p = 0.856	t = –0.470		p = 0.321
e = /ə/	0.13 (0.16)	0.91 (0.37)	0.95 (0.10)	0.84 (0.15)	W = 0	z = –4,599	$p < 0.001$**	t = –1.076		p = 0.151
g = /g/	0.91 (0.15)	0.85 (0.15)	0.92 (0.16)	0.84 (0.15)	W = 4	z = –0.160	p = 0.873	t = –0.313		p = 0.378
g = /ʒ/	0.81 (0.22)	0.80 (0.16)	0.96 (0.11)	0.78 (0.13)	W = 3	z = –2.620	p = 0.009*	t = –1.543		p = 0.068
on, in, an = /ɔ̃/, /ɛ̃/, /ɑ̃/	0.74 (0.24)	0.87 (0.20)	0.94 (0.11)	0.87 (0.17)	W = 0	z = –3.555	$p < 0.001$**	W = 14	z = 0	p = 1.0
on, in, an = /o/-/n/, /i/-/n/, /a/-/n/	0.86 (0.15)	0.90 (0.13)	0.92 (0.11)	0.89 (0.17)	W = 2	z = –2.366	p = 0.018	t = –0.399		p = 0.347
Two sequences of on, in, an . . .	0.84 (0.17)	0.84 (0.17)	0.91 (0.18)	0.79 (0.13)	W = 4	z = –1.414	p = 0.157	t = –2.653		p = 0.007*
Final d, p, s, t	0.50 (0.18)	0.92 (0.19)	0.95 (0.08)	0.91 (0.16)	W = 0	z = –4.563	$p < 0.001$**	t = –0.598		p = 0.277

Note: Mean accuracy (maximum = 1 pt) and RTs (in s) scores (and SD) for pseudowords with inconsistent graphemes as well as for matched control pseudowords, and differences between each pair of pseudowords. With Bonferroni correction, differences are significant at $p < .005$ and tendentially significant at $p < .01$.

inconsistent graphemes. Results are displayed in Table 1, and all data are available at https://osf.io/4a7st/. Regarding accuracy, among the significant differences (with Bonferroni correction) between the pseudowords with inconsistent graphemes and the control pseudowords, the error rates were the highest for the grapheme "e" corresponding to /ə/ ("refapile," "chadelu"; $p < .001$, 87% of unexpected answers relative to 5% for the control pseudowords); for the final consonants meant to be silent ("firédoup," "nujélard"; $p < .001$, 50% of unexpected answers relative to 5% for the control pseudowords); for the grapheme "s" corresponding to /z/ ("posido," "bilusore"; $p < .001$, 36% of unexpected answers relative to 12% for the control pseudowords); and for the graphemes "an," "on," and "in" corresponding to /ɑ̃/, /ɔ̃/, and /ɛ̃/ ("bonfivule," "néfantile"; $p < .001$, 26% of unexpected answers relative to 6% for the control pseudowords). A trend toward significance was found for the grapheme "g" corresponding to /ʒ/ ("bégilor," "tapogé"; $p = .009$, 19% of unexpected answers relative to 4% for the control pseudowords). Conversely, the differences were not significant for the grapheme "s" corresponding to /s/ ("misbudole," "nafuspé"; $p = .04$, 20% of unexpected answers relative to 12% for the control pseudowords); for the sequences of letters "an," "on," and "in" followed by a vowel ("binavule," "futinope"; $p = .018$, 14% of unexpected answers relative to 8% for the control pseudowords); for the two following sequences of letters such as "in," "on," "an," "im," "om," and "am" ("binumate," "lomanube"; $p = .157$, 16% of unexpected answers relative to 9% for the control pseudowords); for the grapheme "e" corresponding to /ɛ/ ("neltari," "chudelpa"; $p = .856$, 9% of unexpected answers relative to 9% for the control pseudowords); and for the grapheme "g" corresponding to /g/ ("mogalou," "tolugar"; $p = .873$, 9% of unexpected answers relative to 8% for the control pseudowords).

Mean RTs were calculated for each scenario on the basis of the participants' RTs for correct answers. Only a trend toward significance was found for the following sequences of "on," "in," and "an" letters. Because latencies are more sensitive to the beginning of items (Sambai et al., 2018), and some targeted graphemes were at the end of the items in the current study, we checked whether differences were significant when considering only the latencies on pseudowords with inconsistencies at the beginning of the items, but this was not the case.

Discussion

The current pilot study was aimed at testing whether French graphemes following inconsistent conversion rules, such as the ambiguous sequences of letters "in," "on," and "an"; the contextual graphemes "g," "s," and "e"; and the final consonants "d," "p," "s," and "t," generated decoding challenges among expert readers.

In line with our hypothesis, many pseudowords with inconsistent graphemes generated more unexpected answers than matched pseudowords without them (Table 1). The scenarios that generated significant differences were, from the highest to the lowest rate of unexpected answers, the grapheme "e" corresponding to /ə/; the final consonants meant to be silent; the grapheme "s" corresponding to /z/; and the graphemes "an," "on," and "in" corresponding to /ɑ̃/, /ɔ̃/, and /ɛ̃/. A trend toward significance was also found for the grapheme "g" corresponding to /ʒ/. Conversely, there was no significant difference for the grapheme "s" corresponding to /s/, for the sequences of letters "an," "on," and "in" followed by a vowel, for the grapheme "e"

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

corresponding to /ɛ/, for the grapheme "g" corresponding to /g/, nor for two following sequences of letters such as "in," "on," and "an." Differences were also not significant for RTs.

Thus, the ambiguous sequences of letters "in," "on," and "an" generated more unexpected answers than their matched pseudowords only when they corresponded to the low-cohesive complex graphemes /ɑ̃/, /ɔ̃/, and /ɛ̃/, not when they were two simple adjacent graphemes. There were not either significantly more unexpected answers when two sequences of such simple graphemes followed each other ("binumate," "lomanube"). Therefore, for these ambiguous adjacent letters that are prominent in French, it is when they appear as low-cohesive complex graphemes that they represent a decoding challenge, not when they correspond to two simple graphemes.

The current results also highlighted that the three contextual graphemes under study ("s," "e," and a trend toward significance for "g") represent decoding challenges. The rate of unexpected answers seems to be explained by dominant rules (Alegria & Mousty, 1996), since the scenarios that generated more unexpected answers were the less frequent GPC. Indeed, the fact that the grapheme "e" corresponding to /ə/ generated more unexpected answers than the grapheme "e" corresponding to /ɛ/ is in line with the fact that this grapheme more often corresponds to /ɛ/ than to /ə/; the fact that the grapheme "s" corresponding to /z/ gave rise to more unexpected answers than the grapheme "s" corresponding to /s/ is in line with the fact that this grapheme more often corresponds to /s/ than to /z/; and the fact that the grapheme "g" tended to generate more unexpected answers when corresponding to /ʒ/ than to /g/ is in line with the fact that this grapheme more often corresponds to /g/ than to /ʒ/. However, the percentages of unexpected answers across the scenarios of the same grapheme are not proportional to their respective frequency. For example, the percentage difference of unexpected answers between the grapheme "s" corresponding to /z/ and their control pseudowords (64% vs 88% = 24%) is three times higher than the percentage difference of unexpected answers between the grapheme "s" corresponding to /s/ and their control pseudowords (80% vs 88% = 8%). Yet, the grapheme "s" is in total almost six times more frequently associated with the phoneme /s/ than with /z/ (Peereman et al., 2007). This lack of proportionality between the percentage of unexpected answers and the frequency of the respective GPC might be explained by the fact that inside words (the place where the graphemes of interest were placed in the current study), the grapheme "s" more often corresponds to /z/ than to /s/. The same applies to the grapheme "g": its correspondence to the phoneme /ʒ/ gives rise to largely more additional unexpected answers (15%) than its correspondence to the phoneme /g/ (1%) relative to the control pseudowords, while this grapheme is only associated 1.2 times more often with the phoneme /g/ than with /ʒ/ in total. This might be explained by the fact that inside words, "g" more often corresponds to /ʒ/ than to /g/. Therefore, it appears that the grapheme-to-phoneme conversions selected by readers firstly depend on the global GPC rule frequencies and that they are secondarily moderated by the GPC rule frequencies depending on the position of the graphemes in the words.

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

Lastly, the pseudowords with final consonants also generated significantly more unexpected answers than their matched pseudowords (50% versus 5%). The final consonants meant to be silent were pronounced in half of the cases, whereas they are silent in this place of the words in more than 95% of cases (Peereman et al., 2007). This result is similar to the 57% of pronounced final consonants found by Perry et al. (2014) on monosyllabic pseudowords and thus extends these results to three-syllable pseudowords. The pseudowords with final consonants that generated the most unexpected answers were, in descending order (see Appendix 1): roudélop (93% of unexpected answers); chépuroid (82% of unexpected answers); vapurit (78% of unexpected answers); léfadis (70% of unexpected answers); julavort (44% of unexpected answers); péfutors (30% of unexpected answers); firédoup (22% of unexpected answers); péluraut (19% of unexpected answers); and nujélard (11% of unexpected answers). So, since the same final consonant can generate very different rates of pronunciation (for example, 93% for the "p" in "roudélop" but 22% for the "p" in "firédoup"), it is not just certain final graphemes that generate more pronunciation than others, but apparently sequences of letters that do (see below).

Contrary to our hypothesis, the latencies between the targeted and control pseudowords were not significantly different. This might be explained by the fact that latencies are overall not variable enough among adults reading pseudowords in French. Differences in latencies have been found in word reading (Hogaboam & Perfetti, 1978), pseudowords in English (Perfetti & Hogaboam, 1975), among young learners (Sprenger-Charolles et al., 2000), between different languages (Paulesu et al., 2001), and between dyslexic and control individuals (Sprenger-Charolles et al., 2000), but latencies might not be a sensitive measure in a design such as ours. It may also be that the small number of items used per category in the present study did not allow us to achieve sufficient statistical power to capture any differences at this level.

While the main aim of this study was to describe the extent to which certain features of the French orthographic system represent decoding challenges, some theoretical implications can be inferred regarding the nature of the sublexical route. In the literature, although nobody would question the special status of graphemes in alphabetic scripts, a common issue is how and when the graphemic stage takes place during the mapping between letters and phonemes (Chetail, 2020). According to the CDP+ (e.g., Perry et al., 2007) and BIA (Diependaele et al., 2010) models, a letter string is first parsed and segmented into grapheme units before any phonological sublexical activation. Conversely, according to the DRC models (e.g., Coltheart et al., 2001), the nonlexical route proceeds letter by letter, not grapheme by grapheme. Letter strings are analyzed serially from left to right, and the system looks for a grapheme-to-phoneme rule to determine the phonemes corresponding to the letters, starting with those furthest to the left. Globally, our results support that the mechanisms at work might partly depend on the GPC scenarios considered and on their position in the words, and that interindividual differences exist. For example, for the adjacent letters "an," "on," and "in," the fact that they generated 26% of unexpected answers when they were low cohesive complex graphemes and 14% of unexpected answers when they were two simple consecutive letters supports the view that these sequences of letters are a little more likely to be read letter by letter rather than as a larger (graphemic) unit. This observation is, however, limited to these sequences of letters since the low percentage of unexpected answers (6%) on

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

pseudowords containing the highly cohesive complex graphemes "ou" or "au" suggests that those are quite spontaneously clustered into graphemes. The results regarding contextual graphemes indicate that in a considerable part of cases, they are converted into a phoneme without taking into account the surrounding letters. For example, for the grapheme "s," the participants read it /s/ rather than /z/ in more than 30% of the cases, that is, without taking into account the surrounding vowels. As developed above, the position in the (pseudo)word may also get into action when it comes to attributing one phoneme rather than another to these contextual graphemes, a position on which the GPCs' frequency rules precisely depend (Peereman et al., 2007). This assignment of a phoneme without taking into account the surrounding letters was especially true for the grapheme "e," which was read /ɛ/ instead of /ə/ in more than 80% of the cases. The GPC rule for this grapheme is, however, less systematic than for the graphemes "s," "c," and "g." Indeed, it corresponds to /ɛ/ in most of the cases where it is followed by two consonants and to /ə/ when followed by a single consonant, but there are some exceptions, such as when the grapheme "e" is part of the prefix "re" (semantic value of repetition), as in "reprendre" (/rəprɑ̃dr/, "take again"), or when it is followed by two "s" (dessin, /dɛsɛ̃/, "drawing"; dessus; /dəsy/, "top"). In short, contextual graphemes do not seem to be analyzed automatically in relation to the surrounding letters and are often associated with the phoneme most frequently linked to the isolated letter, depending in part on its position in the word.

The theoretical implications regarding the final consonant are quite ambivalent. On the one hand, the fact that they were pronounced in 50% of the cases supports the idea that they are read as separate letters rather than being attached to their preceding vowel. In other words, the end of "roudélop" seems to be read "l.o.p" rather than "l.op," as it was supported in the Perry et al. (2014) study. On the other hand, these letters were not pronounced in 50% of the cases, and, as developed above, the same final consonant generated very different rates of pronunciation (77% for the "t" in "vapurit" but 18% in "péluraut"). This suggests that the precedent letters still come into play when deciding whether or not to pronounce the final consonant. Thus, while our results for low-cohesive multiletter graphemes, contextual graphemes, and final consonants globally show that the isolated letter level is quite important, it also appears that the nature of the sublexical route partly depends on the sequences of letters in question and on interindividual differences.

Some limitations of this study must be considered. First, as mentioned earlier, the number of items per category and the number of participants were rather low, and more significant differences might have emerged with more items and/or more participants allowing more statistical power. The significant differences found here in these conditions suggest, however, that in any case, the scenarios which gave rise to significantly more unexpected answers compared to their matched pseudowords are less automatized than the scenarios that did not. Second, to match the targeted pseudowords to the control pseudowords, the sequences of letters that differed between them could not always occur at precisely the same frequency when considering the stability and the pronunciation length of the control grapheme (this was especially the case for the categories s meant to be read /z/, e meant to be read /ə/, and the final consonant). However, when the frequencies could not be comparable, we introduced as far as possible higher frequencies for the sequences of

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

letters of the target pseudowords, so that any disadvantages in the scenarios of interest would not be attributable to a lower frequency. This frequency does not actually appear to be so decisive in the current reading performance, since the percentage of unexpected answers was not aligned with them. For example, the frequency of the three letters ("iso") in the targeted item "nisoulor" was 1601,46 for a mean percentage of expected answers of 44%, while the frequency of the three letters ("ivo") in the control item "nivoulor" was 48,38 for a mean percentage of expected answers of 78% (see Appendix 1). Moreover, asking participants about their bilingual or mother-tongue status would have provided a better description of the sample and allowed them to take into account any effects of these variables on their decoding performance (Papastefanou et al., 2021). Measuring their reading performance through an additional reading task would also have allowed them to take their general reading level into account. Finally, a few of the created pseudowords should be slightly modified after noting certain unexpected recurring responses. For example, the pseudoword "chépuroid," whose end is the same as the very frequent and common word "froid" ("cold"), was sometimes read /ʃepyrɔid/ instead of the expected answer /ʃepyrwa/, possibly under the influence of a word that is much less frequent but of more similar length, "astéroïde" ("asteroid"). In this case, the pronunciation of the final consonant "d" might have been prompted by such an influence.

In sum, this study highlights the fairly large number of inconsistent French graphemes causing decoding difficulties for expert readers, suggesting that the difficulty of decoding French should not be underestimated. This suggests that such inconsistent graphemes might also represent particular challenges in the acquisition phase or for individuals facing reading disorders, although this suggestion should be verified by further studies. Indeed, because the decoding challenges faced here by expert readers can be linked to dominant grapheme-to-phoneme rules, and beginning and struggling readers are less sensitive to frequency effects (Alegria & Mousty, 1996; Mousty & Leybaert, 1999; Sprenger-Charolles et al., 2003), it would be worthwhile to identify the extent to which the various inconsistent grapheme-to-phoneme associations generate difficulties for them. The current results highlight that the grapheme-to-phoneme conversion rules of the inconsistent graphemes studied here are not easy to automate, which suggests that their conversion rules could deserve to be emphasized during instruction. In particular, the identification of the precise scenarios that generate decoding challenges among struggling readers could be pursued, which could lead to additional knowledge on effective ways of dealing with such targeted difficulties.

Replication package. Replication data and materials for this article can be found at https://osf.io/4a7st/.

Competing interests. The authors declare none.

References

Afonso, O., Álvarez, C. J., & Kandel, S. (2015). Effects of grapheme-to-phoneme probability on writing durations. *Memory & Cognition*, **43**, 579–592. https://doi.org/10.3758/s13421-014-0489-8.

Ågren, M. (2016). Apprendre l'orthographe profonde du français langue étrangère. *Revue Française de Linguistique Appliquée*, **2**, 95–108.

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

Alegria, J., & Mousty, P. (1994). On the development of lexical and non-lexical spelling procedures of French-speaking, normal and disabled children. In G. D. A Brown & N. C. Ellis (Eds.), *Handbook of normal and disturbed spelling development: Theory, processes and interventions* (p. 211–226). Chichester: Wiley.

Alegria, J., & Mousty, P. (1996). The development of spelling procedures in French speaking, normal and reading-disabled children: Effects of frequency and lexicality. *Journal of Experimental Child Psychology*, **63**(2), 312–338. https://doi.org/10.1006/jecp.1996.0052

Barra-Jover, M. (2009). Comment évolue un trait grammatical: Le pluriel en français dans une perspective romane. *Romance Philology*, **63**, 25–67.

Burani, C., Barca, L., & Ellis, A. W. (2006). Orthographic complexity and word naming in Italian: Some words are more transparent than others. *Psychonomic Bulletin & Review*, **13**, 346–352. https://doi.org/10.3758/BF03193855

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, **19**, 5–51. https://doi.org/10.1177/1529100618772271

Chetail, F. (2020). Are graphemic effects real in skilled visual word recognition. *Journal of Memory and Language*, **111**, 1–12. https://doi.org/10.1016/j.jml.2019.104085

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**(1), 204–256. https://doi.org/10.1037//0033-295X.108.1.204

Commissaire, E., Besse, A. S., Demont, E., & Casalis, S. (2018). Grapheme coding during sublexical processing in French third and fifth graders. *Journal of Experimental Child Psychology*, **173**, 78–84. https://doi.org/10.1016/j.jecp.2018.03.006

Commissaire, E., & Casalis, S. (2018). The use and nature of grapheme coding during sub-lexical processing and lexical access. *Quarterly Journal of Experimental Psychology*, **71**(6), 1324–1339. https://doi.org/10.1080/17470218.2017.1318294

Diependaele, K., Ziegler, J. C., & Grainger, J. (2010). Fast phonology and the Bimodal Interactive Activation Model. *European Journal of Cognitive Psychology*, **22**(5), 764–778. https://doi.org/10.1080/09541440902834782

Emerson, R. H. (1997). English spelling and its relation to sound. *American Speech*, **72**, 260–288.

Gentaz, E., Sprenger-Charolles, L., & Theurel, A. (2015). Differences in the predictors of reading comprehension in first graders from low socio-economic status families with either good or poor decoding skills. *PLoS ONE*, **10**(3), e0119581. https://doi.org/10.1371/journal.pone.0119581

Hogaboam, T. W., & Perfetti, C. A. (1978). Reading skill and the role of verbal experience in decoding. *Journal of Educational Psychology*, **70**, 717–729. https://doi.org/10.1037/0022-0663.70.5.717

Jaffré, J.-P., & Fayol, M. (2005). Orthography and literacy in French. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy*. Mahwah: L.E.A.

Joubert, S. A., & Roch Lecours, A. (2000). The Role of Sublexical Graphemic Processing in Reading. *Brain and Language*, **72**(1), 1–13. https://doi.org/10.1006/brln.1999.2279

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, **44**, 592–617. https://doi.org/10.1006/jmla.2000.2745

Léon, P., Léon, M., Léon, F., & Thomas, A. (2009). La lettre e. In *Phonétique du FLE*, pp. 27–32. Paris: Armand Colin.

Martinet, C., Valdois, S., & Fayol, M. (2004). Lexical orthographic knowledge develops from the beginning of reading acquisition. *Cognition*, **91**, B11–B22. https://doi.org/10.1016/j.cognition.2003.09.002

Mousty, P., & Leybaert, J. (1999). Évaluation des habiletés de lecture et d'orthographe au moyen de BELEC. Données longitudinales auprès d'enfants francophones testés en 2e et 4e années. *Revue Européenne de Psychologie Appliquée*, **49**(4), 325–342.

Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities in French. *Child Development*, **76**, 324–339

Papastefanou, T., Marinis, T., & Powell, D. (2021). Development of reading comprehension in bilingual and monolingual children—effects of language exposure. *Languages*, **6**(4), 166.

Paulesu, E., Démonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., … **& Frith, U.** (2001). Dyslexia, Cultural diversity and Biological unity. *Science*, **291**, 2165–2167. https://doi.org/10.1126/science.1057179

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

Peereman, R., Lété, B., & Sprenger-Charolles, L. (2007). Manulex-infra: Distributional characteristics of grapheme–phoneme mappings, infra-lexical and lexical units in child directed written material. *Behavior Research Methods*, **39**, 593–603. https://doi.org/10.3758/BF03193029

Perfetti, C. A. (1985). *Reading ability*. Oxford: Oxford University Press.

Perfetti, C. A., & Hogaboam, T. W. (1975). Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology*, **67**, 461–469. https://doi.org/10.1037/H0077013

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**(2), 273–315. https://doi.org/10.1037/0033-295X.114.2.273.

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. https://doi.org/10.1016/j.jml.2014.01.003

Posner, R. (1996). *The Romance languages*. Cambridge: Cambridge University Press.

Rastle, K., & Coltheart, M. (1998). Whammy and double whammy: length effects in nonword naming. *Psychonomic Bulletin and Review*, **5**, 277–282. https://doi.org/10.3758/BF03212951

Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition*, **68**, B71–B80. https://doi.org/10.1016/S0010-0277(98)00051-1

Rey, A., & Schiller, N. O. (2005). Graphemic complexity and multiple print-to-sound associations in visual word recognition. *Memory & Cognition*, **33**, 76–85. https://doi.org/10.3758/BF03195298

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, **75**, B1–B12. https://doi.org/10.1016/S0010-0277(99)00078-5

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**(2), 1–7. https://doi.org/10.4000/cpl.465

Sambai, A., Coltheart, M., & Uno, A. (2018). The effect of the position of atypical character-to-sound correspondences on reading kanji words aloud: Evidence for a sublexical serially operating kanji reading process. *Psychonomic Bulletin & Review*, **25**, 498–513. https://doi.org/10.3758/s13423-018-1434-9

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

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition, *Cognition*, **55**(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Spencer, M., Quinn, J., & Wagner R. (2014). Specific reading disability: Major problem myth or misnomer? *Learning Disabilities Research & Practice*, **29**, 3–9. https://doi.org/10.1111/ldrp.12024

Spinelli, E., Kandel, S., Guerassimovitch, H., & Ferrand, L. (2012). Graphemic cohesion effect in reading and writing complex graphemes. *Language and Cognitive Processes*, **27**(5), 770–791. https://doi.org/10.1080/01690965.2011.586534

Sprenger-Charolles, L., Colé, P., Béchennec, D., & Kipffer-Piquard, A. (2005). French normative data on reading and related skills: from EVALEC, a new computerized battery of tests. *European Review of Applied Psychology*, **55**, 157–186. https://doi.org/10.1016/j.erap.2004.11.002

Sprenger-Charolles, L., Colé, P., Lacert, P., & Serniclaes, W. (2000). On subtypes of developmental dyslexia: Evidence from processing time and accuracy scores. *Canadian Journal of Experimental Psychology*, **54**, 88–104. https://doi.org/10.1037/h0087332

Sprenger-Charolles, L., Siegel, L. S., Béchennec, D., & Serniclaes, W. (2003). Development of phonological and orthographic processing in reading aloud, silent reading and in spelling: A five years longitudinal study. *Journal of Experimental Child Psychology*, **84**, 194–217. https://doi.org/10.1016/s0022-0965(03)00024-9

Stark, C. E., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**(4), 945. https://doi.org/10.1037/0278-7393.26.4.945

Treiman, R., & Kessler, B. (2019). Development of context-sensitive pronunciation in reading: The case of ‹c› and ‹g›. *Journal of Experimental Child Psychology*, **182**, 114–125. https://doi.org/10.1016/j.jecp.2019.02.001

Treiman, R., Kessler, B., & Evans, R. (2007). Anticipatory conditioning of spelling-to-sound translation. *Journal of Memory and Language*, **56**, 229–245.

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

Treiman, R., Kessler, B., Zevin, J. D., Bick, S., & Davis, M. (2006). Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of Experimental Child Psychology*, **93**, 1–24. https://doi.org/10.1016/j.jecp.2005.06.008

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of Reading Ability: Multivariate Evidence for a Convergent Skills Model of Reading Development. *Scientific Studies of Reading*, **11**(1), 3–32. http://doi.org/10.1080/10888430709336632

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, **131**(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, & Computers*, **28**, 504–515. https://doi.org/10.3758/BF03200539

Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: Implications for dyslexia. *Philosophical Transactions of the Royal Society, B*, **369**, 20120397. http://doi.org/10.1098/rstb.2012.0397

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

Appendix 1. Frequencies of the three letters surrounding the grapheme-to-phoneme correspondences (GPCs), mean RTs, and mean percentage of expected answers for each target and control pseudowords

	Target pseudowords	Freq. of 3 letters	Mean RT	Mean % of expected answers	Control pseudowords	Freq. of 3 letters	Mean RT	Mean % of expected answers
s = /s/	chouv**ast**ol	259,88	948,46	96%	chouv**art**ol	2947,93	985,58	96%
	du**tsi**rof	0,19	1084,50	63%	du**tli**rof	0,16	1118,29	67%
	m**isb**udole	1,07	901,92	93%	m**irb**udole	1,1	848,65	89%
	naf**usp**é	27,3	829,64	100%	naf**ulp**é	9,43	793,92	96%
	tavi**bsu**re	1,55	1045,67	33%	tavi**blu**re	1,28	771,48	85%
	v**aspi**lou	82,12	754,67	93%	v**arpi**lou	48,1	773,67	93%
s = /z/	bil**uso**re	8,9	948,43	59%	bil**ufo**re	1,03	932,67	92%
	n**iso**ulor	1601,46	957,17	44%	n**ivo**ulor	48,38	891,29	78%
	p**osi**do	277,54	736,33	67%	p**ori**do	302,38	747,17	100%
	poud**ase**r	1806,43	867,89	33%	poud**ale**r	1277,49	852,24	63%
	r**ésa**bile	63,2	705,30	100%	r**éfa**bile	59,57	813,96	96%
	tor**ési**f	379,26	719,68	81%	tor**éli**f	226,44	739,50	96%
e = /ɛ/	**bes**podir	817,03	939,33	96%	**bas**podir	361,39	808,35	93%
	chu**del**pa	116,42	1016,05	81%	chu**dil**pa	122,44	1263,67	85%
	lo**pes**ta	364,66	788,18	92%	lo**pis**ta	349,59	792,70	93%
	lertibuve	1763,05	893,16	78%	**lor**tibuve	1458,27	868,27	89%
	ma**jer**pule	1,43	748,58	96%	ma**jor**pule	5,41	763,72	93%
	neltari	150,26	762,48	100%	**nol**tari	51,98	765,50	96%

(*Continued*)

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

(*Continued*)

	Target pseudowords	Freq. of 3 letters	Mean RT	Mean % of expected answers	Control pseudowords	Freq. of 3 letters	Mean RT	Mean % of expected answers
e = /ə/	**bet**ouvar	73,59	–	0%	**bét**ouvar	62,21	906,85	100%
	bu**ved**ire	13,42	968,67	15%	bu**vad**ire	3,56	811,32	96%
	cha**del**u	116,42	868,77	52%	cha**dil**u	122,44	791,50	93%
	meladou	109,52	–	0%	**mul**adou	109,44	815,04	93%
	refapile	279,52	1076,00	4%	**rof**apile	218,47	880,20	93%
	va**ter**il	3937,09	756,50	7%	va**tur**il	1624,3	761,04	96%
g = /g/	ch**égu**lofe	176,06	1000,85	89%	ch**étu**lofe	104,53	1017,47	88%
	m**oga**lou	72,6	753,59	100%	m**oba**lou	42,17	783,75	89%
	nuf**igo**	67,98	902,70	89%	nuf**ibo**	70,46	910,05	96%
	tol**uga**r	49,26	837,68	100%	tol**upa**r	74,45	798,33	96%
	vét**agu**	208,78	694,05	74%	vét**adu**	38,91	748,63	93%
	v**ugo**lé	9,05	951,38	92%	v**upo**lé	18,01	881,65	92%
g = /ʒ/	b**égi**lor	128,2	841,61	73%	b**éfi**lor	204,61	886,73	89%
	m**ogi**vale	37,03	749,42	70%	m**ofi**vale	40,93	790,88	96%
	noul**igé**	96,36	925,23	85%	noul**iré**	131,67	776,81	100%
	par**ugi**l	101,88	751,32	70%	par**uvi**l	67,44	707,69	96%
	p**agé**til	49,54	755,96	89%	p**avé**til	16,73	695,74	100%
	tap**ogé**	4,25	817,42	100%	tap**ofé**	1,34	851,95	93%

(*Continued*)

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

(*Continued*)

	Target pseudowords	Freq. of 3 letters	Mean RT	Mean % of expected answers	Control pseudowords	Freq. of 3 letters	Mean RT	Mean % of expected answers
in, on, an = /ɛ̃/, /ɔ̃/, /ɑ̃/	**inv**édire	351,69	920,59	85%	**auv**édire	745,66	864,48	96%
	b**onf**ivule	249,32	790,00	74%	b**auf**ivule	209,73	856,57	96%
	lét**onv**ir	288,74	919,00	78%	lét**auv**ir	745,66	871,56	96%
	néf**ant**ile	10721,24	771,87	96%	néf**out**ile	6618,52	753,17	100%
	r**enl**itoche	96,77	998,00	48%	r**aul**itoche	256,89	944,65	93%
	rét**ind**af	942,68	855,29	63%	rét**oud**af	526,52	974,26	81%
in, on, an = /i/-/n/, /ɔ/-/n/, /a/-/n/	b**ina**vule	766,08	904,13	93%	b**ira**vule	728,74	823,73	100%
	bout**ano**re	169,01	931,77	85%	bout**ajo**re	220,43	873,88	96%
	fut**ino**pe	172,72	1112,29	78%	fut**ilo**pe	207,63	1129,74	89%
	l**oné**doche	12,15	910,59	81%	l**ofé**doche	1,34	1003,95	81%
	mil**onu**f	12,02	833,10	81%	mil**oru**f	2,51	826,35	85%
	t**ani**lode	1308,46	769,48	100%	t**ari**lode	939,59	802,07	100%
Final d, p, s, t = silent	chépur**oid**	326,85	994,80	19%	chépur**oie**	597,8	1105,17	96%
	firéd**oup**	3102,21	1013,14	78%	firéd**oue**	1350,52	1034,38	89%
	julav**ort**	4070,84	844,33	56%	julav**ore**	1397,01	852,38	96%
	léfa**dis**	1371,52	753,63	30%	léfa**die**	504,61	799,00	100%
	nujél**ard**	3053,23	895,17	89%	nujél**are**	1376,04	852,07	100%
	péfut**ors**	1573,6	994,12	70%	péfut**ore**	1397,01	1107,22	93%
	pélur**aut**	4463,73	920,24	81%	pélur**oue**	1350,52	791,71	81%

(*Continued*)

https://doi.org/10.1017/S0142716425100106 Published online by Cambridge University Press

(*Continued*)

	Target pseudowords	Freq. of 3 letters	Mean RT	Mean % of expected answers	Control pseudowords	Freq. of 3 letters	Mean RT	Mean % of expected answers
	roudé**lop**	160,53	708,00	7%	roudé**lue**	78,48	946,35	96%
	vapu**rit**	1505,01	839,50	22%	vapu**rie**	2601,09	710,22	100%
Two in, on, an . . .	l**oman**ube		867,36	52%	l**orav**ube		824,27	81%
	b**inum**ate		914,96	100%	b**ilur**ate		877,35	93%
	v**amin**ole		755,26	100%	v**azir**ole		710,52	100%

Note: Frequencies of the three letters surrounding the GPCs, mean RTs on expected answers (in ms), and mean percentage of expected answers for each target and control pseudowords are displayed. Frequencies for the last category ("Two in, on, an . . . ") are not reported since they are only available for up to three adjacent letters (Peereman et al., 2007).

Cite this article: de Chambrier, A-F., Atzemian, M., Fayol, M., Zesiger, P., & Martinet, C. (2025). French is not so easy to decode: a pilot study. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716425100106