Evaluation of DysWebxia: A Reading App Designed 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

ABSTRACT

In this paper we present the evaluation of DysWebxia, a reading app for iOS devices, specially designed for people with dyslexia. DysWebxia integrates previous results about the best way to present text for people with dyslexia together with a unique feature, the ability to show synonyms on demand for complex words. Although the new algorithm used for this unique feature is language independent, our first prototype is for Spanish. To evaluate DysWebxia we carried out two different user studies. One to evaluate the quality of the synonyms on demand that included 32 participants with dyslexia and 38 strong readers without dyslexia, and another one to evaluate the usability of the app based on 12 participants. Our results show that the quality of the synonyms generated by the new algorithm outperforms a frequency based baseline, and that the participants found DysWebxia very usable. Therefore, we show that this app may have in the future a large impact for people with dyslexia.

Keywords

Dyslexia, reading software, text simplification, synonym generation, text customization.

1. INTRODUCTION

Dyslexia is a *hidden* disability. People with dyslexia cannot perceive whether they are reading or writing correctly. Indeed, dyslexia is a neurological reading disability that 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.[1]

W4A'14 April 07–09, 2014, Seoul, Republic of South Korea.

http://dx.doi.org/10.1145/2596695.2596697

The most frequent way to detect a child with dyslexia is by low-performance in school [8]. In Spain, it is estimated that four out of six cases of school failure are related to this disability.[2] Moreover, dyslexia is frequent. From 10 to 17.5% of the population in the U.S.A. [15] and from 8.6 to 11% of the Spanish speaking population [8, 16] have this cognitive disability. The prevalence of dyslexia and its impact in school failure are the main motivations of our work.

Previous studies have shown that different characteristics of the text presentation such as font size and type or different spacings, make texts more readable for people with dyslexia, for example [26, 31]. Also, it has been shown that people with dyslexia specifically encounter problems with complex words, such as long or infrequent words [14, 28, 34]. Therefore, applying automatic lexical simplification strategies could make texts easier to read and understand for this target group [27]. Thus, our first goal was to integrate all these results in a reading app targeted to read any text, but particularly text present in the Web.

Hence, in this paper we present and evaluate DysWebxia, a reading app for iOS devices such as iPads and iPhones, that was specially designed for people with dyslexia. DysWebxia integrates previous results about the best way to present text for people with dyslexia together with a unique feature, the ability of showing synonyms on demand for complex words. Although the new algorithm used for this unique feature is language independent, our first prototype is, up to now, only for Spanish, and tailored to web text.

To evaluate DysWebxia we carried out two different user studies. The first study was to evaluate the quality of the synonyms on demand, including 32 participants with dyslexia and 38 strong readers without dyslexia. They compared a new algorithm for synonym simplification, CASSA, against a frequency based algorithm for the same task. CASSA (Context Aware Synonym Simplification Algorithm), is a new method to generate simpler synonyms

[1]International Dyslexia Association. Definition of dyslexia: `http://interdys.org/DyslexiaDefinition.htm`

[2]The percentage of school failure is calculated by the number or students who drop school before finishing secondary education (high school). While the average of school failure in the European Union is around 15%, Spain has around 25-30% of school failure, 31% in 2010 [11].

that can be tailored for different target people, in this case people with dyslexia. In the study we measured two variables: (1) the accuracy of the synonyms generated by CASSA, that is, to which extent the synonyms generated preserved the meaning, and (2) how simpler were the synonyms generated. We compared CASSA with the most challenging baseline we could find, Frequency, that selects the most frequent synonyms of the most common sense. Our results show that CASSA outperforms this baseline.

Next, we did a second study performing a usability evaluation of DysWebxia with 12 new participants. Our results show that DysWebxia is very usable, and we collected several ideas for future improvements. Based on these two user studies, we believe that DysWebxia may have a large future impact for people with dyslexia.

This paper is organized as follows. Next section presents the related work. Section 3 describes the DysWebxia app, emphasizing its unique component, the algorithm to show synonyms on demand. Sections 4 and 5 present the evaluation of the simplification algorithm in the context of DysWebxia and the usability evaluation of the app, respectively. In Section 6 we discuss our results, finishing in Section 7 with some concluding remarks and future work.

DysWebxia (presented in Section 3) was demoed at ASSETS 2013 [29] and we plan to offer it for free through the App Store in the near future. The CASSA algorithm (Section 3.4) is summarized in this paper, where we only evaluate its use in the DysWebxia app. The detailed description and the analysis of this algorithm in a general setting will be presented in a forthcoming paper [2].

2. RELATED WORK

Related contributions to our work can be found in (a) accessibility literature regarding reading tools for people with dyslexia, and (b) natural language processing (NLP) literature about lexical simplification.

2.1 Reading Tools for Dyslexia

Santana *et al.* [35] developed the Mozilla Firefox extension *Firefixia*, a tool that allows readers with dyslexia to customize websites to improve readability. They tested *Firefixia* with four users and found that readers with dyslexia appreciate customization.

Dickinson *et al.* [10] asked 12 students with dyslexia to test different colors, sizes, spacings, column widths, and similar letter highlighting to improve the subjective readability of MS Word documents. The best parameters were tested by seven people with dyslexia, which reported a subjective increase in readability. The results of this investigation were included in the tool *SeeWord* for MS Word [12].

The web service *Text4All* [37][3] for web pages and the Android *IDEAL eBook reader*[4] [17] for ebooks are text customization tools developed on the basis of previous research using eye-tracking with people with dyslexia [31].

Text4All also includes other services such as medical language adaptation by explaining terminology, terminology annotation, and a language analyzer. The terminology service was evaluated with 41 participants.

The *IDEAL eBook reader* also includes text-to-speech technology and its usability was tested with 14 participants with dyslexia using the think aloud protocol. The improvements proposed by the participants (mainly to the interface) were subsequently integrated in the application [32].

The project *MultiReader* aimed not only at people with dyslexia but also print-disabled users (visually or hearing impaired). Their system attempts to enrich documents with interface adaptations, text-to-speech, audio description of video material, and multimedia elements such as subtitles and sign language interpretation for audio and video. For its development they used an iterative user-centered design (three iterations). On its final iteration *MultiReader* was tested by 12 people with dyslexia and some usability issues were found [24].

In Table 1, that extends the same table given in [29], we compare the features of the two most popular reading applications –*Amazon's Kindle reading software*[5] and *Apple's iBooks*[6]– with seven specific reading software for people with dyslexia. Although we could not find research papers about *ClaroRead*,[7] we include it because of its broad commercialization among people with dyslexia. The parameters in bold shown in the table are the ones that, to the best of our knowledge, lead to significantly better reading performance (readability and comprehension): font type [26], font size [33], colors for the font and background [13, 31] and character spacing [39].

Synonyms on demand lead to an increase of subjective readability [27], and text-to-speech technology (TTS) have shown gains in word recognition and phonological decoding [23]. We added other parameters because they are included in the British Association of Dyslexia recommendations.[8]

2.2 Generation of Simpler Synonyms

Automatic *text simplification* is an NLP task that transforms a text that is easier to read than the original, preserving the original meaning. *Lexical simplification* is a kind of text simplification which aims at the word level. It can be performed through the substitution of words by simpler synonyms, by adding a definition or by showing simpler synonyms. Most of the approaches aim at the substitution of complex words.

To find appropriate synonyms, Burstain *et al.* [5] use WordNet. Devlin and Unthank [9] use dictionaries. Aluisio and Gasperin [1] use a thesaurus and lexical ontologies. Bott *et al.* [4] make use of the Spanish OpenThesaurus and a simplification corpus.

More recently, the availability of the Simple English Wikipedia (SEW), in combination with the standard English Wikipedia (EW), provided a new generation of text simplification approaches by using machine learning techniques. Yatskar *et al.* [38] use edit histories for the SEW and the combination of SEW and EW in order to create a set of lexical substitution rules. Biran *et al.* [3] also rely on the SEW/EW combination (without the edit history of the SEW), in addition to the explicit sentence alignment between SEW and EW.

The most frequent synonyms are presumed to be the sim-

[3] `http://www.text4all.net/dyswebxia.html`

[4] `https://play.google.com/store/apps/details?id=org.easyaccess.epubreader`

[5] `www.amazon.com/kindle`

[6] `https://itunes.apple.com/en/app/ibooks/id364709193?mt=8`

[7] `http://www.clarosoftware.com/index.php?cPath=333`

[8] British Association of Dyslexia: `http://www.bdadyslexia.org.uk/`.

Software	**Font**	**Size**	Bright.	Color	Spacing				Width	**Show Synonyms**	**TTS**
					Char.	Word	Line	Par.			
Amazon's Kindle	yes	yes	yes	yes	no	no	yes	no	yes	no	no
Apple's iBooks	yes	yes	yes	yes	no	no	no	no	no	no	no
ClaroRead	yes	yes	no	yes	yes	no	yes	yes	no	no	yes
Firefixia	yes	yes	no	yes	yes	no	yes	no	yes	no	no
IDEAL eBook Reader	yes	yes	yes	yes	yes	yes	yes	yes	no	yes*	yes
SeeWord	yes	yes	no	yes	yes	no	yes	no	no	no	no
Text4All	yes	yes	no	yes	yes	no	yes	no	no	yes	no
MultiReader	no	yes	no	yes	no	no	no	no	no	no	yes
DysWebxia	yes	yes	yes	yes	yes	no	no	no	no	yes	no

Table 1: Features comparison for reading tools. The asterisk ('*') means that the feature is under development. The parameters that should improve reading performance are marked in bold.

plest [5, 9], with the exception of [4] that used word frequency and length. In many studies of lexical simplification [3, 4, 9, 38] an algorithm based on the most frequent synonym is a very hard to beat baseline for simpler synonym generation. For instance, in a shared task for English lexical simplification [36], only one system out of nine outperformed the frequency baseline.

The closest algorithm to ours is LexSiS by Bott *et al.* [4] that presents a lexical simplification algorithm for Spanish and also uses the Spanish OpenThesaurus. However, CASSA is conceptually a new algorithm and it differs from LexSiS in: (1) the resources used; (2) the way word complexity is conceived and calculated, and (3) the way CASSA deals with word sense disambiguation, taking into account the word context using the Google Books Ngram Corpus. The later is the major strength of CASSA. Also, CASSA does not aim to do a lexical substitution but to find several simpler synonyms, which can be tailored to different readers by using different word complexity measures. In the case of this paper, we specifically targeted people with dyslexia.

3. A READING APP: DYSWEBXIA

3.1 General Description

The *DysWebxia* reader for iOS combines all the features that, to the best of our knowledge, lead to a significant improvement of the reading performance of people with dyslexia (with the exception of text-to-speech).

For the user interface design, we first performed a competitive analysis of existing reading tools (see Table 1) to understand the user interface and the user-system interaction conventions that prospective users might expect to find in our system, followed by the creation of sketches and mockups.

For the implementation we used the Apple iOS SDK, building the application in Objective-C from the ground up. Given a text file (PDF and Epub formats), we are able to render it to the user and then display synonyms on demand for complex words that may appear in the text. An example of the interface with the configuration options is given in Figure 1.

The features of the app are detailed in the next two subsections.

3.2 Text Presentation

Previous studies have shown that the reading performance of people with dyslexia can be improved when some textual conditions are modified, such as using certain font sizes or font types. Following, we present these features:

- **Font Size**: Sizes ranging from 18 points [33] to 26 points [31] lead to faster reading.
- **Font Type**: *Sans serif*, *roman*, and *monospaced* fonts are good fonts for people with dyslexia, specifically, *Helvetica*, *Courier*, *Arial*, *Verdana*, and *Computer Modern Unicode* [26].
- **Colors**: The color pairs which lead to a better readability were (background/font): cream/black [31], yellow/blue [13, 31], light mucky green/dark brown [31, 13], grey (25%) in the background with white font, and grey font (25%) with white background [31].
- **Character Spacing**: Larger letter spacing was found to lead to faster reading [39].

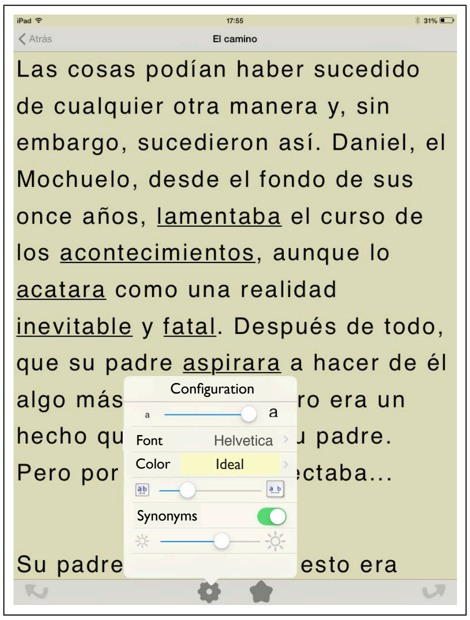


Figure 1: DysWebxia customization interface.

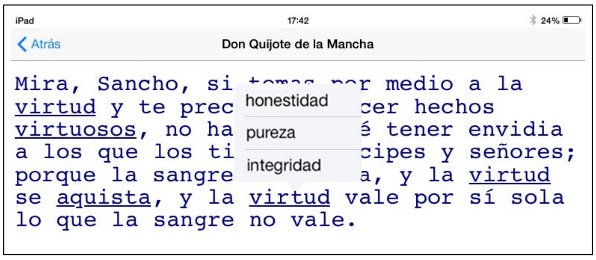


Figure 2: In this example, DysWebxia shows the synonyms *honesty*, *purity*, and *integrity* for the complex word *virtue* after the user clicks in it. Synonyms are available for all underlined words.

3.3 Synonyms on Demand

Regarding the content of the text, previous studies have shown that simpler words lead to a better reading performance. For instance, Hyona *et al.* [14] used eye-tracking to show that low frequency and long words present longer gaze durations and more re-inspections. Rello *et al.* [28] found that frequent words improve readability and short words improve understandability for people with dyslexia. Also, Rüsseler *et al.* [34] showed that it takes more time to recognize infrequent words and this recognition performance is lower in readers with dyslexia.

Since simpler synonyms lead to better readability for people with dyslexia, Rello *et al.* [27] tested an automatic lexical simplification algorithm for Spanish [4]. In this study 47 participants with dyslexia showed that performing automatic lexical simplifications (substituting complex words by simpler synonyms) did not improve the readability of the texts. However, when these synonyms were presented on demand to the user, texts were perceived as significantly simpler.

Even if no significant improvement in readability can be demonstrated, the subjective perception of texts for students with dyslexia seems to be crucial. For example, if texts are perceived as simpler, students with dyslexia might be encouraged to read more. Hence, we avoid the vicious circle that reading less leads them to stay on a lower reading proficiency level.

For this reason, in our reader we include a feature that is unique: we show synonyms on demand for complex words (see Figure 2 for an example). For this we need a synonym simplification algorithm. However, the algorithm that we tested previously in [27] was not designed to present several simpler synonyms for people with dyslexia but to find the best substitution, which is not always the simplest synonym. In fact, only 36.11% of the synonyms substituted were considered simpler by annotators without dyslexia [4]. For that reason we devised an improved algorithm for this task that is presented next.

3.4 CASSA

CASSA (Context Aware Synonym Simplification Algorithm) is a method that generates simpler synonyms of a word. Words can be polysemic,[9] that is, they can have different meanings or senses depending on their context, for instance, the Spanish verb *acostar* can mean either 'to go to bed' or 'to reach coast'. CASSA takes into consideration the context of the complex word for disambiguation in order to find the correct simpler synonyms to show.

Resources. The method is language independent although it was implemented and evaluated for Spanish. It only needs the following two usually freely available resources: (a) a dictionary of synonyms, where we used the Spanish OpenThesaurus;[10] and (b) a large n-gram corpus with frequencies, where we used the Google Books Ngram Corpus [19]. Next we detail these two resources:

- **Spanish OpenThesaurus** (version 2): it is freely available under the GNU Lesser General Public License, to be used with OpenOffice.org. This thesaurus provides 21,378 target words (lemmas) and provides a total of 44,348 different word senses for them. The following is the thesaurus entry for *mono*, which is ambiguous, as it could mean 'ape', 'overall', or the adjective 'cute'.

  (a) mono| 3
  - simio|chimpancé|mandril|mico|macaco|gorila|antropoide
  - overol|traje de faena
  - llamativo|vistoso|atractivo|provocativo|sugerente|resultón|bonito

- **Google Books Ngram Corpus** (2012 edition): The corpus consists of words and phrases (that is, n-grams) and their usage frequency over time. The data is available for download[11] and is derived from 8,116,746 books, over 6% of all books ever published. For Spanish the corpus has 854,649 volumes and 83,967,471,303 tokens [19].

Algorithm. First, we modified and enriched the Spanish OpenThesaurus and created our List of Senses. Instead of having a target word with difference senses, we kept only the list of senses and included the target word in each one.

Then, for each of the words we included their frequency in the Web using a large search engine frequency index. As a result we had a set of lists of synonyms with their frequencies, where each list corresponds to one sense. The Spanish OpenThesaurus contains single-word and multi-word expressions. We only treated single-word units, which represent 98% of the cases, leaving out only 399 multi-word expressions, such as *de esta forma ('in this manner')*.

Second, we use the 5-grams in the Google Books Ngram Corpus, where we use the third token of each 5-gram as our target words. This token is lemmatized and it is included in the Synonyms List as a target word only if it appears in our List of Senses, filtering proper names and stop words (*and, of, at, etc.*). The other four tokens are the context of the target word, enriching it with its frequency in the corpus and the number of times that the contexts appear having different target words. For example:

- *era una* **noche** *oscura de* (it was a dark night of)

Third, we define the complexity of a word taking into account the frequency of the words in the Web, because previous studies have shown that less frequent words were found

[9] Polysemy refers to the coexistence of many possible meanings for a word or phrase.

[10] http://openthes-es.berlios.de

[11] http://books.google.com/ngrams

to be more challenging for people with dyslexia leading to worse reading performance [14, 28, 34]. That is, our definition is tailored to web text. Next, to determine the word complexity we use the relative frequency of the synonyms with the same sense in the List of Senses. If a word is ten or more times less frequent than one or more of its synonyms is considered a complex word. We used ten times as a threshold because worked well in practice (31% of the words have simpler synonyms in this way), but this is a parameter of the algorithm.

Finally, for each complex word and the contexts where it appears, we select as simpler synonyms the three most frequent ones that belong to the sense that appears most frequently for the 5-gram corresponding to that (word,context) pair. That is, to disambiguate the sense, CASSA uses the context where the target word appears.

Next, we present the evaluation of this algorithm in the context of DysWebxia followed by the usability evaluation of the app.

4. EVALUATION OF CASSA

4.1 Methodology

To evaluate CASSA we conducted an experiment with two groups of people, one with and another one without dyslexia. Using online questionnaires, each participant had to read and rate a set of synonyms generated by CASSA and a baseline explained below.

4.1.1 Design

We compared two methods to generate synonyms which served as *independent variables*:

– **Frequency.** As a baseline we use a method that uses the same resources of CASSA, the Google Books Ngram Corpus and the Spanish OpenThesaurus, and whose synonym generation engine is the same as CASSA, with the exception of the last disambiguation step. That is, instead of using the context for disambiguation, the baseline chooses the most used sense and shows the top-3 most frequent synonyms in that sense. The resulting dictionary of synonyms generated by the Frequency method is composed by 135,577 simpler synonyms lists (senses). As we mentioned in Section 2.2, this baseline is hard to outperform.

– **CASSA.** This method is explained in Section 3.4. The resulting resource of CASSA is a dictionary of synonyms and their contexts, that is, 4,229,868 lists of synonyms in context (senses). However, it is worth mentioning that both resources, Frequency and CASSA, have the same number of different target words (43,996). CASSA only has more subgroups of synonyms depending on the context.

The experiment followed a within-subjects design, so every participant contributed to each of the conditions. The order of conditions was counter-balanced to cancel out sequence effects. We aimed to measure two variables with our experiment:

(a) **Synonymy**: to which extent the synonyms generated by the method preserve the meaning, that is, if they are actual *synonyms* of the complex word; and

(b) **Simplicity**: to which extent the synonyms generated by the method are *simpler* than the complex word.

To measure both parameters we use two ratings:

– **Synonymy Rating:** For 40 items on a 10-point Likert scale, we asked the participants to rate the synonymy of the words presented in comparison with the target word.

– **Simplicity Rating:** We asked the participants to rate another 40 items on a 10-point Likert scale, to find out whether the words presented were simpler than the target word.

The rationale behind using a 10-point Likert scales is that our participants were more familiarized with ten points rating systems because half of them (16 participants) were attending Spanish schools or high schools, and in Spain the gradings are given by using a ten-point rating.

4.1.2 Participants

Target Group, People with Dyslexia: We recruited 32 participants with diagnosed dyslexia (18 female, 14 male). They were asked to submit their diagnoses to guarantee that dyslexia was diagnosed in an authorized centre or hospital.[12] They were all native speakers of Spanish, 14 of them were bilingual (twelve in Catalan, one in English, and one in French). Their ages ranged from 6 to 52 years old, with a mean of 23.15 years ($s = 12.07$). Nine of the participants did finish primary school, seven secondary school, and 16 of them had further studies. Only one participant read more than five hours per day, eleven participants read between three and five hours per day, and most of the participants (20) read less than two hours per day.

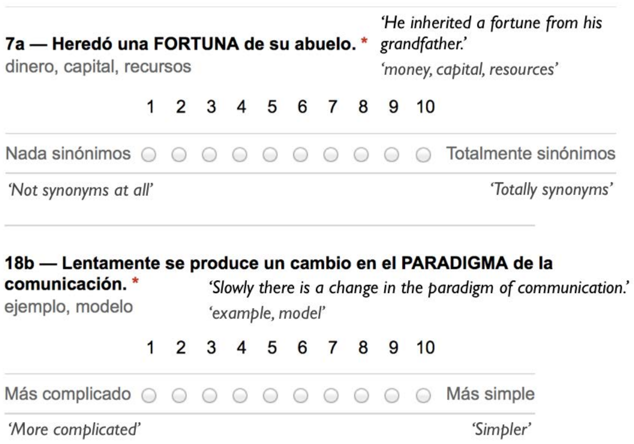
Control Group, Strong Readers: We also recruited 38 participants without dyslexia (24 female, 14 male). We consider them strong readers because they all finished post-compulsory schooling[13] and were frequent readers. They were all native speakers of Spanish, two of them were trilingual and eleven of them were bilingual in Catalan, five in Galician, five in English, three in French and two in German. Their ages ranged from 17 to 69 years old, with a mean of 38.16 years ($s = 13.06$). Of the participants, 14 of them read more than five hours per day, 10 participants read between three and five hours per day, and 14 participants read less than two hours per day.

4.1.3 Materials

To study whether the words shown are actual simpler synonyms to the target word, we need to insert the target word in its context. This is needed because of two reasons. First, depending on the context, words can have different meanings, and second, the comprehension of the text pertain to longer segments, not only words. Following, we describe how we designed the materials that were used in this study.

[12] When dyslexia is diagnosed in Spanish, the different kinds of dyslexia, extensively found in literature about dyslexia in English, are normally not considered.

[13] In Spain post-compulsory schooling corresponds with two year of studies after compulsory secondary education before entering university.


7a — Heredó una FORTUNA de su abuelo. * 'He inherited a fortune from his grandfather.'
dinero, capital, recursos 'money, capital, resources'

1 2 3 4 5 6 7 8 9 10

Nada sinónimos ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Totalmente sinónimos

'Not synonyms at all' 'Totally synonyms'

18b — Lentamente se produce un cambio en el PARADIGMA de la comunicación. * 'Slowly there is a change in the paradigm of communication.'
ejemplo, modelo 'example, model'

1 2 3 4 5 6 7 8 9 10

Más complicado ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Más simple

'More complicated' 'Simpler'

Figure 3: Two items extracted from the test.

Evaluation Dataset. We have two evaluation datasets derived from CASSA and Frequency, respectively. Each data set is composed of:

- **Target Words**: We selected 40 target words which are intended to be complex words for each evaluation dataset, so 80 target words in total. For all the words that are classified as complex according to CASSA, we created two groups: (a) LOW, that includes very low frequency complex words, and (b) HIGH, that contains high frequency complex words. The frequency of the LOW group ranges between 40 to 200 occurrences of the word together with its context in Google Books Ngram Corpus. The frequency range of the HIGH frequency group is between 2,000 to 1,300,000 occurrences. After defining the frequency ranges, we randomly extracted the candidates for target words and selected only polysemic words that have different senses, like *fortuna* (which can mean *'luck'* or *'treasure'*). Monosemous complex words[14] such as infrequent nouns were discarded.

- **Synonyms**: For each target word there is a set of simpler synonyms generated by CASSA and Frequency. The number of synonyms per set ranges from one to eight synonyms.

- **Contexts and sentences**: Each target word is presented within a context in a sentence. The context and their sentences are real instances from books of the 20th and 21st century using Google Books Ngram Corpus. The length of the sentences ranged from 9 to 17 words.

Test. The evaluation dataset was integrated in a online test. The sentence was presented with the complex word in capital letters and the set of synonyms stated below the sentence. For each of the sentences we created two Likert scales items to rate the *Synonymy* and the *Simplicity* of the set of synonyms in comparison with the target word (see Figure 3). There were 160 items and the conditions were counter balanced.

[14] The linguistic property of having only one meaning.

Condition		Synonymy		Simplicity	
		$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
CASSA *Target Group*	LOW	8	7.54 ± 2.80	9	7.67 ± 2.74
	HIGH	8	6.49 ± 3.35	8	6.86 ± 3.19
Frequency *Target Group*	LOW	6	5.86 ± 2.91	6	6.11 ± 2.80
	HIGH	5	4.94 ± 2.96	5	5.12 ± 2.94
CASSA *Control Group*	LOW	9	7.96 ± 2.63	9	8.23 ± 4.45
	HIGH	8	6.64 ± 3.45	8	7.35 ± 2.98
Frequency *Control Group*	LOW	7	6.11 ± 2.79	7	7.02 ± 2.69
	HIGH	5	5.22 ± 2.95	5	5.66 ± 3.01

Table 2: Median, mean, and standard deviation for *Synonymy* and *Simplicity* ratings for CASSA and the baseline for LOW and HIGH frequency complex words.

We also included a set of 20 validation/calibration items to check whether the participants were doing the test correctly (that is, to verify that they were not giving random answers) and to check whether the rating judgements were similar between participants. These items were done manually, containing perfect synonyms or antonyms of the target word, and uniformly interspersed in the test.

Hence, the questionnaire had a total of 180 items. We consider this amount to be quite reasonable to evaluate CASSA because similar studies had smaller or slightly larger evaluation datasets but they were not rated by the target group but by two or three annotators. Yatskar *et al.* [38] used six annotators (three native, three non-native speakers of English) that rated 200 simplification examples in English while Biran *et al.* [3] used 130 examples that were judged by three annotators (native English speakers). In Bott *et. al* [4], three annotators (native speakers of Spanish) rated 69 sentences for each Spanish lexical simplification algorithm evaluated.

4.1.4 Procedure

Both groups performed the test. The main purpose of the control group is to evaluate whether the algorithms are preserving the precise semantic sense of the original words. This way we control that the learning disability does not affect the results of the evaluation. Depending on the participant the test lasted from 30 to 50 minutes. Eleven participants of the target group performed the test at the Madrid for Dyslexia Association[15] supervised by the first author. The rest of the participants undertook the test online at their homes. In that case the first author was also online to ease possible doubts or questions.

First, the participants read the instructions presented in the test and had the opportunity to ask questions if they needed. Then, they began with a questionnaire that was designed to collect demographic information. Third, they started the test and rated the first 90 items, then they took a small break and after judged the last 90 items. Finally, they answered a semi-structured interview to collect feedback about how they used technology, how they found the test, and how the synonyms affected their reading.

[15] Asociación Madrid con la Dislexia: `http://www.madridconladislexia.org/`

4.2 Results

Now we present the analyses of the data from the tests. First, we checked the validation/calibration items. All participants understood the test correctly, so all the answers were valid. The average of the expected low value answers was 1.41 ($s = 0.98$) for the participants with dyslexia and 1.47 ($s = 0.51$) for the control group. The average of the expected high value answers was 8.77 ($s = 0.93$) for the participants with dyslexia and 9.16 ($s = 0.69$) for the control group. This means that the test was well calibrated (if the averages would have been 1 and 10, respectively, that would have implied a perfect agreement).

The debate of what analyses are admissible for Likert scales – parametric or non-parametric tests– is pretty contentious [7]. A Shapiro-Wilk test showed that the results were not normally distributed. Also, a Barlett's test showed that they were homogeneous. Hence, for each experiment we used the Kruskal-Wallis non-parametric test for repeated measures and two conditions, to find significant effects on the participants ratings. To test effects between groups we used the Wilcoxon non-parametric test for repeated measures.

4.2.1 Synonymy

Regarding the candidates generated from both methods, the *Synonymy Rating* of the strong readers ($\tilde{x} = 5$, $\bar{x} = 6.48, s = 3.13$) was significantly higher than the *Synonymy Rating* of the participants with dyslexia ($\tilde{x} = 5$, $\bar{x} = 6.21, s = 3.16$), ($W = 1,740,194, p = 0.013$).

The strong readers' *Synonymy Rating* of the candidates generated by CASSA was significantly higher ($\tilde{x} = 9$, $\bar{x} = 7.30, s = 3.14$) than the *Synonymy Rating* of the target group ($\tilde{x} = 8$, $\bar{x} = 7.02, s = 3.13$), ($W = 430,363, p = 0.029$). Also, the *Synonymy Rating* of strong readers of the candidates generated by Frequency was significantly higher ($\tilde{x} = 6$, $\bar{x} = 5.67, s = 2.90$) than the *Synonymy Rating* of the target group ($\tilde{x} = 5$, $\bar{x} = 5.40, s = 2.97$), ($W = 432,193, p = 0.047$).

CASSA *vs. Frequency*

Target Group: There was a significant effect of the method used on the *Synonymy Rating* ($H(1) = 110.36$, $p < 0.001$). Candidates generated by CASSA were considered to be better synonyms ($\tilde{x} = 8$[16], $\bar{x} = 7.02, s = 3.13$) than candidates generated by Frequency ($\tilde{x} = 5$, $\bar{x} = 5.40, s = 2.97$).
Control Group: There was a significant effect of the method used on the *Synonymy Rating* ($H(1) = 198.72$, $p < 0.001$). Candidates generated by CASSA were considered to be better synonyms ($\tilde{x} = 9$, $\bar{x} = 7.30, s = 3.14$) than candidates generated by Frequency ($\tilde{x} = 6$, $\bar{x} = 5.67, s = 2.90$).

LOW *vs.* HIGH *Frequency Complex Words*

Target Group: There was a significant effect of the frequency of the complex word used on *Synonymy Rating* ($H(1) = 35.77$, $p < 0.001$). The synonyms presented for LOW frequency complex words were considered to be better synonyms ($\tilde{x} = 7$, $\bar{x} = 6.70, s = 2.98$) than the candidates generated for HIGH frequency complex words ($\tilde{x} = 6$, $\bar{x} = 5.72, s = 3.25$).
Control Group: Similarly, there was a significant effect of the frequency of the complex word used on *Synonymy Rating* ($H(1) = 100.19$, $p < 0.001$). The synonyms presented for LOW frequency complex words were considered to be better synonyms ($\tilde{x} = 8$, $\bar{x} = 6.90, s = 2.91$) than the candidates generated for HIGH frequency complex words ($\tilde{x} = 6$, $\bar{x} = 5.84, s = 3.27$).

In Table 2 we show the results for the all the subgroups.

[16]Recall that $\tilde{x}$ represents the median of the distribution.

4.2.2 Simplicity

Taking into account all the candidates from both methods, the *Simplicity Rating* of the strong readers ($\tilde{x} = 8$, $\bar{x} = 7.07, s = 2.94$) was significantly higher than the *Simplicity Rating* of the participants with dyslexia ($\tilde{x} = 7$, $\bar{x} = 6.44, s = 3.07$), ($W = 1,605,891, p < 0.001$).

The strong readers' *Simplicity Rating* of the synonyms generated by CASSA was significantly higher ($\tilde{x} = 9$, $\bar{x} = 7.79, s = 2.77$) than the *Simplicity Rating* of the target group ($\tilde{x} = 8$, $\bar{x} = 7.26, s = 3.00$), ($W = 407,465, p < 0.001$). Also, the strong readers' *Simplicity Rating* of the synonyms generated by Frequency was significantly higher ($\tilde{x} = 7$, $\bar{x} = 6.34, s = 2.93$) than the *Simplicity Rating* of the target group ($\tilde{x} = 6$, $\bar{x} = 5.62, s = 2.91$), ($W = 390,338, p < 0.001$).

CASSA *vs. Frequency*

Target Group: There was a significant effect of the method used on the *Simplicity Rating* ($H(1) = 131.76$, $p < 0.001$). Candidates generated by CASSA were considered simpler ($\tilde{x} = 8$, $\bar{x} = 7.26, s = 3.00$) than candidates generated by Frequency ($\tilde{x} = 6$, $\bar{x} = 5.62, s = 2.91$).
Control Group: There was a significant effect of the method used on the *Simplicity Rating* ($H(1) = 179.82$, $p < 0.001$). Candidates generated by CASSA were considered simpler ($\tilde{x} = 9$, $\bar{x} = 7.79, s = 2.77$) than candidates generated by Frequency ($\tilde{x} = 7$, $\bar{x} = 6.34, s = 2.93$).

LOW *vs.* HIGH *Frequency Complex Words*

Target Group: There was a significant effect of the frequency of the complex word used on *Synonymy Rating* ($H(1) = 30.66$, $p < 0.001$). The synonyms presented for LOW frequency complex words were considered to be simpler ($\tilde{x} = 8$, $\bar{x} = 6.89, s = 2.88$) than the candidates generated for HIGH frequency complex words ($\tilde{x} = 6$, $\bar{x} = 5.99, s = 3.19$).
Control Group: In addition, there was a significant effect of the frequency of the complex word used on *Synonymy Rating* ($H(1) = 102.18$, $p < 0.001$). The synonyms presented for LOW frequency complex words were considered to be simpler ($\tilde{x} = 8$, $\bar{x} = 7.33, s = 2.76$) than the candidates generated for HIGH frequency complex words ($\tilde{x} = 6$, $\bar{x} = 6.30, s = 3.15$).

In Table 2 we show the results for the all the subgroups.

5. USABILITY EVALUATION

In this section we present the usability evaluation of DysWebxia, explaining first the methodology, followed by the results obtained.

5.1 Methodology

5.1.1 Design

In a within-subject design, all the participants had to perform some tasks with the tool using the think aloud protocol [18]. They had to choose and customize a text to later

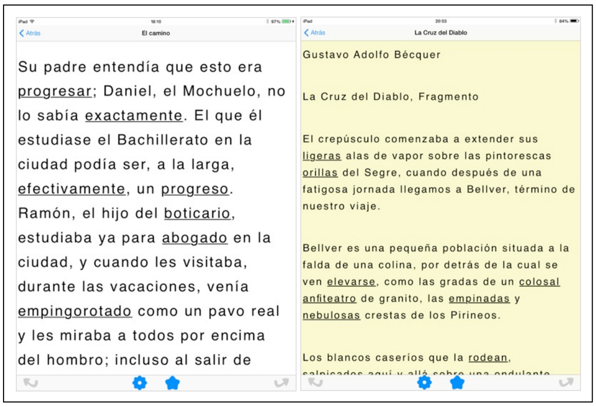


Figure 4: Preferred settings of two participants.

read it. They also undertook a questionnaire and a semi-structured interview.

5.1.2 *Participants*

We recruited 12 new participants with diagnosed dyslexia (9 male, 3 female). We believe this number of participants is enough to discover most of the usability problems since Nielsen and Landauer [21] showed that only five users are enough to find 80% of the usability problems. Eleven were native speakers of Spanish and one was learning Spanish as a second language (English native). Ten of them were bilingual in Catalan. Their ages ranged from 9 to 34 years old, with a mean of 18.25 years ($s = 7.77$). Two of the participants were at primary school, five were at secondary school, one was at high school, two were studying at the university, and two had finished their university degree. Except from one participant they were all familiarized or had tablets at home. Three participants were frequent users of eBook readers.

5.1.3 *Materials*

Texts. For the evaluation we used three texts processed by CASSA. The texts were fragments of compulsory readings[17] from high school in Spain. They had similar length (350 to 360 words) and an average of 31.7 complex words for which simpler synonyms could be demanded (See Figure 2).

Questionnaire. The items of the questionnaire were inspired by the WACG 2.0 [6] and usability principles [22]. It included open questions as well as a 7-point Likert scale items regarding: (a) the language used in the application, (b) navigation and control, (c) functionalities of the application, and (d) personal opinions.[18]

Interview. The interview contains questions about their daily difficulties as a person with dyslexia, their use of technology, and how the application of their dreams would be.

[17]These are: *The ingenious gentleman Don Quixote of La Mancha*, Second Part, beginning of chapter 42, by Miguel de Cervantes; *The cross of the devil*, beginning of the short story, by Gustavo Adolfo Béquer; and *El camino*, beginning of Chapter 1, by Miguel Delibes.

[18]The survey is available at `http://luzrello.com/DysWebxia/usability-test.html`

5.1.4 *Procedure*

The sessions lasted around 30 minutes and were conducted in a quiet room prepared for the study at Universitat Pompeu Fabra. In each session the participant was alone (or with their parents in case they requested it) with the interviewer (first author). 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 from the bookshelf of the application. Then, each participant was asked to use the application to customize the text until s/he found the options they preferred (see the customization settings in Figure 1). In this step each participant was asked to think aloud while exploring the application and finding her/his favorite settings, while the interviewer wrote down her/his comments. Next, they read the text in silence. When they finished, the participants were asked to complete the questionnaire on paper, ending with the personal interview.

5.2 Results

Following we describe the observations collected while the participants were performing the task, and the relevant data extracted from the questionnaires.

During the *customization task*, all the participants but one, decided to turn the synonyms option on, so complex words in the text were underlined. The participants reactions while discovering the tool functionalities were very positive. They seemed to be positively surprised about the synonyms and the letter spacing options, maybe because these options are not frequently found in other reading softwares. We believe that the positive attitude towards the tool have impacted the answers of the survey we present next. During the *reading* task only eight participants (67%) actually made use of the synonyms on demand option and only two participants changed the customization settings more than once. In Figure 4 we show two of the preferred settings.

All the participants found that the *language* used in the application was descriptive ($\bar{x} = 6.66 \pm 0.65$, on a 7-point Likert scale) and they were familiar with it ($\bar{x} = 6.58 \pm 1.44$). The exception was a young participant, with nine years old, who did not know yet what the word "synonym" meant. The symbols used (a star for favorites, a knob for the settings button, a sun for the brightness, etc.) were also found understandable ($\bar{x} = 6.33 \pm 0.65$). The *navigation* through the bookshelf and the text was found easy ($\bar{x} = 6.83 \pm 0.39$) as well as the *customization* of the text ($\bar{x} = 6.75 \pm 0.62$). Some participants proposed further text customizations options such as more font sizes, smaller font types, more colors and more spacing alternatives. Other additions proposed were adding a text-to-speech engine or providing the possibility of having folders to organize the books in the bookshelf.

Regarding the *synonyms* option (Figure 2), most of the participants found the option very helpful for reading ($\bar{x} = 6.42 \pm 0.79$). The main objection of them was the coverage of the *synonyms* option since there were complex (monosemous) words in the texts with no synonyms, such as the names of birds *rendajo (goldfinch)* or *jilguero (redbreast)*. Two participants missed not finding definitions for the complex words in addition to the synonyms and two others found underlining words confusing and would have preferred to see the complex words in boldface or in a different color. Finally, one participant would have liked to remember the synonyms which he found more readable for future readings.

Overall, the application was found easy to use ($\bar{x} = 6.58 \pm 0.66$), and people with dyslexia considered that they could read better by customizing the presentation of the text ($\bar{x} = 6.83 \pm 0.39$) as well as accessing synonyms ($\bar{x} = 6.50 \pm 0.67$).

6. DISCUSSION

The summary of the results is that for the participants with dyslexia, CASSA was found to generate more accurate and simpler synonyms than Frequency, a baseline which is challenging to outperform [3, 4, 9, 38]. When the complex word has a lower frequency, both algorithms, CASSA and Frequency, gave better results for meaning preservation as well as for producing simpler synonyms.

We believe that the high scores for both algorithms, in comparison with previous studies [3, 4, 27], are so because we only show alternative synonyms instead of substituting the best synonym in the original sentence. The substitution task requires better meaning preservation in order to not generate inaccurate or unusual sentences. Also notice that substituting synonyms and showing them on demand are different tasks so any comparison shall be taken with care.

As we specifically tested the examples where Frequency and CASSA gave different synonyms candidates for the complex word, because the hardest case for any simplification algorithm are polysemic words. Most probably, if we only included monosemous words, the output of both algorithms would have been more similar.

Also, we only evaluated the synonyms within a sentence. Even if the sentence is the largest text part that have been used for evaluating lexical simplification in previous literature [4], some synonyms may need a larger context than a full sentence for their disambiguation. However, those would represent very few cases plus we did not find any of these cases in the evaluation data set.

It is worth mentioning that in this study the algorithms were only evaluated with the target group, people with dyslexia. The results cannot be extended to other target groups or the general population because the perception of word complexity is very particular in the case of dyslexia. For example, words with typographical errors do not impede the text comprehension as they do for people without dyslexia [25], or the frequency of the word has a larger effect on reading difficulty for people with dyslexia than for people without dyslexia [28].

7. CONCLUSIONS

We have evaluated DysWebxia, a reading app for iOS specifically designed for people with dyslexia, through two user studies for two orthogonal dimensions: first, the quality of synonyms generated on demand, a unique feature of our app; and second, the usability of the app. In the first user study we tested how accurate and simpler were the synonyms generated by CASSA, a new synonym simplification algorithm, and Frequency, a well known baseline. Our results show that CASSA generates better synonyms than this baseline. As CASSA seems to outperform the algorithm presented in [27], the negative results found earlier for readability might be different for this new algorithm. Studying this is part of our future research.

Future work includes the refinement of CASSA by tailoring the detection of lexical complexity. We will consider the orthographic and phonetic similarity of words, because these language features makes words more difficult to recognize for people with and also without dyslexia [20]. This implies defining a new measure of word complexity that takes into account these features. Regarding the resource generated by CASSA (a dictionary of synonyms with its contexts), we plan to publish it through the Web in the near future, giving the opportunity to other researchers and developers to enrich their tools. Also, we will include a module to transform numerical expressions written with words into digits [30].

In the second user study we did a usability evaluation of the app. Our participants found that the app was very usable and gave very good feedback to improve our next prototype. Based on them, we will add to each suggested synonym a link to search the complex word in Wikipedia and an option to read its definition. We also plan to add a module with hyperonyms for targeting complex specific words which have no synonyms, such as names of animals or plants. We will also add further options to add other file formats, in particular HTML, and to customize the highlighting of the complex words with boldface, colors, or different kinds of underlinings.

Acknowledgments

We thank Clara Bayarri for implementing the App prototype using the synonyms generated by CASSA. We also thank María Sanz-Pastor Moreno de Alborán for helping in the organization of the synonym evaluation at Madrid for Dyslexia Association. Finally, we are indebted to all the participants who volunteered in our experiments, their fast disposition and their invaluable feedback.

8. REFERENCES

[1] 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.

[2] R. Baeza-Yates, L. Rello, and J. Dembowski. CASSA (Context Aware Synonym Simplification Algorithm), 2014. In preparation.

[3] O. Biran, S. Brody, and N. Elhadad. Putting it simply: a context-aware approach to lexical simplification. In *Proc. ACL'11*, pages 496–501, Portland, Oregon, USA, 2011.

[4] S. Bott, L. Rello, B. Drndarevic, and H. Saggion. Can Spanish be simpler? LexSiS: Lexical simplification for Spanish. In *Proc. Coling '12*, Mumbay, India, 2012.

[5] J. Burstein, J. Shore, J. Sabatini, Y.-W. Lee, and M. Ventura. The automated text adaptation tool. (demo). In *Proc. NAACL'07*, pages 3–4, 2007.

[6] B. Caldwell, M. Cooper, L. G. Reid, and G. Vanderheiden. Web content accessibility guidelines (WCAG) 2.0. *WWW Consortium (W3C)*, 2008.

[7] J. Carifio and R. Perla. Resolving the 50-year debate around using and misusing Likert scales. *Medical education*, 42(12):1150–1152, 2008.

[8] M. S. Carrillo, J. Alegría, P. Miranda, and S. Pérez. Evaluación de la dislexia en la escuela primaria: Prevalencia en español (Evaluation of dyslexia in primary school: The prevalence in Spanish). *Escritos de Psicología (Psychology Writings)*, 4(2):35–44, 2011.

[9] S. Devlin and G. Unthank. Helping aphasic people

process online information. In *Proc. ASSETS '06*, pages 225–226. ACM, 2006.

[10] A. Dickinson, P. Gregor, and A. Newell. Ongoing investigation of the ways in which some of the problems encountered by some dyslexics can be alleviated using computer techniques. In *Proc. ASSETS'02*, pages 97–103, Edinburgh, Scotland, 2002.

[11] M. F. Enguita, L. M. Martínez, and J. R. Gómez. *Fracaso y abandono escolar en España (School Failure in Spain)*. Obra Social, Fundación la Caixa, 2010.

[12] 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.

[13] P. Gregor and A. F. Newell. An empirical investigation of ways in which some of the problems encountered by some dyslexics may be alleviated using computer techniques. In *Proc. ASSETS'00*, ASSETS 2000, pages 85–91, New York, NY, USA, 2000. ACM Press.

[14] 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.

[15] Interagency Commission on Learning Disabilities. *Learning Disabilities: A Report to the U.S. Congress*. Government Printing Office, Washington DC, U.S., 1987.

[16] J. E. Jiménez, R. Guzmán, C. Rodríguez, and C. Artiles. Prevalencia de las dificultades específicas de aprendizaje: La dislexia en español (the prevalence of specific learning difficulties: Dyslexia in Spanish). *Anales de Psicología (Annals of Psychology)*, 25(1):78–85, 2009.

[17] G. Kanvinde, L. Rello, and R. Baeza-Yates. IDEAL: a dyslexic-friendly e-book reader (poster). In *Proc. ASSETS'12*, pages 205–206, Boulder, USA, October 2012.

[18] C. Lewis. *Using the "thinking-aloud" method in cognitive interface design*. IBM TJ Watson Research Center, 1982.

[19] Y. Lin, J.-B. Michel, E. L. Aiden, J. Orwant, W. Brockman, and S. Petrov. Syntactic annotations for the Google books ngram corpus. (demonstration). In *Proc. ACL'12*, pages 169–174, 2012.

[20] 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, 2009.

[21] J. Nielsen and T. K. Landauer. A mathematical model of the finding of usability problems. In *Proc. INTERCHI'93*, pages 206–213. ACM, 1993.

[22] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proc. SIGCHI'90*, pages 249–256, 1990.

[23] R. K. Olson and B. W. Wise. Reading on the computer with orthographic and speech feedback. *Reading and Writing*, 4(2):107–144, 1992.

[24] H. Petrie, G. Weber, and W. Fisher. Personalization, interaction, and navigation in rich multimedia documents for print-disabled users. *IBM Systems Journal*, 44(3):629–635, 2005.

[25] L. Rello and R. Baeza-Yates. Lexical quality as a proxy for web text understandability (poster). In *Proc. WWW '12*, pages 591–592, Lyon, France, 2012.

[26] L. Rello and R. Baeza-Yates. Good fonts for dyslexia. In *Proc. ASSETS'13*, Bellevue, Washington, USA, 2013. ACM Press.

[27] L. Rello, R. Baeza-Yates, S. Bott, and H. Saggion. Simplify or help? Text simplification strategies for people with dyslexia. In *Proc. W4A '13*, Rio de Janeiro, Brazil, 2013.

[28] 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.

[29] L. Rello, R. Baeza-Yates, H. Saggion, C. Bayarri, and S. D. J. Barbosa. An iOS reader for people with dyslexia (demo). In *Proc. ASSETS'13*, Bellevue, Washington, USA, 2013. ACM Press.

[30] 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.

[31] 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.

[32] L. Rello, G. Kanvinde, and R. Baeza-Yates. A mobile application for displaying more accessible ebooks to dyslexics. In *DSAI 2012*, Douro, Portugal, 2012.

[33] L. Rello, M. Pielot, M. C. Marcos, and R. Carlini. Size matters (spacing not): 18 points for a dyslexic-friendly Wikipedia. In *Proc. W4A '13*, Rio de Janeiro, Brazil, 2013.

[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] V. F. Santana, R. Oliveira, L. Almeida, and M. Ito. Firefixia: An accessibility web browser customization toolbar for people with dyslexia. In *Proc. W4A '13*, Rio de Janeiro, Brazil, 2013.

[36] L. Specia, S. K. Jauhar, and R. Mihalcea. Semeval-2012 task 1: English lexical simplification. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, pages 347–355, 2012.

[37] V. Topac. The development of a text customization tool for existing web sites. In *Text Customization for Readability Symposium*, November 2012.

[38] M. Yatskar, B. Pang, C. Danescu-Niculescu-Mizil, and L. Lee. For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia. In *Proc. ACL'10*, pages 365–368, Uppsala, Sweden, 2010.

[39] M. Zorzi, C. Barbiero, A. Facoettia, I. Lonciari, M. Carrozzi, M. Montico, L. Bravar, F. George, C. Pech-Georgel, and J. Ziegler. Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences*, 109:11455–11459, 2012.