

Simplify or Help? Text Simplification Strategies for People with Dyslexia

Luz Rello

Web Research Group &
NLP Research Group
Universitat Pompeu Fabra
Barcelona, Spain
luzrello@acm.org

Ricardo Baeza-Yates

Yahoo! Labs Barcelona &
Web Research Group
Universitat Pompeu Fabra
Barcelona, Spain
rbaeza@acm.org

Stefan Bott
Horacio Saggion

NLP Research Group
Universitat Pompeu Fabra
Barcelona, Spain
{stefan.bott,horacio.saggion}@upf.edu

ABSTRACT

We present a user study for two different automatic strategies that simplify text content for people with dyslexia. The strategies considered are the standard one (replacing a complex word with the most simpler synonym) and a new one that presents several synonyms for a complex word if the user requests them. We compare texts transformed by both strategies with the original text and to a gold standard manually built. The study was undertaken by 96 participants, 47 with dyslexia plus a control group of 49 people without dyslexia. To show device independence, for the new strategy we used three different reading devices. Overall, participants with dyslexia found texts presented with the new strategy significantly more readable and comprehensible. To the best of our knowledge, this is the largest user study of its kind.

Keywords

Text simplification, dyslexia, readability, understandability, eye-tracking, laptop, tablet, smartphone.

1. INTRODUCTION

Dyslexia is a reading disability which affects from 10 to 17.5% of the population in the U.S.A. [20] and from 8.6 to 11% of the Spanish speaking population [29]. This condition makes accessing written information more difficult, particularly in the Web.

Previous findings have shown that people with dyslexia specifically encounter problems with complex words, such as long or infrequent words [10, 19, 30, 34]. Therefore, applying automatic lexical simplification strategies, that is, substituting complex words by simpler synonyms, could make texts easier to read and understand for people with dyslexia. However, previous applications for people with dyslexia [7, 17, 22] modify only the text presentation but not its content.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

W4A2013 - Technical May 13-15, 2013, Rio de Janeiro, Brazil. Co-Located with the 22nd International World Wide Web Conference.
Copyright 2013 ACM 978-1-4503-1844-0 ...\$5.00.

With this idea in mind, we used an automatic lexical simplification system, LexSiS [4], with two different strategies: one that substitutes each complex word for a simpler one and another one that allows the user to see several synonyms for a complex word when needed. For the later strategy, we tested implementations in different devices (laptop, tablet, smartphone) to make sure that the strategy is device independent, integrating it in *DysWebxia* [31], an application that helps people with dyslexia to read text in the Web.

The goal of this paper is to evaluate the impact of our lexical simplification strategies on the readability,¹ understandability,² and easiness ratings for native Spanish speakers with dyslexia. We study readability and understandability separately because readability has been found to be independent of comprehension for people with dyslexia [28].

To the best of our knowledge, this is the first time that an automatic lexical simplification system is evaluated for end-users with dyslexia. In addition, this is the largest user study of its kind. Indeed, for this study, 96 people (47 with dyslexia) participated in our experiments, which combined eye-tracking, questionnaires, and the use of different devices. This paper presents the following main contributions:

- An evaluation of an automatic lexical simplification system, LexSiS, replacing a complex word with the best simpler synonym, SUBSBEST, analyzing its impact on readability, comprehension and easiness ratings in comparison with the original text without lexical simplification and a gold standard manually simplified.
- A new strategy, SHOWSYNS, which adapts LexSiS, and allows users to interactively request simpler synonyms for complex words.
- An evaluation of SHOWSYNS using three different devices (laptop, tablet, smartphone), to analyze the impact on comprehension and easiness ratings in comparison with SUBSBEST, the original text, and the gold standard.
- That participants with dyslexia found that texts presented with the new strategy were significantly more

¹It refers to legibility, the ease with which text can be read.

²It refers to comprehension, the ease with which text can be understood.

readable and comprehensible while participants without dyslexia found it significantly more comprehensible.

Our findings can have an impact on interactive systems that rely on text since applying our suggested lexical simplification strategy, these systems could make texts more appealing for people with dyslexia, which may lower the subjective barrier of engaging in text reading activities.

Our results not only impact a relative large population but also are extensible to other groups and to general usability problems, since dyslexia-related difficulties are shared by other people with special needs [16] and dyslexia symptoms are common to varying degrees among most people [13].

The rest of the paper is organized as follows. Next section presents the language related-problems of dyslexia while section 3 covers related work. In Section 4 we present the two simplification strategies and in Section 5 the evaluation methodology. Section 6 presents the results, which are subsequently discussed in Section 7. Conclusions and future challenges are given in Section 8.

2. LANGUAGE PROBLEMS OF DYSLEXIA

Dyslexia is a neurological reading disability which is characterized by difficulties with accurate and/or fluent word recognition as well as by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unrelated to other cognitive disabilities. Secondary consequences include problems in reading comprehension and reduced reading experience that can impede vocabulary growth and background knowledge [21].

People with dyslexia encounter problems, not only with some text presentation conditions, such as small font size [24, 11], but also with language-related conditions [10, 26]:

- (a) Phonology: Irregular words, *vase*³; homophonic words or pseudo homophonic words, *weather* and *whether*; and foreign words.
- (b) Orthography: Orthographically similar words, *addition* and *audition*; number and letter recognition and recollection; poor spelling, such as letter reversals, *trail* for *trial*.
- (c) Morphology: Derivational errors, *discomfortable*.
- (d) Lexicon and Syntax: New words, *fantabulous*; pseudo-words and non-words,⁴ *happisfaction*; less frequent words, *pristine*; long words, *prestidigitation*; word additions and omissions; word recognition and recollection; substitutions of functional words,⁵ *of* by *for*; confusions of small words, *in* by *is*; and error recognition.
- (e) Discourse: Fixation problems; punctuation recognition; long sentences and paragraphs; and poor comprehension.

³Words with no consistent correspondence between grapheme and phoneme, e.g. *vase* pronounced as /vāz/.

⁴A non-word is a word that has no meaning, is not known to exist, or is disapproved.

⁵Functional words are words that have little lexical meaning, but instead serve to express grammatical relationships with other words within a sentence.

3. RELATED WORK

Given that dyslexia is a disability that affects language, we can assume that accessibility can be approached not only from the text presentation, but also from the text content. Even though, the use of complicated language has been extensively pointed out as one of the key problems for this target group [24, 11], all the existing applications at the moment only alter the design of the text [17, 7, 22], but not its content. To the best of our knowledge, this is the first attempt, to design and evaluate automatic text simplification strategies for people with dyslexia.

Related to our contributions, we divide related work in three areas: (a) work on natural language processing about lexical simplifications algorithms, (b) work on experimental psychology about how people with dyslexia read and comprehend under different language conditions, and (c) accessibility studies about people with dyslexia.

Natural language processing and lexical simplification: Automatic text simplification is an NLP task that transforms a text into an equivalent which is easier to read than the original, preserving the original meaning.

Lexical simplification is a kind of text simplification which aims at the substitution of words by simpler synonyms. Lexical simplification requires, at least, two things: a way of measuring lexical complexity and a way of finding synonyms. Most of the approaches to lexical simplification use word frequency [6, 12, 2] and word length [3] as a measure of lexical complexity. To find appropriate word substitutions they use different resources such as WordNet [6], dictionaries [12], thesaurus and lexical ontologies [2], and synonym dictionaries [3].

Experimental psychology and word processing: One of the most studied language conditions is the effect of frequent words and long words on readability and comprehension of people with dyslexia because word frequency and word length are related to the word's processing time [27], and because people with dyslexia specifically encounter problems with less frequent words and long words [19, 30, 34, 36]. Since our lexical simplification strategies are based on frequency and length we give an special attention to these studies.

Using eye-tracking, Hyona *et al.* [19] show that low frequency and long words present longer gaze durations and more re-inspections in both groups. Also Rello *et al.* [30] find that frequent words improve readability and short words improve understandability for people with dyslexia. Also, Rüsseler *et al.* [34] show that it takes more time to recognize infrequent words and this recognition performance is lower in readers with dyslexia. Simmons and Singleton [36] measured comprehension of people with dyslexia who performed significantly poorer on the inferential questions.

Accessibility studies about people with dyslexia: If we compare our study with other accessibility studies, our study differs in its goal and has the greatest number of participants with dyslexia. In [1], 10 participants tested Web navigation using semi structured interviews. In [23], 27 participants did assignments after reading texts with different presentations. In [9], interviews, questionnaires, log sheets and focus groups are used to explore user behavior and usability issues relating to the use of web-based resources by people with disabilities (9 participants with dyslexia); and in [37] 6 participants performed tasks in a website to explore its design. Hence, our number of participants is much larger.

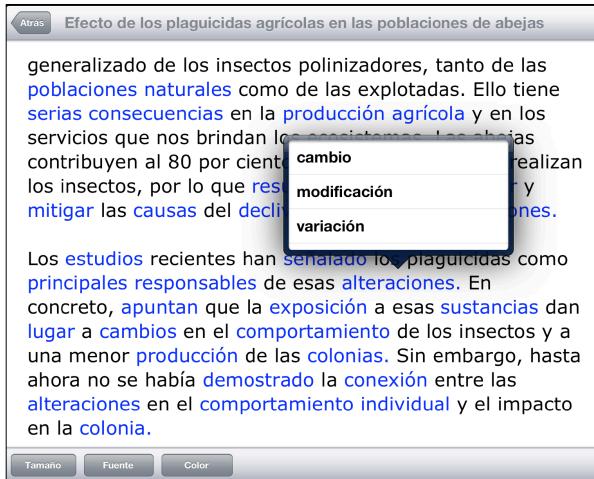


Figure 1: SHOWSYNS mock-up for iPad.

4. SIMPLIFICATION STRATEGIES

In this work we evaluate two lexical simplification strategies based on the LexSiS algorithm [4]. LexSiS is the first system for the lexical simplification of Spanish text and is being developed in the context of the Simplext project [35]. It aims to improve text accessibility for people with cognitive impairments. The performance of LexSiS is similar to the state of the art of other lexical simplifications systems for English, overcoming the baseline of substituting a word by the most frequent synonym.⁶

The first lexical simplification strategy substitutes complex words by simpler synonyms. We call this strategy SUBSBEST, since substitution is the original goal of LexSiS. The second simplification strategy is called SHOWSYNS and instead of substituting a word, provides simpler synonyms for a complex word.

For instance, in the text “*responsables de estas alteraciones*” (*responsible for these alterations*), SUBSBEST would substitute the plural of the word *alteración* (*alteration*) by the plural of the word *cambio* (*change*), while SHOWSYNS would pop-up up to three synonyms if the user chooses to do so (See Figure 1).

LexSiS uses (i) a word vector model to find possible substitutes for a target word using available resources such as the free OpenThesaurus and a corpus of Spanish documents from the Web, and (ii) a simplicity computation procedure grounded on a corpus study and implemented as a function of word length and word frequency.

LexSiS works in two steps: First it selects a set of synonyms and then it ranks those synonyms according to a simplicity criterion. To select potential synonyms, the system consults OpenThesaurus for Spanish.⁷ The following is an example of an entry in OpenThesaurus:

- (a) hoja|3
 - |acero|espada|pual|arma blanca
 - |bráctea|hojilla|hojuela|bractéola
 - |lámina|plancha|placa|tabla|rodaja|película|chapa|lata|viruta|loncha|lonja|capa|laminilla

⁶Out of the synonyms that LexSiS generates, 65% are simpler than the target word [4].

⁷There is no English version, see <http://openthes-es.berlios.de>.

The word *hoja* is semantically ambiguous and can mean ‘blade’, ‘leaf’ or ‘layer’. The first line of the entry represents the target word and states that there are three different meanings. The three lines that follow list synonyms for the three word meanings. For each word to be substituted, LexSiS first uses a distributional semantic model to identify the list with the correct meaning. For that LexSiS extracts the typical contexts of each word using a 9-word window (4 words, to both, the left and the right side of the target word) from an 8 million word corpus of Spanish Web news. LexSiS uses this model to construct vectors which represent a given word meaning by aggregating the vectors of all words listed for this meaning. Then it extracts a vector for the target context in which we want to replace a given complex word, using again a 9-word window, and compares it to the vector for each word meaning. The word meaning whose vector has the minimal cosine distance to the context vector is taken to be the correct sense.

Once selected the word sense, LexSiS assigns a simplicity score to each word, combining word frequency and word length. LexSiS also applies a series of filters: (i) it does not try to simplify already frequent words, (ii) it does not use words with a frequency score which is only slightly higher than the score for the original word, and (iii) a simplicity score difference threshold, that is, it also discards words whose vector has a high distance to the context vector (which indicates that it probably does not fit into the given context). The synonym with best simplicity score is then used for the SUBSBEST strategy.

In SHOWSYNS the way to detect more complex words differs from SUBSBEST. It detects more complex words because we disable the simplicity score difference threshold including words that have a lower simplicity score than the original (more details can be found in [4]). The rationale behind is that substituting a word in a text may damage the meaning of the text if the substitution is not accurate enough. Since in SHOWSYNS there are no substitutions, we can present more synonyms to the user. The list of synonyms used in SHOWSYNS is the list of words with the highest simplicity score and if the list contains more than three synonyms, only the three top scoring alternatives are shown. Moreover, SHOWSYNS only shows these synonyms upon the user’s request. Depending on the interaction methods, the user has to tap on (touch screen) or click (mouse) a word to open the synonyms pop-up.

5. METHODOLOGY

To study the effect of the two text simplification strategies, we conducted an experiment with 96 participants (47 with dyslexia) using eye-tracking, questionnaires, and different reading devices. Each of them had to read one text that was either in its original state, automatically simplified by SUBSBEST or SHOWSYNS, or manually simplified (gold standard).

5.1 Design

The lexical simplification strategy serves as independent variable with four levels:

- [ORIG]: the text without any alterations,
- [SUBSBEST]: automatically simplified text using the best substitution computed by LexSiS,

- [SHOWSYNS]: a reading assistant with on-demand synonyms presentation based on LexSis, and
- [GOLD]: a manually simplified text serving as gold standard.

We used a between-subject design, that is, each participant contributed to one condition only. For the ORIG, SUBSBEST, and GOLD conditions we used an eye-tracker to record the readings. For SHOWSYNS was not possible to use the eye-tracker as the interaction needed for this strategy was not available. Then, we could not record the readings for this condition. Hence, for SHOWSYNS we implemented mock-ups on three different devices: smartphone, tablet, and laptop. In this way we made sure that our measures were device independent. To cancel out possible effects of a device, we rotated the use of the devices amongst participants.

For quantifying readability and understandability, we took the following dependent measures coming from the eye-tracker and the questionnaires:

Reading Time: Shorter reading durations are preferred to longer ones since faster reading is related to more readable texts [38]. Therefore, we use *Reading Time*, that is, the time it takes for a participant to completely read one text, as a measure of readability. This measure is extracted from the eye-tracking data.

Fixation Duration: When reading a text, the eye does not move contiguously over the text, but alternates saccades and visual fixations, that is, jumps in short steps and rests on parts of the text. *Fixation duration* denotes how long the eye rests still on a single place of the text. Fixation duration has been shown to be a valid indicator of readability. According to [27], shorter fixations are associated with better readability while longer fixations can indicate that processing loads are greater. Hence, we use fixation duration as a readability measure in addition to the reading time.

Comprehension Score: To measure text comprehension, we used multiple-choice questions with three possible choices, one correct choice, one partially correct choice, and one wrong choice. To compute the text comprehension score, the choices counted 100%, 50%, and 0%, respectively.

Easiness Rating: In addition to the quantitative measures, we asked the participants to rate three items on a five-point Likert scale, regarding how easy the text is to read, to understand, and to remember.

5.2 Participants

To check for differences between people with and without dyslexia, we recruited two groups of participants. First, 47 Spanish speakers (28 females, 19 males) with a confirmed diagnosis of dyslexia (group D). They were asked to bring their diagnoses to the experiment to guarantee that dyslexia was diagnosed in an authorized centre or hospital.⁸ Their ages ranged from 13 to 50, with a mean age of 24.36 years ($s = 10.19$).

A control group of 49 Spanish speakers (29 females, 20 males) without dyslexia also participated in the study (group N). Their ages ranged from 13 to 54, with a mean age of 28.24 years ($s = 7.24$). That is, overall, we had 96 participants (57 females, 39 males).

⁸In the Catalonian protocol of dyslexia diagnosis [8], the different kinds of dyslexia, extensively found in literature, are not considered.

Except from 3 participants with dyslexia and 2 without dyslexia, all of the participants were attending school or high school (18 participants with dyslexia and 16 participants without dyslexia), or they were studying or had already finished university degrees (26 participants with dyslexia and 31 participants without dyslexia).

5.3 Materials

To study the effects of the simplification strategies, we need to study them in context, that is, as part of a text. The rationale behind this is that readability and understandability pertain to longer segments of texts [18]. To isolate the effects of the different strategies, the texts need to be comparable in complexity. Hence, in this section, we describe how we designed the texts that were used in this study.

5.3.1 Base Texts

As basis for our tests, we picked two texts from a scientific dissemination magazine called *Investigación y Ciencia*, the Spanish edition of *Scientific American*. To meet the comparability requirements among the texts, we adapted the base texts maintaining the original text as much as possible. We matched the readability of the texts by making sure that the parameters commonly used to compute readability [14] yielded the same or similar values:

- Within each experiment, the texts use the same genre, scientific articles.
- They are about similar topics: reports from the *Nature* journal on new findings, one about the decline of the population of bees and another about a type of stars. In the following, we denote these texts with *Star* and *Bee*.
- They have the same number of words: 302 words.
- They have a similar discourse structure: title, the first paragraph presents a summary of the article, the second paragraph an introduction of the finding, the third paragraph explains the background of the finding, and the last paragraph explains more details of the findings.
- They contain the same number of sentences: eleven.
- They do not contain acronyms or numerical expressions since numerical expressions are processed differently by people with and without dyslexia [32]. Both texts have the same foreign word (*Nature*).

5.3.2 Lexical Simplifications

The base texts, *Star* and *Bee*, were altered by human experts who performed lexical simplification on the text, and by our systems giving as a result eight texts to be used in our experiments, two for each case: [ORIG], [SUBSBEST], [SHOWSYNS], and [GOLD]. All the texts have a similar word length, with an average length ranging from 4.89 to 5.50 letters.

The SUBSBEST strategy made the same numbers of substitutions in both texts: 34. SHOWSYNS provided 100/110 synonyms for 50/55 words in *Star/Bee*, respectively. For the gold standard, two language experts substituted 40/44 words in *Star/Bee*, respectively. Examples of these alterations are shown in Table 1. Please see the Appendix for the complete lists of lexical simplifications performed for the *Bee* text.

ORIG	SUBSBEST	SHOWSYNS	GOLD
alteración	cambio	cambio, modificación, variación	cambio
'alteration'	'change'	'change, modification, variation'	'change'

Table 1: Example of lexical simplifications.

5.3.3 Text Presentation

Since for people with dyslexia the presentation of a text has an effect on the readability [33], we followed the recommendations of previous research. As font type we chose *Arial, sans serif* and left-justified text [5]. Each line did not exceed 62 characters/column, the font size was 20 points, and the colors used were black font on creme background and an almost black font (10% grey scale) on white background [33].

5.3.4 Comprehension Questionnaires

Each of the questionnaires was composed of three multiple-choice inferential questions, that is, questions that require a deep understanding of the content because the question cannot be answered straight from the text (see example in Figure 2). We made sure that the questions did not include a synonym that may benefit a particular strategy.

5.4 Equipment

The eye-tracker used was a Tobii T50, which has a 17-inch TFT monitor with a resolution of 1024×768 pixels. The time measurements of the eye-tracker has a precision of 0.02 seconds. The eye-tracker was calibrated for each participant and the light focus was always in the same position. The distance between the participant and the eye-tracker was constant (approximately 60 cm. or 24 in.) and controlled by using a fixed chair.

Now we detail the devices used for SHOWSYNS. As smartphone we used a Samsung Galaxy Ace S5830 with a 3.5 inches touch screen and a resolution of 320×480 pixels running the Android operating system; for the tablet we used an iPad 2 with a 9.7 inches multi touch screen and a resolution of 1024×768 pixels running the iOS operating system; and for the laptop we used a MacBook Air with a 11 inches screen and a resolution of 1366×768 pixels running the Mac OS X 10.7.4 operating system. We used the native Web browsers for the first two devices and Firefox 16.0.2 for the laptop.

5.5 Procedure

The sessions were conducted at Universitat Pompeu Fabra and lasted around 30 minutes. In each session, the participant was alone with the interviewer (first author) in the quiet room prepared for the study, and performed the following five steps.

First, we began with a questionnaire that was designed to collect demographic information. Second, to assure the engagement of the participant while reading, s/he chose the text to read. For this, on a piece of paper, we presented the participant the title and a brief summary of both scientific articles, *Star* and *Bee*, so the participant could select the more appealing text. Third, the participants were asked to read the texts in silence. Next, when they finished, the participants were asked to complete the comprehension tests,

El texto trata sobre ‘The text is about.’

- Un artículo científico sobre el origen de la luz infrarroja. ‘About a scientific article about the origin of the infrared light’.
- La desestimación de dos teorías sobre las estrellas errantes. ‘On the dismissal of two theories of the wandering stars’.
- Las primeras galaxias del universo y la luz infrarroja que desprenden. ‘On the first galaxies in the universe and their infrared light’.

Figure 2: Example of an inferential question.

which were issued on paper. Finally, each participant was asked to provide his/her easiness ratings. After finishing the experiment, some participants (14 with dyslexia and 14 without dyslexia) wanted to read the other scientific text and so they undertook the experiment again reading that text.

6. RESULTS

In this section we present the analyses of the data from the eye-tracker (reading time and fixation duration), the comprehension tests, and the easiness ratings. For [ORIG] we had 16 samples for group D and 15 for group N; for [SUBSBEST] we had 16 samples for group D and 17 for group N; for [SHOWSYNS] we had 14 samples for group D and 14 for group N; and for [GOLD] we had 15 samples for group D and 14 for group N.

First, we analyzed the differences among groups and then the effect of the conditions within each group.

A Shapiro-Wilk test showed that the datasets were normally distributed. Also, a Barlett’s test showed that they were homogeneous. Hence, for each experiment we used:

- Student’s independent two tailed t-test to show effects on reading time, fixation duration, and comprehension score among groups D and N.
- One-way analysis of variance (ANOVA) to show effects of the conditions on reading time, fixation duration, and comprehension score within groups.
- Kruskal-Wallis and Mann-Whitney tests for post-hoc comparison to show effects on the easiness participants’ ratings.
- Pearson correlation coefficient to assess the relationship between groups and the comparisons between the quantitative data (reading time, fixation duration and comprehension score) with the qualitative data (easiness ratings).

6.1 Reading Time

Considering all the conditions, we found a significant difference between the groups regarding reading time ($t(67.657) = 4.417, p < 0.001$). Participants with dyslexia had significantly longer reading times ($\mu = 132.08, s = 51.17$ seconds) than the participants without dyslexia ($\mu = 95.25, s = 26.02$ seconds).

We did not find a significant effect of any of the conditions on reading time in group D ($F(2, 44) = 0.174, p = 0.841$) or in group N ($F(2, 43) = 2.247, p = 0.117$). Also,

Simplification Condition	Reading time ($\mu \pm s$ in sec.)		Fixation duration ($\mu \pm s$ in sec.)		Comprehension ($\mu \pm s$ in %)	
	Group D	Group N	Group D	Group N	Group D	Group N
[ORIG]	134.79 \pm 63.03	90.24 \pm 20.33	0.24 \pm 0.07	0.20 \pm 0.03	57.00 \pm 47.39	63.89 \pm 45.99
[SUBSBEST]	135.77 \pm 53.65	105.77 \pm 32.12	0.24 \pm 0.04	0.20 \pm 0.04	50.00 \pm 45.83	50.83 \pm 47.38
[SHOWSYNS]	—	—	—	—	61.90 \pm 43.91	63.10 \pm 42.85
[GOLD]	125.86 \pm 37.16	89.08 \pm 21.44	0.24 \pm 0.04	0.20 \pm 0.03	50.19 \pm 42.76	65.39 \pm 45.54

Table 2: Means of the reading time and fixation duration in seconds and the comprehension score (%).

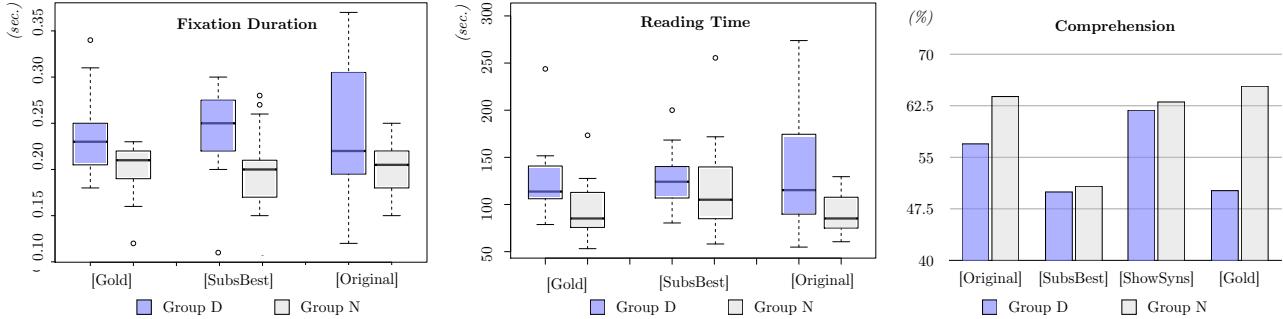


Figure 3: Average reading time and fixation duration in seconds, and average of the comprehension score.

there was a strong positive correlation between groups ($r = 0.625$). Both groups read faster under the same condition, [GOLD]. In Table 2 and in Figure 3 we show the averages of the reading times.

6.2 Fixation Duration

Pooling the data together for all the conditions, there was a significant difference between the groups fixation duration ($t(77.161) = 4.078$, $p < 0.001$). Participants with dyslexia had significantly longer fixation times ($\mu = 0.24$, $s = 0.05$ seconds) than the participants without dyslexia ($\mu = 0.20$, $s = 0.03$ seconds).

We did not find a significant effect of any of the conditions on fixation time in group D ($F(2, 44) = 0.062$, $p = 0.94$) or in group N ($F(2, 43) = 0.101$, $p = 0.904$). Again, there was strong positive correlation between groups ($r = 0.994$). See Table 2 and Figure 3 for the average of fixation durations.

6.3 Comprehension Score

Considering all the conditions, participants with dyslexia answered less questions correctly ($\mu = 54.5\%$, $s = 45.0\%$) than participants without dyslexia ($\mu = 59.9\%$, $s = 45.9\%$). However, the difference between the groups was not statistically significant ($t(389.355) = -1.180$, $p = 0.239$).

We did not find a significant effect of text simplification on the comprehension score in group D ($F(3, 186) = 0.741$, $p = 0.529$) or in group N ($F(1, 198) = 1.163$, $p = 0.325$). Again, there was a strong positive correlation between groups ($r = 0.429$). See Table 2 and Figure 3 for the averages of the comprehension scores.

6.4 Easiness Ratings

There was no correlation between both groups on the ratings about the understandability of the text ($r = -0.085$), and there was a small correlation between both groups on the readability ($r = 0.241$) and the ease of remembering the text ($r = 0.160$). In Figure 4 we show the histograms of the

easiness ratings and in Table 3 we show their averages.

For the participants with dyslexia, we found a significant effect of the simplification strategy on readability ratings ($H(3) = 8.275$, $p = 0.041$). Pairwise comparisons showed that the participants found [SHOWSYNS] significantly easier to read than [GOLD] ($p = 0.034$) and [ORIG] ($p = 0.015$).

For group D, we found a significant effect of the simplification strategy on understandability ratings ($H(3) = 12.197$, $p = 0.007$). Pairwise comparisons showed that the participants found [SHOWSYNS] significantly easier to understand than [ORIG] ($p = 0.001$) and [SUBSBEST] ($p = 0.013$).

For the participants without dyslexia, we found a significant effect of the simplification strategy on understandability ratings ($H(3) = 9.595$, $p = 0.022$). Pairwise comparisons showed that the participants found [SUBSBEST] significantly more difficult to understand than [ORIG] ($p = 0.003$), [SHOWSYNS] ($p = 0.047$) and [GOLD] ($p = 0.049$).

For group N, we found a significant effect of the simplification strategy on memorability ratings ($H(3) = 9.020$, $p = 0.029$). Pairwise comparisons showed that in the [SUBSBEST] condition, the participants found texts significantly more difficult to remember than in the [GOLD] condition ($p = 0.003$).

6.5 Comparisons

Comparing our quantitative and qualitative data we found that there is a medium positive correlation of the easiness ratings and the comprehension score for group D ($r = 0.459$) and a strong positive correlation for group N ($r = 0.928$). The options with a higher comprehension score, [SHOWSYNS] and [ORIG], were also perceived as more comprehensible conditions by both groups. For readability in group D we found a strong positive correlation between reading time and easiness rating for readability ($r = 0.637$) and a medium positive correlation between fixation duration and easiness rating for readability ($r = 0.469$). For group N, we found strong negative correlations between the easiness rating for readability and reading time ($r = -0.999$) and fixation duration

Simplification Condition	Group D (ave. \pm std. dev.)			Group N (ave. \pm std. dev.)		
	Readability	Understandability	Memorability	Readability	Understandability	Memorability
[ORIG]	3.65 \pm 0.61	3.24 \pm 0.83	3.29 \pm 0.92	4.22 \pm 0.65	4.44 \pm 0.51	3.89 \pm 0.90
[SUBSBEST]	3.88 \pm 0.49	3.65 \pm 0.70	3.59 \pm 0.51	3.90 \pm 0.72	3.80 \pm 0.62	3.70 \pm 0.47
[SHOWSYNS]	4.29 \pm 0.73	4.36 \pm 0.74	4.14 \pm 0.77	4.36 \pm 0.74	4.29 \pm 0.73	4.14 \pm 0.77
[GOLD]	3.63 \pm 0.89	3.75 \pm 1.06	3.44 \pm 1.03	4.25 \pm 0.68	4.25 \pm 0.68	4.38 \pm 0.72

Table 3: Results of the average easiness ratings.

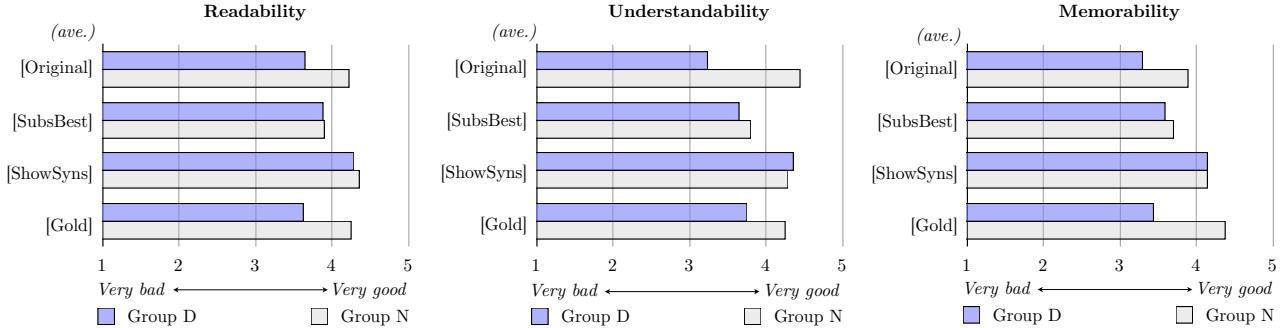


Figure 4: Average of the easiness ratings on readability, understandability, and memorability.

($r = -0.554$). Regarding readability, people with dyslexia perceived as more readable the options that they read faster. However, for people without dyslexia we found the opposite situation, the options that they read faster were perceived as the less readable by group N.

7. DISCUSSION

In this section, we discuss the results, first among groups, and then within each group for each of the measures.

Groups: In general, participants without dyslexia read significantly faster and had shorter fixation durations than participants with dyslexia. However, no significant differences were found in the comprehension of the texts between the groups. The analysis of the quantitative data shows strong positive correlations between the groups, that is, both groups read faster and understood better for the same conditions. However, both groups did not agree or only slightly agree in their easiness ratings of the simplification strategies. The objectively more readable options, [GOLD] and [SHOWSYNS], were perceived as more readable by people with dyslexia and less readable by people without dyslexia. The objectively more comprehensible options, [SHOWSYNS] and [ORIG], were perceived as more comprehensible by both groups. Regarding the differences between the groups, our quantitative results for readability are consistent with other eye-tracking studies that found statistical differences among the two populations [15]. However, our comprehension results are not consistent with [36] because our participants with dyslexia did not have a significantly poorer understanding of the texts using inferential items.

Readability: As expected, the lowest reading and fixation durations were observed for the manual simplifications, [GOLD]. However, this condition does not lead to significant faster readings for any of the groups. Previous findings [19, 30] have shown that participants with dyslexia read significantly faster and have significantly shorter fixation durations when reading texts with more frequent words. One

possible reason for not finding significant effects in our conditions is that the lexical simplification was performed on texts published in the Web, instead of using manually designed texts, which allows to control more variables related to word complexity, such as frequency and length [30].

Another possible explanation is that only a relatively small percentage of the words in the text was modified. For instance, with [SUBSBEST] only 10% of the words in the test were substituted. This relatively small text variation makes it difficult to identify existing significant effects, compared to previous studies which only focussed on target words [19]. We analyzed the eye fixation duration and the reading time of the whole text and not target words only as in Hyona *et al.* [19] because we aim to measure text readability and the readability is related to longer text segments [18].

Comprehension: The tested lexical simplification strategies had no positive effect on the comprehension of the text. In fact, it seems that the modification of the text is counterproductive for improving comprehension because the best scores for it are using [SHOWSYNS] for group D and [ORIG] for group N, that is, options which do not include any lexical substitutions in the text. For participants with dyslexia, the possibility of quick access to simpler synonyms may improve the comprehension score. One possible reason to these results is that the comprehension of the text depend on longer segments of texts [18], that is, it does not depend on single words but on the relations between words. One of the main learning strategies for understanding new words is paying attention to the context of the word. Even if [SUBSBEST] substitutes words by a synonym that also appear in that context with high frequency, the resulting text may lead to misunderstandings or strange word combinations. For instance, *las poblaciones explotadas de abejas*, ‘the exploited populations of bees’ does not mean the same as *los pueblos explotados de abejas*, ‘exploited people of bees’.

Easiness Ratings: Within groups, the only significant effects were found on the easiness ratings. Participants with dyslexia found texts with [SHOWSYNS] significantly more readable than the original text and the gold standard; and easier to understand than the original text and than using [SUBSBEST]. On the other hand, participants without dyslexia found [SUBSBEST] significantly more difficult to comprehend than the other options; and more difficult to remember than the gold standard. The correlations between the quantitative results and the easiness ratings show that people with dyslexia perceived as more readable and comprehensible the options that they actually read faster and understood better. Surprisingly, people without dyslexia, perceived as the most readable and comprehensible, the options which took them longer to read and where the comprehension was poorer.

8. CONCLUSIONS

We tested the effect of two lexical simplification strategies on readability, comprehension and easiness ratings. We did not find significant effects of the lexical simplification strategy on readability and comprehension. But, we found significant effects on the participants' easiness ratings. For the participants without dyslexia, automatic lexical simplification by LexSiS (SUBSBEST) caused the resulting texts to be subjectively more difficult to understand than all other strategies, and more difficult to remember than the manually simplified text. Participants with dyslexia found texts presented with SHOWSYNS significantly more understandable than texts modified by SUBSBEST, and more readable than the original text and the gold standard. Therefore, a system like SHOWSYNS which displays the synonyms on demand without modifying the text may benefit the comprehension of people with dyslexia. These results indicate that the current state-of-the-art of automatic lexical simplification through word substitution might negatively affect the reading experience.

On the other hand, students with dyslexia can easily run into a vicious circle where they read less because they are slower readers and reading less leads them to staying on a lower reading proficiency level. Therefore, anything which might help them to subjectively perceive reading as being easier, can potentially help them to avoid this vicious circle, even if no significant improvement in readability can be demonstrated. Therefore, these findings can have an impact on interactive systems that rely on text as the main information medium, such as web browsers, PDF viewers, or eBook readers. By applying our suggested lexical simplification strategy, namely offering simpler synonyms on demand, these systems could make texts more appealing and easier to understand for people with dyslexia. In addition, our results may also imply that the user interface might be more important than lexical simplification, as just the perception that on-demand help is available, can make reading less challenging.

Future work includes the refinement of the SHOWSYNS algorithm for the specific characteristics of people with dyslexia. In particular, we will tailor the detection of lexical complexity considering writing errors of people with dyslexia and orthographic and phonetic similarity of words since these language features makes words more difficult to recognize for people with [10] and also without dyslexia [25].

Acknowledgments

We thank Clara Bayarri, Roberto Carlini and Gaurang Kanvinde for their brilliant –last minute– mock-up implementations. We also thank Mari-Carmen Marcos for her assistance with the eye-tracker hardware, Martin Pielot for his help with statistics, and Sara Rodríguez Fernández for reviewing the questionnaires. We also thank Susana Bautista and Mónica López Ferrado for the dissemination of the participation call for this study, thanks to which we had many anonymous persons to whom we are also indebted.

9. REFERENCES

- [1] A. Al-Wabil, P. Zaphiris, and S. Wilson. Web navigation for individuals with dyslexia: an exploratory study. *Universal Access in Human Computer Interaction. Coping with Diversity*, pages 593–602, 2007.
- [2] S. M. Aluísio and C. Gasperin. Fostering digital inclusion and accessibility: the PorSimples project for simplification of Portuguese texts. In *Proc. NAACL HLT '10 Workshop YIWCALA '10*, pages 46–53, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
- [3] S. Bautista, C. León, R. Hervás, and P. Gervás. Empirical identification of text simplification strategies for reading-impaired people. In *European Conference for the Advancement of Assistive Technology*, Maastricht, the Netherlands, 2011.
- [4] S. Bott, L. Rello, B. Drndarevic, and H. Saggion. Can Spanish be simpler? LexSiS: Lexical simplification for Spanish. In *Proc. Coling '12*, Mumbai, India, December 2012.
- [5] British Dyslexia Association. Dyslexia style guide, January 2012. <http://www.bdadyslexia.org.uk/>.
- [6] J. Carroll, G. Minnen, D. Pearce, Y. Canning, S. Devlin, and J. Tait. Simplifying text for language-impaired readers. In *Proc. EACL '09*, pages 269–270, 1999.
- [7] ClaroSoftware. ClaroSoftware, 2012 2012. <http://www.clarosoftware.com/index.php?cPath=348>.
- [8] Col·legi de Logopèdes de Catalunya. *PRODISCAT Protocol de detecció i actuació en la dislèxia. Àmbit Educativo*. Departament d'Ensenyament de la Generalitat de Catalunya, 2011.
- [9] J. Craven and H. Booth. Putting awareness into practice: practical steps for conducting usability tests. *Library review*, 55(3):179–194, 2006.
- [10] F. Cuetos and F. Valle. Modelos de lectura y dislexias (Reading models and dyslexias). *Infancia y Aprendizaje (Infancy and Learning)*, 44:3–19, 1988.
- [11] V. de Santana, R. de Oliveira, L. Almeida, and M. Baranauskas. Web accessibility and people with dyslexia: a survey on techniques and guidelines. In *Proc. W4A '12*, page 35. ACM, 2012.
- [12] S. Devlin and G. Unthank. Helping aphasic people process online information. In *Proc. ASSETS '06*, pages 225–226. ACM, 2006.
- [13] M. Dixon. Comparative study of disabled vs. non-disabled evaluators in user-testing: dyslexia and first year students learning computer programming. *Universal Access in Human Computer Interaction*.

- Coping with Diversity*, pages 647–656, 2007.
- [14] B. Drndarevic and H. Saggion. Towards automatic lexical simplification in Spanish: an empirical study. In *Proc. NAACL HLT '12 Workshop PITR '12*, 2012.
 - [15] G. Eden, J. Stein, H. Wood, and F. Wood. Differences in eye movements and reading problems in dyslexic and normal children. *Vision Research*, 34(10):1345–1358, 1994.
 - [16] L. Evett and D. Brown. Text formats and web design for visually impaired and dyslexic readers-clear text for all. *Interacting with Computers*, 17:453–472, 2005.
 - [17] P. Gregor, A. Dickinson, A. Macaffer, and P. Andreasen. Seeword: a personal word processing environment for dyslexic computer users. *British Journal of Educational Technology*, 34(3):341–355, 2003.
 - [18] M. Huenerfauth, L. Feng, and N. Elhadad. Comparing evaluation techniques for text readability software for adults with intellectual disabilities. In *Proc. ASSETS '09*, pages 3–10. ACM, 2009.
 - [19] J. Hyönä and R. Olson. Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6):1430, 1995.
 - [20] Interagency Commission on Learning Disabilities. *Learning Disabilities: A Report to the U.S. Congress*. Government Printing Office, Washington DC, U.S., 1987.
 - [21] International Dyslexia Association. Definition of dyslexia:
<http://interdys.org/DyslexiaDefinition.htm>, 2011. Based in the initial definition of the Research Committee of the Orton Dyslexia Society, former name of the IDA, done in 1994.
 - [22] G. Kanvinde, L. Rello, and R. Baeza-Yates. IDEAL: a dyslexic-friendly e-book reader (poster). In *The 14th International ACM SIGACCESS Conference of Computers and Accessibility (ASSETS 2012)*, Boulder, USA, October 2012. ACM Press.
 - [23] S. Kurniawan and G. Conroy. Comparing comprehension speed and accuracy of online information in students with and without dyslexia. *Advances in Universal Web Design and Evaluation: Research, Trends and Opportunities*, Idea Group Publishing, Hershey, PA, pages 257–70, 2006.
 - [24] J. E. McCarthy and S. J. Swierenga. What we know about dyslexia and web accessibility: a research review. *Universal Access in the Information Society*, 9:147–152, June 2010.
 - [25] R. Mitkov, L. A. Ha, A. Varga, and L. Rello. Semantic similarity of distractors in multiple-choice tests: extrinsic evaluation. In *Proc. EACL Workshop GeMS '09*, pages 49–56. Association for Computational Linguistics, 2009.
 - [26] A. Newell and L. Booth. The use of lexical and spelling aids with dyslexics. *Computers and Literacy*, pages 35–44, 1991.
 - [27] K. Rayner and S. Duffy. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3):191–201, 1986.
 - [28] L. Rello and R. Baeza-Yates. Lexical quality as a proxy for web text understandability (poster). In *Proc. WWW '12*, Lyon, France, 2012.
 - [29] L. Rello and R. Baeza-Yates. The presence of English and Spanish dyslexia in the Web. *New Review of Hypermedia and Multimedia*, 8:131–158, 2012.
 - [30] L. Rello, R. Baeza-Yates, L. Dempere, and H. Saggion. Frequent words improve readability and short words improve understandability for people with dyslexia. In *Proc. INTERACT '13*, Cape Town, South Africa, 2013.
 - [31] L. Rello, R. Baeza-Yates, H. Saggion, S. Bott, R. Carlini, C. Bayarri, A. Górriz, S. Gupta, G. Kanvinde, and V. Topac. Dyswebxia 2.0! accessible text for people with dyslexia (demo). In *Proc. W4A '13, The Paciello Group Web Accessibility Challenge*, Rio de Janeiro, Brazil, 2013.
 - [32] L. Rello, S. Bautista, R. Baeza-Yates, P. Gervás, R. Hervás, and H. Saggion. One half or 50%? An eye-tracking study of number representation readability. In *Proc. INTERACT '13*, Cape Town, South Africa, 2013.
 - [33] L. Rello, G. Kanvinde, and R. Baeza-Yates. Layout guidelines for web text and a web service to improve accessibility for dyslexics. In *Proc. W4A '12*, Lyon, France, 2012. ACM Press.
 - [34] J. Rüsseler, S. Probst, S. Johannes, and T. Münte. Recognition memory for high-and low-frequency words in adult normal and dyslexic readers: an event-related brain potential study. *Journal of clinical and experimental neuropsychology*, 25(6):815–829, 2003.
 - [35] H. Saggion, E. Gómez-Martínez, E. Etayo, A. Anula, and L. Bourg. Text Simplification in Simplext: Making Text More Accessible. *Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural*, 47, 2011.
 - [36] F. Simmons and C. Singleton. The reading comprehension abilities of dyslexic students in higher education. *Dyslexia*, 6(3):178–192, 2000.
 - [37] S. Swierenga, J. E. Porter, S. Ghosh, and J. E. McCarthy. Dyslexia and website design: The importance of user-based testing. In *Fourth International Congress of Qualitative Inquiry*, Urbana-Champaign, IL, May 2008.
 - [38] S. Williams, E. Reiter, and L. Osman. Experiments with discourse-level choices and readability. In *Proc. ENLG '03*, Budapest, Hungary, 2003.

APPENDIX

In Table 4 we show a list of the morphologically unique lexical simplification alterations for Spanish (separated by a semicolon) in the text *Bee, Efecto de los plaguicidas agrícolas en las poblaciones de abejas, ‘Effect of agricultural pesticides in bee populations’*.

ORIG	SUBSBEST	SHOWSYNS	GOLD
agrícola	—	agrario	—
alteraciones	cambios	cambios, modificaciones, variaciones	cambios
alteran	transforman; cambian	transforman; cambian, varían	cambian
amortiguan	atenúan	atenúan, mitigan	disminuyen
apenas	—	—	casi
apuntan	comentan	comentan, mencionan	indican
aumenta	—	adelanta	—
cambios	—	mejora, variedad	—
campo	—	ámbito, dominio	—
causas	procesos	procesos	—
colonia	provincia	provincia	—
combinación	composición	composición	unión
combinados	mezclados	mezclados	juntos
comportamiento	—	actuación; conducta	—
concentraciones	aplicaciones	aplicaciones, entregas, constancias	cantidades
conexión	—	enlace, correspondencia	relación
consecuencias	—	repercusiones	efectos
cultivos	—	labores	—
declive	descenso	descenso, bajada	pérdida
demostrado	señalado; probado	señalado, mostrado, apuntado; probado	probado; encontrado
desarrollo	—	crecimiento	—
distintos	diferentes	diferentes	—
entender	—	coger, comprender	—
estudios	—	investigaciones	—
examinado	probado	probado	analizado
exposición	—	muestra, demostración; exhibición	—
expuestas	—	descubiertas, destapadas	—
extensión	—	referencia	—
forrajero	—	—	alimenticio
funcionamiento	—	—	actividad
hallan	—	—	están
impacto	—	—	efecto
importante	—	considerable; crucial	—
individual	personal	personal	—
informando	—	avisando	—
investigado	estudiado	estudiado	estudiado
leves	—	—	pequeños
lugar	—	sitio, paraje	—
mitigar	relajar	relajar	combatir
modo	forma	forma, manera	forma
naturales	—	autóctonos	—
numerosos	—	—	muchos
obreras	—	trabajadoras	—
observado	señalado	señalado, indicado	visto
plaguicidas	—	—	insecticidas
poblaciones	pueblos	pueblos	grupos —
principales	—	primeros	—
producción	—	filme	—
produce	—	causa, ocasiona	—
propensión	—	—	tendencia
publicado	—	escrito	—
reducción	—	disminución	—
responsables	—	cabezas	—
resulta	sigue	sigue, sucede	es
señalado	—	indicado	—
serios	—	graves, peligrosos	graves
sustancias	centros	centros, corazones, núcleos	químicos
tiempo	—	momento, época	—
tipos	—	excéntricos; elementos	—
viene	—	llega	—

Table 4: Examples of lexical simplifications in Spanish.