CrossMark

Silex: A database for silent-letter endings in French words

Maxime Gingras[1] · Monique Sénéchal[1]

Abstract Silent-letter endings are often claimed to be a major source of inconsistency in the French orthography. In this report, we introduce Silex, a database designed to facilitate the study of spelling performance in general, and silent-letter endings in particular. It was derived from two large and recent corpora based on child- and adult-targeted material. Silex consists of three kinds of Excel workbooks: a set of Stimuli Selector workbooks that allow researchers to select words based on a variety of statistics and word characteristics; a Table Generator workbook that allows researchers to build consistency distribution tables by selecting specific phonological or orthographic units; and a Master File workbook, from which all statistics were derived, and that allows researchers to compute other statistics. Silex is different from existing databases in the manner that silent-letter endings were coded and how consistency indices were computed. Importantly, Silex provides unconditional- and conditional-consistency indices for silent-letter endings. To demonstrate the utility of Silex, we first described the silent-letter phenomenon in French. We found that, at minimum, 28 % of French words end with a silent letter. Moreover, silent-letter endings are usually t, e, s, x, or d, and the occurrence of these letters is conditioned by the phonological ending of words. Second, we showed how Silex could prove useful for the development of theoretical models and for empirical studies. The novel information provided in Silex as well as the flexibility of this database should enable researchers to advance our understanding of developing and skilled spelling performance.

✉ Monique Sénéchal
Monique.Senechal@carleton.ca

1   Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada

Keywords Lexical databases · French orthography · Silent-letter endings · Spelling

Introduction

Descriptions of the statistical properties of words (Content, Mousty, & Radeau, 1990; Peereman & Content, 1999; Peereman, Lété, & Sprenger-Charolles, 2007; Peereman, Sprenger-Charolles, & Messaoud-Galusi, 2013; Ziegler, Jacobs, & Stone, 1996) have proven to be essential to the study of how children and adults read and spell words. Such descriptions have enabled researchers to establish frequency, consistency, and contextual effects on reading across a number of languages, including French (e.g., Content, 1991; Ferrand, Brysbaert, Keuleers, New, Bonin, Méot, et al., 2011; Perry, Ziegler, & Zorzi, 2014). Given that lexical databases have proven their utility for the study of reading, it is not surprising that researchers interested in spelling have used these databases as their starting point (e.g., Lété, Peereman, & Fayol, 2008). Certainly, word characteristics such as the frequency of occurrence in written texts and the number of letters are the same whether one studies reading or spelling, but consistency indices are necessarily calculated differently across modalities. Whereas it is the probability for an orthographic unit to correspond to a given pronunciation that is quantified for reading (i.e., spelling-to-sound consistency), the perspective is reversed for spelling: It is the probability for a phonological unit to correspond to a given spelling (i.e., sound-to-spelling consistency). A complete definition of sound-to-spelling consistency must also include the probability of occurrence of silent letters – that is, orthographic units without any phonological value. In French, different segmentation rules have been used to establish indices of sound-to-spelling consistency (e.g., Peereman et al., 2007; Ziegler et al.,

1996). Herein, we propose a theory-driven approach to the quantification of sound-to-spelling consistency, and this approach was applied to the words found in two existing French corpora: one for children (Manulex-infra: Peereman et al., 2007) and one for adults (Lexique 3.80: New, Pallier, Brysbaert, & Ferrand, 2004). The resulting unified database, Silex, is a flexible tool that should prove useful to the study of developing and skilled spelling performance.

Spelling is of interest to researchers because it provides insight into the quality of orthographic representations. According to the Fuzzy Representation Model (Sénéchal, Gingras, & L'Heureux, 2016), orthographic representations for words are initially constructed as a frame in which the graphemes for consistently spelled phonemes are clearly specified, while those for inconsistently spelled phonemes and silent letters are more likely to be underspecified, if represented at all. Fuzzy orthographic representations are not likely to lead to the correct spelling of specific words. For instance, marking silent-letter endings accurately is particularly difficult for French spellers (Fayol, Totereau, & Barrouillet, 2006; Sénéchal, 2000). In Silex, the term *grapheme* was defined as an orthographic unit that corresponds to a phoneme or, if silent, that plays a distinctive or morphological role.

In contrast to English where silent letters can occur anywhere in words (e.g., *knee*, *island*, *dumb*), the French orthography is a system that favors the occurrence of silent letters at the end of words (Catach, 1995; Jaffré, 2005). Two morphological factors contribute to the frequency of silent-letter endings in French. First, they play an important role in marking inflectional morphology (e.g., the silent *s* in plural nouns). Second, derivational morphology is such that many words end with silent letters that are pronounced in derivatives (e.g., the silent *d* in *bavard* /bavaʀ/ [chatterbox] is pronounced in *bavarder* /bavaʀde/ [to chat]). In addition to these inflectional and derivational silent letters, other words end with silent letters that are part of their idiosyncratic spellings (e.g., *foulard* /fulaʀ/ [scarf], a masculine singular noun with no derivatives). Finally, silent-letter endings are a way to distinguish between homophones (e.g., *sans* /sɑ̃/ [without] and *sang* /sɑ̃/ [blood]). It should be noted that qualitative descriptions of this kind are not new in the literature.

The psycholinguistic research on silent-letter endings in French has focused on spelling rather than reading because silent letters are likely to cause more difficulties in spelling than reading (Treiman & Kessler, 2014). While Fayol and colleagues assessed how spellers learn and apply the rules of inflectional morphology (Fayol, Hupet, & Largy, 1999; Fayol et al., 2006), Sénéchal and colleagues examined how the awareness of morphological families could be useful to spellers (Sénéchal, 2000; Sénéchal, Basque, & Leclaire, 2006; see also Pacton & Deacon, 2008). Taken together, these studies highlighted the difficulty of spelling words with silent-letter endings in general, whatever the role of these letters.

This difficulty was confirmed by Jubenville, Sénéchal, and Malette (2014). In this study, Grade 3 children were incidentally exposed to pseudowords that were relatively consistent, except that they ended with a silent letter (e.g., *pocrat* /pokʀa/). At post-test, children had little difficulty spelling the consistent portions of the words, but they had a silent-letter error in 95 % of their misspellings. The morphological role of silent-letter endings could not be invoked at all to explain the results obtained in this study. Indeed, to determine a plausible silent-letter ending in the pseudowords, the participants could only rely on their orthographic knowledge. Even with real words that had derivatives (e.g., *bavard* /bavaʀ/ [chatterbox]), Sénéchal et al. (2016) found that children in Grades 1–3 had a silent-letter error in 94 % of their misspelings. Therefore, the Silex database was developed with the goal of providing a clear picture of the orthographic patterns that spellers could be expected to represent in memory, irrespective of the role of silent-letter endings in derivational morphology.

According to the Fuzzy Representation Model (Sénéchal et al., 2016), silent letters are always a source of inconsistency because they are orthographic marks without phonological counterparts (e.g., *gt* in *doigt* /dwa/ [finger]). Previous quantifications of consistency, however, included silent-letter endings in sound-to-spelling correspondences in one way or another. On the one hand, Ziegler et al. (1996) calculated consistency indices based on the phonological rimes of French monosyllabic words. The rime /o/, for example, can be spelled *au*, *aud*, *aut*, *aux*, *eau*, *o*, *oc*, *op*, *os*, *ot*, or *ô*, and for each spelling, a probability of occurrence was computed. However, this did not distinguish the spelling of the phoneme /o/ (*au*, *eau*, *o*, and *ô*) from the silent-letter endings (*c*, *d*, *s*, *p*, *t*, and *x*), neither did it provide the unconditional (i.e., non-contextual) probability of occurrence for each silent letter.

On the other hand, Peereman et al. (2007, 2013) were initially interested in grapheme-phoneme consistency for the study of reading. Thus, they used a coding that is valid for the study of reading, but that posed two main problems for the study of spelling. In the Silex database, the goal was to circumvent these problems for researchers interested in spelling. The first problem was that Peereman et al. treated silent letters as special cases of phoneme-grapheme consistency by mapping them to a null phoneme (e.g., #-t in *debout* /dəbu/ [standing up], as opposed to /t/-t in *mazout* /mazut/ [oil]). Due to this coding, the consistency of final phonemes was computed on *null* phonemes for words with silent-letter endings and on *pronounced* phonemes in all other cases. For instance, the word *saint* /sɛ̃/ [holy], with a silent *t*, and the word *pains* /pɛ̃/ [bread, plural form], with a silent *s*, were deemed more consistent than the word *pain* /pɛ̃/ [bread, singular form] as regards the final phoneme, even though these words end with the same *pronounced* phoneme /ɛ̃/. In Silex, words without silent-letter endings were coded with a final null *grapheme* corresponding to the absence of silent letters. As such, the

consistency of final phonemes could be computed on *pronounced* phonemes in all words (e.g., /ɛ̃/ spelled *ain* in *saint*, *pains*, and *pain*), while additional probabilities could be computed for the orthographic endings in all words (e.g., the silent *t* in *saint*, the silent *s* in *pains*, and the final null grapheme in *pain*).

The second problem in Peereman et al.'s (2007, 2013) reading-oriented approach was that their segmentation rules did not allow multi-letter silent endings. Take the example of the silent *gt* in *doigt* (/dwa/ [finger]), where the silent *g* corresponded to a null phoneme in the middle section of the word and the silent *t*, to the final null phoneme. This rule was also applied to inflected words. For the plural form *doigts*, it is the plural marker *s* that corresponded to the final null phoneme, while *g* and *t* were in the middle section of the word. From a spelling perspective, however, it is necessary to recognize that silent endings are a specific locus of difficulty and that they can vary in difficulty depending on the number of letters in them (e.g., *gt* in *doigt* compared to *t* in *toit* /twa/ [roof]). The Silex database used an alternative approach to phonographic segmentation whereby silent-letter endings, regardless of the number of letters, were treated as orthographic units without any phonological value. In sum, the benefits of our approach for the study of spelling is that it better reflects the fact that, other things being equal, words without silent-letter endings (e.g., *toi* /twa/ [you]) are more consistent than words ending with one silent letter (e.g., *toit*), and that these are more consistent than words ending with more than one silent letter (e.g., *doigt*).

In computational terms, our treatment of silent-letter endings had four main consequences: (1) the calculation of unconditional probabilities of occurrence of silent-letter endings based on all words, including those that do not end with silent letters (e.g., the silent *t* in *saint* is less likely to occur than the final null grapheme in *pain*); (2) the calculation of consistency indices for the last *pronounced* phoneme in all words, including those that end with silent letters (e.g., *saint* and *pain* have the same consistency index for the last pronounced phoneme /ɛ̃/), resulting in the availability of a phoneme-grapheme consistency index for all words; (3) the calculation of whole-word consistency indices by averaging all the unconditional probabilities for a given word (e.g., the mean consistency for /s/, /ɛ̃/ and the silent *t* in *saint*, and the mean consistency for /p/, /ɛ̃/ and the final null grapheme in *pain*); and (4) the calculation of conditional probabilities of occurrence of silent-letter endings and null graphemes given the preceding letter (e.g., *n* in *saint* and *pain*) or the spelling of the last pronounced phoneme (e.g., /ɛ̃/ spelled *ain* in *saint* and *pain*), among other possible types of contexts (e.g., the spelling of the last phonological rime). As a result of coding silent-letter endings from a spelling perspective, the mean whole-word consistency indices in Silex (*M*s = .66 and .66 based respectively on the number of different words, i.e., type count, and on the frequency per

million words, i.e., token count) are more conservative than those found in Manulex-infra (*M*s = .72 and .74 for type and token counts, respectively) when only using the nouns found in the Grades 1–5 corpus of both databases. Remember that Silex also includes consistency indices for the words found in Lexique 3.80, for which none were previously provided in published materials. The impact of silent-letter endings on whole-word consistency, as well as the effect of the different conditioning contexts, are shown in the appendix.

Another theoretical assumption of the Fuzzy Representation Model that has guided the construction of Silex is that, during acquisition, the parts of words that are inconsistent but stable are more likely to be represented in memory than inconsistencies that are unstable. In French, the unstable parts of words correspond to inflectional morphology (e.g., the silent *s* in *amis* /ami/ [friends], which is unstable given the existence of the singular form *ami* /ami/ [friend]). In accord with this view, consistency indices for words should be based on uninflected words only, because the endings in these words are stable (e.g., the silent *s* in *radis* /ʀadi/ [radish], which is stable given that the word *radi* does not exist), contrary to those found in inflected words. Although Manulex-morpho (Peereman et al., 2013) did identify inflectional markers for 20 % of the words found in Manulex (Lété, Sprenger-Charolles, & Colé, 2004), consistency indices for all uninflected words were not computed. The coding in Manulex-morpho was on the silent letters themselves, and a different set of codes was used (e.g., *s* coded as the plural marker in *amis*; *s* coded as a verbal letter in *finis* /fini/ [finish]; *s* coded as a derivational silent letter in *gris* /gʀi/ [gray]; *s* coded as a non-derivational silent letter in *radis*). In constrast, all words included in Silex are either inflected (e.g., *amis*, *finis*) or uninflected (e.g., *ami*, *gris*, *radis*), whether they end with a silent letter or not. This allows researchers to select consistency indices that were computed when including all words, uninflected words only, or inflected words only, for both Manulex-infra and Lexique 3.80. The impact of removing inflected words from the calculations is particularly striking, as shown in the appendix.

In the next sections, the Silex database was described and its utility for psycholinguistic research illustrated with two examples. In the first example, we showed how Silex could be used to describe the structure of the French orthography. In the second example, we revisited the analysis of silent-letter errors in Sénéchal et al. (2016) to show different orthographic factors that are at play when spelling words with silent-letter endings. In doing so, we demonstrated that Silex could be used to investigate how statistical learning (e.g., Negro, Bonnotte, & Lété, 2014) and grain-size theory (e.g., Ziegler & Goswami, 2005) apply to spelling (see also Treiman & Kessler, 2014). That is, Silex can be used to explore the effects of letter frequency (e.g., Pacton, Borchardt, Treiman, Lété, & Fayol, 2014) and preceding orthographic context (e.g.,

Pacton, Fayol, & Perruchet, 2005; Treiman, Kessler, & Bick, 2002) because it provides unconditional- and conditional-consistency indices for silent-letter endings. In sum, the richness of the information provided in Silex is such that it should enable researchers to test a number of specific hypotheses about spelling acquisition and skilled performance.

Description of the Silex database

Silex consists of a set of freely downloadable Excel workbooks (http://carleton.ca/cllr/silex/). First, the Stimuli Selector workbooks allow users to select words based on a variety of statistics and word characteristics. Second, the Table Generator workbook allows users to build different consistency distribution tables by selecting specific phonological or orthographic units. Third, the Master File workbook allows interested users to compute other statistics than those already provided.

Silex was derived from two existing databases, namely Manulex-infra (Peereman et al., 2007; http://leadserv.u-bourgogne.fr/bases/manulex/manulex_infra/index.htm) and Lexique 3.80 (New et al., 2004; http://lexique.org). Given this origin, any reference to Silex must include a reference to Manulex-infra, Lexique 3.80, or both. These databases were selected because they are based on large and recent corpora targeted to completely different audiences. The Manulex-infra corpus comprises 54 elementary schoolbooks and is divided into grade-specific subcorpora, while Lexique 3.80 includes a corpus of 218 French novels for adults and a corpus of film subtitles. To obtain words that actually occurred in written texts, the corpus of film subtitles was not used to create Silex. The word entries from both databases, along with their syntactic class, gender, number of letters, and frequency per million words, were imported to Silex, with the exception of proper names (e.g., *Paris*), interjections and onomatopoeia (e.g., *oh*, *atchoum*), abbreviations and symbols (e.g., *svp*, *m*), compound words (e.g., *aujourd'hui*, *abat-jour*, *à tâtons*), 24 words that can be pronounced in different ways depending on their unavailable lexical meaning (e.g., *jet* /ʒɛ/ [throw] or /dʒɛt/ [jet plane]), as well as 36 loanwords with inconsistencies that cannot be satisfyingly treated in our approach to phonographemic segmentation (e.g., *sunlight*, *breitschwanz*). The phonological coding, syllabification, and phonographemic segmentation of words in Silex followed linguistic principles that are described in the Silex User Guide.

As a result, Silex is based on a total of 119,664 word entries coming from Manulex-infra, Lexique 3.80, or both corpora. These entries are not lemmas (i.e., dictionary entries). They include inflected and uninflected words (e.g., *ami* and its plural form *amis*) that can be found in actual texts. It should also be noted that the number of entries is inflated by homophonous homographs (e.g., *anglais* noun and *anglais* adjective).

When homophonous homographs are counted only once, Silex comprises a total of 105,926 words, of which 37 % are found in Manulex-infra and 98 % in Lexique 3.80.

One novelty in Silex is that five types of word characteristics relevant to the study of silent-letter endings were coded for all words from Manulex-infra and Lexique 3.80. The coding was carried out by the first author, who is a linguist, with the assistance of computer programs written in Perl. To minimize errors, the automatic coding was manually verified by the first author. The coded information is briefly presented here, illustrated in the next section, and described in detail in the Silex documentation.

The first type of coding concerned word endings: Words without silent-letter endings received a code for the null grapheme, and the other words had their silent-letter endings specified. The second type of coding was the phonographemic segmentation of all words, based on an original spelling-oriented approach. Contrary to the reading-oriented approach, whereby "segmentation must highlight inconsistencies in the pronunciation of orthographic strings" (Peereman et al., 2007), the spelling-oriented approach must highlight inconsistencies in the spelling of phonological strings. This principle resulted in differential segmentation decisions, the most important of which being the treatment of silent letters described above, but there are other examples of units segmented differently in the two approaches (e.g., *oua* treated as a graphemic unit corresponding to /wa/ in spelling, but divided into *ou* and *a* in reading, respectively corresponding to /w/ and /a/). The third type of coding concerned the orthographic context for word endings. Specifically, the last non-silent letter was identified, be it the final letter of the word (e.g., *u* in *bijou* /biʒu/ [jewel]) or the letter before the silent-letter ending (e.g., *u* in *debout* /dəbu/ [standing up]). The grapheme corresponding to the final phoneme in the word was also identified (e.g., *ou* in both *bijou* and *debout*), and so were the graphemes corresponding to the phonemes of the final phonological rime (e.g., *our* corresponding to /uʀ/ in *amour* /amuʀ/ [love] and *lourd* /luʀ/ [heavy]). The fourth type of coding concerned the phonology of the word's ending, be it the final phoneme (e.g., /ʀ/), rime (e.g., /uʀ/), or type of rime (i.e., oral vowel, nasal vowel, /ʀ/ preceded by a vowel, /ʀ/ preceded by a consonant, or any other consonant). The fifth type of coding was whether a word was inflected or not. If so, the type of inflection was specified (i.e., plural, feminine, or verbal). Other types of codes were included, such as the final letter of each word, whether it is silent or not, and the letter that precedes it. All this coded information is found in the Master File workbook, and much of it is also found in the Stimuli Selector and Table Generator workbooks. The Master File workbook was used to compute consistency indices for individual words in the Stimuli Selector workbooks and for linguistic units in the Table Generator workbook. Further explanations on the workbooks and their contents are available in the Silex User Guide.

Users looking for words with specific characteristics can select the appropriate corpus and variables in the Stimuli Selector workbooks. For example, a selection of words could be sorted by increasing spelling difficulty, based on the information provided. To do so, words could be filtered within a range of word length and frequency, and then sorted by decreasing whole-word consistency. Another possible usage of filtering could be to neutralize the effect of the preceding orthographic context on the spelling of word endings by selecting only the words with a given grapheme for the final phoneme (e.g., *au* corresponding to /o/ in words like *étau*, *chaud*, *saut*, and *taux*). In contrast, the Table Generator workbook contains thousands of probability distributions that can be useful to linguists who want to describe the French orthography and to psycholinguists who want to design new studies or create pseudowords based on the orthographic consistency of words. It consists of a series of worksheets, of which two are of special interest: one for unconditional- or conditional-consistency distributions for word endings (i.e., silent-letter endings and null graphemes), and the other for position-independent or position-dependent consistency distributions for the graphemes that correspond to actual phonemes. Position-dependent consistency distributions are provided for phonemes in initial, medial, and final positions (e.g., /o/ spelled *au* in *aube*, *taupe*, and *saut*, respectively; the *t* is silent in *saut* and does not correspond to any phoneme). Users of the Stimuli Selector or the Table Generator workbooks can obtain information as a function of corpus (i.e., Grades 1, 2, 3–5, or 1–5 of Manulex-infra, or Lexique 3.80), word set (i.e., all words, inflected words only, or uninflected words only), type or token count, etc. For instance, in the present report, all tables were built with the Table Generator.

Examples of the utility of Silex

This section is divided into two parts. The first part shows how Silex can be used to describe the silent-letter phenomenon in French. In the second part, the focus shifts from the written language itself to the study of spelling acquisition and performance. All statistics presented in this section are from Silex or calculated using Silex. As a reminder, the words included in Silex were from two existing corpora – Manulex-infra (Peereman et al., 2007) and Lexique 3.80 (New et al., 2004) – and new calculations were performed separately for each corpus. The statistics for the Manulex-infra corpus (Grades 1–5) are reported first in the text, and those for the Lexique corpus are given in parentheses. As shown in the present section, the statistics for the child- and adult-targeted corpora are strikingly similar, indicating that general properties of the French orthography might be identified.

Silent-letter endings in French

When describing silent-letter endings in French, there are linguistic decisions that cannot be avoided, and these considerably affect how the phenomenon is portrayed. One such decision concerns the segmentation of two types of word endings: units made of *e* followed by a consonant (e.g., *et* in *ballet* /balɛ/ [ballet]), where the consonant could be considered as a silent letter (e.g., Catach, 1995); and units made of *e* preceded by a consonant other than *r* (e.g., *ce* in *place* /plas/ [place]), where *e* could be considered as a silent letter (e.g., Peereman et al., 2007) or as an optional silent letter (e.g., Peereman et al., 2013). In Silex, the default option is to treat these units as multi-letter graphemes without silent letters, because removing the final letter in these units usually results in impossible spellings (e.g., *e* and *c* alone are not possible spellings for /ɛ/ and /s/ at the end of words). Users of Silex can decide otherwise by removing the default option. In the present report, we treated final units such as *et* and *ce* in *ballet* and *place* as multi-letter graphemes without silent letters. As for the final *e* after *r*, it is treated as a silent-letter ending given that *r* is a possible spelling for a final /ʀ/, and that it is the only consonant grapheme that allows a variety of silent-letter endings in uninflected words, including *e* (e.g., *caviar* /kavjaʀ/ [caviar], *avare* /avaʀ/ [miser], *part* /paʀ/ [part], *canard* /kanaʀ/ [duck]). Further explanations are given in the Silex User Guide.

Another important decision is whether inflected words (i.e., plural, feminine, and verbal forms) should be included or not in the calculations. As shown in Table 1, the overall percentage of words with silent-letter endings in French ranges from a maximum of 56 % (61 %) to a minimum of 29 % (28 %) depending on the treatment of inflected words. The maximum is obtained when inflected words are included and the most conservative estimate is obtained when inflected words are excluded. These analyses confirm that silent-letter endings play an important role in marking inflectional morphology in French. Consider, for instance, that 25 (34) different forms of the verb *aimer* are found in Silex, and that at least 17 (23) of these forms end with silent letters (e.g., *aimes* /ɛm/, *aiment* /ɛm/, *aimais* /ɛmɛ/, *aimait* /ɛmɛ/, *aimée* /ɛme/, *aimés* /ɛme/, *aimées* /ɛme/, *aimons* /ɛmɔ̃/, *aimant* /ɛmɑ̃/).

For all subsequent analyses, we excluded inflected words, which represent 67 % (71 %) of the words in Silex. We posit that from a linguistic perspective it is necessary to distinguish between word-specific (e.g., *s* in *radis* /ʀadi/ [radish]) and rule-based silent-letter endings (e.g., *s* in *amis* /ami/ [friends], plural of *ami* /ami/ [friend]). One question that could be asked is whether the percentage of uninflected words with silent-letter endings is conditioned by the type of phonological ending. First, as shown in Table 1, silent-letter endings are found in words ending with three types of phonological endings: oral vowels, nasal vowels, and /ʀ/ (e.g., the silent *t* in *saut* /so/ [jump], *gant* /gɑ̃/ [glove], and *part* /paʀ/ [part]).

Table 1 Unconditional and conditional probabilities of occurrence of a silent-letter ending in French words as a function of database (Manulex-infra or Lexique 3.80) and words included (all words or uninflected words)

		Unconditional probabilities		Conditional probabilities by type of phonological ending							
				Vowel				Consonant			
				Oral vowel (*radis*)		Nasal vowel (*gant*)		/ʀ/ (*dard*)		Other (*legs*)	
		M	L	M	L	M	L	M	L	M	L
All words	*p*	.56	.61	.54	.61	.79	.81	.78	.80	.40	.45
	N	39,197	103,843	17,829	49,147	5,892	15,700	4,079	9,386	11,397	29,610
Uninflected words	*p*	.29	.28	.36	.36	.54	.53	.60	.58	< .01	< .01
	N	12,974	29,833	3,680	8,147	2,727	6,322	1,707	3,606	4,860	11,758

M = words found in a corpus of Grades 1–5 schoolbooks (Manulex-infra: Peereman et al., 2007); L = words found in a corpus of adult novels (Lexique 3.80: New et al., 2004). All statistics are based on type counts (i.e., number of different words).

Orthographically, this means that silent-letter endings are preceded by a vowel letter, by *n* or *m* marking nasality, or by *r*. Strikingly, only 12 words ending with phoneme consonants other than /ʀ/ have silent-letter endings, and their silent letter is always *s* (e.g., *legs*, *aurochs*, *certes*). Second, in words ending with a nasal vowel or /ʀ/, words with silent-letter endings slightly outnumber words without. In words ending with an oral vowel, however, there are less words with silent-letter endings than words without. Third, considering all words with these three types of phonological endings shows that 47 % (46 %) of them end with a silent letter. That is, almost half of the words ending with a vowel or /ʀ/ have a silent-letter ending. The silent-letter phenomenon being circumscribed, an examination of specific silent letters follows.

As shown in Table 2, the five most frequent silent-letter endings overall are *t*, *e*, *s*, *x*, and *d*, representing a total of 98 % (98 %) of uninflected words with silent-letter endings. The occurrence of these silent letters, however, is conditioned by the type of phonological ending. Only *t*, *s*, and *d* are compatible with all three types, although with low levels of probability in some cases. While *t* is by far the most frequent silent-letter ending overall, it is found mainly in words ending with a nasal vowel, and, secondarily, in words ending with an oral vowel. The situation of *t* with nasal vowels is in fact a quasi-monopoly. The letter *e* is by far the second most frequent silent-letter ending overall, but it cannot occur in words ending with a nasal vowel. Indeed, adding *e* to a nasal vowel results in denasalization (e.g., *brun* /bʀœ̃/ [brown] and *brune* /bʀyn/ [brown, feminine form]). Yet, *e* is the most likely silent-letter ending in the other two types of phonological endings. As for *s* and *x*, they are almost equivalent in their overall frequency and they are found mainly in words ending with an oral vowel. In fact, *x* is never a silent-letter ending in words with other phonological endings. Finally, *d* is slightly less frequent than *s* and *x* overall, and it usually occurs in words ending with /ʀ/. The less frequent silent-letter endings,

for their part, can be single letters (e.g., *c*, *g*, *p*) or two-letter sequences (e.g., *ct*, *ds*, *ps*). Together, these rare silent-letter endings never represent in more than 3 % (2 %) of the words with a given type of phonological ending.

What is striking about Table 2 is the distinction between unconditional and conditional probabilities of occurrence. Although the unconditional probability distribution is the generalization of the three conditional ones, it differs from these, and these differ among themselves. For example, oral vowels are compatible with a diversity of frequent silent-letter endings, but not nasal vowels and /ʀ/. The three contexts included in Table 2, however, are broad generalizations. In Silex, users can actually find the probability distributions for each final phoneme (e.g., /ã/) or phonological rime (e.g., /aʀ/); for the graphemes corresponding to the final phoneme (e.g., for /ã/: *an*, *en*, *am*, *em*) or phonological rime (e.g., for /aʀ/: *ar*, *arr*); and for the last non-silent letter (e.g., for /ã/ and /aʀ/: *m*, *n*, *r*). Given the high number of possible contexts, more detailed data from Silex could not be presented here. We return to conditional probabilities in the following section.

The study of spelling

In this section, we demonstrate the utility of Silex for developing models of spelling acquisition and, more generally, for predicting spelling performance (accuracy and errors) by children and adults. To do so, we show how Silex was used to test and refine theoretical assumptions underlying the Fuzzy Representation Model introduced earlier (Sénéchal et al., 2016).

Regarding silent-letter endings, the key elements of Sénéchal et al.'s (2016) model are: (1) the view that, during acquisition, spellers first have to learn whether or not a given word ends with a silent letter, and then specify what that letter is; (2) the fact that both unconditional and conditional probabilities of occurrence of silent-letter endings can influence this

Table 2 Unconditonal and conditional probabilities of occurrence of specific silent-letter endings in French uninflected words that end with silent letters as a function of database (Manulex-infra or Lexique 3.80)

| | Unconditional probabilities[a] | | Conditional probabilities by type of phonological ending | | | | | |
| | | | Oral vowel (*craie, trait*) | | Nasal vowel (*pont, fond*) | | /ʀ/ (*pore, nord*) | |
	M	L	M	L	M	L	M	L
N	3,812	8,416	1,310	2,967	1,469	3,363	1,030	2,074
t	.448	.460	.198	.175	.956	.974	.043	.037
e	.360	.366	.393	.471	-[b]	-[b]	.834	.811
s	.072	.063	.179	.156	.006	.004	.027	.018
x	.062	.057	.180	.161	-	-	-	-
d	.036	.039	.016	.011	.018	.011	.087	.126
c	.005	.002	.007	.003	.004	.002	.003	.001
g	.004	.002	.001	< .001	.007	.005	.002	.001
p	.003	.002	.006	.004	.001	.001	-	-
l	.003	.002	.008	.005	-	-	-	-
h	.001	.002	.002	.005	-	-	-	-
Other SL	.002	.002	.003	.003	.002	.001	.002	.001
Two SLs	.005	.004	.007	.003	.005	.003	.002	.004

M = words found in a corpus of Grades 1–5 schoolbooks (Manulex-infra: Peereman et al., 2007); L = words found in a corpus of adult novels (Lexique 3.80: New et al., 2004); SL = silent letter; Other SL = b, ë, f, r, w, z; Two SLs = ch, cs, ct, ds, es, gt, he, ls, lt, ps, pt, rs, th, ts. All statistics are based on type counts (i.e., number of different words). A hyphen indicates the absence of a specific characteristic in the database.

[a] Unconditional probabilities include words with silent-letter endings that end with a phoneme consonant other than /ʀ/. Conditional probabilities for these words were not provided in the table, however, given that there are only three of them in Manulex-infra (12 in Lexique 3.80). In these rare words, the silent-letter ending is always s (e.g., *legs*, *aurochs*, *certes*).

[b] A silent e is not possible because it results in the denasalization of nasal vowels (e.g. *gain* /gɛ̃/ [winning] vs. *gaine* /gɛn/ [sheath]).

acquisition process as well as spelling performance. Sénéchal et al. illustrated the value of this model by examining how French-speaking children in Grades 1–3 spelled words ending with the frequent silent *t* (e.g., *début* /deby/ [start]) and words ending with the less frequent silent *d* (e.g., *bavard* /bavaʀ/ [chatterbox]). The preceding orthographic context was defined as the spelling of the phonemes in the phonological rime (e.g., /y/ spelled *u* in *début*; /aʀ/ spelled *ar* in *bavard*). Using Silex, they calculated that although the phonological rimes differed across *t*- and *d*-words, their mean consistency did not, nor did the mean conditional consistency of the correct letter (*t* or *d*). As predicted, they found an effect of unconditional probabilities on spelling accuracy: *t*-words were easier to spell than *d*-words, with the locus of difficulty being, without a doubt, the silent letter.

Then Sénéchal et al. (2016) went on to analyze silent-letter errors. They first examined the proportion of omission errors (e.g., *débu* for *début*), and then the different types of substitutions (e.g., *débue* and *débus* for *début*). Using Silex, they calculated that both unconditional and conditional probabilities for omission errors were higher in *t*-words than in *d*-words. The mean proportions of omissions made by children, however, were almost constant across word types and grades. In other terms, children's omission errors did not reflect the

differential pattern predicted by the statistical properties of French. Sénéchal et al. interpreted this finding by suggesting that silent-letter endings might be difficult to encode, or their representations difficult to activate, as children are developing their orthographic knowledge.

To analyze silent-letter substitution errors, they used the Table Generator in Silex to obtain unconditional and conditional probability distributions for alternative spellings of the silent letter. These distributions were labeled theoretical distributions, because they reflected the structure of the French orthography. They then compared the theoretical distributions to the observed error distributions for *t*- and *d*-words separately to acknowledge the difference in the frequency of *t* and *d* as silent-letter endings. Although the theoretical and observed distributions were significantly different in all cases, the conditional distributions were relatively better predictors than the unconditional ones. Indeed, children showed some sensitivity to the preceding orthographic context when making substitution errors.

As shown, Silex was central to the development and test of a theoretical model of spelling acquistion. In fact, this database could also be used to refine any existing model or theory. For instance, Sénéchal et al. (2016) made a series of

assumptions about the orthographic factors influencing spelling acquisition and performance. In what follows, we revisit their analysis of substitution errors to illustrate how these assumptions are actually testable using Silex. We chose to focus on two of their assumptions: their rime-based definition of the preceding orthographic context, and the exclusion of inflected words.

In Sénéchal et al. (2016), the conditional probability distributions used to predict errors were based on the spelling of the phonemes in the phonological rime. However, at least two other options are available in Silex, namely, the preceding letter and the preceding grapheme. In Silex, letter-based probabilities are entirely orthographic (e.g., the letter *u* in graphemes such as *u*, *eu*, *ou*, *au*, and *eau*). In contrast, grapheme-based probabilities represent a phoneme (e.g., *u* and *au* pronounced /y/ and /o/, respectively), and rime-based probabilities represent one or more phonemes. To illustrate these theoretical distributions, we selected six words from Sénéchal et al. for which letter-based, grapheme-based, and rime-based theoretical distributions differed (i.e., *bavard*, *lourd* /luʀ/ [heavy], *nord* /nɔʀ/ [north], *regard* /ʀəgaʀ/ [look], *renard* /ʀənaʀ/ [fox], and *tard* /taʀ/ [late]). The observed and theoretical distributions based on the exclusion or inclusion of inflected words are reported in Table 3.

A chi-square goodness-of-fit approach was used to compare the theoretical distributions with the observed error distributions. When these analyses were possible, they always yielded significant chi-squares, indicating that the theoretical and observed distributions were not the same. Examination of the chi-square values for the distributions that excluded inflected words, however, revealed that the theoretical distribution that came the closest to predicting children's types of substitutions was the rime-based distribution. The grapheme-based and the letter-based distributions predicted almost no other errors than substitutions with *e*. As for the unconditional distribution, it led to predictions that were contrary to the observed data.

Regarding inflectional morphology, Sénéchal et al. (2016) assumed that children's substitution errors would not reflect the silent-letter endings that are often found in inflected words. The rationale was that children might be less likely to encode silent-letter endings that are unstable in words (e.g., the *s* in *bavards*, because the word *bavard* also exists) than silent-letter endings that are stable in words (e.g., the *d* in *bavard*, because the word *bavar* does not exist). Consistent with this prediction, the inclusion of inflected words generally worsened the predictive value of all theoretical distributions in Table 3. Indeed, children never produced *ent* endings, produced a *ds* ending once, and made relatively few substitutions with *s*. These illustrative examples highlight the richness of the novel information included in Silex as well as its flexibility, making it a useful tool to model the development of orthographic representations.

Table 3 Observed and theoretical distributions of the types of substitution errors ($N = 110$ errors) as a function of the exclusion or inclusion of inflected words

Error distributions	Types of substitution errors				$\chi^2(3)$
	e	t	s	Other[a]	
Observed	.58	.24	.07	.11	-
Theoretical distributions excluding inflected words					
Conditional: Rime	.62	.20	.14	.04	15.6
Conditional: Grapheme	.91	.05	.03	.01	-[b]
Conditional: Letter	.77	.04	.03	.15	108.8
Unconditional	.37	.46	.07	.09	25.7
Theoretical distributions including inflected words					
Conditional: Rime	.24	.07	.18	.51	141.3
Conditional: Grapheme	.38	.02	.14	.45	319.5
Conditional: Letter	.44	.02	.24	.30	247.8
Unconditional	.14	.22	.38	.26	197.6

The rime distribution is the sum of the distributions obtained for each specific word, because the rimes varied across words. Given that participants could use different preceding orthographic contexts with the same pronunciation for grapheme and rime distributions, we used conditional probabilities based on the word's final phoneme and phonological rime, respectively. The letter distribution is based on conditional probabilities of occurrence of final letters, silent or not, given the preceding *r* (e.g., a final *c* is predicted by words like *porc* and *parc*, although it is silent in *porc* and pronounced in *parc*). All theoretical distributions were constructed using the Silex Table Generator, after removal of the probability of occurrence of the correct letter (i.e., *d*). The options selected for all distributions were: (1) the corpus of Grades 1–5 schoolbooks (Manulex-infra: Peereman et al., 2007); and (2) probabilities based on type counts (i.e., number of different words).

[a] The types of observed errors (and their frequency) included in the Other category were: ed (3); de, es (2 each); b, é, z, ds, td (1 each). As for the types of predicted errors, they varied as a function of theoretical distributions.

[b] χ^2 could not be computed because there were two cells with less than five predicted errors.

Discussion

The Silex database consists of three kinds of Excel workbooks: a set of Stimuli Selector workbooks that allow researchers to select words based on a variety of statistics and word characteristics; a Table Generator workbook that allows researchers to build consistency distribution tables by selecting specific phonological or orthographic units; and a Master File workbook, from which all statistics were derived, and that allows interested researchers to compute other statistics. Silex was based on the words found in two French corpora: one consisting of adult novels (Lexique 3.80: New et al., 2004), and the other of elementary-school readers (Manulex-infra: Peereman et al., 2007). Hence, Silex presents in a unified database statistics based on and provided for a total of 119,664 word entries.

Silex was created in response to the absence of a lexical database that reflected our theoretical views on spelling (e.g., Sénéchal et al., 2016). As shown in the present report, the novelty of Silex resides in its original spelling-oriented approach to the quantification of consistency, which had a considerable impact on the indices computed and might better reflect the specificity of spelling acquisition (for a comparison of spelling and reading, see Treiman & Kessler, 2014). The major difference, compared to previous works (e.g., Peereman et al., 2007, 2013), is our theoretically-driven treatment of silent-letter endings that has impacted the computation of all consistency indices provided. Another important novelty of Silex is that users can also find a variety of unconditional- and conditional-consistency indices for word endings, enabling researchers to investigate how grain-size theory (e.g., Ziegler & Goswami, 2005) and statistical learning (e.g., Negro et al., 2014) apply to spelling. Specifically, Silex provides conditional-consistency indices based on orthographic contexts that vary in nature and size. This could allow researchers not only to show the effect of orthographic context on spelling (e.g., Pacton et al., 2005; Treiman et al., 2002), but also to test the predictive value of different definitions of this orthographic context.

Another element of theoretical importance that is central to Silex is our approach to the treatment of inflectional morphology. In our view, the distinction between inflected and uninflected words is fundamental to the quantification of consistency, because inflected words contain unstable word endings (e.g., the silent *s* in *bavards*) contrary to the stable endings of uninflected words (e.g., the silent *d* in *bavard*). Silex is the first database to provide consistency indices based on uninflected words only, in which the word endings are more likely to be represented in memory due to their stability. Consistency indices based on all words and on inflected words only, however, are also available, thus allowing researchers to select and test models of spelling development. All these elements are combined in a database designed to be as flexible as possible, with the aim of providing a diversity of statistics for researchers interested in spelling.

In the present report, we demonstrated the utility of Silex to describe the silent-letter phenomenon in French. We found that silent-letter endings are usually *t*, *e*, *s*, *x*, or *d*, and that the occurrence of these letters is conditioned by the phonological ending of words. We identified the types of phonological endings where a diversity of silent-letter endings is possible (i.e., vowels and /ʀ/). Importantly, the statistics reported here for the corpus of elementary schoolbooks (Peereman et al., 2007) and for the corpus of French adult novels (New et al., 2004) were strikingly similar, suggesting that they might be generalizable to the French orthographic lexicon. We also showed that silent-letter endings in French needed to be quantified and analyzed independently from phoneme-grapheme consistency. Compared to previous quantifications of consistency

(Peereman et al., 2007, 2013; Ziegler et al., 1996), Silex has the advantage of providing unconditional and conditional probabilities of occurrence of specific silent-letter endings.

As shown, Silex can prove useful for empirical studies of spelling. First, we argue that models for predicting spelling accuracy (e.g., Lété et al., 2008) in children and adults should include as distinct predictors the presence (or absence) of a silent-letter ending, as well as probabilities of occurrence of specific silent-letter endings. Silent-letter endings are indeed the cause of specific spelling difficulties (e.g., Sénéchal, 2000), and we found these letters to be a major cause of spelling inconsistency. Second, Silex provides researchers with the means to control for the statistical properties of silent-letter endings in French. It could be of interest, then, to study the effects of derivational morphology (e.g., Sénéchal et al., 2006) and spelling inconsistency (e.g., Jubenville et al., 2014) on spelling accuracy while controlling for unconditional and conditional probabilities of occurrence of silent-letter endings.

As mentioned, the content of Silex is important not only for empirical but also for theoretical reasons. The inclusion of unconditional- and conditional-consistency indices can enable researchers to find the nature and size of the context that could influence spelling acquisition and performance (e.g., Treiman & Kessler, 2014; Ziegler & Goswami, 2005). The analyses of silent-letter errors presented here showed how Silex could be used to this end. Another important feature of Silex is the distinction between inflected and uninflected words. Our analyses suggested that learning and applying the rules of inflectional morphology (e.g., Fayol et al., 2006) might involve different processes than remembering and memorizing word-specific silent-letter endings (e.g., Sénéchal et al., 2006).

All things considered, the main theoretical contribution of Silex could be to shed new light on how statistical learning (e.g., Negro et al., 2014) and grain-size theory (Ziegler & Goswami, 2005) apply to spelling. First, as shown by Sénéchal et al. (2016), a specific type of silent-letter error made by children (i.e., omission errors) did not seem to reflect the predictions based on the statistical properties of French, while the types of substitution errors did. Second, as regards contextual effects, the analyses presented here suggested that children, while reading, might extract orthographic knowledge from phonologically processed strings of letters, and not from mere visual processing of these letters. Third, our analyses also suggested that children, when extracting orthographic knowledge from words, might be less likely to encode the unstable silent-letter endings marking inflectional morphology than the lexically stable ones. Naturally, further research is needed for each of these issues, and the statistics necessary to this end are available in Silex. In sum, the novel information provided in Silex as well as the flexibility of this database should enable researchers to advance our understanding of developing as well as skilled spelling performance.

Appendix

Table 4 Means (and standard deviations) for six variables in the Silex Stimuli Selectors as a function of corpus and inclusion or exclusion of inflected words

	Whole-word consistency	Phoneme-grapheme consistency	Consistency for silent orthographic endings			
			Unconditional	Conditional		
				Letter	Grapheme	Rime[a]
Manulex-infra						
Including inflected words	.65	.72	.26	.37	.44	.39
N = 39,197	(.12)	(.14)	(.16)	(.22)	(.26)	(.19)
Excluding inflected words	.70	.74	.53	.68	.81	.79
N = 12,974	(.12)	(.14)	(.28)	(.30)	(.29)	(.30)
Lexique 3.80						
Including inflected words	.65	.72	.24	.36	.42	.40
N = 103,843	(.11)	(.13)	(.14)	(.21)	(.25)	(.17)
Excluding inflected words	.71	.74	.54	.70	.82	.81
N = 29,833	(.12)	(.13)	(.28)	(.29)	(.28)	(.28)

Two Stimuli Selector workbooks were used to construct this table: The one for the Grades 1–5 readers included in Manulex-infra (Peereman et al., 2007), and the one for the novels included in Lexique 3.80 (New et al., 2004). In each workbook, two worksheets were used – the one for all words, and the one for uninflected words only. In these worksheets, the six following variables were used: WW.ME.TY.P, PG.PD.PW.ME.TY.P, OE.AL.UN.TY.P, OE.AL.CD.LE.TY.P, OE.AL.CD.G1.TY.P, and OE.AL.CD.G2.TY.P. All these variables are based on type counts (i.e., number of different words). See the Silex User Guide for definitions of the abbreviations in the variable labels.

[a] For the Rime variable, only the words ending with /R/ preceded by a vowel are included in the calculations. Consequently, for Manulex-infra, N = 3,450 (including inflected words) and 1,480 (excluding inflected words), whereas for Lexique 3.80, N = 8,477 (including inflected words) and 3,255 (excluding inflected words).

References

Catach, N. (1995). *L'orthographe française*. Paris: Éditions Nathan.

Content, A. (1991). The effect of spelling-to-sound regularity on naming in French. *Psychological Research, 53,* 3–12.

Content, A., Mousty, P., & Radeau, M. (1990). Brulex: Une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique, 90,* 551–566.

Fayol, M., Hupet, M., & Largy, P. (1999). The acquisition of subject-verb agreement in written French: From novices to experts' errors. *Reading and Writing: An Interdisciplinary Journal, 11,* 153–174.

Fayol, M., Totereau, C., & Barrouillet, P. (2006). Disentangling the impact of semantic and formal factors in the acquisition of number inflections: Noun, adjective and verb agreement in written French. *Reading and Writing, 19,* 717–736. doi:10.1007/s11145-005-1371-7

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex. *Frontiers in Psychology, 2,* 1–10. doi:10.3389/fpsyg.2011.00306

Jaffré, J. (2005). Introduction: The orthography of French. *L1 – Educational Studies in Language and Literature, 5,* 353–364. doi:10.1007/s10674-005-4493-6

Jubenville, K., Sénéchal, M., & Malette, M. (2014). The moderating effect of orthographic consistency on oral vocabulary learning in monolingual and bilingual children. *Journal of Experimental Child Psychology, 126,* 245–263. doi:10.1016/j.jecp.2014.05.002

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers, 36,* 156–166.

Lété, B., Peereman, R., & Fayol, M. (2008). Consistency and word-frequency effects on spelling among first- to fifth-grade French children: A regression-based study. *Journal of Memory and Language, 58,* 952–977.

Negro, I., Bonnotte, I., & Lété, B. (2014). Statistical learning of past participle inflections in French. *Reading and Writing, 27,* 1255–1280. doi:10.1007/s11145-013-9485-9

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers, 36,* 516–524.

Pacton, S., & Deacon, S. H. (2008). The timing and mechanisms of children's use of morphological information in spelling: A review of evidence from English and French. *Cognitive Development, 23,* 339–359. doi:10.1016/j.cogdev.2007.09.004

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

Pacton, S., Borchardt, G., Treiman, R., Lété, B., & Fayol, M. (2014). Learning to spell from reading: General knowledge about spelling patterns influences memory for specific words. *The Quarterly Journal of Experimental Psychology, 67,* 1019–1036. doi:10.1080/17470218.2013.846392

Peereman, R., & Content, A. (1999). LEXOP: A lexical database providing orthography-phonology statistics for French monosyllabic

words. *Behavior Research Methods, Instruments, & Computers, 31,* 376–379.

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

Peereman, R., Sprenger-Charolles, L., & Messaoud-Galusi, S. (2013). The contribution of morphology to the consistency of spelling-to-sound relations: A quantitative analysis based on French elementary school readers. *L'Année Psychologique, 113,* 3–33.

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. doi:10.1016/j.jml.2014.01.003

Sénéchal, M. (2000). Morphological effects in children's spelling of French words. *Canadian Journal of Experimental Psychology, 54,* 76–85.

Sénéchal, M., Basque, M. T., & Leclaire, T. (2006). Morphological knowledge as revealed in children's spelling accuracy and reports of spelling strategies. *Journal of Experimental Child Psychology, 95,* 231–254. doi:10.1016/j.jecp.2006.05.003

Sénéchal, M., Gingras, M., & L'Heureux, L. (2016). Modeling spelling acquisition: The effect of orthographic regularities on silent-letter representations. *Scientific Studies of Reading, 20,* 155–162. doi:10.1080/10888438.2015.1098650

Treiman, R., & Kessler, B. (2014). *How Children Learn to Write Words.* Oxford: Oxford University Press.

Treiman, R., Kessler, B., & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language, 47,* 448–468. doi:10.1016/S0749-596X(02)00010-4

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

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.