



HAL
open science

Reading with maculopathy: the inhibitory effect of word neighborhood size is modulated by word predictability and reading proficiency

Lauren Sauvan, Natacha Stolowy, Carlos Aguilar, Thomas François, Nuria Gala, Frederic Matonti, Eric Castet, Aurelie Calabrese

► To cite this version:

Lauren Sauvan, Natacha Stolowy, Carlos Aguilar, Thomas François, Nuria Gala, et al.. Reading with maculopathy: the inhibitory effect of word neighborhood size is modulated by word predictability and reading proficiency. 2020. hal-02920126v1

HAL Id: hal-02920126

<https://inria.hal.science/hal-02920126v1>

Preprint submitted on 24 Aug 2020 (v1), last revised 16 Nov 2020 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Title: Reading with maculopathy: the inhibitory effect of word neighborhood**
2 **size is modulated by word predictability and reading proficiency**

3

4 **Authors:** Lauren Sauvan¹, Natacha Stolowy¹, Carlos Aguilar², Thomas François³,
5 Núria Gala⁴, Frédéric Matonti⁵, Eric Castet⁶, Aurélie Calabrèse^{7*}

6

7 **Affiliations:**

8 ¹North Hospital, Marseille France

9 ²Mantu Lab, Amaris Research Unit, Sophia Antipolis, France

10 ³UCLouvain, IL&C, CENTAL, Belgique

11 ⁴Aix-Marseille Univ., CNRS UMR 7309, Aix-en-Provence, France

12 ⁵Centre Paradis Monticelli, Marseille, France

13 ⁶Aix-Marseille Univ., CNRS UMR 7290, Marseille, France

14 ⁷Université Côte d'Azur, Inria, Sophia Antipolis, France

15

16 **Corresponding author:**

17 Aurélie Calabrèse (AC)

18 aurelie.calabrese@inria.fr

19 2004 Route des Lucioles, 06902 Valbonne, France

20

21

22 **Word count:**

23 Abstract = 324 words

24 Main text (not including Methods) = 3451 words

25 *Intro = 1080; Methods = 1324; Results = 936; Discussion = 1435*

26 References = 55; Figures = 6; Tables = 5

27

28 **Funding:**

29 This work was supported by the Fondation de France and by the Belgian FNRS.

30

31 **Declarations of interest:**

32 None

33

34

35 ABSTRACT

36 **Background:** For normally sighted readers, word neighborhood size (i.e., the total
37 number of words that can be formed from a single word by changing only one letter)
38 has a facilitator effect on word recognition. When reading with central field loss
39 (CFL), however, individual letters may not be correctly identified, leading to possible
40 misidentifications and a reverse neighborhood size effect. Here we investigate this
41 inhibitory effect of word neighborhood size on reading performance and whether it is
42 modulated by word predictability and reading proficiency.

43 **Methods:** Nineteen patients with binocular CFL from 32 to 89 years old (mean \pm SD
44 = 75 ± 15) read short sentences presented with the self-paced reading paradigm.
45 Accuracy and reading time were measured for each target word read, along with its
46 predictability, i.e., its probability of occurrence following the two preceding words in
47 the sentence using a trigram analysis. Linear mixed effects models were then fit to
48 estimate the individual contributions of word neighborhood size, predictability,
49 frequency and length on accuracy and reading time, while taking into account
50 patients' reading proficiency.

51 **Results:** For the less proficient readers who quit reading on a daily basis, we found
52 that the effect of neighborhood size was reversed compared to normal readers and of
53 higher amplitude than the effect of frequency. Furthermore, this inhibitory effect is of
54 greater amplitude (up to 50% decrease in reading speed) when a word is not easily
55 predictable because its chances to occur after the two preceding words in a specific
56 sentence are rather low.

57 **Conclusion:** Severely impaired patients with CFL often quit reading on a daily basis
58 because this task becomes simply too exhausting. Based on our results, we envision

59 lexical text simplification as a new alternative to promote decisive rehabilitation in
60 these patients. By increasing reading accessibility for those who struggle the most,
61 text simplification might be used as an efficient readaptive tool and daily reading
62 assistive technology, fostering overall reading ability and fluency through increased
63 practice.

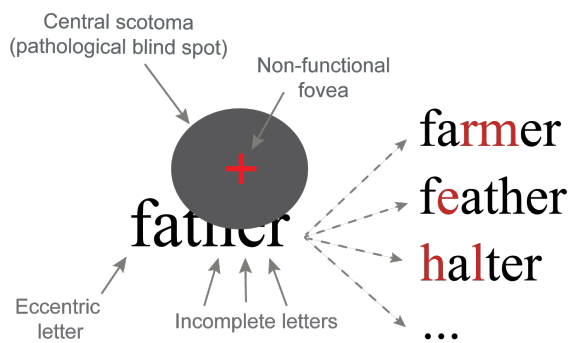
64 INTRODUCTION

65

66 Individuals with Central Field Loss (CFL) induced by maculopathy experience
67 severely impaired functional vision, leading to major incapacitating reading deficits
68 [1,2,3]. Reading speed being a strong determinant of overall quality of life in low-
69 vision patients of all ages, there is a real societal need to help promote their reading
70 performance through readaptive interventions and assistive reading technology [4,5].
71 Such initiative requires a better characterization of the underlying factors involved in
72 their overall reading deficit, which remain poorly understood. The influence of
73 psycholinguistic factors, for instance, have been generally overlooked in the context
74 of low vision. In 2016, Calabrèse et al. investigated eye movement characteristics
75 when low-vision patients with CFL read short sentences and showed that specific
76 disruptions in fixation patterns occurred in the presence of low-frequency words [6].
77 This result alone suggests that psycholinguistic factors may be an important
78 determinant of reading performance for low vision, opening the way for further
79 investigations and potential use of text simplification for the visually impaired.

80 Word frequency and word neighborhood size are some of the most important lexical
81 factors known to affect reading in normal vision [7,8]. However, whether findings
82 from normal readers can be applied to low-vision individuals is not obvious. In the
83 case of CFL, visual input is deteriorated because of blur, distortion and/or missing
84 parts [9] and access to text is only partial (Fig. 1). For instance, when reading the
85 word “father” with a central blind spot, some letters are totally or partially occluded
86 by the scotoma. In addition, complete letters that fall out of the scotoma must be
87 identified with eccentric portions of the retina, yielding degraded letter identification
88 due to low acuity and strong crowding [10]. Therefore, incomplete and/or eccentric

89 letters may not be correctly identified, leading to many possible misidentifications
90 (“farmer”, “feather”, “halter”, etc.). Since bottom-up visual input is not reliable,
91 patients must rely much more on top-down linguistic inference than normally sighted
92 readers [11,12,13]. For this reason, the effect of lexical factors on reading
93 performance should be rather different in visually impaired readers than reported
94 before with normally sighted individuals.



95
96 **Figure 1: The presence of a central scotoma occludes part of the word “father”, leading to**
97 **possible misidentifications (e.g., farmer, feather, halter) and a greater need for linguistic**
98 **inference**

99 A first attempt to characterize this effect has been made recently by Stolowy et al.
100 who inspected the effect of word frequency on the reading performance of CFL
101 readers [14]. The clear-cut frequency effects they reported on word reading time
102 validated the hypothesis that, as for normal readers, low-frequency words tend to
103 decrease reading speed with CFL patients. However, the amplitude of this effect was
104 much larger (differences in the range of seconds) than it had been reported before
105 with normal vision (range of milliseconds). More interestingly, the same effect of
106 frequency was not observed for all pairs of synonyms. For instance, the high-
107 frequency word “*utiliser*” (*use*) was read on average more slowly than its lower
108 frequency synonym “*employer*” (*employ*). Such observation suggests that frequency
109 cannot be the only predictive linguistic factor of reading speed with CFL. This

110 reasoning was a cornerstone for the present work that aims at investigating two other
111 psycholinguistic factors extensively studied in normal reading: orthographic similarity
112 and word predictability.

113 The most common measure of orthographic similarity in the psychological literature
114 is Coltheart's orthographic neighborhood size metric 'N' [15]. N measures the
115 number of close orthographical neighbors of a stimulus word and is defined as the
116 total number of words that can be formed by changing one letter, while preserving
117 letter positions. For example, "shore" has many neighbors, including chore, score and
118 share, while "neighbor" has 0 neighbor. For skilled readers with normal vision, N
119 (i.e., word neighborhood size) seems to have a facilitator effect on word recognition:
120 the more neighbors, the faster a word is identified - although this effect is now often
121 assumed to be task dependent and language dependent (see [16,17] for reviews). It is
122 possible that the visual constraint imposed by the presence of a central scotoma,
123 hiding portions of the text (*i.e.*, letters) and forcing to use eccentric vision, may lead
124 to a reverse effect of neighborhood size. The lack of high resolution coupled with
125 missing visual information, would lead CFL readers to confuse one word with its
126 orthographic neighbors, creating even more uncertainty for large word neighborhood
127 size. Therefore, we hypothesize that, unlike normal vision, word neighborhood size
128 has a negative effect on reading speed with CFL (Hypothesis 1). This first hypothesis
129 has recently received support through some preliminary work of ours [18].

130 In addition to frequency and orthographic similarity, word predictability counts as one
131 of the most influential processing factors during word recognition with normal vision
132 [19]. The predictability of a target word in sentence (based on the context induced by
133 the preceding words) has been shown to influence processing speed by enabling

134 readers to make forward inferences (i.e. predictions about the probable upcoming
135 word). Thus, as shown by eye movement studies, highly predictable words: (1) are
136 skipped more often, (2) are more likely to be identified within a single fixation and (3)
137 yield overall shorter fixations during sentence reading [20,21]. In the context of CFL,
138 word predictability should also play an additional role by reducing uncertainty when
139 identifying a word with many confusable neighbors. In the sentence “You should go
140 for a walk along the shore” for instance, even if the word “shore” has many
141 potentially confusable neighbors, most of them (such as chore, score and share) may
142 be easily ruled out based on semantic context and forward inferences. Therefore, we
143 hypothesize that the amplitude of the neighborhood size effect is influenced by word
144 predictability: the more predictable a word is (thanks to sentence context), the smaller
145 the effect of neighborhood size should be (Hypothesis 2).

146 Finally, for normally sighted adult readers, word predictability does influence reading
147 processing differently depending on individual reading skills [21,22]. Simply put, less
148 skilled readers rely more on context for word identification than highly skilled readers
149 do. Based on this result, we assume that results of CFL individuals should also be
150 influenced by their reading proficiency. Therefore, we hypothesize that the interaction
151 between neighborhood size and word predictability (see Hypothesis 2) depends on
152 patients’ reading proficiency. Our prediction is that this interaction will be more
153 pronounced for less proficient readers than for proficient ones (Hypothesis 3).

154 In short, the purpose of this work is two-fold: (1) to test the hypothesis that word
155 neighborhood size exerts an inhibitory effect on reading performance with CFL
156 (Hypothesis 1; Analyses 1 & 2) and (2) to test whether this effect is modulated by
157 word predictability and reading proficiency (Hypothesis 2 & 3; Analysis 3).

158 **METHODS**

159

160 **Participants**

161 19 patients (13 females) were recruited from the Low-Vision Clinic of the La Timone
162 Hospital (Marseille, France) between March and June 2017. All presented a bilateral
163 central scotoma with a monocular acuity of 0.4 logMAR or worse in their better eye.
164 Patients with ophthalmologic disorders other than maculopathy (e.g. glaucoma) or
165 cognitive disorders were not included. The following information was collected for
166 each participant: age, etiology, lens status, duration of maculopathy, field loss (central
167 only vs. central + peripheral), low-vision rehabilitation history and daily reading time.
168 Recruited participants ranged in age from 32 to 89 years (mean \pm SD = 75 \pm 15).
169 Mean best-corrected visual acuity was 0.81 \pm 0.28 logMAR. Details of the
170 participants' demographic, visual and reading characteristics are given in Table 1. The
171 research was carried out in accordance with the Code of Ethics of the World Medical
172 Association (Declaration of Helsinki). Informed consent was obtained from all
173 participants after complete explanation of the nature and possible consequences of the
174 study.

Partici pants	Gen der	Age (years)	Etiology	Disease onset (years)	Low- vision rehabili tation	Visual acuity (logMAR)	Lens status	Daily reading time (min/day)
1	M	76	Atrophic AMD	2	Yes	1	Phakic	0
2	F	77	Atrophic AMD	7	No	0.4	Pseudophakic	45
3	F	80	Atrophic AMD	7	Yes	1	Pseudophakic	0
4	F	80	Atrophic AMD	5	Yes	0.8	Phakic	0
5	F	83	Atrophic AMD	12	Yes	0.5	Phakic	60
6	F	83	Atrophic AMD	2	Yes	0.5	Pseudophakic	0
7	F	85	Atrophic AMD	15	No	0.7	Pseudophakic	60
8	M	86	Atrophic AMD	8	Yes	0.7	Pseudophakic	60
9	M	88	Atrophic AMD	20	Yes	0.7	Phakic	60
10	F	78	Exudative AMD	3	No	1.1	Pseudophakic	0
11	F	87	Exudative AMD	2	Yes	0.5	Pseudophakic	10

12	M	89	Exudative AMD	3	Yes	0.4	Pseudophakic	30
13	F	48	Stargardt's disease	35	Yes	1.3	Phakic	25
14	F	72	Stargardt's disease	16	No	1	Phakic	120
15	F	32	Cone dystrophy	30	No	1	Phakic	120
16	M	73	Diabetic retinopathy	10	No	0.7	Pseudophakic	0
17	M	59	Myopic retinopathy	15	Yes	1	Phakic	0
18	F	70	Myopic retinopathy	20	Yes	1.3	Pseudophakic	120
19	F	70	Myopic retinopathy	70	Yes	0.8	Pseudophakic	30

175

176 **Table 1: Participants' demographic, visual and reading characteristics.**

177 **Visual acuity and lens status are given for the tested eye only. Participants with central scotoma**
178 **only are highlighted in red. Participants with central scotoma and peripheral lesions are**
179 **highlighted in blue.**

180

181 **Apparatus and stimuli**

182 Experiments were run at 60 Hz with an LCD HP LE2201W monitor (full display area:
183 1680 x 1050 pixels; 47.4 x 29.6 cm). Stimuli (i.e. sentences) were generated with the
184 PsychoPy library [23,24] and presented on a 1400 x 1050 pixel window that
185 subtended 56° x 42° at 40 cm. Sentences were aligned to the left and displayed in
186 Courier (non-proportional font) in black on a white background. Print size was chosen
187 optimally for each participant as the value of his/her critical print size, measured
188 before testing with a French computerized version of MNREAD [25,26]. Reading was
189 monocular (eye with better visual acuity) with an appropriate correction for near
190 vision (wide-field Metrovision lenses).

191

192 **Reading material**

193 Reading material was created in French using ReSyf, a French lexicon with
194 disambiguated and graded synonyms [27] and Lexique3, a lexical database with word
195 frequencies of standard written and oral French [28]. First, we created a pool of target
196 words (Figure 2 – Step 1), by selecting 32 pairs of synonyms matching the following

197 criteria: (1) equal length within a pair (from 3 to 8 characters); (2) difference in
198 number of orthographic neighbors (N) comprised between 5 and 10 within a pair; (3)
199 frequency ratio between a large neighborhood word and a its small neighborhood
200 synonym could be either <1 or >1 . Second, 32 pairs of matching sentences were
201 created so that each word from a pair could fit within either sentence of the
202 corresponding sentence pair (Figure 2 – Step 2). Three criteria were used: (1) all
203 sentences had similar length (42 to 65 characters; mean \pm SD = 54 ± 6), with a
204 maximum difference of 5 characters within a pair; (2) pairs of sentences were
205 specifically designed to fit the single and most frequent common sense of a synonym
206 pair; (3) within each sentence, comprised of ‘n’ words, the target word could be
207 located in any of these three locations: ‘n-1’, ‘n-2’, or ‘n-3’. Target words were never
208 presented in position ‘n’ to avoid any interference from a possible wrap-up effect
209 [14,29]. At last, we created our final reading material by combining sentence pairs
210 with their matching synonym pairs (Figure 2 - Step 3). In Condition 1, the first word
211 of a pair was assigned to the first sentence of the corresponding pair, while the second
212 word was assigned to the second sentence, thus creating 64 full sentences. In
213 Condition 2, the “sentence – word” pairing was reversed to create a different set of 64
214 full sentences. These two experimental conditions allowed us to counterbalance any
215 potential effect of the sentence structure and complexity (by randomly assigning
216 participants to Condition 1 or 2), while providing two different measures of
217 predictability for a single target word.

218

Step 1: Selection of 32 pairs of synonyms

Synonym 1

shore - 13 neighbors (share, score, short, ...)

Synonym 2

coast - 3 neighbors (roast, toast, boast)

Length = 5 characters
Diff. = 10 neighbors

Selection criteria within a pair

- 1- Equal length - 3 to 8 characters
- 2- Difference in neighbors number - 5 to 10
- 3- No restriction on the frequency ratio

Step 2: Creation of 32 matching pairs of sentences

Sentence A

You should go for a walk along the [...] to relax
(44 characters / synonym position = n-2)

Sentence B

My parents have worked by the [...] for many years
(45 characters / synonym position = n-3)

Creation criteria within a pair

- 1- Equivalent length - 5 characters difference max.
- 2- Must fit the **same sense** with both synonyms
- 3- Must include synonym in position n-1, n-2 or n-3

Length diff. = 1 character
Unique sense = 'seaboard'

Step 3: Creation of the final corpus of sentences

Condition 1 - 64 sentences

Sentence A + **Synonym 1**
*You should go for a walk along the **shore** to relax.*

Sentence B + **Synonym 2**
*My parents have worked by the **coast** for many years*

Condition 2 - 64 sentences

Sentence A + **Synonym 2**
*You should go for a walk along the **coast** to relax*

Sentence B + **Synonym 1**
*My parents have worked by the **shore** for many years*

219

220 Figure 2: Creation of the reading material.

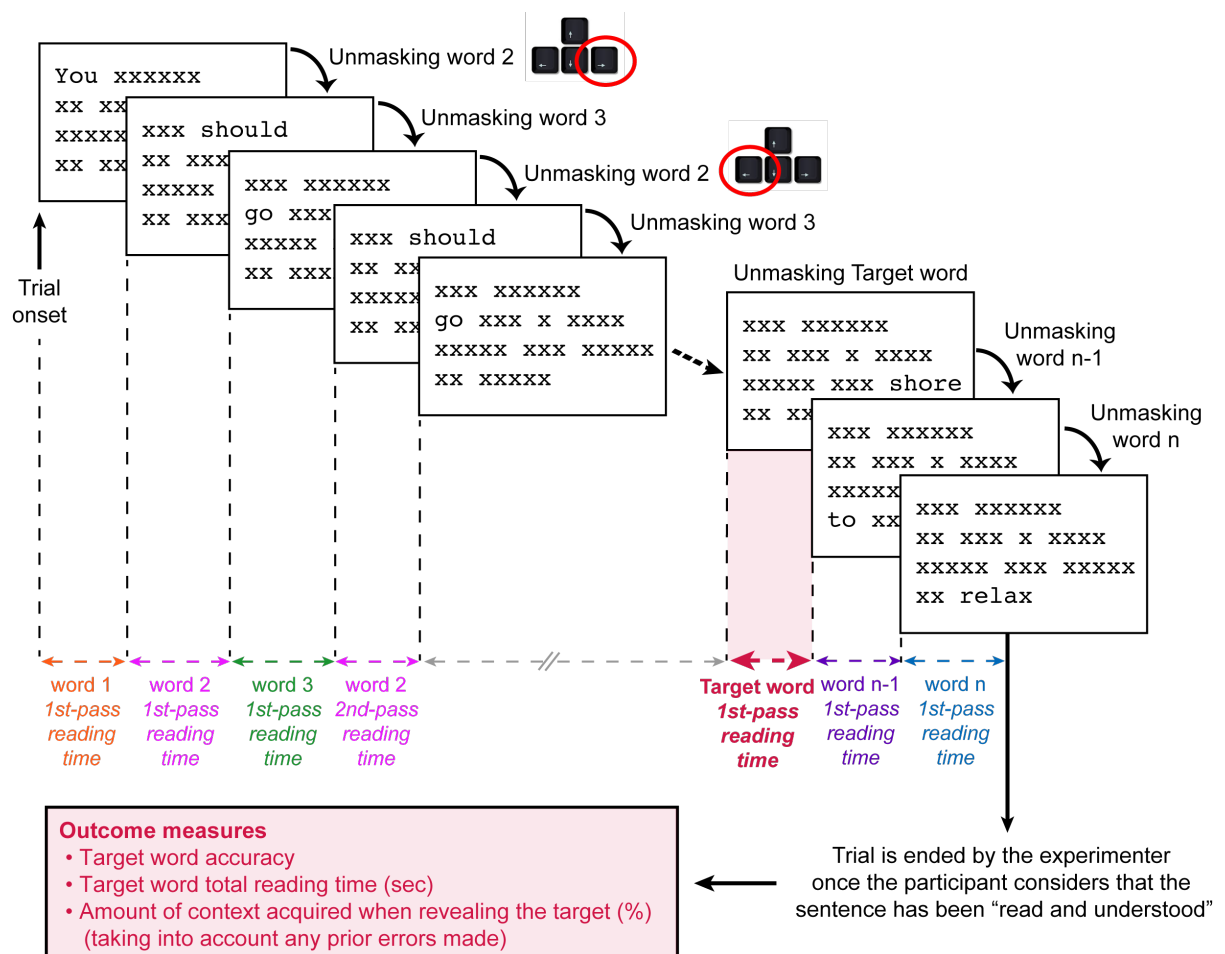
221 **Illustration of the reading material creation, using the synonym pair “shore / coast”. Both words**
222 **in this example contain 5 characters. Their difference in number of neighbors is 10. Sentences A**
223 **and B were created so either word of this synonym pair can fit in both sentences. In Condition 1,**
224 **sentence A contains the target word 1 (‘shore’) and sentence B contains the target word 2**
225 **(‘coast’). In Condition 2, the pairing is reversed. Note that this example is given in English to ease**
226 **understanding. See Supplementary Material for the complete set of French sentences created.**

227

228 **Reading procedure and experimental design**

229 Sentences were presented within 4 blocks of 16 trials (8 pairs of sentences) each.
230 Participants were randomly assigned to Condition 1 or 2 and read between two to four
231 blocks, depending on their reading speed and level of fatigue. Sentences were
232 displayed randomly within each block with non-cumulative self-paced reading, where
233 sentences appear as a whole but with all words masked by strings of “x” [30,31,32]
234 (Figure 3). As opposed to other reading paradigms specifically designed for low
235 vision, such as “word mode” [33], self-paced reading allows to present words

236 individually while still maintaining a whole sentence view and therefore to remain
 237 closer to the visual constraints of natural reading (e.g., crowding). Participants were
 238 instructed to read each sentence aloud as quickly and accurately as possible while
 239 revealing each word one at a time using keyboard presses, with the possibility to
 240 unmask backward as many times as they wanted. When participants considered they
 241 had finished reading the sentence accurately (no matter which word was unmasked at
 242 that moment), they said the word “stop” and the experimenter stopped the trial. Prior
 243 testing, a training phase with short French proverbs was performed to familiarize the
 244 participant with the task and protocol. Reading accuracy (correct vs. incorrect) and
 245 total reading time (in seconds) were recorded for each target word. For words
 246 unmasked several times, reading time was obtained by summing the duration of all
 247 the unmasked periods.



248

249 **Figure 3: Self-paced Reading paradigm.**

250 Words are revealed one at a time by the participant, using keyboard presses. A press on the right
251 arrow key will unmask the upcoming word (while masking the currently visible word) and a
252 press on the left arrow key will unmask the previous word (while masking the currently visible
253 word). For words unmasked only once (e.g. word 1), the total reading time equals the 1st-pass
254 reading time. For words unmasked more than once (e.g. word 2), the total reading time is the
255 sum of each individual reading time pass.

256

257 **Measure of word predictability**

258 For each target word, we estimated its percentage of occurrence following the two
259 preceding words in the sentence based on a large corpus of French texts. To do so, we
260 selected in our reading material all series of three consecutive words (i.e., 3-grams or
261 trigrams) ending with a target word. Using the Google Books Ngram Viewer resource
262 [34], we extracted for each of these 128 trigrams (64 sentences x 2 conditions) its
263 percentage of occurrence in the ‘French 2012’ corpus, a corpus of 792 118 digitized
264 books published in French between 1800 and 2008. This measure will be referred to
265 as the ‘trigram occurrence’.

266

267 **Measure of reading proficiency**

268 The amount of daily reading (reported subjectively by each patient in minutes per
269 day; cf. Table 1) was used to create the binary variable “Daily reading” (*yes or no*),
270 later used as a “proxy” for reading proficiency. Patients who reported to read at least
271 10 minutes each day were categorized as Daily reading – *yes*; The other patients who
272 reported to read between 0 and 10 minutes daily were categorized as Daily reading –
273 *no*.

274

275 **Statistical analysis**

276 Statistical analyses were carried out using R [35]. Reading accuracy (i.e. binary
277 variable) was analyzed by fitting a generalized linear mixed-effects model (GLME;
278 Analysis 1), while reading time (i.e. continuous variable) was analyzed with linear
279 mixed-effects models (LME; Analyses 2 & 3), both allowing the modeling of
280 experimental designs with unbalanced repeated measures [36,37]. Models were
281 constructed with either target word accuracy or target word reading time as the
282 dependent variable. Several factors of interest (Table 2) were included as independent
283 variables to inspect their effect on accuracy and reading time, as well as their potential
284 interaction with each other. The random structure of all models included a random
285 intercept for participants, assuming a different “baseline” performance level for each
286 individual, as well as random intercept for each target word. Before analysis,
287 variables of interest were inspected and transformed as follow to satisfy the
288 assumptions of parametric statistical tests [38,39]: reading time and word frequency
289 were transformed in natural logarithm (ln) units, trigram occurrence was transformed
290 with the ordered quantile normalization and context amount was square-root
291 transformed. Word frequency and word length were centered around their mean.
292 Optimal model structures were assessed using the Akaike Information Criterion (AIC)
293 and likelihood-ratio tests [40]. Significance of the fixed effects was estimated using z-
294 values for the GLME model and t-values for the LME models. Z- and t- values larger
295 than 2 were considered significant, corresponding to a 5% significance level in a two-
296 tailed test [41,42]. In the Results section fixed-effects estimates are reported along
297 with their z- and t-values and 95% confidence intervals [43].

298

Variable name	Unit / Transformation	Centered	Description
Target word characteristics			
Neighborhood size (N)	--	--	Number of neighbors
Frequency (ln)	Natural log	Yes	Occurrences/million
Length	--	Yes	Number of characters
Sentence characteristics			
Trigram occurrence	Ordered quantile normalization	--	Percentage of occurrence for each trigram ending with a target word
Participant's individual characteristics			
Age	Years	--	--
Visual Acuity	LogMAR	--	--
Type of field loss	<ul style="list-style-type: none"> • Central only • Central + peripheral 	--	--
Still reading daily	<ul style="list-style-type: none"> • Yes • No 	--	--

299

300 **Table 2: Description of the different factors included as independent variables in the linear**
301 **mixed-effects models.**

302

303 **RESULTS**

304

305 **Analysis 1 – Effect of neighborhood size on reading accuracy**

306 On average, target words were read accurately 94% of the time, with individual
307 variations ranging from 62 to 100% depending on patients. When all implemented in
308 a GLME model, N, word frequency, word length and word predictability (expressed
309 as trigram occurrence) showed no significant effect on accuracy (Table 3). As
310 estimated by the model, percentage of accuracy for patients who continue reading on
311 a daily basis was 99.1% ($\exp(4.716) / (1 + \exp(4.716)) * 100$) ($z = 7.694$; $p < 0.001$;
312 $95\%CI = [3.65; 6.17]$). For patients who quit daily reading activities, percentage of
313 accuracy was 97.3% ($\exp(4.716-1.126) / (1 + \exp(4.716-1.126)) * 100$). This 1.8%
314 difference barely reached significance ($z = -2.064$; $p = 0.039$; $95\%CI = [-2.31; -$

315 0.03]). Figure 4 shows the null effect of N on accuracy, for these two groups of
 316 participants.

	Estimate	Standard Error	z-value	p-value	95% Confidence Interval
Intercept	4.716	0.613	7.694	<0.001	[3.65; 6.17]
Word neighborhood size (N)	-0.058	0.054	-1.069	0.285	[-0.17; 0.05]
Word frequency (ln)	0.147	0.147	1.003	0.316	[-0.15; 0.46]
Word length	-0.028	0.213	-0.13	0.897	[-0.47; 0.41]
Trigram occurrence	0.361	0.292	1.236	0.216	[-0.21; 0.98]
Daily reading No	-1.126	0.545	-2.064	0.039	[-2.31; -0.03]

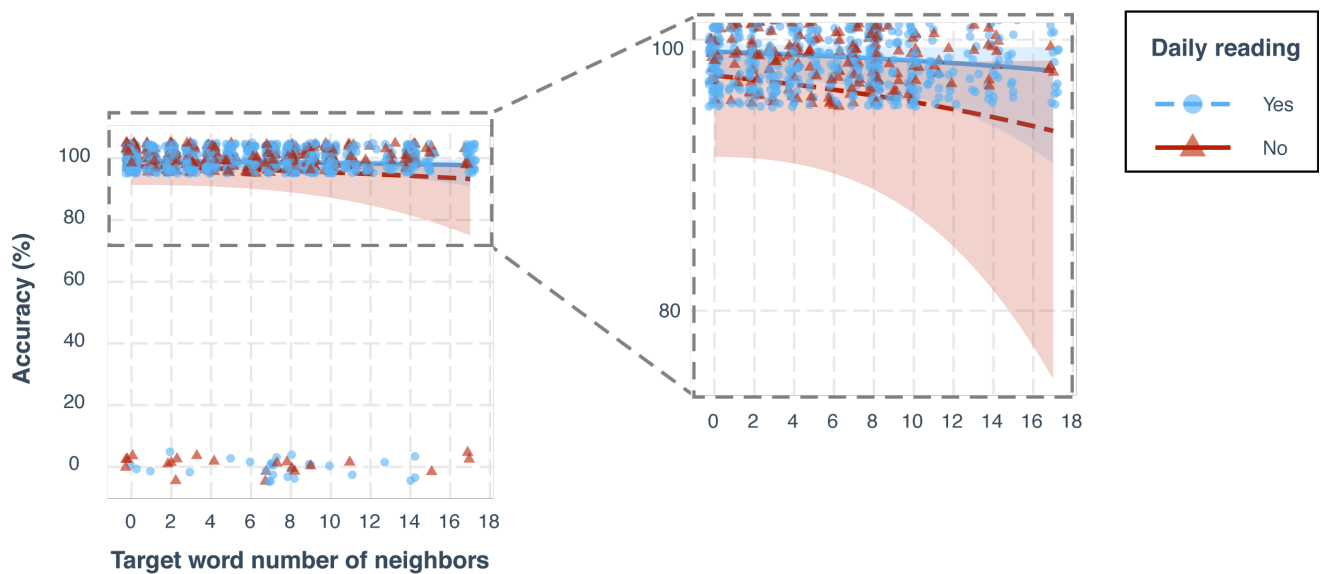
317

318 **Table 3: Fixed-effects estimates from the GLME model (Analysis 1).**

319 **The dependent variable is ‘Target word reading accuracy’. The intercept estimate is the**
 320 **predicted value of the dependent variable when all independent variables are at 0 (continuous**
 321 **variables) or at their reference level (categorical variable). Reference level for the binary factor**
 322 **”Daily reading” (representing reading proficiency) is ‘yes’. Factors showing a significant effect**
 323 **on reading accuracy are in bold font and highlighted in grey.**

324

325



326

327

328 **Figure 4: Effect of the neighborhood size (N) on word reading accuracy.**

329 Each data point represents a target word. Raw data points and fitted curves estimated by the
 330 GLME model are grouped according to reading proficiency: blue for participants reading daily,
 331 red for participants who quit daily reading activities. The right side panel shows a zoomed view
 332 of the left plot between 70 and 100% accuracy.

333

334 **Analysis 2 – Effect of neighborhood size on reading time**

335 In this first LME model (Table 4), the respective effects of neighborhood size, word
 336 predictability and reading proficiency are estimated individually, without any
 337 interaction term, to test our Hypothesis 1. According to this simple model, words with
 338 zero neighbors were read on average in 2.3 seconds ($\exp(0.841)$). Increasing the
 339 number of neighbors by 1 did increase reading time significantly, but moderately, by
 340 a factor of 1.01 ($\exp(0.013)$; $t = 2.507$, $p = 0.013$, $95\%CI = [0.003; 0.023]$; Figure 5).
 341 In other words, increasing neighborhood size from 0 to 6 (the mean value in our pool
 342 of target words), increases reading time by a factor of 1.08 ($\exp(0.013)^6$), i.e. an 8%
 343 increase. Similarly, increasing neighborhood size from 0 to 10 (where most of our
 344 values lie), increases reading time by a factor of 1.14 ($\exp(0.013)^{10}$), i.e. a 14%

345 increase. Word predictability (expressed as trigram occurrence) also showed a
 346 significant effect on reading time ($t = -4.129$, $p < 0.001$, $[-0.17; -0.06]$). Age, acuity
 347 and type of field loss (which were dropped from the final model) showed no
 348 significant effect on reading time.

349

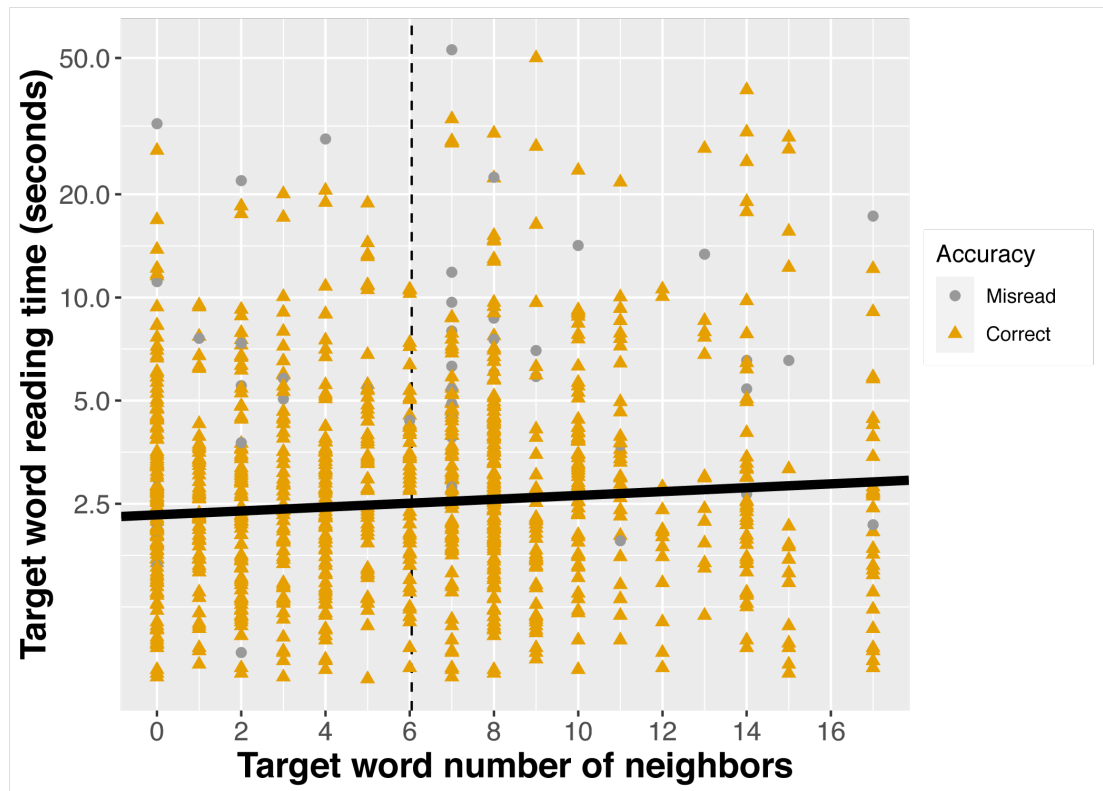
	Estimate	Standard Error	t-value	p-value	95% Confidence Interval
Intercept	0.841	0.143	5.855	<0.001	[0.55; 1.13]
Word neighborhood size (N)	0.013	0.005	2.507	0.013	[0.003; 0.023]
Word frequency (ln)	-0.047	0.013	-3.419	<0.001	[-0.07; -0.021]
Word length	0.016	0.030	0.546	0.589	[-0.04; 0.08]
Trigram occurrence	-0.115	0.027	-4.129	<0.001	[-0.17; -0.06]
Daily reading <i>No</i>	0.392	0.217	1.809	0.087	[-0.09; 0.87]

350

351 **Table 4: Fixed-effects estimates from the simple LME model (Analysis 2).**

352 **The dependent variable is log-transformed target word total reading time. The intercept estimate**
 353 **represents the log-transformed reading time when all factors included in the model are at their**
 354 **reference level (categorical variable) or at 0 (continuous variables). Reference level for the binary**
 355 **factor "Daily reading" is 'yes'. Factors showing a significant effect on reading time are in bold**
 356 **font and highlighted in grey.**

357



358

359 **Figure 5: Effect of neighborhood size (N) on word reading time.**

360 Data for both well-read words (yellow triangles) and misread words (grey circles) are plotted but
 361 the fitted line extracted from the LME model (solid black) only took into account the words
 362 correctly read. The black dotted line marks N mean value.

363

364 **Analysis 3 – Effect of the interaction between neighborhood size, word**
 365 **predictability and reading proficiency on reading time**

366 In this second LME model (Table 5), a 3-way interaction between neighborhood size,
 367 word predictability and reading proficiency was added to test our Hypotheses 2 and 3.
 368 According to this complex model, when trigram occurrence is at 0 (implying a highly
 369 infrequent trigram and low predictability; Figure 6A), words with zero neighbors were
 370 read on average in 3.0 seconds ($\exp(1.099)$) by patients who practice reading daily
 371 (Figure 6A – blue dashed line). For this same group of readers, increasing the number
 372 of neighbors did not have a significant effect on reading time ($t = -1.043$, $p = 0.298$;
 373 $95\%CI = [-0.02; 0.01]$; Figure 6A – blue dashed line). For patients who quit daily

374 reading activities, average reading time of words with zero neighbors was 3.14
 375 seconds ($\exp(1.099+0.044)$) and was not significantly different from the ‘daily
 376 reading’ group estimate ($t = 0.188$, $p = 0.853$, $95\%CI = [-0.42; 0.52]$; Figure 6A – red
 377 solid line). However, for these participants who stopped reading on a daily basis,
 378 increasing the number of neighbors by 1 did increase reading time significantly by a
 379 factor of 1.07 ($\exp(0.07)$; $t = 5.22$, $p < 0.001$, $95\%CI = [0.04; 0.10]$; Figure 6A – red
 380 solid line). In other words, for low predictability, increasing neighborhood size from 0
 381 to 6 (the mean value in our pool of target words), increases reading time by a factor of
 382 1.52 ($\exp(0.07)^6$), i.e. a 52% increase. Similarly, increasing neighborhood size from 0
 383 to 10 (where most of our values lie), increases reading time by a factor of 2.01
 384 ($\exp(0.07)^{10}$), i.e. a 101% increase.

385

	Estimate	Standard Error	t-value	p-value	95% Confidence Interval
Intercept	1.099	0.136	8.09	<0.001	[0.84; 1.37]
Word neighborhood size (N)	-0.009	0.008	-1.043	0.298	[-0.02; 0.01]
Word frequency (ln)	-0.046	0.014	-3.39	<0.001	[-0.07; -0.02]
Word length	0.013	0.028	0.48	0.633	[-0.04; 0.07]
Trigram occurrence	-0.162	0.049	-3.24	0.002	[-0.26; -0.07]
Daily reading <i>No</i>	0.044	0.234	0.188	0.853	[-0.42; 0.52]
N : Daily reading <i>No</i>	0.07	0.013	5.22	<0.001	[0.04; 0.10]
Trigram occurrence : Daily reading <i>No</i>	0.093	0.081	1.15	0.248	[-0.06; 0.25]
N : Trigram occurrence : Daily reading <i>Yes</i>	0.011	0.006	1.85	0.066	[-4e-04; 0.02]
N : Trigram occurrence : Daily reading <i>No</i>	-0.017	0.008	-2.03	0.043	[-0.03; -8e-04]

386

387 **Table 5: Fixed-effects estimates from the complex LME model (Analysis 3).**

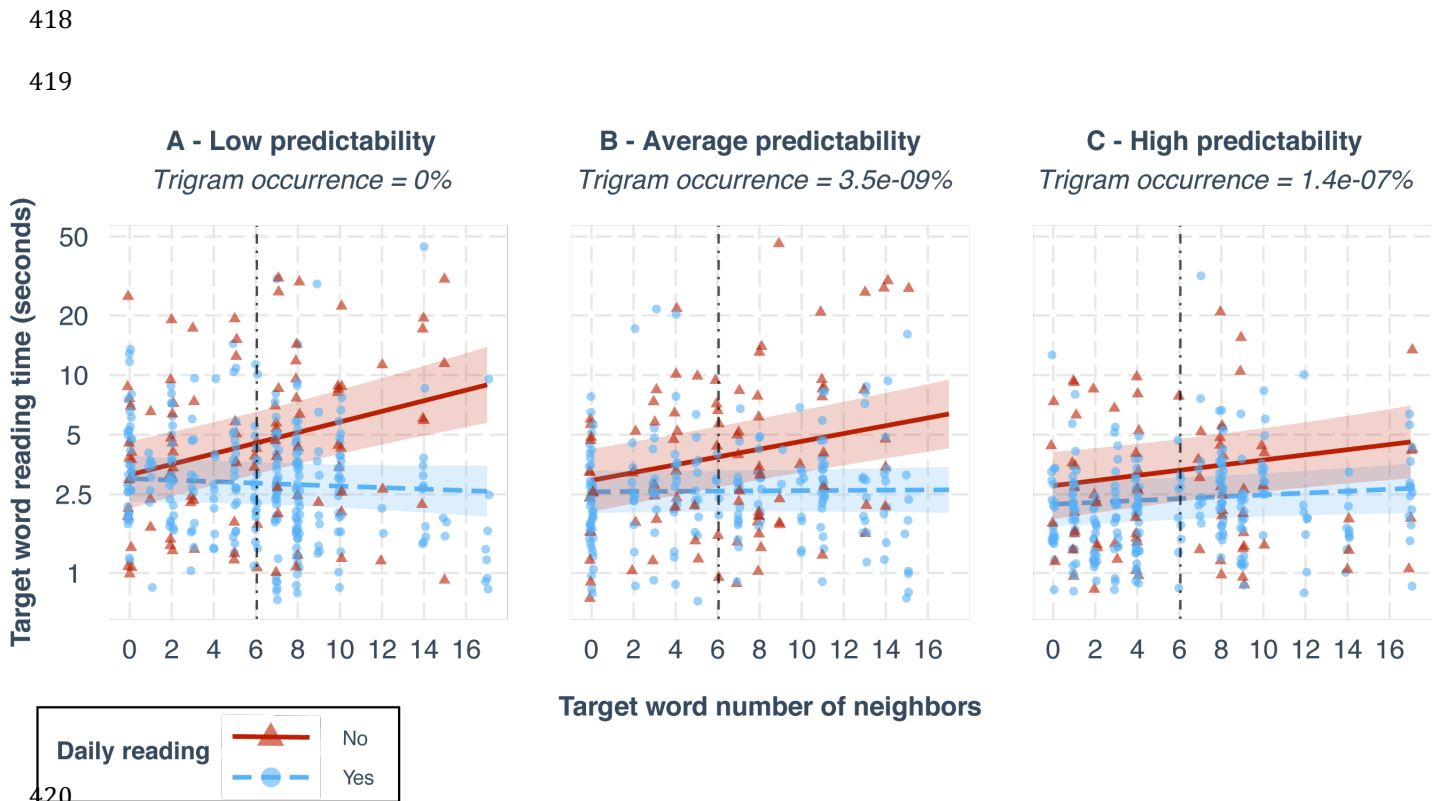
388 **The dependent variable is log-transformed target word total reading time. The intercept estimate**
389 **represents the log-transformed reading time when all factors included in the model are at their**
390 **reference level (categorical variable) or at 0 (continuous variable). Reference level for the binary**
391 **factor "Daily reading" is 'yes'. Interactions are represented by the symbol ":". Factors showing a**
392 **significant effect on reading time are in bold font and highlighted in grey.**

393

394 As trigram occurrence increases to an average value (Figure 6B) and a high value
395 (Figure 6C), the effect of neighborhood size remains null for the 'daily reading' group
396 ($t = 1.85$; $p = 0.066$; $95\%CI = [-4e-04; 0.02]$; blue dashed lines). However, for the
397 other group of patients who stopped practicing reading daily (red solid lines), the
398 amplitude of the neighborhood size effect is significantly reduced by a factor of 1.02
399 ($\exp(0.017)$; $t = -2.03$; $p = 0.043$; $95\%CI = [-0.03; -8e-04]$) every time trigram
400 occurrence increases by one unit. As given by a slopes post-hoc analysis, the
401 amplitude of the effect for the group who quit daily reading (red solid line) was of
402 1.05 ($\exp(0.05)$; $t = 5.39$, $p < 0.001$) for average trigram occurrence values (Figure
403 6B) and of 1.03 ($\exp(0.03)$; $t = 2.77$, $p = 0.01$) for highly frequent trigrams (Figure
404 6C). In other words, for fairly frequent trigrams yielding average predictability,
405 increasing neighborhood size from 0 to 10, increases reading time significantly by a
406 factor of 1.65 ($\exp(0.05)^{10}$), i.e. a 65% increase, while the same increase in
407 neighborhood size for highly frequent trigrams yielding high predictability, increases
408 reading time significantly by a factor of only 1.35 ($\exp(0.03)^{10}$), i.e. a 35% increase.

409 Besides this significant 3-way interaction, word frequency also had a significant effect
410 on reading time with a regression coefficient estimate of -0.046 ($t = -3.39$, $p < 0.001$,
411 $95\%CI = [-0.07; -0.02]$). As both reading time and frequency are expressed in natural
412 log units, multiplying frequency (in original units) by 10 multiplies reading time (in

413 original units) by 0.90 ($10^{-0.046}$), i.e. a 10% decrease. We found no significant
 414 interaction between frequency and the ‘daily reading’ factor, nor between frequency
 415 and the number of neighbors. Word length had no significant effect on reading time (t
 416 = 0.48, $p = 0.633$, 95%CI = [-0.04; 0.07]). Similarly, age, acuity and type of field loss
 417 (all dropped from the final model) showed no significant effect on reading time.



420

421 **Figure 6: Effect of neighborhood size (N) on reading time, depending on word predictability and**
 422 **reading proficiency.**

423 Raw data points and fitted lines estimated by the LME model are grouped by reading
 424 proficiency: red for participants who quit reading on a daily basis, blue for participants still
 425 reading daily. Each subplot represents the effect of N for a different value of the trigram
 426 occurrence distribution, split into three equal-sized groups (i.e., terciles): (A) median of the lower
 427 tercile, (B) median of the middle tercile, (C) median of the upper tercile (C).

428

429

DISCUSSION

430

431

432 The first goal of the present work was to test the hypothesis that the visual constraint
433 imposed by the presence of a central scotoma leads to an inhibitory effect of
434 neighborhood size during sentence reading. Therefore, we assessed the effect of word
435 neighborhood size (N) on the reading performance of 19 patients with CFL, namely
436 reading accuracy (Analysis 1) and reading time (Analysis 2). Our results show that N
437 has no significant effect on accuracy, which ceils around 94% overall. On the other
438 hand, we found a moderate inhibitory effect of N on reading time, with a 14%
439 increase in word reading time when N goes from 0 to 10 neighbors (i.e., the range
440 where most of our values lie). This result confirms our first hypothesis and builds up
441 the recent report of a reversed neighborhood size effect for visually impaired
442 individuals compared to normal readers [18].

443 Effects of orthographic neighbors on word identification have been extensively
444 explored in readers with normal vision. Despite the many contradictory findings
445 accumulated over the years, it is now accepted that the neighborhood size effect
446 depends on the task and is modulated by the frequency of the neighbors themselves
447 (i.e. the neighborhood frequency effect). Indeed, large neighborhood has consistently
448 been reported to facilitate responses in a variety of tasks such as word naming [44]
449 and lexical decision [45,46], but this facilitative effect seems to be restricted to low-
450 frequency neighbors [47]. Despite their great interest to help understand the
451 underlying mechanisms of lexical processing, these results are difficult to interpret in
452 the context of our work with natural reading since they are restrained to isolated word
453 identification.

454 Because semantic context may help decide between visually similar words, it is
455 necessary to examine the effects of neighborhood during natural reading in context to
456 assess their influence on reading performance. To this end, Pollatsek et al. examined
457 the pattern of eye movements during both a lexical decision task and silent reading
458 when target words varying in neighborhood size were embedded in neutral sentences
459 [48]. Their overall conclusion was that, for silent reading, increasing the number of
460 higher frequency neighbors had a clear inhibitory effect on word identification,
461 whereas increasing the number of lower frequency neighbors may have a weak
462 facilitative effect. In the present work, reading material was created without
463 controlling for neighborhood frequency, but a post-hoc analysis revealed that most of
464 our target words (82%) had a majority of low-frequency neighbors (from 60% to
465 100%; mean = $83 \pm 14\%$). Based on this distribution, results from normal vision
466 would predict a weak facilitative effect of neighborhood size on reading performance
467 [48]. On the contrary, we found a reverse effect of neighborhood size, with a weak
468 inhibitory effect. This result confirms our assumption that under degraded visual
469 conditions, the lack of complete stimulus information will have more of an effect on
470 words that are visually similar to many others than on words with few neighbors. It is
471 likely that in the interest of time performance, readers may infer a word by guessing
472 one of its high-frequency neighbors. However, because of the incongruousness
473 between this guess and the overall sentence meaning, processing would persist until a
474 better match, that fits both the visually identified letters and the meaning, is found.
475 This would explain why reading accuracy remained very high among our participants,
476 at the expense of reading time.

477 The second goal of the present work was to test the hypotheses that, with CFL, the
478 effect of neighborhood size is modulated by both word predictability and reading

479 proficiency. Therefore, we assessed the word neighborhood size effect through a 3-
480 way interaction, including a measure of predictability (trigram occurrence) and
481 proficiency (daily reading - *yes or no*) (Analysis 3). First, we found that the inhibitory
482 effect of N is modulated by word predictability: the more familiar a sequence of
483 words is, the weaker the effect of neighborhood size on the last word identification
484 time. This result confirms our second hypothesis that the amplitude of the
485 neighborhood size effect is influenced by word predictability.

486 However, we found that this is only true for the less proficient readers, who have
487 stopped reading on a daily basis, confirming our third hypothesis that the interaction
488 between neighborhood size and word predictability depends on patients' reading
489 proficiency. In short, we found that for the less proficient readers reaching a low-
490 predictable word in a sentence, reading speed decreases by up to 50% (101% increase
491 in reading time) when the number of neighbors goes from 0 to 10. As predictability
492 increases, the amplitude of this effect lessens gradually, with 39% decrease in reading
493 speed for average predictability (65% increase in reading time) and 26% decrease in
494 reading speed for high predictability (35% increase in reading time). For proficient
495 readers who reported to retain a daily leisure activity of reading however, even for a
496 few minutes each day, the effect of N remains null, regardless of the word
497 predictability. This result, close to what has been reported with normally sighted
498 readers [48], leads us to the conclusion that practice can help minimize the adverse
499 effect of ambiguity induced by orthographic similarity when visual input is degraded
500 and access to text is only partial because of maculopathy.

501 These results are particularly relevant in the context of low-vision rehabilitation, as
502 they reinforce the need to provide patients with individualized readaptive care of
503 functional vision, in order to help maintain daily reading practice. More importantly,

504 our results suggest that text simplification might be a powerful way to leverage text
505 accessibility for low-vision patients. Text simplification is the process of reducing the
506 linguistic complexity of a text, while still retaining the original information and
507 meaning [49,50]. Its main goal is to make a text more accessible to people with low
508 literacy [51] or individuals with reading disorders (e.g., aphasia [52], dyslexia [53]).
509 For the first time, our results show that it could also be used efficiently to improve
510 low-vision readaptation, especially for the most impaired patients, who have stopped
511 reading on a daily basis (36.8% in our population sample). For these less proficient
512 readers, substituting complex words (i.e, words with many orthographic neighbors)
513 with synonyms that have less neighbors and equal or higher frequency should reduce
514 reading difficulty. Therefore, using simplified texts with increased accessibility as
515 rehabilitation training material might help (1) improve overall reading ability and
516 fluency, while (2) fostering the long-term motivation necessary to resume daily
517 reading practice. Text simplification could then be used daily as an efficient reading
518 aid (made available through tablets, e-readers, or web plugins) to keep practicing
519 reading at home and enhance everyday reading performance.

520 As a side note, we would like to point to the fact that our results were obtained using
521 the occurrence percentage of trigrams (3-gram, i.e., sequence of three words) in the
522 French literature. It is worth mentioning that we also ran the analysis with other n-
523 gram values, namely 2- and 4-gram and that neither of them showed significant effect.
524 Since most of our target words were common nouns (89%; against 9.4% of adjectives
525 and 1.6% of verbs), they were most likely preceded by an article that did not convey
526 meaningful information. Therefore an analysis based on 2-gram prediction was not
527 likely to be meaningful. On the other hand, we expected 4-gram analysis to be highly
528 significant. However, because we created our reading material so that target words

529 were not too predictable, 77% of our 4-grams were highly infrequent (less than 40
530 occurrences across the ‘French 2012’ corpus) and rated at 0 percent occurrence by
531 Google Ngram. We suspect that the absence of significant effect with 4-grams is due
532 to this highly skewed distribution towards 0.

533 Overall, the present work presents some limitations that should be taken into account
534 in future investigations of neighborhood effects on low vision. First, neighborhood
535 frequency should be gauged thoroughly when designing the reading material in order
536 to better control for its effect when assessing the effect of neighborhood size. Second,
537 the definition of orthographic neighbor used in this work is letter-position-specific
538 and length-dependent [15]. Given that letter position uncertainty is a crucial factor
539 limiting peripheral word recognition, and reading without central vision in general
540 [54], Coltheart’s definition should probably be extended to include letter transposition
541 (e.g., trial and trail), addition (e.g. trial and tribal) and deletion (e.g. trial and rial) in
542 order to encompass a larger number of highly similar words. Finally, the results
543 presented here should be interpreted cautiously in the context of reading under natural
544 conditions, since they were obtained with a paradigm that does not allow word
545 skipping, forcing participants to read each single word of a sentence [55].

546

547

548

549

550

551

REFERENCES

552

553

554 1. Brown JC, Goldstein JE, Chan TL, Massof R, Ramulu P, Low Vision Research
555 Network Study Group (2014) Characterizing functional complaints in patients seeking
556 outpatient low-vision services in the United States. *Ophthalmology* 121: 1655-62.e1.

557

558 2. Kanonidou E (2011) Reading performance and central field loss. *Hippokratia* 15:
559 103-8.

560

561 3. Chung S (2020) Reading in the presence of macular disease: a mini-review.
562 *Ophthalmic and Physiological Optics*.

563

564 4. Murro V, Sodi A, Giacomelli G, Mucciolo DP, Pennino M, Virgili G, et al. (2017)
565 Reading Ability and Quality of Life in Stargardt Disease. *Eur J Ophthalmol* 27: 740-
566 745.

567

568 5. Pondorfer SG, Terheyden JH, Heinemann M, Wintergerst MWM, Holz FG, Finger
569 RP (2019) Association of Vision-related Quality of Life with Visual Function in Age-
570 Related Macular Degeneration. *Scientific Reports* 9: 15326.

571

572 6. Calabrèse A, Bernard J, Faure G, Hoffart L, Castet E (2016b) Clustering of Eye
573 Fixations: A New Oculomotor Determinant of Reading Speed in Maculopathy. *Invest*
574 *Ophthalmol Vis Sci* 57: 3192-202.

575

576 7. Leroy G, Kauchak D (2014) The effect of word familiarity on actual and perceived
577 text difficulty. *J Am Med Inform Assoc* 21: e169-72.

578

579 8. Adelman JS, Brown GDA (2007) Phonographic neighbors, not orthographic
580 neighbors, determine word naming latencies. *Psychonomic Bulletin & Review* 14:
581 455--459.

582

583 9. Taylor DJ, Edwards LA, Binns AM, Crabb DP (2018) Seeing it differently: self-
584 reported description of vision loss in dry age-related macular degeneration.
585 *Ophthalmic Physiol Opt* 38: 98-105.

586

587 10. Chung ST (2007) Learning to identify crowded letters: Does it improve reading
588 speed?. *Vision Research* 47.

589

590 11. Bullimore M, Bailey I (1995) Reading and eye movements in age related
591 maculopathy. *Optometry and Vision Science*.

592

593 12. Fine EM, Peli E (1996) The role of context in reading with central field loss.
594 *Optom Vis Sci* 73: 533-9.

595

596 13. Legge GE, Klitz TS, Tjan BS (1997) Mr. Chips: an ideal-observer model of
597 reading. *Psychol Rev* 104: 524-53.

598

599 14. Stolowy N, Calabrèse A, Sauvan L, Aguilar C, François T, Gala N, et al. (2019)
600 The influence of word frequency on word reading speed when individuals with
601 macular diseases read text. *Vision Research* 155: 1 - 10.

602

603 15. Coltheart M, Davelaar E, Jonasson JE, Besner D (1977) Access to the internal
604 lexicon. In: Dornio S, editor. *Attention and performance VI*. Hillsdale, NJ: Erlbaum:
605 pp. 535--555.

606

607 16. Andrews S (1997) The effect of orthographic similarity on lexical retrieval:
608 Resolving neighborhood conflicts. *Psychonomic Bulletin & Review* 4: 439--461.

609

610 17. Perea M, Martínez E (2000) The effects of orthographic neighborhood in reading
611 and laboratory word identification tasks. *Psicológica* 21.

612

613 18. Sauvan L, Stolowy N, Aguilar C, François T, Gala N, Matonti F, et al. (2020)
614 Text Simplification to Help Individuals With Low Vision to Read More Fluently.
615 Workshop Tools and Resources to Empower People with Reading Difficulties
616 (READI) at International conference on Language Resources and Evaluation (LREC
617 2020), oral presentation.: 27-32. Marseille, France.

618

619 19. Rayner K (1998) Eye movements in reading and information processing: 20 years
620 of research. *Psychol Bull* 124: 372-422.

621

622 20. Balota DA, Pollatsek A, Rayner K (1985) The interaction of contextual
623 constraints and parafoveal visual information in reading. *Cogn Psychol* 17: 364-90.

624

625 21. Hawelka S, Schuster S, Gagl B, Hutzler F (2015) On forward inferences of fast
626 and slow readers. An eye movement study. *Sci Rep* 5: 8432.

627

628 22. Ashby J, Rayner K, Clifton C (2005) Eye movements of highly skilled and
629 average readers: differential effects of frequency and predictability. *Q J Exp Psychol*
630 *A* 58: 1065-86.

631

632 23. Peirce JW (2007) PsychoPy--Psychophysics software in Python. *J Neurosci*
633 *Methods* 162: 8-13.

634

635 24. Peirce JW (2009) Generating stimuli for neuroscience using PsychoPy. *Frontiers*
636 *in Neuroinformatics* 2: 10.

637

638 25. Calabrèse A, Bernard J, Faure G, Hoffart L, Castet E (2014) Eye movements and

639 reading speed in macular disease: the shrinking perceptual span hypothesis requires
640 and is supported by a mediation analysis. *Invest Ophthalmol Vis Sci* 55: 3638-45.
641

642 26. Calabrèse A, Mansfield JS, Legge GE (2019) mnreadR, an R package to analyze
643 MNREAD data. version 2.1.3 <https://CRAN.R-project.org/package=mnreadR>.
644

645 27. Billami M, François T, Gala N (2018) ReSyf: a French lexicon with ranked
646 synonyms. In *Proceedings of the 27th Conference on Computational Linguistics*
647 (COLING 2018), Santa Fe, USA: 2570-2581. <https://cental.uclouvain.be/resyf/>.
648

649 28. New B, Ferrand L, Pallier C, Brysbaert M (2006) Reexamining the word length
650 effect in visual word recognition: new evidence from the English Lexicon Project.
651 *Psychon Bull Rev* 13: 45-52.
652

653 29. Just MA, Carpenter PA (1980) A theory of reading: from eye fixations to
654 comprehension. *Psychol Rev* 87: 329-54.
655

656 30. Aaronson D, Scarborough HS (1976) Performance theories for sentence coding:
657 Some quantitative evidence. *Journal of Experimental Psychology: Human Perception*
658 *and Performance* 2: 56-70.
659

660 31. Mitchell DC, Green DW (1978) The effects of context and content on immediate
661 processing in reading. *Q. J. Exp. Psychol.* 30: 609-636.
662

663 32. Just MA, Carpenter PA, Woolley JD (1982) Paradigms and processes in reading
664 comprehension. *J Exp Psychol Gen* 111: 228-38.
665

666 33. Wallis S, Yang Y, Anderson SJ (2018) Word Mode: a crowding-free reading
667 protocol for individuals with macular disease. *Sci Rep* 8: 1241.
668

669 34. Michel J, Shen YK, Aiden AP, Veres A, Gray MK, Google Books Team, et al.
670 (2011) Quantitative analysis of culture using millions of digitized books. *Science* 331:
671 176-82.
672

673 35. R Core Team (2018) R: A Language and Environment for Statistical Computing.
674 Vienna, Austria: R Foundation for Statistical Computing. , [https://www.R-](https://www.R-project.org/)
675 [project.org/](https://www.R-project.org/).
676

677 36. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MH, et al.
678 (2009) Generalized linear mixed models: a practical guide for ecology and evolution.
679 *Trends Ecol Evol* 24 (3): 127-135.
680

681 37. Cheng J, Edwards LJ, Maldonado-Molina MM, Komro KA, Muller KE (2010)
682 Real longitudinal data analysis for real people: building a good enough mixed model.

683 Stat Med 29: 504-20.
684
685 38. Tabachnick BG, Fidell LS, Ullman JB (2007) Using multivariate statistics.
686 Pearson Boston, MA, .
687
688 39. Howell DC (2009) Statistical methods for psychology. Cengage Learning, .
689
690 40. Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid
691 common statistical problems. *Methods in Ecology and Evolution* 1: 3-14.
692
693 41. Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed
694 random effects for subjects and items. *Journal of Memory and Language* 59: 390-412.
695
696 42. Gelman A, Hill J (2007) Data analysis using regression and
697 multilevel/hierarchical models. Cambridge University Press, New York.
698
699 43. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects
700 Models Using lme4. *Journal of Statistical Software* 67: 1--48.
701
702 44. Andrews S (1989) Frequency and neighborhood effects on lexical access:
703 Activation or search? *Journal of Experimental Psychology: Learning, Memory, and*
704 *Cognition* 15: 802--814.
705
706 45. Sears CR, Hino Y, Lupker SJ (1995) Neighborhood size and neighborhood
707 frequency effects in word recognition. *Journal of Experimental Psychology: Human*
708 *Perception and Performance* 21: 876-900.
709
710 46. Forster KI, Shen D (1996) No enemies in the neighborhood: absence of inhibitory
711 neighborhood effects in lexical decision and semantic categorization.. *J Exp Psychol*
712 *Learn Mem Cogn* 22: 696--713.
713
714 47. Carreiras M, Perea M, Grainger J (1997) Effects of orthographic neighborhood in
715 visual word recognition: cross-task comparisons.. *J Exp Psychol Learn Mem Cogn*
716 *23: 857--871.*
717
718 48. Pollatsek A, Perea M, Binder KS (1999) The effects of "neighborhood size" in
719 reading and lexical decision. *J Exp Psychol Hum Percept Perform* 25: 1142-58.
720
721 49. Siddharthan A (2014) A survey of research on text simplification. *Int. J. Appl.*
722 *Linguist.* 165: 259-298.
723
724 50. Saggion H (2017) Automatic Text Simplification. *Synthesis Lectures on Human*
725 *Language Technologies* 10: 1-137.
726

727 54. Bernard J, Castet E (2019) The optimal use of non-optimal letter information in
728 foveal and parafoveal word recognition. *Vision Res* 155: 44-61.

729

730 55. Albregues C, Lavigne F, Aguilar C, Castet E, Vitu F (2019) Linguistic processes
731 do not beat visuo-motor constraints, but they modulate where the eyes move
732 regardless of word boundaries: Evidence against top-down word-based eye-movement
733 control during reading. *PLOS ONE* 14: 1-47.

734

735

736 ACKNOWLEDGEMENTS

737 This work was supported by the Fondation de France and by the Belgian FNRS.

738

739 AUTHOR CONTRIBUTIONS

740 LS, TF and NG prepared the linguistic material used for data collection. CA
741 developed the program used for data collection. LS and NS collected the data. AC
742 performed the data analysis and wrote the main manuscript text. All authors reviewed
743 the manuscript.

744

745

746 **FIGURE LEGENDS**

747 Figure 1: The presence of a central scotoma occludes part of the word “father”,
748 leading to possible misidentifications (e.g., farmer, feather, halter) and a greater need
749 for linguistic inference

750

751 Figure 2: Creation of the reading material.

752 Illustration of the reading material creation, using the synonym pair “*shore / coast*”.
753 Both words in this example contain 5 characters. Their difference in number of
754 neighbors is 10. Sentences A and B were created so either word of this synonym pair
755 can fit in both sentences. In Condition 1, sentence A contains the target word 1
756 (‘shore’) and sentence B contains the target word 2 (‘coast’). In Condition 2, the
757 pairing is reversed. Note that this example is given in English to ease understanding.
758 See Supplementary Material for the complete set of French sentences created.

759

760 Figure 3: Self-paced Reading paradigm.

761 Words are revealed one at a time by the participant, using keyboard presses. A press
762 on the right arrow key will unmask the upcoming word (while masking the currently
763 visible word) and a press on the left arrow key will unmask the previous word (while
764 masking the currently visible word). For words unmasked only once (e.g. word 1), the
765 total reading time equals the 1st-pass reading time. For words unmasked more than
766 once (e.g. word 2), the total reading time is the sum of each individual reading time
767 pass.

768

769 Figure 4: Effect of the neighborhood size (N) on word reading accuracy.

770 Each data point represents a target word. Raw data points and fitted curves estimated
771 by the GLME model are grouped according to reading proficiency: blue for
772 participants reading daily, red for participants who quit daily reading activities. The
773 right side panel shows a zoomed view of the left plot between 70 and 100% accuracy.
774

775 Figure 5: Effect of neighborhood size (N) on word reading time.

776 Data for both well-read words (yellow triangles) and misread words (grey circles) are
777 plotted but the fitted line extracted from the LME model (solid black) only took into
778 account the words correctly read. The black dotted line marks N mean value.
779

780 Figure 6: Effect of neighborhood size (N) on reading time, depending on word
781 predictability and reading proficiency.

782 Raw data points and fitted lines estimated by the LME model are grouped by reading
783 proficiency: red for participants who quit reading on a daily basis, blue for
784 participants still reading daily. Each subplot represents the effect of N for a different
785 value of the trigram occurrence distribution, split into three equal-sized groups (i.e.,
786 terciles): (A) median of the lower tercile, (B) median of the middle tercile, (C) median
787 of the upper tercile (C).

788

789

790

791 TABLES

792

793

Partici pants	Gen der	Age (years)	Etiology	Disease onset (years)	Low-vision rehabili tation	Visual acuity (logMAR)	Lens status	Daily reading time (min/day)
1	M	76	Atrophic AMD	2	Yes	1	Phakic	0
2	F	77	Atrophic AMD	7	No	0.4	Pseudophakic	45
3	F	80	Atrophic AMD	7	Yes	1	Pseudophakic	0
4	F	80	Atrophic AMD	5	Yes	0.8	Phakic	0
5	F	83	Atrophic AMD	12	Yes	0.5	Phakic	60
6	F	83	Atrophic AMD	2	Yes	0.5	Pseudophakic	0
7	F	85	Atrophic AMD	15	No	0.7	Pseudophakic	60
8	M	86	Atrophic AMD	8	Yes	0.7	Pseudophakic	60
9	M	88	Atrophic AMD	20	Yes	0.7	Phakic	60
10	F	78	Exudative AMD	3	No	1.1	Pseudophakic	0
11	F	87	Exudative AMD	2	Yes	0.5	Pseudophakic	10
12	M	89	Exudative AMD	3	Yes	0.4	Pseudophakic	30
13	F	48	Stargardt's disease	35	Yes	1.3	Phakic	25
14	F	72	Stargardt's disease	16	No	1	Phakic	120
15	F	32	Cone dystrophy	30	No	1	Phakic	120
16	M	73	Diabetic retinopathy	10	No	0.7	Pseudophakic	0
17	M	59	Myopic retinopathy	15	Yes	1	Phakic	0
18	F	70	Myopic retinopathy	20	Yes	1.3	Pseudophakic	120
19	F	70	Myopic retinopathy	70	Yes	0.8	Pseudophakic	30

794

795 **Table 1: Participants' demographic, visual and reading characteristics.**796 **Visual acuity and lens status are given for the tested eye only. Participants with central scotoma**797 **only are highlighted in red. Participants with central scotoma and peripheral lesions are**798 **highlighted in blue.**

799

800

801

Variable name	Unit / Transformation	Centered	Description
Target word characteristics			
Neighborhood size (N)	--	--	Number of neighbors
Frequency (ln)	Natural log	Yes	Occurrences/million
Length	--	Yes	Number of characters
Sentence characteristics			
Trigram occurrence	Ordered quantile normalization	--	Percentage of occurrence for each trigram ending with a target word
Participant's individual characteristics			
Age	Years	--	--
Visual Acuity	LogMAR	--	--
Type of field loss	<ul style="list-style-type: none"> • Central only • Central + peripheral 	--	--
Still reading daily	<ul style="list-style-type: none"> • Yes • No 	--	--

802

803 **Table 2: Description of the different factors included as independent variables in the linear**

804 **mixed-effects models.**

805

	Estimate	Standard Error	z-value	p-value	95% Confidence Interval
Intercept	4.716	0.613	7.694	<0.001	[3.65; 6.17]
Word neighborhood size (N)	-0.058	0.054	-1.069	0.285	[-0.17; 0.05]
Word frequency (ln)	0.147	0.147	1.003	0.316	[-0.15; 0.46]
Word length	-0.028	0.213	-0.13	0.897	[-0.47; 0.41]
Trigram occurrence	0.361	0.292	1.236	0.216	[-0.21; 0.98]
Daily reading <i>No</i>	-1.126	0.545	-2.064	0.039	[-2.31; -0.03]

807

808 **Table 3: Fixed-effects estimates from the GLME model (Analysis 1).**

809 **The dependent variable is ‘Target word reading accuracy’. The intercept estimate is the**
810 **predicted value of the dependent variable when all independent variables are at 0 (continuous**
811 **variables) or at their reference level (categorical variable). Reference level for the binary factor**
812 **”Daily reading” (representing reading proficiency) is ‘yes’. Factors showing a significant effect**
813 **on reading accuracy are in bold font and highlighted in grey.**

814

	Estimate	Standard Error	t-value	p-value	95% Confidence Interval
Intercept	0.841	0.143	5.855	<0.001	[0.55; 1.13]
Word neighborhood size (N)	0.013	0.005	2.507	0.013	[0.003; 0.023]
Word frequency (ln)	-0.047	0.013	-3.419	<0.001	[-0.07; -0.021]
Word length	0.016	0.030	0.546	0.589	[-0.04; 0.08]
Trigram occurrence	-0.115	0.027	-4.129	<0.001	[-0.17; -0.06]
Daily reading <i>No</i>	0.392	0.217	1.809	0.087	[-0.09; 0.87]

816

817 **Table 4: Fixed-effects estimates from the simple LME model (Analysis 2).**

818 **The dependent variable is log-transformed target word total reading time. The intercept estimate**
819 **represents the log-transformed reading time when all factors included in the model are at their**
820 **reference level (categorical variable) or at 0 (continuous variables). Reference level for the binary**
821 **factor "Daily reading" is 'yes'. Factors showing a significant effect on reading time are in bold**
822 **font and highlighted in grey.**

	Estimate	Standard Error	t-value	p-value	95% Confidence Interval
Intercept	1.099	0.136	8.09	<0.001	[0.84; 1.37]
Word neighborhood size (N)	-0.009	0.008	-1.043	0.298	[-0.02; 0.01]
Word frequency (ln)	-0.046	0.014	-3.39	<0.001	[-0.07; -0.02]
Word length	0.013	0.028	0.48	0.633	[-0.04; 0.07]
Trigram occurrence	-0.162	0.049	-3.24	0.002	[-0.26; -0.07]
Daily reading <i>No</i>	0.044	0.234	0.188	0.853	[-0.42; 0.52]
N : Daily reading <i>No</i>	0.07	0.013	5.22	<0.001	[0.04; 0.10]
Trigram occurrence : Daily reading <i>No</i>	0.093	0.081	1.15	0.248	[-0.06; 0.25]
N : Trigram occurrence : Daily reading <i>Yes</i>	0.011	0.006	1.85	0.066	[-4e-04; 0.02]
N : Trigram occurrence : Daily reading <i>No</i>	-0.017	0.008	-2.03	0.043	[-0.03; -8e-04]

823

824 **Table 5: Fixed-effects estimates from the complex LME model (Analysis 3).**

825 **The dependent variable is log-transformed target word total reading time. The intercept estimate**
826 **represents the log-transformed reading time when all factors included in the model are at their**
827 **reference level (categorical variable) or at 0 (continuous variable). Reference level for the binary**
828 **factor "Daily reading" is 'yes'. Interactions are represented by the symbol ":". Factors showing a**
829 **significant effect on reading time are in bold font and highlighted in grey.**

830 SUPPLEMENTARY MATERIAL

831

832 Mon ami nous parlait de son **métier (boulot)** passionnant

833 Il ne s'ennuyait pas car c'était un **boulot (métier)** prenant

834 Il était aimé de tous, c'était le **chef (boss)** de l'usine

835 Antoine a toujours voulu être le **boss (chef)** du groupe

836 Je suis resté bloqué car le **loquet (verrou)** ne s'ouvrait plus

837 Si j'étais toi je ferais installer un **verrou (loquet)** sur la porte

838 Tu devrais aller faire une **marche (balade)** pour te détendre

839 Nous avons aimé cette belle **balade (marche)** dans les bois

840 La semaine prochaine il faudra faire un **arrêt (halte)** chez mamie

841 En rentrant nous avons fait une **halte (arrêt)** à la boulangerie

842 Il a finalement réussi à trouver une **faille (brèche)** dans le système

843 Ils sont entrés par une petite **brèche (faille)** à travers le mur

844 On marchait dans de la **boue (vase)** c'était désagréable

845 Il fallait traverser de la **vase (boue)** pour y accéder

846 Elle a longtemps travaillé avec un autre **type (gars)** du bureau

847 Chaque samedi il vendait des poulets avec un **gars (type)** du village

848 Avec sa voiture décapotable il aime **frimer (crâner)** en ville

849 Elle est toujours en train de **crâner (frimer)** avec sa robe

850 Ne me dis pas que le chien a encore mangé ma **godasse (soulier)** en cuir

851 Ce matin en rangeant j'ai trouvé mon **soulier (godasse)** sous le lit

852 Tu devrais aller te mettre au **repos (calme)** dans la chambre

853 Ça m'a fait du bien ce séjour au **calme (repos)** chez ma mère

854 Il a travaillé de longues années pour devenir un **expert (savant)** réputé

855 Je lui ai demandé son avis car c'est un **savant (expert)** dans ce domaine

856 Nous pensons qu'il s'agit d'un **germe (virus)** plutôt agressif

857 Elle n'est pas sortie car elle avait un **virus (germe)** dangereux

858 Faites attention, vous devez marcher sur le **bord (côté)** de la route

859 Les enfants se mettaient sur le **côté (bord)** pour laisser passer

860 J'étais la seule déguisée hier soir, c'était **marrant (amusant)** mais gênant

861 Ils se sont revus par hasard, c'est **amusant (marrant)** cette coïncidence

862 Ce serait gentil de vous mettre bien au **centre (milieu)** de l'écran

863 Il tira de toutes ses forces depuis le **milieu (centre)** du terrain

864 Tous les vendredis midi je vais au **bar (pub)** du village

865 La semaine dernière il m'a invitée au **pub (bar)** du coin

866 Si tu vas jusqu'à la falaise tu vois une **baie (anse)** sublime en bas

867 J'aimais marcher jusqu'à cette **anse (baie)** à l'abri des touristes

868 Au musée nous avons vu un ancien **char (tank)** très abîmé

869 On nous a laissé grimper dans un **tank (char)** de l'époque

870 Au milieu de la cour il y avait une **botte (gerbe)** de blé

871 Pour la décoration je mettrai une **gerbe (botte)** de fleurs

872 Elle portait toujours un **large (grand)** manteau en hiver

873 Devant la maison il y avait un **grand (large)** patio
874 Avec le brouillard ils n'ont pas vu la **jetée (digue)** devant eux
875 Chaque dimanche il marchait jusqu'à la **digue (jetée)** et revenait
876 Mon frère ne travaille pas, c'est un **feignant (flemmard)** depuis toujours
877 Il ne fallait pas se plaindre ni être un **flemmard (feignant)** pour réussir
878 Le patron n'a pas pu obtenir un **délai (durée)** acceptable
879 Ils peuvent me rembourser avec une **durée (délai)** convenable
880 Il a freiné d'un coup pour éviter le **bus (car)** qui passait
881 Je l'ai rencontré dans le **car (bus)** la semaine dernière
882 Pour valider ce certificat il faut le **cachet (tampon)** du médecin
883 Vous êtes sensé mettre un **tampon (cachet)** sur ces documents
884 Ça me brûlait depuis trois jours, j'avais une **saleté (crasse)** dans l'œil
885 Lorsque nous sommes entrés, il y avait de la **crasse (saleté)** partout
886 Je te préviens car cet objet est **cassant (fragile)** et j'y tiens
887 Je l'ai hérité de ma mère il est **fragile (cassant)** mais très utile
888 Je souhaite refaire ma cuisine avec cette **nuance (teinte)** de bleu
889 On distingue sur tous ses tableaux une **teinte (nuance)** particulière
890 Ils ont passé la journée dans la même **salle (pièce)** sans se parler
891 Pouvez-vous déposer ces affaires dans la **pièce (salle)** du fond
892 Pierre est énervant, il est **avare (radin)** comme son père
893 Il devenait de plus en plus **radin (avare)** avec le temps
894 Olivier a fait refaire sa cuisine et il est **content (heureux)** du résultat
895 Il l'a trouvé sur une brocante, il était **heureux (content)** de sa trouvaille
896
897
898
899
900
901
902
903