

# Text Simplification to Help Individuals With Low Vision Read More Fluently

Lauren Sauvan, Natacha Stolowy, Carlos Aguilar, Thomas François, Núria Gala, Frédéric Matonti, Eric Castet, Aurelie Calabrese

► **To cite this version:**

Lauren Sauvan, Natacha Stolowy, Carlos Aguilar, Thomas François, Núria Gala, et al.. Text Simplification to Help Individuals With Low Vision Read More Fluently. LREC 2020 - Language Resources and Evaluation Conference, May 2020, Marseille, France. pp.11 - 16. hal-02946900

**HAL Id: hal-02946900**

**<https://hal.inria.fr/hal-02946900>**

Submitted on 23 Sep 2020

**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.

# Text Simplification to Help Individuals With Low Vision Read More Fluently

Lauren Sauvan<sup>1\*</sup>, Natacha Stolowy<sup>1\*</sup>, Carlos Aguilar<sup>2</sup>, Thomas François<sup>3</sup>, Núria Gala<sup>4</sup>,  
Frédéric Matonti<sup>5</sup>, Eric Castet<sup>6</sup>, Aurélie Calabrèse<sup>7</sup>

<sup>1</sup>North Hospital, Marseille France; <sup>2</sup>Mantu Lab, Amaris Research Unit, Sophia Antipolis, France; <sup>3</sup>UCLouvain, IL&C, CENTAL, Belgique; <sup>4</sup>Aix-Marseille Univ. Laboratoire Parole et Langage, CNRS UMR 7309, Aix-en-Provence, France; <sup>5</sup>Centre Paradis Monticelli, Marseille, France; <sup>6</sup>Aix-Marseille Univ., Laboratoire de Psychologie Cognitive, CNRS UMR 7290, Marseille, France; <sup>7</sup>Université Côte d’Azur, Inria, Sophia Antipolis, France.

\* These authors contributed equally  
aurelie.calabrese@inria.fr

## Abstract

The objective of this work is to introduce text simplification as a potential reading aid to help improve the poor reading performance experienced by visually impaired individuals. As a first step, we explore what makes a text especially complex when read with low vision, by assessing the individual effect of three word properties (frequency, orthographic similarity and length) on reading speed in the presence of Central visual Field Loss (CFL). Individuals with bilateral CFL induced by macular diseases read pairs of French sentences displayed with the self-paced reading method. For each sentence pair, sentence *n* contained a target word matched with a synonym word of the same length included in sentence *n+1*. Reading time was recorded for each target word. Given the corpus we used, our results show that (1) word frequency has a significant effect on reading time (the more frequent the faster the reading speed) with larger amplitude (in the range of seconds) compared to normal vision; (2) word neighborhood size has a significant effect on reading time (the more neighbors the slower the reading speed), this effect being rather small in amplitude, but interestingly reversed compared to normal vision; (3) word length has no significant effect on reading time. Supporting the development of new and more effective assistive technology to help low vision is an important and timely issue, with massive potential implications for social and rehabilitation practices. The end goal of this project will be to use our findings to custom text simplification to this specific population and use it as an optimal and efficient reading aid.

**Keywords:** low vision, lexical simplification, word frequency, word neighborhood size, word length

## 1. Introduction

Age-related macular degeneration (AMD) accounts for 8.7% of all blindness worldwide and is the most common cause of blindness in developed countries. Older adults suffering from AMD often lose the ability to use central vision after developing a central scotoma. Despite advances in the treatment of AMD (Miller, 2013), central vision cannot be restored and difficulty with reading is often the primary complaint of patients with central field loss (CFL) (Brown et al., 2014), who have to use their eccentric vision for reading. The number of Europeans with AMD being expected to reach 60 million by 2030 (Wong et al., 2014), there is a real societal need to understand the reading deficit of these patients in order to help restore their functional reading.

Over the past twenty years, different approaches have been taken to explore this still unresolved matter. First, great effort has been invested in determining whether manipulations of text display (magnification, line spacing, etc.) could improve reading performance (Calabrèse et al., 2010). However, no modification of text presentation has proven to significantly increase reading speed for people with central vision loss. Another approach, extensively explored recently, is to optimize the capabilities of the remaining peripheral vision for reading through perceptual learning. Unfortunately, studies investigating training benefits in people with AMD show a very wide and uneven range of reading speed improvement (Calabrèse et al., 2017). A third approach lies in the development of cutting-edge reading aids targeted towards central vision loss to increase reading accessibility. The current work falls directly within this scope with the innovative idea to use text simplification as a new reading aid for individuals with CFL.

Text simplification is a growing domain in the field of Natural Language Processing (NLP), combining computer science, psycholinguistics and computational linguistics. Given a text, its main objective is to identify difficult linguistic forms for a given population and then remove or substitute them with simpler equivalents, customized to the needs of this specific population. The aim is to produce an equivalent version while keeping the meaning unchanged (Saggion, 2017). Text simplification has been used to make texts more accessible to various populations: people with low-literacy (Watanabe et al., 2009), second language learners (Crossley et al., 2014), deaf people (Inui et al., 2003), autistic readers (Barbu et al., 2013) or individuals with reading disorders, such as dyslexia (Rello et al., 2013; Ziegler et al., 2015). Text simplification can be achieved through: (1) the addition of information (definitions, explanations, etc.), (2) the deletion of unnecessary information or (3) the reduction of linguistic complexity using simpler equivalents (Shardlow, 2014). These three types of linguistic simplification can be carried at different linguistic levels: lexical (through synonym substitution), morpho-syntactic (through word-form variation, sentence splitting, clause deletion, among others), or discursive (expliciting pronouns by their referent, expliciting discourse relations through discourse markers, etc.). Although very promising in its current application fields, text simplification has never been applied to low vision before.

The general objective of the present work is to investigate whether text simplification can promote higher reading performance with AMD by reducing the linguistic complexity of text for individuals with low vision. By investigating which lexical transformation(s) can most benefit reading with central field loss, our long-term goal is to provide a first set of useful guidelines to design

reading aids using text simplification that will promote reading performance improvement for this population.

Word frequency, word length and word neighborhood size are some of the most important linguistic factors known to affect text complexity and reading performance in normal vision (Adelman & Brown, 2007; Leroy & Kauchak, 2014). The neighborhood of a given word (*e.g.*, FIST) being defined as all the words of the same length varying from it by only one letter (*e.g.*, GIST, FAST, etc. (Coltheart et al., 1977)). In the case of CFL, visual input is deteriorated and access to text is only partial (Taylor et al., 2018). When reading the word “halte” with a scotoma (Figure 1), eccentric or incomplete letters may not be properly identified, leading to many possible misidentifications (“halle”, “balte”, “balle”, “batte”, etc.). Because bottom-up visual input is less reliable, CFL individuals must rely much more on top-down linguistic inference than readers with normal vision (Bullimore, 1995; Fine & Peli, 1996). Thus, we hypothesize that the effect of linguistic factors on reading performance should be much different in CFL patients than reported before with normally sighted readers. In a recent work, we started investigating this hypothesis by inspecting the effect of word frequency on the reading performance of 28 readers with CFL (Stolowy et al., 2019). As expected, results showed that low-frequency words significantly decrease reading speed. However, the amplitude of this effect was much larger for the visually impaired (differences in the range of seconds) than reported before for normal readers (range of milliseconds) (Khelifi et al., 2019; Schuster et al., 2016).

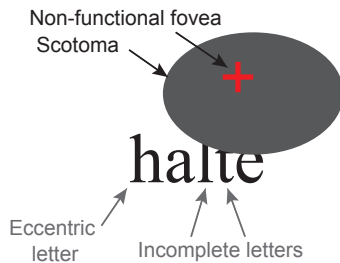


Figure 1: Partial access to text in the presence of a scotoma leads to a greater need for linguistic inference

In the current paper, we increment on our previous work by inspecting the respective effects of frequency, length and the number of neighbor words on reading speed in CFL individuals. In the following sections we will describe the methodology of our experiment (Section 2), presents its outcome results (Section 3) and discuss these results while proposing some future work directions (Section 4).

## 2. Methods

### 2.1 Participants

31 participants (18 women) were recruited from the Low-Vision Clinic of La Timone Hospital (Marseille, France). We selected our patients on three criteria: (1) presence of a bilateral central scotoma with monocular acuity of 4/10 (0.4 logMAR) or worse in their better eye; (2) absence of eye pathology other than maculopathy; (3) be fluent French readers. A total of six pathologies inducing CFL

were present in our sample: atrophic AMD (n = 15), exudative AMD (n = 4), Stargardt’s disease (n = 4), diabetic retinopathy (n = 1), cone dystrophy (n = 1) and myopic retinopathy (n = 6). Recruited participants ranged in age from 32 to 89 years.

### 2.2 Apparatus & stimuli

Sentences were displayed on an LCD monitor and presented on a window that subtended 56° x 42° at 40 cm. Sentences were aligned to the left and displayed in Courier (non-proportional font) in black on a white background. Print size was chosen optimally for each participant as the value of his/her critical print size, measured before testing with a French computerized version of MNREAD (Calabrèse et al., 2014; Calabrèse et al., 2019). Reading was monocular (eye with better visual acuity) with an appropriate correction for near vision.

### 2.3 Reading material

Reading material was created in French using ReSyf, a French lexicon with graded synonyms (Billami et al., 2018) and Lexique3, a lexical database providing word frequencies (in occurrences / million) and word neighborhood size (Coltheart’s N) of standard written and oral French (New et al., 2001). The whole material was created in three steps, in order to generate pairs of synonyms with constrained linguistic properties (*i.e.* target words) embedded within pairs of interchangeable sentences. An example (in English) is given in Table 1.

Synonym pair	<b>coast</b> <i>characters = 5 / frequency = 48 / neighbors = 3</i> <b>shore</b> <i>characters = 5 / frequency = 24 / neighbors = 13</i>
Sentence pair	<b>You should go for a walk along the [...] to relax</b> <i>44 characters / target word = n-2</i> <b>My parents have worked by the [...] for many years</b> <i>45 characters / target word = n-3</i>
Cond. 1	<b>You should go for a walk along the <b>coast</b> to relax</b>
	<b>My parents have worked by the <b>shore</b> for many years</b>
Cond. 2	<b>You should go for a walk along the <b>shore</b> to relax</b>
	<b>My parents have worked by the <b>coast</b> for many years</b>

Table 1: Reading material example

First, we created a pool of target words, by selecting 32 pairs of synonyms matching the following criteria: (1) equal number of characters within a pair, with a length comprised between 3 and 8 characters; (2) frequency ratio between a high-frequency word and its low-frequency synonym comprised between 2 and 10; (3) difference in number of orthographic neighbors between the two synonyms comprised between 5 and 10.

Second, 32 pairs of short matching sentences were created so that each word from a pair could fit within either sentence of the corresponding sentence pair. Three criteria were used: (1) within a pair, sentences could have a maximum difference of 5 characters. Overall, sentences ranged in length from 42 to 65 characters (mean ± SD = 54 ± 6); (2) within each sentence, comprised of ‘n’ words,

the target word could be located in any of these four locations: ‘n’, ‘n-1’, ‘n-2’, or ‘n-3’; (3) pairs of sentences were specifically designed to fit the single and most frequent common sense for both words of a synonym pair.

Third, we generated our final reading material by combining sentence pairs with their matching pairs of synonym. In Condition 1, the first word of a pair was assigned to the first sentence of the corresponding pair, while the second word was assigned to the second sentence, thus creating 64 full sentences. In Condition 2, the “sentence – word” pairing was reversed to create a different set of 64 full sentences. These two experimental conditions allowed us to counterbalance any potential effect of the sentence itself (structure, complexity, predictability) by randomly assigning participants to Condition 1 or 2 (Steen-Baker et al., 2017).

### 2.4 Reading procedure & experimental design

Sentences were presented within 4 blocks of 16 trials (8 pairs of sentences) each. Participants were randomly assigned to Condition 1 or 2 and read between two to four blocks, depending on their reading speed and level of fatigue. Sentences were displayed randomly within each block with non-cumulative self-paced reading, where sentences appear as a whole but with all words masked by strings of “x” (Aaronson & Scarborough, 1976; Just et al., 1982). Participants were instructed to read each sentence aloud as quickly and accurately as possible while revealing each word one at a time using keyboard presses. Reading accuracy (correct vs. incorrect) and total reading time (in seconds) were recorded for each target word.

### 2.5 Statistical analysis

Statistical analyses were carried out in R (R Core Team, 2018). Reading accuracy (*i.e.*, binary variable) was analyzed by fitting a generalized linear mixed-effects model (GLME). Reading time (*i.e.*, continuous variable) was analyzed with a linear mixed-effects model (LME). In each model, two kinds of independent variables were included: (1) characteristics of the target word, *i.e.* their frequency, their length and their number of orthographic neighbors; (2) individual characteristics of the participants, *i.e.* their age and daily reading habits. The random structure of both models included a random intercept for participants, assuming a different “baseline” performance level for each individual. Reading time and word frequency were transformed in natural logarithm (ln) units to satisfy the assumptions of parametric statistical tests (Howell, 2009; Tabachnick et al., 2007). All continuous variables were centered around their mean. Optimal model structures were assessed using the Akaike Information Criterion (AIC) and likelihood-ratio tests (Zuur et al., 2010). In the Results section, fixed-effects estimates are reported along with their p-values and 95% confidence intervals (Bates et al., 2015).

## 3. Results

### 3.1 Effect of frequency, length and neighborhood size on reading accuracy

On average, target words were read accurately 94% of the time, with individual variations ranging from 62 to 100% depending on participants. When all implemented in a single GLME model, word frequency (in

occurrences/million), word length (in number of characters) and number of orthographic neighbors showed no significant effect on accuracy.

### 3.2 Effect of frequency, length and neighborhood size on reading time

Word frequency, word length and number of orthographic neighbors were all included in a single LME model in order to assess the individual influence of each factor (when partialling out the effect of the other two) on word reading time. Fixed effects results from this model are presented in Table 2.

	Estimate	SE	t-value	p-value	95% CI
Intercept ( <i>ln(seconds)</i> )	1.308	0.156	8.39	<0.001	[0.99; 1.62]
<b>Frequency</b> ( <i>ln(occurrences/million)</i> )	<b>-0.088</b>	<b>0.010</b>	<b>-9.03</b>	<b>&lt;0.001</b>	<b>[-0.11; -0.07]</b>
<b>Number of neighbors</b>	<b>0.011</b>	<b>0.005</b>	<b>2.09</b>	<b>0.03</b>	<b>[0.001; 0.020]</b>
Length ( <i>characters</i> )	0.006	0.021	0.26	0.79	[-0.04; 0.05]
Age ( <i>years</i> )	0.003	0.006	0.45	0.66	[-0.01; 0.01]
Still reading <i>No</i>	-0.228	0.181	-1.26	0.22	[-0.60; 0.15]

Table 2: Results from the LME model; SE stands for Standard Error; CI stands for Confidence Interval. Factors showing a significant effect are highlighted in bold font.

As given by the model, average reading time when all factors are at their mean value is 3.7 seconds ( $\exp(1.31)$ ). Word frequency has a significant effect with a regression coefficient estimate of -0.088 ( $t = -9.03$ ,  $p = <0.001$ , 95% CI = [-0.11; -0.07]; Figure 2). This means that multiplying frequency (in original units) by 10 multiplies reading time (in original units) by 0.82 ( $10^{-0.088}$ ), *i.e.*, a 18 % decrease. Similarly, multiplying frequency (in original units) by 1000 (*i.e.*, from 0.5 to 500, where most of our values lie) multiplies reading time (in original units) by 0.54 ( $1000^{-0.088}$ ), *i.e.*, a 56 % decrease.

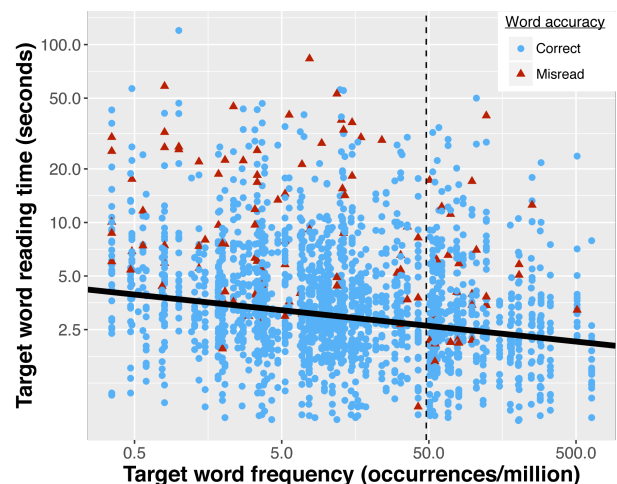


Figure 2: Scatterplot of target word reading time as a function of word frequency

The number of neighbors also has a significant effect on reading time but of smaller amplitude: when increasing neighborhood size by one neighbor, reading time is multiplied by 1.01 ( $\exp(0.011)$ ), representing a 1.01% increase (estimate = 0.011,  $t = 2.09$ ,  $p = 0.03$ , 95% CI = [0.001; 0.020]; Figure 3). In other words, increasing the number of neighbors by 10 (*i.e.*, where most of our values lie from, 0 to 10) increases reading time by 10%.

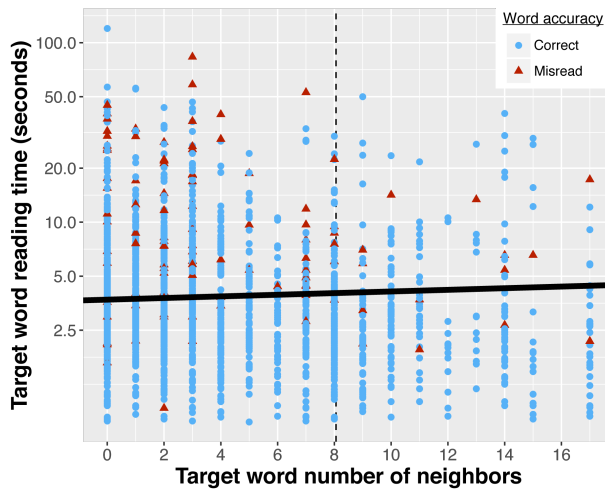


Figure 3: Scatterplot of target word reading time as a function of word number of neighbors

Word length has no significant effect on reading time (estimate = 0.006,  $t = 0.26$ ,  $p = 0.79$ , 95% CI = [-0.04; 0.05]). It is notable that low-frequency words are on average longer than high-frequency words (Kliegl et al., 1982; Rayner & Duffy, 1986). Plus, within the specific range of word length represented in our experiment (3 to 8 characters), word neighborhood size and word length covary, with longer words having less neighbors ( $r = -0.44$ ,  $t = -15$ , 95% CI = [-0.48, -0.39]). Therefore, the effects of frequency and neighborhood size reported above could have been induced by the confound between word length and each of these factors. However, including all three variables in our model enabled to rule out this possibility by partialling out the effect of word length. The age of the participant and the fact that they still maintain a daily reading activity shows no significant effect on reading time either.

#### 4. Conclusions

In this work, we investigated for the first time the respective effects of multiple lexical factors on the reading performance of low-vision individuals with CFL.

Our first result is the facilitator effect of word frequency on reading time: the more frequent in the language, the faster a word is read. This result confirms what has been reported recently with CFL readers (Calabrèse et al., 2016b; Stology et al., 2019) and is also in line with the effect reported in the literature for subjects with normal vision (Kliegl et al., 2006). However, the effect of word frequency found in our CFL population is much greater than what has been reported before for normally sighted subjects, suggesting that low-vision individuals do rely more on lexical inference to support eccentric reading (Legge et al., 2001).

Our second result is the small but significant effect of word neighborhood size on reading time with CFL: the more neighbors, the slower a word is read. For normal readers however, word neighborhood size has long been known to have a facilitator effect on word recognition (the more neighbors, the easier to identify) (Vergara-Martínez & Swaab, 2012). We hypothesize that the reversed effect we report in CFL individuals is due to the visual constraint imposed by the presence of a central scotoma, hiding portions of the text (*i.e.*, letters) and forcing to use eccentric vision. The lack of high resolution coupled with missing visual information, would lead CFL readers to confuse one word with its orthographic neighbors, creating even more uncertainty for large word neighborhood size. Here are a couple of examples using the word “salle” that has 14 neighbors in French: “*Je vois la salle de ma fenêtre*” could be confused with “*Je vois le sable de ma fenêtre*”; “*J’ai loué une salle pour demain*” could be confused with “*J’ai loué une selle pour demain*”. Although one could argue that the context normally helps the reader choose the correct statement, it is important to keep in mind that if a lot of relevant words are misidentified, there is no meaningful context to rely on, therefore leading to confusion.

Our third result is the absence of significant effect of word length on word reading time. For normal vision however, this effect is commonly reported in eye movement research as word reading time increasing with word length, mainly as a result of the increasing number of “refixations” (Kliegl et al., 2004; Vitu et al., 2001).

Trying to reduce reading deficits in AMD is a hot topic in the low-vision community. To our knowledge, the present project is the first one to propose the investigation of the linguistics aspects of this reading deficit by combining psycholinguistics, psychophysics of reading and ophthalmology. The long-term challenge of this work is to investigate what aspects of a text make it specifically complex for individuals with CFL (*e.g.*, visual, lexical, syntactic, etc.) in order to provide simplification guidelines to promote reading performance improvement for this population. For instance, the present results suggest that when simplifying a text by substituting a “complex” word with a simpler synonym, one should preferably choose a synonym with higher frequency rather than one with few neighbors, no matter what length they are. Furthermore, despite its small amplitude, the reverse effect of word neighborhood size that we report is of great fundamental interest, as it confirms that the characteristics of text complexity differs when reading with CFL and should be investigated, rather than extrapolated from results with normal vision.

In the future, other aspects of text complexity, namely syntactic and discursive, should be investigated with CFL readers to build upon this work. The long-term objective will be to provide full comprehensive guidelines to design reading aids using text simplification tailored to low vision users. Furthermore, recent advances in the domain of natural language processing should allow a large-scale implementation of such reading aids, using automated text simplification algorithms. Assistive technology could be developed (in the form of web plug-ins or dedicated software) and used by individuals with visual impairment to enhance daily reading performance on computers,



tablets, e-readers, etc. Optometrists in charge of visual readaptation in eye clinics could also benefit from this approach. In this context, text simplification would be used to train reading under the optometrist's supervision and advice. Our hope is that reducing the complexity of lexical units in text, without changing their meaning, should improve overall reading performance of low-vision readers.

## 5. Acknowledgements

This work was supported by the Fondation de France and by the Belgian FNRS. The authors thank Dr. Frédéric Chouraqui and Dr. Alban Comet for their help in recruiting the participants.

## 6. Bibliographical References

- Aaronson D, Scarborough HS (1976) Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance* 2: 56-70.
- Adelman JS, Brown GDA (2007) Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review* 14: 455--459.
- Barbu E, Martín-Valdivia MT, Ureña-López LA (2013) Open Book: a tool for helping ASD users' semantic comprehension. *Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility*: 11--19.
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67: 1--48.
- Brown JC, Goldstein JE, Chan TL, Massof R, Ramulu P, Low Vision Research Network Study Group (2014) Characterizing functional complaints in patients seeking outpatient low-vision services in the United States. *Ophthalmology* 121: 1655-62.e1.
- Bullimore BI M. (1995) Reading and eye movements in age related maculopathy. *Optometry and Vision Science*.
- Calabrèse A, Bernard J, Faure G, Hoffart L, Castet E (2014) Eye movements and reading speed in macular disease: the shrinking perceptual span hypothesis requires and is supported by a mediation analysis. *Invest Ophthalmol Vis Sci* 55: 3638-45.
- Calabrèse A, Bernard J, Faure G, Hoffart L, Castet E (2016) Clustering of Eye Fixations: A New Oculomotor Determinant of Reading Speed in Maculopathy. *Invest Ophthalmol Vis Sci* 57: 3192-202.
- Calabrèse A, Bernard J, Hoffart L, Faure G, Barouch F, Conrath J, et al. (2010) Small effect of interline spacing on maximal reading speed in low-vision patients with central field loss irrespective of scotoma size. *Invest Ophthalmol Vis Sci* 51: 1247-54.
- Calabrèse A, Liu T, Legge GE (2017) Does Vertical Reading Help People with Macular Degeneration: An Exploratory Study. *PLoS One* 12: e0170743.
- Calabrèse A, Mansfield JS, Legge GE (2019) mnreadR, an R package to analyze MNREAD data. version 2.1.3 <https://CRAN.R-project.org/package=mnreadR>.
- Coltheart M, Davelaar E, Jonasson JE, Besner D (1977) Access to the internal lexicon. In: Dornio S, editor. *Attention and performance VI*. Hillsdale, NJ: Erlbaum: pp. 535--555.
- Crossley SA, Yang HS, McNamara DS (2014) What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language* 26: 92-113.
- Fine EM, Peli E (1996) The role of context in reading with central field loss. *Optom Vis Sci* 73: 533-9.
- Howell DC (2009) *Statistical methods for psychology*. Cengage Learning.
- Inui K, Fujita A, Takahashi T, Iida R, Iwakura T (2003) Text Simplification for Reading Assistance: A Project Note. *Proceedings of the Second International Workshop on Paraphrasing* 16: 9--16.
- Just MA, Carpenter PA, Woolley JD (1982) Paradigms and processes in reading comprehension. *J Exp Psychol Gen* 111: 228-38.
- Khelifi R, Sparrow L, Casalis S (2019) Is a Frequency Effect Observed in Eye Movements During Text Reading? A Comparison Between Developing and Expert Readers. *Scientific Studies of Reading* 0: 1-14.
- Kliegl R, Grabner E, Rolfs M, Engbert R (2004) Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology* 16: 262-284.
- Kliegl R, Nuthmann A, Engbert R (2006) Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *J Exp Psychol Gen* 135: 12-35.
- Kliegl R, Olson RK, Davidson BJ (1982) Regression analyses as a tool for studying reading processes: comment on Just and Carpenter's eye fixation theory. *Mem Cognit* 10: 287-96.
- Legge GE, Mansfield JS, Chung ST (2001) Psychophysics of reading XX - Linking letter recognition to reading speed in central and peripheral vision. *Vision Res* 41: 725-43.
- Leroy G, Kauchak D (2014) The effect of word familiarity on actual and perceived text difficulty. *J Am Med Inform Assoc* 21: e169-72.
- Miller JW (2013) Age-related macular degeneration revisited--piecing the puzzle: the LXIX Edward Jackson memorial lecture. *Am J Ophthalmol* 155: 1-35.e13.
- Rayner K, Duffy SA (1986) Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Mem Cognit* 14: 191-201.
- R Core Team (2018) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. , <https://www.R-project.org/>.
- Saggion, H. (2017) *Automatic Text Simplification*. *Synthesis Lectures in Human Language Technologies*. California, Morgan & Claypool Publishers.
- Schuster S, Hawelka S, Hutzler F, Kronbichler M, Richlan F (2016) Words in Context: The Effects of Length, Frequency, and Predictability on Brain Responses During Natural Reading. *Cereb Cortex* 26: 3889-3904.
- Shardlow M (2014) A survey of automated text simplification. *Int. J. Adv. Comput. Sci. Appl. Special Issue on Natural Language Processing 2014*: doi:10.14569.
- Steen-Baker AA, Ng S, Payne BR, Anderson CJ, Federmeier KD, Stine-Morrow EAL (2017) The effects of context on processing words during sentence reading

- among adults varying in age and literacy skill. *Psychol Aging* 32: 460-472.
- Stolowy N, Calabrèse A, Sauvan L, Aguilar C, François T, Gala N, et al. (2019) The influence of word frequency on word reading speed when individuals with macular diseases read text. *Vision Research* 155: 1 - 10.
- Tabachnick BG, Fidell LS, Ullman JB (2007) Using multivariate statistics. Pearson Boston, MA.
- Taylor DJ, Edwards LA, Binns AM, Crabb DP (2018) Seeing it differently: self-reported description of vision loss in dry age-related macular degeneration. *Ophthalmic Physiol Opt* 38: 98-105.
- Vergara-Martínez M, Swaab TY (2012) Orthographic neighborhood effects as a function of word frequency: an event-related potential study. *Psychophysiology* 49: 1277-89.
- Vitu F, McConkie GW, Kerr P, O'Regan JK (2001) Fixation location effects on fixation durations during reading: an inverted optimal viewing position effect. *Vision Res* 41: 3513-33.
- Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng C, et al. (2014) Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health* 2: e106-16.
- Ziegler JC, Gala N, Brunel A, Combes M (2015) Text Simplification to increase Readability and facilitate Comprehension: a proof-of-concept pilot study. Workshop Brain and Language Cargèse.
- Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1: 3-14.

## 7. Language Resource References

- Billami M, François T, Gala N (2018) ReSyf: a French lexicon with ranked synonyms. In Proceedings of the 27th Conference on Computational Linguistics (COLING 2018), Santa Fe, USA.: 2570-2581 <https://cental.uclouvain.be/resyf/>
- New B, Brysbaert M, Veronis J, Pallier C (2007) The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics* 28: 661-677.