Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-021-02029-1

BRIEF REPORT

Diversity matters: The sensitivity to sublexical orthographic regularities increases with contextual diversity

Fabienne Chetail[1] · **Karinne Sauval**[1]

Accepted: 3 October 2021

Abstract

Readers capture statistics about letter co-occurrences very rapidly. This has been demonstrated with artificial lexicons and/or with restricted sets of orthographic regularities. The aim of the study was twofold: To examine the learning of new orthographic regularities in a more incidental exposure paradigm, and to investigate the impact of the diversity of letter contexts in which new orthographic regularities appear. For 2 months, participants played detection games for 20 min per day and were exposed to a large set of pseudowords, some of which included new bigrams (e.g., GK). Half of the new bigrams occurred in eight different items (high contextual diversity) and the other half were presented in only two items (low context diversity). At six time points, the participants performed a "wordlikeness" task in which they chose between two new pseudowords the one that was more similar to the items previously exposed (e.g., PUGKALE vs. PUGZALE). The results showed that the participants very rapidly developed a preference for items with a frequent new bigram and that this sensitivity increased steadily over the 2 months. Furthermore, the sensitivity to these new orthographic regularities was higher in cases of high letter contextual diversity. The latter result parallels what is observed at a lexical level with semantic contextual diversity.

Keywords Visual word recognition · Contextual diversity · Frequency · Orthographic regularities · Bigram frequency

Introduction

We are able to pick up regularities from the flow of stimulation very efficiently and rapidly, whatever the inputs (see Conway & Christiansen, 2005; Perruchet & Pacton, 2006). When reading, people extract regularities in letter sequences, allowing them to learn which regularities are more likely without being explicitly taught. The term "orthographic regularities" is used to refer to the kind of regularities learned incidentally throughout print exposure. These regularities correspond to facts about the distribution of single letters or letter sequences, without direct reference to higher-order levels such as phonological or morphological units (Chetail, 2017). For example, the letters S and A co-occur more frequently in English words than the letters J and A, the letter R is more often doubled than the letter D. Many studies showed that readers rapidly capture regularities occurring in their orthography (e.g., Cassar & Treiman, 1997; Chetail, 2017; Gingras & Sénéchal, 2019; O'Brien, 2014; Pacton et al., 2001). For example, after only a few months of exposure to print, young readers are sensitive to statistical facts about consonant doublets. In a "wordlikeness" task (i.e., choosing between two items which one is more wordlike), children consider pseudowords such as *ommera* to be more wordlike than pseudowords such as *ovvera*, which is consistent with the fact that the letter M is frequently doubled (in French or English) whereas the letter V is never doubled (e.g., Cassar & Treiman, 1997; O'Brien, 2014; Pacton et al., 2001). When sensitivity to finer orthographic regularities is tested (e.g., position and frequency of bigrams within visual stimuli), only a few minutes of exposure to new regularities is sufficient for adult readers to become sensitive to such properties of bigrams (e.g., Chetail, 2017; Samara & Caravolas, 2014).

The development of sensitivity to orthographic regularities is typically explained in terms of frequency of exposure: Because readers are repeatedly exposed to some patterns of letter co-occurrences, they find them more typical of their orthography (e.g., Cassar & Treiman, 1997; Pacton et al.,

---

✉ Fabienne Chetail
fabienne.chetail@ulb.be

[1] Laboratoire Cognition Langage, & Développement (LCLD), Centre de Recherche Cognition et Neurosciences (CRCN), Université Libre de Bruxelles (ULB), Av. F. Roosevelt, 50 / CP 191, 1050 Brussels, Belgium

2001; Samara & Caravolas, 2014) and process them more easily (e.g., Chetail, 2017; Pitchford et al., 2008; Samara & Caravolas, 2014). This account is in line with the interpretation of the word frequency effect (e.g., Stanners et al., 1975; Whaley, 1978): Because readers are repeatedly exposed to some words when reading texts, these items become more familiar than others, are more accessible in the lexicon, and are processed more rapidly in tasks such as lexical decision (e.g., Balota et al., 2001; Connine et al., 1990). However, this interpretation of word frequency effects has been called into question (e.g., Adelman et al., 2006; Baayen, 2010; Brysbaert & New, 2009; Perea et al., 2013; Plummer et al., 2014). The accessibility of words in memory would not be determined by a mere count of past presentations but would depend on the number of contexts in which the words occurred (referred to as "contextual diversity"). Adelman et al. (2006) operationalized contextual diversity in terms of the proportion of texts in which a given word occurs. They showed in adults that word frequency effects were eliminated when contextual diversity was taken into account whereas contextual diversity facilitated word processing in the lexical decision task, even when word frequency was controlled (see also Keuleers et al., 2010). Similar results were obtained in developing readers. Fourth graders exhibited a facilitative effect of contextual diversity on lexical decision latencies, with words appearing in many contexts being recognized faster than words appearing in the same contexts (Perea et al., 2013). Such results have been replicated with more direct measures of contextual diversity that take into account the degree of similarity between the different contexts in which a word appears (e.g., Hoffman et al., 2013; Hsiao & Nation, 2018; Johns et al., 2016; Jones et al., 2012). Especially, Hoffman et al. (2013) showed that a measure of semantic diversity based on the linguistic contexts in which words occur is a strong predictor of performance in a semantic judgment task in adults, accounting for unique variance beyond variables such as lexical frequency or word imageability. Hence, more than the number of word occurrences in texts, it would be the diversity of passages in which words occur throughout print exposure that would be important.

Interestingly, the notion of "contextual diversity" can be extended to sublexical processing. Indeed, just as words are generally not encountered in isolation but in context (i.e., embedded in sentences), letter clusters such as bigrams are not encountered in isolation but within letter strings. Hence, when readers are exposed to new bigrams, they are necessarily exposed to the letter contexts in which these bigrams occur. One may thus assume that letter contextual diversity is critical for new bigrams to become salient orthographic regularities rather than mere repeated exposure, just as semantic and syntactic diversity is essential for words to become salient. The present study aimed to test this hypothesis. If familiarity with letter clusters depends on the letter context in which they occur rather than on mere repeated exposure, this would enable us to specify the conditions under which sensitivity to orthographic regularities develops spontaneously.

To examine this issue, we presented new orthographic regularities to readers while manipulating the frequency and letter contextual diversity of these regularities. More precisely, participants were exposed to pseudowords that included new bigrams (i.e., illegal in their orthography). After the exposure phase, the participants performed a wordlikeness task in which they had to choose between two items which was more wordlike (i.e., more similar to the items of the exposure phase). One of the items entailed a new exposed bigram (critical item), the other was a random item with similar orthographic properties to the critical item but with no new bigram. Overall, if learning of new regularities takes place, we expected participants to select the critical item above chance level in such a two-alternative forced-choice task (e.g., Cassar & Treiman, 1997; Chetail, 2017; Gingras & Sénéchal, 2019; Pacton et al., 2001). Furthermore, during the exposure phase, the new bigrams were either repeated a lot (high-frequency bigrams) or not (low-frequency bigrams). According to frequency accounts, the selection rate of critical items in the wordlikeness task should be higher if they entail a high-frequency new bigram rather than a low-frequency one. Critically, bigram frequency was manipulated orthogonally with contextual diversity: New bigrams could be presented almost always in the same pseudowords (i.e., low letter contextual diversity) or in different pseudowords (high letter contextual diversity). If contextual diversity plays a role at a sublexical level, items with high-contextual diversity bigrams should be preferred in the wordlikeness task over the low-contextual diversity ones, and the bigram frequency effect should disappear in case of low contextual diversity.

To test these hypotheses, we used an ambitious paradigm that, on the one hand, allowed us to track the development of sensitivity to orthographic regularities over a 3-month period, and, on the other hand, maximized the incidental aspect of exposure to new regularities. For 2 months, participants played a series of games on their computers, at home. During each daily session, they were exposed to pseudowords entailing new bigrams, while their goal was simply to achieve the highest score in each game. The presence of pseudowords was explained as a necessity for the games to be played. The stimuli were inspired by the Basque language to keep a linguistically plausible material. At several time points, the participants came to the lab to perform a wordlikeness task (before exposure, and after 5, 10, 20, 60 days, and 1 month after the end of exposure). This allowed us to examine the emergence of sensitivity to new orthographic regularities that are linguistically plausible and how

Table 1 Characteristics of bigrams and pseudowords used in the exposure phase according bigram frequency and contextual diversity

	Bigram frequency			
	High		Low	
	High contextual diversity	Low contextual diversity	High contextual diversity	Low contextual diversity
New bigrams	fv, jx, tk, gv, kz, vj	bp, jn, xg, df, vh, xz	fx, mg, pv, jh, pb, zx	bk, kj, vf, lp, lx, xk
Mean letter frequency[a]	30,141	36,370	33,749	26,318
Number of unique pseudowords entailing a new bigram (contextual diversity)	8	2	8	2
Number of unique pseudowords per condition[b,c]	48 (6 × 8)	12 (6 × 2)	48 (6 × 8)	12 (6 × 2)
Number of repetitions of unique pseudowords per daily session[c]	2	8	0.5	2
Frequency of the new bigrams per daily session	16 (8 × 2)	16 (2 × 8)	4 (8 × 0.5)	4 (2 × 2)
Frequency of the new bigrams during the whole exposure phase (60 days)	960	960	240	240

[a] In number of occurrences per million

[b] The sum of the row corresponds to the 120 experimental pseudowords

[c] The sum pairwise multiplications of the two rows correspond to the 240 items exposed each day

it is maintained over time. We expected participants to be at chance in the wordlikeness task before exposure (i.e., no preference between the critical bigram and the random item with the same characteristics), and then to show a clear preference for the critical item, which strengthens over time (sessions 2–5) and is maintained without exposure (session 5).

Experiment

Method

Participants

Twenty-four volunteers participated in the entire study (46% were females, mean age: 23.04 years, range 19–36).[1] They were all native French speakers and reported having a normal or corrected-to-normal vision. None of them reported being familiar with the Basque language. They received financial compensation for their participation.

Materials

Exposure phase To test the development of sensitivity to new bigrams, we used 24 illegal bigrams in French, partly inspired by existing bigrams in Basque (see Table 1). These bigrams contained either two frequent letters in French (e.g., *bp*, *df*) or a frequent letter with a low- to very low-frequency letter (e.g., *vj*, *fx*). The mean letter frequency was 31,110 occurrences per million (SD = 22,877).[2] In the exposure phase, these bigrams were presented to the participants according to the orthogonal manipulation of two factors. The bigrams were exposed either only four times a day or 16 times a day (low- vs. high-frequency bigrams) and were either always embedded in the same two pseudowords or distributed in eight different pseudowords (low vs. high contextual diversity) (see Table 1). With these bigrams, we devised 120 pseudowords of five to eight letters, derived from the Basque lexicon by changing one or two letters of Basque words (see Table 2). The use of Basque meant that the participants' attention was not drawn specifically to the target bigrams, as Basque spelling contains many bigrams that are either non-existent or very infrequent for French speakers. The position of the critical bigrams in the pseudowords was between the first and the sixth letters (mean = 3.31; SD = 0.97) and the mean bigram frequency of the pseudowords was 4,329 occurrences per million (SD = 2,424) (see Table 3, in the Appendix).

---

[1] Initially, 30 participants were enrolled, but six of them stopped or were excluded in the first weeks.

[2] All letter and bigram frequencies were computed from the Lexique database (New et al., 2004) as token frequencies with no positional constraint within words and no word length constraint.

Table 2 Stimuli used in the exposure phase

Bigram frequency							
High				Low			
High-contextual diversity		Low-contextual diversity		High-contextual diversity		Low-contextual diversity	
Word	Bigram	Word	Bigram	Word	Bigram	Word	Bigram
xafva	fv	ebpetu	bp	sufxo	fx	eubkal	bk
befve	fv	kabpelo	bp	pefxo	fx	jaubkor	bk
suafva	fv	lidfa	df	horfxe	fx	Ifpilu	fp
ifvoki	fv	jadfalta	df	umefxo	fx	ofpetsu	fp
defvio	fv	kojnea	jn	tafxan	fx	Mokje	kj
agefvin	fv	pejnil	jn	guanfxe	fx	Larakja	kj
goufvet	fv	suvho	vh	pelufxe	fx	Falxa	lx
fofvazio	fv	karivhio	vh	leziofxo	fx	balxiar	lx
begva	gv	xixgar	xg	bujhe	jh	Buvfa	vf
magvu	gv	jaxgune	xg	pejha	jh	ikevfan	vf
ogvatu	gv	ixzitu	xz	ajhibe	jh	Atexka	xk
bigvia	gv	oixzeko	xz	nujheo	jh	luzexko	xk
eragvi	gv			jharin	jh		
kogvatu	gv			zijhiko	jh		
zabagvo	gv			esjhaba	jh		
azagvune	gv			deisajho	jh		
sejxa	jx			lemga	mg		
pajxu	jx			ikamgu	mg		
exajxo	jx			ijemgu	mg		
tijxak	jx			amgeta	mg		
ajxore	jx			babemgo	mg		
brajxea	jx			amgopos	mg		
lejxibo	jx			ihemgin	mg		
obidujxu	jx			akamgabe	mg		
mokzu	kz			dapba	pb		
ikzai	kz			epbor	pb		
idakze	kz			kapbio	pb		
zikzor	kz			opbide	pb		
okzopo	kz			kopbexu	pb		
bilukze	kz			urbopba	pb		
anikzar	kz			azupbre	pb		
lakzanka	kz			ipbitatu	pb		
zetko	tk			kopva	pv		
ostka	tk			zapva	pv		
tkabut	tk			tipve	pv		
motkiz	tk			apvike	pv		
neutko	tk			zapvaka	pv		
atkofia	tk			lanpvan	pv		
litkoka	tk			dipvoma	pv		
jukutkia	tk			supvizio	pv		
tavje	vj			kizxa	zx		
savju	vj			tizxo	zx		
avjiko	vj			etezxa	zx		
livjor	vj			jaizxo	zx		
movjal	vj			soizxi	zx		
elauvja	vj			jazxeta	zx		
lanuvje	vj			tazxeta	zx		
sovjegun	vj			imazxeta	zx		

Experimental phase We devised 144 pairs of new pseudowords that were not used in the exposure phase (see Table 3, in the Appendix). In the experimental condition, 72 pairs were used to test the development of sensitivity to new regularities. The first stimulus in pairs entailed one of the 24 new bigrams exposed in the daily sessions (critical item). Three different pseudowords were devised with each new bigram yielding 18 items in the four experimental conditions (bigram frequency * contextual diversity). The second stimulus of the pair had the same letters, except that the critical new bigram was replaced by another bigram which was either illegal or of very low frequency in French and which had never been presented in the exposure phase (control item). These bigrams could be repeated among the 72 control items but never in association with the same critical new bigram. We also created a control condition with 72 pairs. In each pair, both items contained at least one illegal or very low-frequency bigrams in French, which were never exposed in the exposure phase. The mean bigram frequency of the pseudowords was very similar between the items of the pairs in both conditions (5,073 and 5,197 occurrences per million in the experimental pairs, 5,621 and 5,669 in the control pairs).

Procedure

Exposure phase Figure 1 presents a summary of the exposure phase. For 60 days, the participants played a series of four small computer-based games for 20 min per day. Each daily session started with a probe detection task. In this task, a probe was presented (a letter, a bigram, or an open-bigram, e.g., *i*, *ko*, and *d_a* respectively) with a list of 20 lowercase pseudoword targets on a grid. The participants had to click on the targets containing the probe as quickly as possible within a maximum of 30 s. The probes were devised so that the single letters or bigrams did not correspond to the critical new bigrams. Each day, the participants played this game four times with three different probes and grids each time, leading to the exposure of 240 items in one session (three grids of 20 targets times four). Being good at this game allowed the participants to earn extra lives for the following game. The next game was a brick-breaking game specifically designed for the experiment. The participants had to break as many bricks as possible with a ball and could catch or avoid falling items, leading to positive or negative effects (e.g., increase or decrease of the ball speed, of the pad size). We designed 400 levels of increasing difficulty. In a daily session, the participants had four runs of brick-breaking, each lasting 2.5 min. They were instructed to do their best to achieve a high score. A game session usually ended with two other short games, namely a 1-min lexical decision task and a 1-min letter detection task. In the lexical decision task, the participants were presented with 60 items, including 30 pseudowords previously exposed and 30 new pseudowords never exposed. The participants' task was to decide as rapidly as possible whether they had already encountered these items during the daily sessions. Participants had a maximum of 1,500 ms to respond. During the exposure phase, we used four different lists, each repeated every 4 days. In the letter detection task, the participants had to decide whether a given letter briefly presented was present in a carrier item subsequently presented. There were 52 trials per day and we used ten different lists, each repeated every

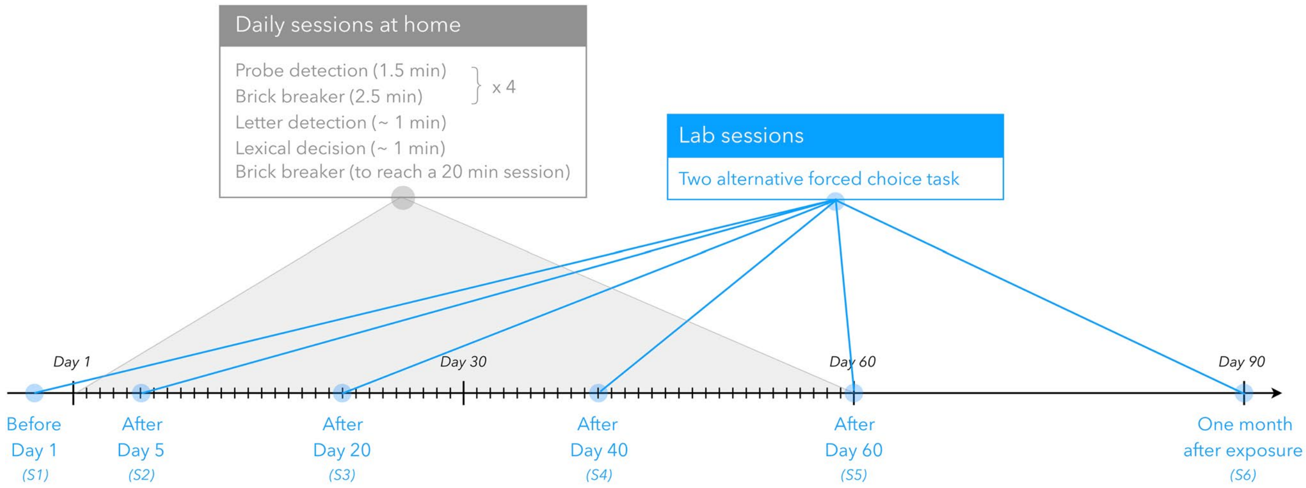

Fig. 1 Summarized procedure of the experiment with the exposure phase in grey and the test phase in blue. Vertical ticks correspond to the daily game sessions

10 days. The carrier items were pseudowords that had never been exposed. Again, the participants had a maximum of 1,500 items to respond.

In a daily session, the central task was the probe detection task, which allowed us to expose the participants to the 24 new bigrams. The other tasks had different roles (as filling tasks, adding challenge, and maintaining motivation by varying the tasks). Further, to increase motivation throughout the experiment, the participants were challenged to achieve the highest scores. They were informed that at the end of the experiment, the participants with the highest scores in the tasks would receive higher financial compensation.

Test phase At six different time points, we tested the participants' sensitivity to the new regularities using a wordlikeness task. They did the task before starting exposure (S1, more or less 10 days before), after 5 days (S2), after 20 days (S3), after 40 days (S4), after 60 days (S5, i.e., the day after the end of the exposure phase), and 1 month after the end (S6). The task was programmed with PsychoPy (Peirce et al., 2019). There were 144 trials. In each trial, a pair of uppercase stimuli was presented on the screen. The position of the two items (right or left) was randomly determined on each trial. At S2–S6, the participants were asked to choose, between the two items, which one looked the most familiar to the words they encountered during their daily sessions. At S1, the participants were simply asked to choose the one they preferred. The material used was the same in each session. There was no time constraint, but the participants were asked to be as spontaneous as possible.

Results

In all the experiments, the statistical analyses were run with R packages (R Core Team, 2018) under the RStudio environment (RStudio Team, 2016). Raw data and script analyses are available here: https://osf.io/8uvhs/.

Control tasks First, we used the participants' performance in two tasks (probe detection and lexical decision) to ensure they were properly exposed to bigrams throughout the study. In the probe-detection task, we analyzed the evolution of scores over the sessions. Out of the 1,440 daily sessions (24 participants times 60 days), data were missing for two sessions due to a technical problem. For the remaining sessions, we decided to group scores in 12 periods of five consecutive days. Overall, the participants correctly detected the probes, whether they were single letters (M = 92%, SD = 0.89%), bigrams (M = 90%, SD = 1.00%), or open bigrams (M = 79%, SD = 1.79%). Unsurprisingly, a repeated-measures ANOVA showed that the performance varied with the type of probes, $F1(2, 46) = 164.02$, $p < .001$ and $F2(2,251) = 83.12$, $p < .001$, with open-bigrams being the most difficult to detect. Moreover, as presented in Fig. 2, the performance of the participants increased across sessions as shown by a linear mixed model analysis, $\beta = 0.70$, $t = 13.58$, $p < .001$. This increase varied according to probes, the improvement being the highest for open bigrams: $\beta = 0.57$, $t = 10.19$, $p < .001$ for letters, $\beta = 0.74$, $t = 9.56$, $p = .05$ for bigrams, and $\beta = 1.39$, $t = 11.00$, $p < .001$ for open bigrams.

In the lexical decision task, we also analyzed the evolution of scores over the sessions. As the participants' task was to recognize items that were repeatedly presented in the exposure phase, we expected the performance to be better with time if they were properly exposed to the pseudowords in the probe detection task. Out of the 1,440 daily sessions, the data for one session were missing due to a technical problem. Extremely short reaction times (< 300 ms) were excluded from the data, as well as trials reaching the 1,500 ms deadline (7.94% of the data). For the analyses, we again decided to group the scores into 12 periods of five consecutive days. As presented in Fig. 3, a linear mixed-model analysis showed that the reaction times and error rates varied across sessions, with a general decreasing trend, $\beta = -8.83$, $t = -10.91$, $p < .001$ and $\beta = -1.05$, $t = -14.07$, $p < .001$, respectively. Note that individual analyses suggest that two participants did not perform the task correctly as they had very high error rates close to chance (45% and 49%) as well as a high proportion of responses given in less than 300 ms (respectively 51%

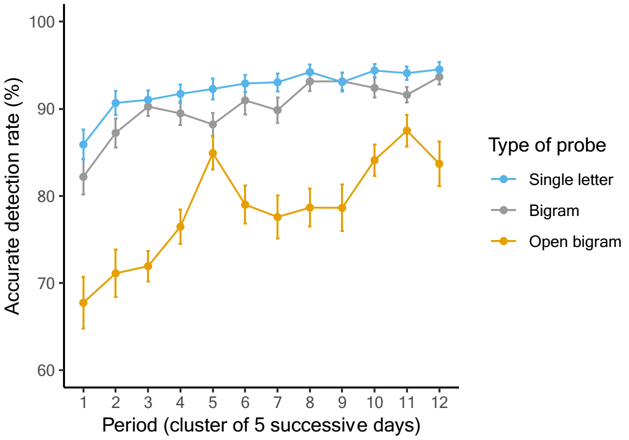

Fig. 2 Percentage of correct responses (with standard errors) in the probe detection task as a function of probes and time (five consecutive day periods)

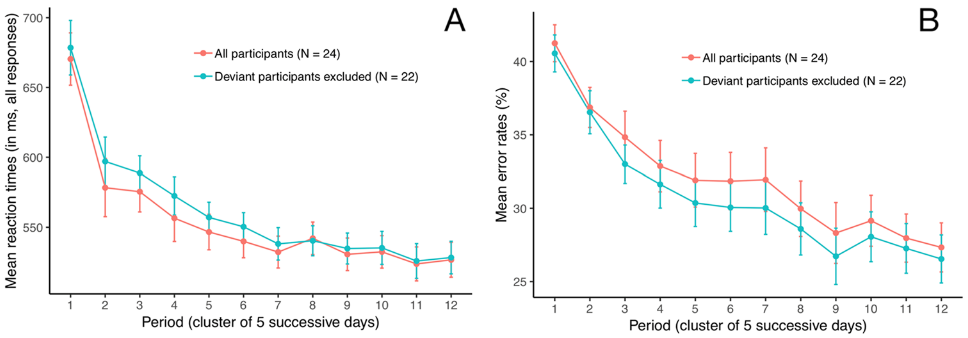

Fig. 3 Mean reaction times (**A**) and error rates (**B**) in the lexical decision task (with error bars), with or without the exclusion of two deviant participants, as a function of time (five consecutive day periods)

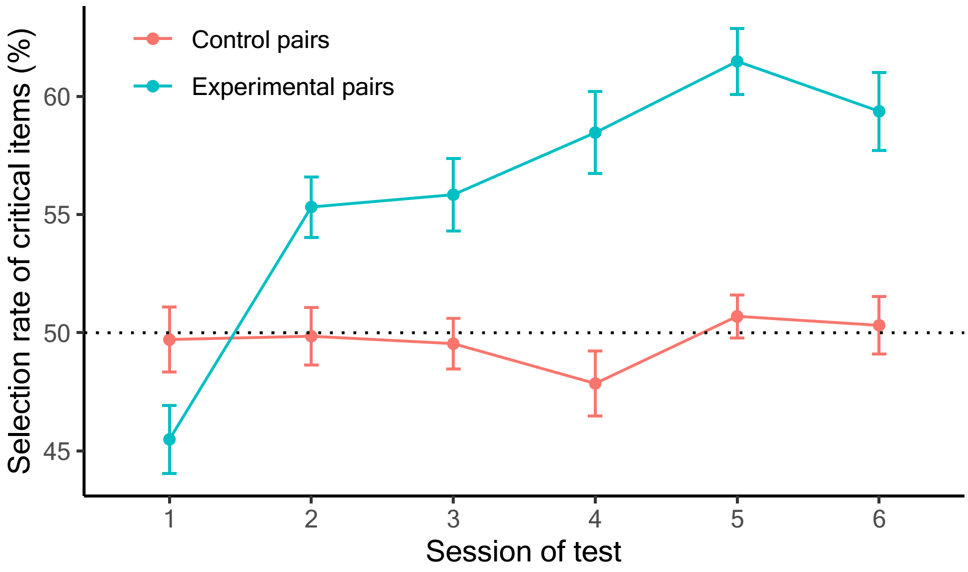

Fig. 4 Selection rate of critical items in control and experimental pairs in the wordlikeness task (with standard errors). The dotted line represents chance level

and 47%). However, as presented in Fig. 3, the general pattern of results for the whole sample does not seem to be affected by these two deviant participants.

Wordlikeness task Trials with extremely short (< 200 ms) or long (> 10,000 ms) reaction times were excluded (0.03 % of the data). As shown in Fig. 4, the choice of a specific item in pairs was at chance in the control pairs.[3] In experimental pairs, the selection rate of the critical items (i.e., including an exposed new bigram) varied across sessions. More precisely, one-sample t-test (with comparison to chance level) showed that the selection rate was significantly lower than chance in session 1 (45%), $t(23) = -3.12$, $p < .001$, while it was significantly higher than chance in the following sessions (sessions 2–6), $t(23) = 8.79$, $p < .001$. Between sessions 2 and 5, a linear contrast showed that the selection rate increased significantly, $F(3,69) = 11.94$, $p < .001$, and even after 1 month without exposure (session 6), it was still significantly higher than chance, $t(23) = 5.68$, $p < .001$.

Second, we conducted an ANOVA with the two manipulated factors (bigram frequency and contextual diversity) in sessions 2–6. The analyses showed that the selection rate of critical items was higher if they entailed a new bigram frequently exposed rather than a less frequent new bigram,

---

[3] One item in the pair was arbitrarily designed as the "critical" one.

$F1(1, 23) = 13.95$, $p < .001$, $F2(1, 68) = 4.2$, $p = .04$. An effect of contextual diversity was also obtained, showing that the selection rate of critical items was higher if they entailed a bigram presented in many different pseudowords rather than in a few pseudowords, $F1(1 ,23) = 29.51$, $p < .001$, $F2(1, 68) = 4.8$, $p = .03$. Moreover, the interaction between the two factors was significant, $F1(1 ,23) = 27.12$, $p < .001$, $F2(1, 68) = 6.2$, $p = .01$. As shown in Fig. 5, this is due to the fact that the effect of contextual diversity was clearly present for high-frequency bigrams, but not for low-frequency bigrams.

Discussion

The aim of the present study was to examine how the sensitivity to new orthographic regularities develops over time and whether letter contexts surrounding new regularities influence their processing. To do this, we asked participants to play a series of games on their computer for 2 months. During each session, they were exposed to pseudowords with new plausible bigrams. Control analyses showed that the participants were correctly exposed to these pseudowords as the scores in the games were very high. Before, during, and after the exposure phase, we used a wordlikeness task to test the extent to which the participants became familiar with the new regularities.

Overall, the results showed that the participants actually became sensitive to the repetition of the new bigrams within pseudowords, very rapidly. Indeed, before exposure (session 1), the results for the experimental pairs showed that the participants did not prefer the critical item (i.e., the one with a new regularity) over the control item (i.e., a random item with the same orthographic characteristics but without an exposed new bigram). On the contrary, we found an opposite effect: The participants chose the control items more than the critical items, above what chance would have predicted. This unexpected effect can be explained by the way these items were devised in the experimental pairs: The control pseudowords entailed a bigram that was either illegal or of very low frequency in French, whereas the target bigram of the critical items was systematically illegal in the French orthography. Since in session 1 the participants were asked to choose the item the most similar to the words they knew, their implicit knowledge of the French orthography may have led them to choose the control items with very low-frequency bigrams (probably seen previously) over the critical items with illegal bigrams (never seen previously). This initial bias suggests that the reverse effect we found in the next session is underestimated.

From the beginning of the exposure phase, a clear preference for the critical items entailing the new bigrams was found in the wordlikeness task. The fact that the effect was already present after 5 days of exposure showed that sensitivity to orthographic regularities develops extremely rapidly (less than 2 h of exposure), as previously reported (e.g., Cassar & Treiman, 1997; Chetail, 2017; Pacton et al., 2001), and that it can be maintained over time despite the absence of stimulation (effect still present in session 6). Moreover, the fact that the effect became stronger over time suggests that the number of times a new regularity is exposed influences its subsequent processing. This is confirmed by the main effect of bigram frequency: Bigrams repeatedly presented

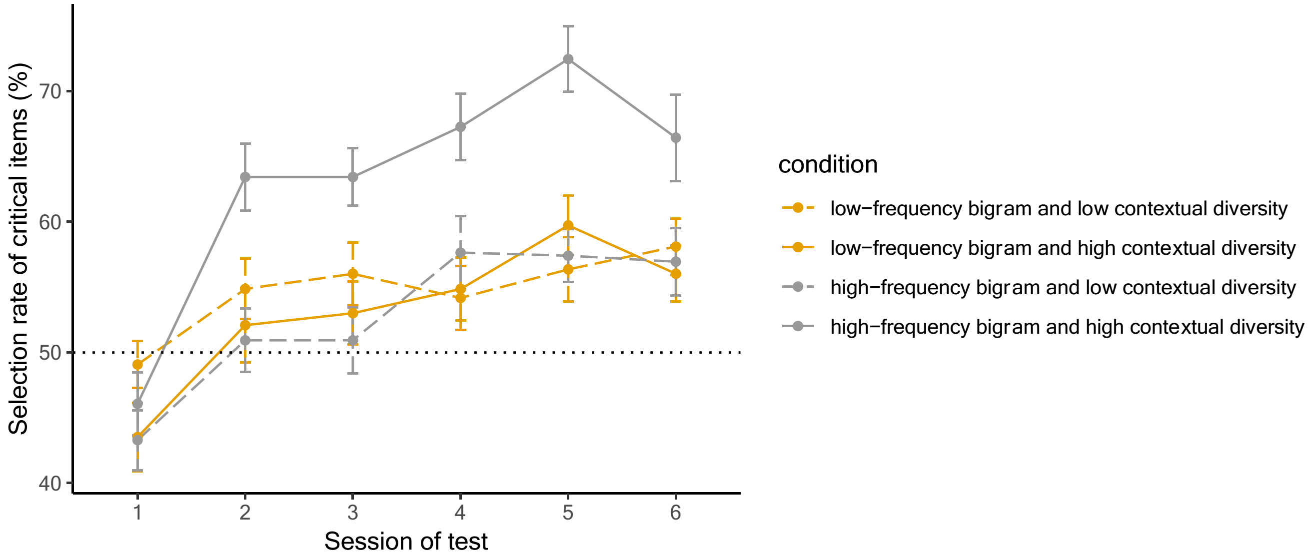

Fig. 5 Selection rate of critical items in experimental pairs in the wordlikeness task (with standard errors) as a function of bigram frequency and contextual diversity. The dotted line represents chance level

in pseudowords were more selected in the wordlikeness task than those that were less repeated.

Importantly, the results also led to a significant effect of contextual diversity. Items with high contextual diversity bigrams were preferred to items with low contextual diversity ones, and this occurred only in high-frequency bigrams. This effect is a direct demonstration that contextual diversity plays a role at a sublexical level. Encoding new regularities in memory is thus not the result of mere repeated exposure. The number of repetitions does play a role but the salience of a new bigram, frequently repeated, is increased if this bigram occurs in different items with various letter contexts, as suggested by the significant interaction between bigram frequency and contextual diversity. This result parallels what was found regarding word frequency: High-frequency words are processed more rapidly if they usually appear in many different texts rather than in the same passages (e.g., Adelman et al., 2006; Perea et al., 2013; Plummer et al., 2014).

To explain these results, we draw on the study by Jones et al. (2012) that examined semantic contextual diversity. Initially, this study addressed the limits of previous operationalization of contextual diversity. As the authors pointed out, contextual diversity is usually calculated as the number of documents in which a given word appears in a corpus of texts (e.g., Adelman et al., 2006). Jones et al. (2012) argued that this type of operational definition leads to an invalid measure as it does not necessarily capture changes in contexts. For example, the word *bank* could appear in distinct documents (thus increasing contextual diversity if operationalized as document count) but the documents could not truly provide distinct contextual uses of the word if these texts are all related to financial topics (see also Hoffman et al., 2013). Hence, Jones et al. (2012) provided a new way of measuring contextual diversity that directly takes into account the overlap of information between the linguistic contexts of a word (the so-called "semantic context" per se). This measure is a weighted sum of the number of documents in which a given word appears. The weighting is based on the distinctiveness of the contexts in the documents: The lower the overlap between words in two documents, the more distinct contexts these two documents provide to a given word, and thus, the higher the measure of semantic contextual diversity.

In a first corpus-based experiment of word identification times in the lexical decision and naming tasks, Jones et al. (2012) showed that this new measure of contextual diversity was a better predictor of latencies than pure measures of lexical frequency or document count of contextual diversity. Words that were experienced in a larger number of semantically distinct texts led to shorter identification times. Interestingly, the authors confirmed this finding with an experimental design very similar to the one we used in the present study. Participants were exposed to an artificial language through written three-word "sentences" composed of one-syllable "words" corresponding to pronounceable nonwords (e.g., *plurt gluds leuts*). Each sentence was accompanied by an unfamiliar image and the task of the participants was to learn this new language. The sentence set was devised according to the orthogonal manipulation of word frequency and contextual diversity. Low-frequency words appeared 45 times during the training session compared to 180 times for high-frequency words. Further, the words in the low-diversity condition always appeared in the same semantic context (i.e., the same sentence with the same image), whereas the words in high-diversity context appeared in eight different contexts. It is worth noting that this manipulation is very similar to what we did, despite the level of analysis ("bigram/word level" for us vs. "word/sentence level" in Jones et al.). Indeed, in our experiment, low-frequency bigrams also appeared four times less than high-frequency bigrams, and our manipulation of contextual diversity was very similar to that of Jones et al.'s, with a ratio of two versus eight different contexts. After the training session, the participants performed a lexical decision task during which the new words were exposed. Jones et al. reported an interaction strikingly similar to the one we found: High-diversity words resulted in better processing compared to low-diversity ones, but only if the words were very frequently exposed. According to the authors, this demonstrates that increasing the number of repetitions of a pseudoword does not facilitate its processing if the contexts in which the item occurs are unchanged. In other words, "processing savings occurred only if the increase of frequency was accompanied by a change in contexts across learning" (p. 121). Our results allow us to extend this conclusion to the sublexical level: The processing of new bigrams is facilitated (e.g., selection of critical items with newly exposed bigrams) only if the increase in frequency is accompanied by a change in contexts during learning. Here, "context" refers to letter strings in which the bigrams occur.

To account for the influence of contextual diversity in the organization of lexical knowledge in memory, Jones et al. (2012) proposed the semantic distinctiveness model (SDM). In this model, the representation of a word in the lexical memory develops gradually, with each new experience. When a word is encountered in context, its representation in memory is compared to the current episodic context. If the current context is highly consistent with the contents stored in memory (low contextual diversity), the context is encoded at a weaker magnitude, decreasing access to the word in future processing. On the contrary, if the information in the current context is new compared to what is stored in memory (high contextual diversity), it is encoded at a much stronger magnitude, facilitating access to the word in future processing. This corresponds to the implementation of a weighted sum of the context.

Although the SDM has been proposed to account for semantic contextual diversity effects at the lexical level, it can be extended to the sublexical level. Indeed, the mechanisms described do not rely on a semantic analysis per se (i.e., analysis of the existing semantic or associative relationships between, e.g., *milk* and *cow*) but on finding patterns of word-by-context co-occurrence (i.e., the overlap between the words forming the context of *milk* in a new document and the content of the current memory representation of *milk*). Because words appear in texts and sentences that are semantically constructed, the SDM ultimately captures the semantic contents of words and contexts, but strictly speaking, the model is based on a purely formal analysis mechanism. This can be observed in the results of Jones et al. (2012). In Experiment 2 (artificial language learning), minimal semantic content was provided to the participants (images of pseudo-objects), but the impact of contextual diversity was quite similar to that obtained with "real" texts (Experiment 1). Further, the SDM was able to generate the same pattern of results while receiving no semantic information (simulation 2).

Thus, the model implements a general mechanism for handling contextual diversity processing that is not limited to words in texts or meaningful sentences. This mechanism, based on the distinctiveness of the contexts in which a given element is repeated, accounts for our results obtained at a sublexical level. The representation of a bigram in memory develops progressively, with each new experience. Overall, the more it is encountered, the higher the access to its representation (bigram frequency effect). Furthermore, when a bigram is newly encountered, its representation in memory is compared to the current episodic context (i.e., letter string in which it is embedded). If the current context is very consistent with the contents stored in memory (low contextual diversity), the context is encoded at a weaker magnitude, but if the information in the current context is novel compared to what is stored in memory (high contextual diversity), it is encoded at a much stronger magnitude. Note that when a bigram is little experienced (low-frequency bigram), the difference in context weighing (between high and low contextual diversity) would not be strong enough to generate a difference in processing efficiency (see the results for low-frequency bigrams here and for low-frequency words in Jones et al., 2012). As Jones et al. (2012) pointed out, the interaction between contextual diversity and frequency is "a natural consequence of a mechanism that encodes words relative to their information overlap with what has already been stored" (p. 122). This point is in line with the idea that contextual diversity is not a by-product of frequency, as already supported by the ERP study by Vergara-Martinez, Comesaña, and Perea (2017) showing different electrophysiological responses for contextual diversity (operationalized in terms of document count) and word frequency. The interpretation of our results in the SDM thus showed that, as with words in sentences, the accumulation of new contexts when encountering a bigram would strengthen the representation of the bigram, making it more accessible in future processing. This conclusion suggests that the effects of semantic diversity previously reported (e.g., Hoffman et al., 2013; Hsiao & Nation, 2018; Johns et al., 2016; Jones et al., 2012) rely more on a formal analysis than on semantics, and reflect a general mechanism of visual feature processing.

The parallel between the effects of letter contextual diversity, as observed in the present study, and the effects of semantic contextual diversity as previously reported, is in line with recent demonstrations that processes at work at the lexical or sublexical level are also present at a multi-word sentence level. For example, as confirmed by the present study, repeated exposure to print leads to the statistical learning of sublexical regularities, which can lead to faster processing of words entailing these regularities (e.g., shortest fixations on words including frequent initial trigram; Lima & Inhoff, 1985). At a multi-word level, Snell and Theeuwes (2020) also reported fewer and shorter fixations on sentence structures that are repeatedly encountered in texts. Another similar parallel can be found with superiority effects. The fact that a letter is better detected in a word than in a pseudoword (i.e., word superiority effect; Reicher, 1969) is typically interpreted as a feedback effect of a higher level of orthographic word-forms on letter detection (e.g., McClelland & Rumelhart, 1981). The fact that a word is better detected in grammatically correct short sentences than in scrambled agrammatical sequences (i.e., syntactic sentence superiority effect) is also interpreted as the result of a feedback effect of higher sentence-level representations on single-word form processing (Snell & Grainger, 2017). Hence, some of the processes assumed to be at work at the level of single word processing would fall under more general visual and linguistic processes also present at the level of sentence processing. At least, the present study demonstrates that this is the case for contextual diversity effects.

In summary, using an incidental exposure paradigm, the present study confirms that readers grasp subtle new orthographic regularities very rapidly. This sensitivity remains after 1 month without exposure. Importantly, contextual diversity is determinant for the development of such sensitivity since frequent bigrams become more salient if they occur in different words. The repetition of bigrams is beneficial to processing only if it is accompanied by a change of contexts, the contexts referring to the letter strings in which bigrams are embedded. The latter result parallels what is observed at a lexical level and can be explained by a model of contextual diversity based on an item-by-context co-occurrence analysis in which distinct contexts are given more weight than similar contexts.

Appendix

Table 3 Stimuli used in the wordlikeness task

Condition (critical items)		Experimental pairs		Control pairs		
	Critical items	Critical bigrams	Control items	Control bigram	First items	Second items
High-frequency and high contextual diversity bigrams	idefvo	fv	idefso	fs	uhatnixu	uhacdixu
	obafvort	fv	obacmort	cm	uhotnan	uhovnan
	olifvede	fv	olinpede	np	kratniit	kravniit
	ogegvuto	gv	ogemhuto	mh	tevlain	tegbain
	rogvuia	gv	rofsuia	fs	unevlinu	unekrinu
	rugvia	gv	rutgia	tg	jeivliki	jeitniki
	lajxet	jx	ladbet	db	amivnafo	amicdafo
	sijxul	jx	siczul	cz	klovnixo	klorxixo
	ziujxeta	jx	ziucfeta	cf	afovnoia	afoznoia
	daikzuku	kz	daidluku	dl	leoxhies	leobcies
	koekziel	kz	koetfiel	tf	kauxhagi	kaubmagi
	okukzi	kz	okuzti	zt	joixharf	joimsarf
	blutkona	tk	blumvona	mv	azuznulo	azubnulo
	ezatki	tk	ezazti	zt	jaizni	jaivni
	meitko	tk	meizto	zt	grezniti	grezsiti
	bievjato	vj	biekgato	kg	jazsur	jacdur
	davjeke	vj	damheke	mh	ioizsuza	ioicnuza
	plevjail	vj	plemdail	md	dozsien	dovlien
High-frequency frequency and low contextual diversity bigrams	akibpida	bp	akizpida	zp	amsezu	acnezu
	isebpudu	bp	iseztudu	zt	ahamsa	ahakma
	kebpant	bp	kexsant	xs	dripfaja	dricdaja
	bodfujo	df	bodbujo	db	deipfuzi	deifguzi
	odfiru	df	ozmiru	zm	pepfum	pekpum
	ugadfe	df	ugacfe	cf	lorjeze	lobceze
	ezujnal	jn	ezutgal	tg	isorje	isorxe
	kajne	jn	kalze	lz	osorjir	ososrir
	zajnemi	jn	zakvemi	kv	duirxohi	duibnohi
	egavhizu	vh	egagzizu	gz	klurxial	klurjial
	eluvhart	vh	elumdart	md	exarxun	exasrun
	oevhelo	vh	oeczelo	cz	kiarxoak	kiavloak
	amuxge	xg	amuxse	xs	nosjei	nogbei
	ekuxgaze	xg	ekukfaze	kf	teosji	teokmi
	enexgi	xg	enekfi	kf	irusjixa	irurjixa
	nuxzour	xz	nuczour	cz	utisroil	utibnoil
	ogixzore	xz	ogizpore	zp	masraka	makpaka
	oxzeba	xz	oxseba	xs	mosriez	moxhiez

Table 3 (continued)

Condition (critical items)		Experimental pairs		Control pairs		
	Critical items	Critical bigrams	Control items	Control bigram	First items	Second items
Low-frequency and high contextual diversity bigrams	afxana	fx	akgana	kg	hafjuin	habcuin
	oifxutz	fx	oizmutz	zm	fofjore	fogbore
	piafxiro	fx	piapkiro	pk	zeigbeti	zeifgeti
	esojhipa	jh	esolzipa	lz	kegbon	kekhon
	ijhufe	jh	iztufe	zt	etagbon	etaxhon
	ohajhu	jh	ohapku	pk	ekhuen	ebnuen
	eramga	mg	eradla	dl	isakhehe	isafjehe
	glomgiez	mg	glomdiez	md	kuakhuta	kuaxhuta
	spomgabo	mg	spopkabo	pk	djikhu	djizsu
	apbuez	pb	agduez	gd	rakmong	rafgong
	maopbie	pb	maogzie	gz	ohikmadi	ohisradi
	xiipbor	pb	xiitfor	tf	siukmita	siuznita
	ixopviru	pv	ixokfiru	kf	hokpu	hofju
	ozepvoro	pv	ozemhoro	mh	epekpe	epepfe
	upvofe	pv	ulzofe	lz	ahokro	ahogbo
	guezxame	zx	guetvame	tv	urokramu	uropfamu
	puazxitx	zx	puaxsitx	xs	eikrone	eivlone
	zezxoi	zx	zetgoi	tg	olomsain	olobmain
Low-frequency and low contextual diversity bigrams	elibkual	bk	elikvual	kv	zubca	zufja
	ogabkus	bk	ogafsus	fs	agabcipo	agatnipo
	ubkioz	bk	uzmioz	zm	inabca	inaxha
	efpupi	fp	etvupi	tv	ubmios	umsios
	ohefpohi	fp	ohepkohi	pk	viebma	viemsa
	usifpo	fp	usigzo	gz	ebmide	eznide
	epokjune	kj	epotfune	tf	bnuku	bmuku
	izakju	kj	izamhu	mh	hebnazu	hekhazu
	uzikjuja	kj	uzicfuja	cf	zuibnasi	zuikhasi
	ekalxuki	lx	ekalzuki	lz	omicdaki	omikmaki
	failxido	lx	faikvido	kv	fecdoxi	fekroxi
	iralxigu	lx	iragdigu	gd	ilecdizo	ilemsizo
	huevfia	vf	hueczia	cz	acnogo	abmogo
	psivfolo	vf	psicmolo	cm	ubecni	ubekri
	zivfoi	vf	zidboi	db	nocnahi	novlahi
	axixkia	xk	axikgia	kg	loafger	loakher
	izoxkere	xk	izozpere	zp	ozifgexa	ozikpexa
	pixkadu	xk	pigzadu	gz	ihufgarg	ihuznarg

Authors' notes The work reported here has been supported by a research grant from the Université Libre de Bruxelles. Karinne Sauval had a post-doctoral position at the LCLD at the time of the study. All the raw data and scripts for analyses are available on the Open Science Framework: https://osf.io/8uvhs/.

References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual Diversity, Not Word Frequency, Determines Word-Naming and Lexical Decision Times. *Psychological Science*, *17*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Baayen, H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*(3), 436–461. https://doi.org/10.1075/ml.5.3.10baa

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, *29*(4), 639–647. https://doi.org/10.3758/BF03200465

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Conway, C. M., & Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*, *89*(4), 631–644. https://doi.org/10.1037/0022-0663.89.4.631

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, *163*, 103–120. https://doi.org/10.1016/j.cognition.2017.02.015

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1084–1096. https://doi.org/10.1037/0278-7393.16.6.1084

Gingras, M., & Sénéchal, M. (2019). Evidence of Statistical Learning of Orthographic Representations in Grades 1–5: The Case of Silent Letters and Double Consonants in French. *Scientific Studies of Reading*, *23*(1), 37–48. https://doi.org/10.1080/10888438.2018.1482303

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, *103*, 114–126. https://doi.org/10.1016/j.jml.2018.08.005

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, *23*(4), 1214–1220. https://doi.org/10.3758/s13423-015-0980-7

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(2), 115–124. https://doi.org/10.1037/a0026727

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. https://doi.org/10.3758/BRM.42.3.643

Lima, S. D., & Inhoff, A. W. (1985). Lexical access during eye fixations in reading: Effects of word-initial letter sequence. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(3), 272–285. https://doi.org/10.1037/0096-1523.11.3.272

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516–524. https://doi.org/10.3758/BF03195598

O'Brien, B. A. (2014). The Development of Sensitivity to Sublexical Orthographic Constraints: An Investigation of Positional Frequency and Consistency Using a Wordlikeness Choice Task. *Reading Psychology*, *35*(4), 285–311. https://doi.org/10.1080/02702711.2012.724042

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General*, *130*(3), 401–426. https://doi.org/10.1037/0096-3445.130.3.401

Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology*, *116*(1), 37–44. https://doi.org/10.1016/j.jecp.2012.10.014

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Pitchford, N. J., Ledgeway, T., & Masterson, J. (2008). Effect of orthographic processes on letter position encoding. *Journal of Research in Reading*, *31*(1), 97–116. https://doi.org/10.1111/j.1467-9817.2007.00363.x

Plummer, P., Perea, M., & Rayner, K. (2014). The influence of contextual diversity on eye movements in reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(1), 275–283. https://doi.org/10.1037/a0034058

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, . URL https://www.R-project.org/

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*(2), 275–280. https://doi.org/10.1037/h0027768

RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., URL http://www.rstudio.com/.

Samara, A., & Caravolas, M. (2014). Statistical learning of novel graphotactic constraints in children and adults. *Journal of Experimental Child Psychology*, *121*, 137–155. https://doi.org/10.1016/j.jecp.2013.11.009

Snell, J., & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition*, *168*, 217–221. https://doi.org/10.1016/j.cognition.2017.07.003

Snell, J., & Theeuwes, J. (2020). A story about statistical learning in a story: Regularities impact eye movements during book reading. *Journal of Memory and Language*, *113*, 104-127. https://doi.org/10.1016/j.jml.2020.104127

Stanners, R. F., Jastrzembski, J. E., & Westbrook, A. (1975). Frequency and visual quality in a word-nonword classification task. *Journal of Verbal Learning and Verbal Behavior*, *14*(3), 259–264. https://doi.org/10.1016/S0022-5371(75)80069-7

Vergara-Martínez, M., Comesaña, M., & Perea, M. (2017). The ERP signature of the contextual diversity effect in visual word recognition. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(3), 461–474. https://doi.org/10.3758/s13415-016-0491-7

Whaley, C. P. (1978). Word—Nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 143–154. https://doi.org/10.1016/S0022-5371(78)90110-X