DysWebxia

A Text Accessibility Model for People with Dyslexia

Luz Rello

PhD Thesis – June 2014

Supervisors:
Ricardo Baeza-Yates and Horacio Saggion
Department of Information and Communication Technologies

Universitat
Pompeu Fabra
Barcelona

A las personas con dislexia

To people with dyslexia

Abstract

Worldwide, 10% of the population has dyslexia, a cognitive disability that reduces readability and comprehension of written information. The goal of this thesis is to make text more accessible for people with dyslexia by combining human computer interaction validation methods and natural language processing techniques. In the initial phase of this study we examined how people with dyslexia identify errors in written text. Their written errors were analyzed and used to estimate the presence of text written by individuals with dyslexia in the Web. After concluding that dyslexic errors relate to presentation and content features of text, we carried out a set of experiments using eye tracking to determine the conditions that led to improved readability and comprehension. After finding the relevant parameters for text presentation and content modification, we implemented a lexical simplification system. Finally, the results of the investigation and the resources created, lead to a model, *DysWebxia*, that proposes a set of recommendations that have been successfully integrated in four applications.

Resumen

Un 10 % de la población mundial tiene dislexia, una dificultad de aprendizaje de origen cognitivo que reduce la legibilidad y la comprensión de la información escrita. El objetivo de esta tesis es mejorar la accesibilidad textual de las personas con dislexia, combinando métodos de validación de interacción persona-ordenador y técnicas de procesamiento del lenguaje natural. Primero estudiamos como las personas con dislexia identifican los errores en un texto y analizamos sus errores escritos. Esto permitió estimar la presencia en la Web de textos escritos por personas con dislexia. Al descubrir que los errores disléxicos están relacionados tanto con la presentación como con el contenido del texto, llevamos a cabo una serie de experimentos utilizando seguimiento automático de la vista (*eye tracking*) para averiguar qué condiciones mejoraban la legibilidad y la comprensión. Después de determinar los parámetros relevantes para la presentación del texto y la modificación de su contenido, implementamos un sistema de simplificación léxica. Los resultados obtenidos y los recursos generados han dado lugar a un modelo, *DysWebxia*, que propone recomendaciones que ya han sido integradas satisfactoriamente en cuatro aplicaciones.

Resum

Un 10% de la població mundial té dislèxia, una dificultat cognitiva que redueix la llegibilitat i la comprensió de la informació escrita. L'objectiu de la tesi és millorar l'accessibilitat textual de les persones amb dislèxia, combinant mètodes de validació de la interacció persona-ordinador i tècniques de processament del llenguatge natural. Vam estudiar com identifiquen els errors d'un text les persones amb dislèxia i vam analitzar els seus errors d'escriptura; això va servir per a estimar la presència a la web de textos escrits per persones amb dislèxia. Un cop establert que els errors dislèctics responen a trets relacionats amb la presentació i el contingut del text, vam realitzar experiments mitjançant el seguiment de la mirada per a descobrir quines condicions milloren la llegibilitat i la comprensió. Havent determinat els paràmetres rellevants per a la presentació del text i la modificació dels continguts, vam implementar un sistema de simplificació lèxica. Els resultats obtinguts i els recursos creats han donat lloc a un model, *DysWebxia*, en el qual es proposen recomanacions que s'han integrat satisfactòriament en quatre aplicacions.

Acknowledgments

This journey started almost four years ago when Ricardo Baeza-Yates thought that I could use my personal and academic background to help people like me. Thank you Ricardo for proposing the right PhD topic for me, which also has fulfilled me personally. Thank you for teaching me professionalism and research soundness. I have the impression I could have learnt much more from you and, at the same time, I think I will understand and learn more in the future from my memories of this process.

A year later Horacio Saggion joined us as co-advisor. I cannot think of better advisors, they always supported me when I needed, were direct and honest with my mistakes, trusted me, and gave me freedom to develop my own ideas.

The first year I also met Gaurang Kanvinde and Vasile Topac, who very early incorporated the first results of this thesis into practical applications. Thanks to them, our research began to become a reality. Seeing its usefulness gave me the strength to keep working. Thank you, Gaurang and Vasile, for making things real.

In the first year was specially difficult to find participants for our studies. It was my friend Joaquim Llisterri who helped me here. He is the only person who has told me openly how terrible I write –apologies for my writing to the readers of this thesis–. Thank you Joaquim, for welcoming me in Barcelona, teaching me perfectionism, and helping me finding participants in a new city for me. Special thanks to Mari-Carmen Marcos for allowing me to use the eye tracker for the experiments.

After the second year, results started to take shape and I went to Pontifícia Universidade Católica do Rio de Janeiro, Brazil, where I had the opportunity of working with Simone Barbosa to start designing the interface of the *DysWebxia Reader*. Thank you Simone for introducing me to interaction techniques and usability evaluations. I am still learning from your recommendations.

The *DysWebxia Reader* could not have been possible without my friends, Clara Bayarri and Azuki Gòrriz. Thank you Clara for offering your time and implementing a robust useful tool, and thank you Azuki for the most beautiful logo (the blue hippo) and graphics.

Also deep thanks to Julia Dembowski, who in the last months improved and optimized the simpler synonym generation algorithm used by *DysWebxia*.

Four years is a long time. There are moments when things get very stressful, results make no sense, programming gets so complicated, things seems impossible. Thank you, Martin Pielot, for being with me in the worse moments. Also, thank you, for the most enriching and patient statistical discussions.

Thank you to my Chilean friends, Eduardo Graells and Diego Saez, for their good disposition to help me with my implementation weaknesses throughout this thesis. Thank you for all those afternoons where the *asados* became the best place to discuss research ideas. I cannot imagine having better PhD classmates.

There are people who believe in a better world and empower others to make good things happen. Thank you Luis Miret for your generosity introducing our research to the right public administrations and companies. Thanks to you, this work can have a greater impact, and hopefully reach more people with dyslexia.

The protagonists of this research are the participants of our studies. Thank you to the more than two hundred people who came to our lab to undertake our experiments: wonderful children and adults, concerned and responsible families with children with dyslexia. Some of the families made a great effort to come to the lab. Their positive feedback full of hope and words of encouragement, motivated me to carry on this work in a way that is difficult to express with words. Thank you for believing in what we were doing. Among many others, I am deeply thankful to Yolanda Otal de la Torre and María Sanz-Pastor Moreno de Alborán, two strong and driven mothers with children with dyslexia, who have specially supported me with real world feedback, and pointing out more families with children with dyslexia.

All this could not have been possible without the people mentioned before whom generously invested their time with the only reward of helping people with dyslexia.

Finally, at a personal level, I thank my parents, Alfonso and Pilar, who, since I was little, sacrificed themselves to support my education, the best present I could ever have. They taught me to work hard and kept being surprised with each one of my achievements. If something went really wrong, I called them as they always helped me to put things in perspective, and made me felt loved. Thank you, Clara, for being a incredible friend and for listening to me speaking about the same topic all these years. Thank you, Miguel, for taking care of me by all possible means during this last period. I especially thank you for bringing to me an optimistic vision of the future.

This research was conducted with funding from a doctoral fellowship (FI-DGR) of the *Generalitat de Calaluyna* (Government of Catalonia) and the HIPERGRAPH grant (TIN2009-14560-C03-01) of the *Ministerio de Economía y Competitividad* (Ministry of Economy and Competitiveness) of Spain.

Barcelona, April 10th, 2014

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

There are three reasons why we decided to start this work, all of them related to its social relevance: dyslexia is frequent and universal; accessibility practices are not only good for dyslexics but also useful for all users; and it contributes to the democratization of information.

Dyslexia. Dyslexia is a cognitive disability that reduces readability and comprehension of written information (see a detailed definition in Section 2.1.1). Even if dyslexia is popularly identified with brilliant famous people, such as Steve Jobs or Steven Spielberg, dyslexia is a extremely serious problem affecting to school performance of a great amount of children, among other problems. Dyslexia is not easy to detect, it is a *hidden* disability. People with dyslexia cannot perceive if they are reading or writing correctly. Dyslexia is characterized by difficulties with accurate word recognition, poor spelling, and poor decoding abilities. This implies that people with dyslexia have more difficulty accessing written information and, as side effect, this impedes the growth of vocabulary and background knowledge [177]. Moreover, there are brain structure, brain function, and genetics studies confirming the biological foundations of dyslexia [398]. Not only this, the most frequent way to detect a child with dyslexia is by low-performance at school [71]. In Spain, approximately four out of six cases of school failure are related to dyslexia [132].[1] Furthermore, dyslexia is frequent. From 10 to 17.5% of the population in the U.S.A. [176], and from 8.6% [183] to 11.8% [71] of the Spanish speaking population have this cognitive disability. In short, dyslexia is universal, frequent and cause serious problems related to school failure.

Good for dyslexics, useful for all. Dyslexia-related difficulties are shared by other groups with special needs, such as people with low vision [130]. Moreover, symptoms of dyslexia are common to varying degrees among most people [114]. For example, Dixon [114] tested a piece of educational software with dyslexic and non–dyslexic readers and the results suggest that the symptoms of dyslexia are common to varying degrees among most people. Also, Pollak [290] showed how students with and without dyslexia benefit from using multimodal documents. According to Zarach [423], the guidelines to enhance readability for people with dyslexia also benefit people without dyslexia. As a matter of fact, dyslexic-accessible practices overlap with general textual accessibility recommendations [228], and with other guidelines for groups with special needs such as blind or low vision readers [130]. Hence, we believe that our work is also extensible to general usability problems and to other target groups.

Information Democratization. Access to information and communication technologies is recognized as a basic human right by United Nations [392]. For this reason, United Nations recommend that all public information services and documents should be accessible to the widest possible readership [247]. At the same time, the essential property of the Web is its universality.[2] The Web is fundamentally designed to work for all people and it removes communication and interaction barriers that many people face in the physical world. Moreover, the Web is an increasingly important resource in many aspects of life such as education, employment or health care. Therefore, web access by

[1] 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 [127].

[2] "The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect." Quote by Tim Berners-Lee, W3C Director and inventor of the World Wide Web [33].

everyone regardless of disabilities is an essential aspect [418].[3] Indeed, the development of this kind of work could improve the ability of people with dyslexia to read and access a wider range of information content. This could help democratize the Web, increasing their participation and improving the overall quality of the content by diversifying the pool of contributors. We started this work focusing on the Web, later we extended our focus to general digital text. Now is the right time to pay attention on making digital text more accessible. For instance, in January 2011 the Association of American Publishers reported that ebook sales increased by 115.8 percent [15].

1.2 Goals

The main goal of this PhD thesis is to improve the textual accessibility for people with dyslexia when they read on a screen. There are four secondary goals:

- To have a deeper understanding of dyslexia by analyzing how people with dyslexia read and write, using their misspelling errors as a starting point (Part II).

- To find out the best text presentation parameters which benefit the reading performance –readability and comprehension– of people with dyslexia (Part III).

- To find out the text content modifications that benefit the reading performance of people with dyslexia (Part IV).

- To propose a set of recommendations combining the positive results, and integrate them in reading applications for people with dyslexia (Part V).

1.3 Approach

To pursuit these goals, we first reviewed the cognitive neuroscience literature to understand the difficulties that people with dyslexia face. We also reviewed the accessibility literature and the applications done

[3] There are four principles that provide the foundation for Web accessibility: the content must be perceivable, operable, understandable, and robust [63].

so far to find out what was missing. From the cognitive neuroscience literature we learnt the language difficulties of dyslexia, and found out that none of the experiments done with language modification included automatic language modification techniques. From the previous accessibility literature we learnt that changing the presentation of the text could ameliorate problems of people with dyslexia. Even if cognitive neuroscience literature is mainly focused on the text content, none of the applications we found modified it.

Then, to find out the language modifications that could be done automatically, we reviewed the natural language processing (NLP) literature. We mapped the linguistic difficulties related to dyslexia with the NLP methods that could be potentially developed. What we found from previous NLP literature is that none of the previous methods have been specifically adapted for this target reading group. For a deeper understanding of dyslexia we also analyzed the reading and writing of people with dyslexia using linguistic knowledge, and measured the presence of dyslexic errors in the Web.

Based on the literature review, we chose the text presentation and content conditions to study. We decided to use an eye tracking methodology because it would give us the possibility of measuring –directly and objectively– how people with dyslexia read under certain textual conditions. Then, we reviewed the previous eye tracking literature where we learnt what to expect from the eye movements of people with dyslexia when they read. However, we did not find any evaluation of accessibility nor of NLP techniques using eye tracking with this target group.

From all conditions tested in the first round of experiments, the most positive results –surprisingly– came from the presentation of the text. We integrated these design recommendations in the first tools, *AccessibleNews DysWebxia* and *IDEAL eBook reader*, and carried on with the second –and final– round of experiments and the implementation of the most promising text content modification technique: automatic lexical simplification. In the evaluation of the algorithm we found that how to present simpler synonyms is more important than the lexical simplification itself. That is, substituting complex words by simpler synonyms was not useful –even counterproductive–, but when they were shown on demand, the text became subjectively easier to read and comprehend. This is why we implemented a new algorithm that would rank simpler synonyms of a word to show the simplest ones for people with dyslexia.

After all the experiments we combined the positive results in a set of recommendations and created a language resource of simpler synonyms to be integrated in other applications. As a result, *Text4all* and *DysWebxia Reader* are currently the only two applications for people with dyslexia that also include text content modifications. Currently we are working on the integration of the recommendations in the educational system and other accessibility projects.

1.4 Main Results and Publications

We summarize the main results of this thesis below. The detailed contributions can be found at the beginning of each chapter.

- The presence of errors written by people with dyslexia in the text does not impact the reading performance of people with dyslexia, while it does for people without dyslexia (Chapter 4).

- Normal –correctly written– texts present more difficulties for people with dyslexia than for people without dyslexia. To the contrary, texts with jumbled letters present similarly difficulties, for both, people with and without dyslexia (Chapter 4).

- Lexical quality is a good indicator for text readability and comprehensibility, except for people with dyslexia (Chapter 4).

- Written errors by people with dyslexia are phonetically and visually motivated. The most frequent errors involve the letter without a one-to-one correspondence between grapheme and phone.[4] Most of the substitution errors share phonetic features and the letters tend to have certain visual features, such as mirror and rotation features (Chapter 5).

- The rate of dyslexic errors is independent from the rate of spelling errors in web pages. Around 0.67% and 0.43% of the errors in

[4]Phone is a speech sound, the smallest discrete segment of sound in a stream of speech.

the Web are dyslexic errors for English and Spanish, respectively. This rates are smaller than expected probably due to spelling correction aids (Chapter 6).

– The different kind of error percentages in the Web are growing. This could be explained by the growing number of new users. However, the lexical quality of the social media –user generated content, mainly– is not as good as the one from the overall Web (Chapter 6).

– Larger font sizes improve the readability, especially for people with dyslexia. (Chapter 7).

– Larger character spacing improve readability for people with and without dyslexia (Chapter 7).

– For reading text based web pages, font size of 18 points ensures good subjective and objective readability and comprehensibility (Chapter 8) .

– *Sans serif*, *monospaced*, and *Roman* font types increase the readability of people with and without dyslexia, while *italic* fonts decrease it. Good fonts for people with dyslexia are *Helvetica*, *Courier*, *Arial*, *Verdana* and *CMU*, taking into consideration both, reading performance and subjective preferences (Chapter 9).

– Frequent words improve readability while shorter words may improve comprehensibility, especially in people with dyslexia (Chapter 10).

– Numbers represented as digits instead of words, as well as percentages instead of fractions, improve readability of people with dyslexia (Chapter 11).

– Graphical schemes improve the subjective readability and comprehensibility of people with dyslexia (Chapter 13).

– Highlighted keywords increases the objective comprehension by people with dyslexia, but not the readability (Chapter 13).

– Lexical simplification via automatic substitution of complex words by simpler synonyms is not helpful. However, showing synonyms on demand improves the subjective readability and comprehensibility of people with dyslexia (Chapter 14).

– A new model called *DysWebxia*, that combines all our results and that has been integrated so far in four reading tools (Chapter 16).

– As a by-product, there are two new available language resources. *DysList*, a list of dyslexic errors annotated with linguistic, phonetic and visual features (Chapter 5); and *CASSA List*, a new resource for Spanish lexical simplification composed of a list of disambiguated complex words, their context, and their corresponding simpler synonyms, ranked by complexity (Chapter 15).

These contributions were presented in the following main publications:

Rello L., Baeza-Yates R., and Llisterri, J. DysList: An Annotated Resource of Dyslexic Errors. In: *Proc. LREC'14.* Reykjavik, Iceland; 2014. p. 26–31 (Chapter 5 [323]).

Rello L., and Llisterri, J. There are Phonetic Patterns in Vowel Substitution Errors in Texts Written by Persons with Dyslexia. In: *21st Annual World Congress on Learning Disabilities (LDW 2012).* Oviedo, Spain; 2012. p. 327–338 (Chapter 5 [332])

Rello L., and Baeza-Yates R. The Presence of English and Spanish Dyslexia in the Web. *New Review of Hypermedia and Multimedia.* 2012;8. p. 131–158 (Chapter 6 [317]).

Rello, L., Kanvinde, G., Baeza-Yates, R. Layout guidelines for web text and a web service to improve accessibility for dyslexics. In: *Proc. W4A 2012.* Lyon, France: ACM Press; 2012 (Chapter 7 [331]).

Rello L., Pielot M., Marcos, MC., and Carlini R. Size Matters (Spacing not): 18 Points for a Dyslexic-friendly Wikipedia. In: *Proc. W4A '13.* Rio de Janeiro, Brazil: ACM Press; 2013 (Chapter 8, [333])

Rello, L. and Baeza-Yates, R. Good Fonts for Dyslexia. *Proc. ASSETS'13.* Bellevue, Washington, USA: ACM Press; 2013 (Chapter 9 [319]).

Rello, L., Baeza-Yates, R., Dempere, L. and Saggion, H. Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia. *Proc. INTERACT '13.* Cape Town, South Africa: IFIP Press; 2013, p. 203–219 (Chapter 10 [322]).

Rello, L., Bautista, S., Baeza-Yates, R., Gervás, P., Hervás, R. and Saggion, H. One Half or 50%? An Eye-Tracking Study of Number Representation Readability. *Proc. INTERACT '13.* Cape Town, South Africa: IFIP Press; 2013, p. 229-245 (Chapter 11 [329])

Rello, L., Baeza-Yates, R., Bott, S. and Saggion, H. Simplify or Help? Text Simplification Strategies for People with Dyslexia. *Proc. W4A '13.* Rio de Janeiro, Brazil: ACM Press; 2013 (Chapter 14 [321], best paper award).

Rello, L. and Baeza-Yates, R. Evaluation of DysWebxia: A Reading App Designed for People with Dyslexia. *Proc. W4A '14.* Seoul, South Korea: ACM Press; 2014 (Chapter 15 [320], best paper nominee).

1.5 Content Structure

After the introduction, Part I, the background (Chapter 2) covers first an explanation of dyslexia and the language difficulties that it implies extracted from the cognitive neuroscience literature. Second, it shows the state of the art focusing on eye tracking research, accessibility studies, NLP literature, and recommendations for dyslexia. While accessibility studies are focused on the presentation of the text, NLP studies focus on the content of the text. The current recommendations for dyslexia which cover both text presentation and content. This part ends explaining the methodologies used in the thesis (Chapter 3).

Part II shows the studies we performed to understand dyslexia. By using eye tracking, we measured how dyslexic written errors and words with transposed interior letters *(e.g. Do you hvae dexiysla?)* have an effect on the reading performance and comprehension of people with and without dyslexia (Chapter 4). We also analyzed the writing errors of people with dyslexia (Chapter 5), and measured the presence of dyslexic written errors in the English and the Spanish Web (Chapter 6).

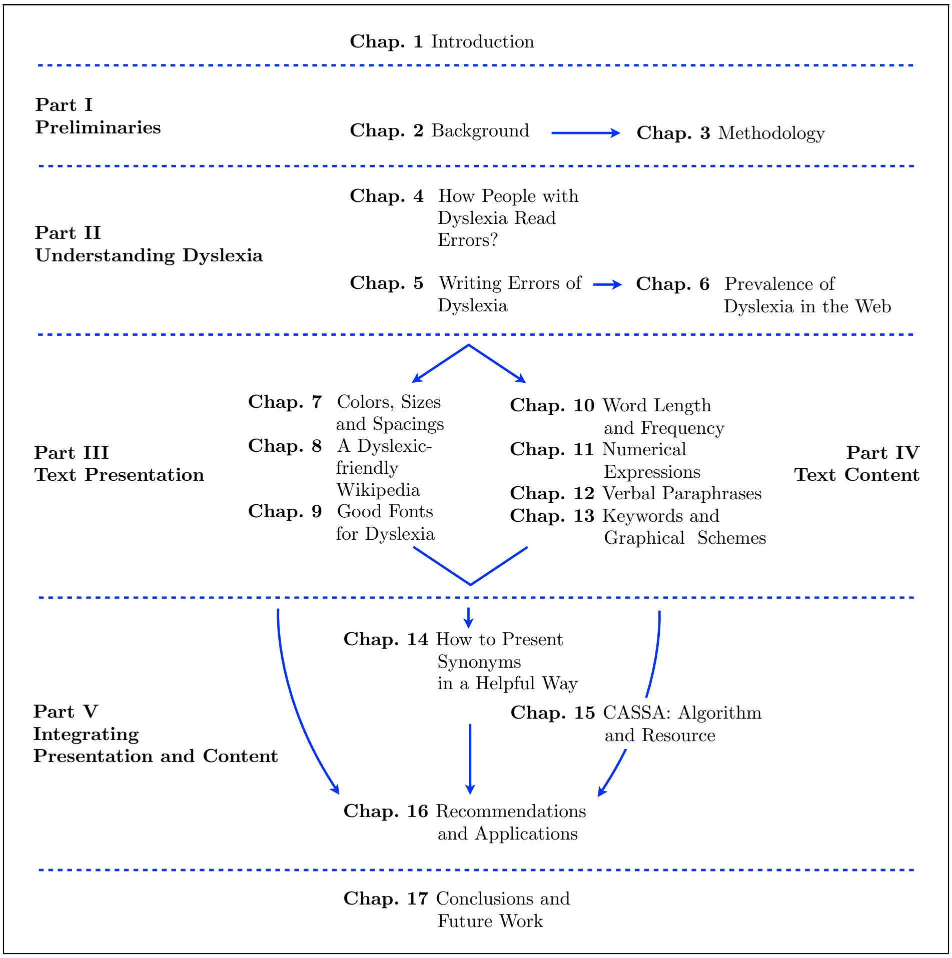


Figure 1.1: Content structure of the thesis.

In Part III we present the eye tracking studies on the textual conditions that have an effect on the reading performance by modifying the presentation of the text. We analyzed the effect of different text and background color, text and background grey scales, font size, column width and spacing between paragraphs, lines and letters (Chapter 7). Then, we measured the effect and the interactions between font size and line spacing in the context of the Web (Chapter 8). Finally, we carried out an experiment which measured the impact of font type[5] (Chapter 9).

Part IV presents the eye tracking studies to find out the textual conditions that have an impact on the reading performance by modifying the content of the text. We tested the effect on readability and comprehension of word frequency and word length (Chapter 10), the type of numerical representations (Chapter 11), verbal simplification via verbal paraphrases (Chapter 12), and the presence of graphical schemes (mind maps) and highlighted keywords in the text (Chapter 13).

In Part V we present an eye tracking experiment to address which is the best strategy to show simpler synonyms for people with dyslexia (Chapter 14). Then, we evaluate a lexical simplification algorithm for people with dyslexia (Chapter 15). Finally, based on the results of parts III and IV, we merge all our results in a set of recommendations to make texts more accessible for people with dyslexia, and show four tools that integrate the *DysWebxia* model.

Finally, this thesis ends with our conclusions for each of the chapters and future lines of research. In Figure 1.1 we summarize the structure of the thesis.

[5]We use font type as a synonym of typeface throughout this thesis.

Part I

PRELIMINARIES

In this first part of the thesis we introduce everything the reader needs to know to understand the rest of the thesis: the background and the methodology used. The background in Chapter 2 covers dyslexia and the state of the art. Since literature about dyslexia is very vast and multidisciplinary, we chose to focus on the definition on dyslexia, the different estimations about the prevalence of dyslexia in different languages, the universality and specificity of dyslexia across languages, the different types of dyslexia, and the language difficulties that people with dyslexia have. The state of the art covers previous studies relevant to the topic of this thesis. These are studies on eye tracking studies for dyslexia readability, accessibility about the text presentation conditions that improve the reading performance of people with dyslexia; natural language processing studies in relation with the language difficulties of people with dyslexia, and previous recommendations on how to make texts friendlier for people with dyslexia by changing the presentation and the context of the text. The last part collects recommendations from readability studies, cognitive science, associations of dyslexia, educational studies, as well as designers. In chapter 3 we explain the methodology of the eye tracking studies that will be followed in most of the chapters of this thesis. Also, we briefly explain the other methodologies used which are further explained in the corresponding chapters.

Chapter 2

Background

In this chapter we explain what is dyslexia, its prevalence, and the difficulties that people with dyslexia face, taking into consideration cognitive neuroscience literature (Section 2.1), as well as previous studies and recommendations on improving the reading of people with dyslexia on a screen (Section 2.2).

2.1 Dyslexia

2.1.1 Definition

Dyslexia is a specific learning disability with neurological origin. It is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede growth of vocabulary and background knowledge [177, 220, 222].

In some literature, dyslexia is referred to as a specific reading disability [398] and dysgraphia as its writing manifestation only [337].[1] However, this thesis follows the standard definitions of the *International Statistical Classification of Diseases (ICD-10)* [417] and the *Di-*

[1] Dysgraphia refers to a writing disorder associated with the motor skills involved in writing, handwriting, and sequencing, but also orthographic coding [34]. It is comorbid with dyslexia [249].

agnostic and Statistical Manual of Mental Disorders (DSM-IV) [9], where dyslexia is listed as a reading and spelling disorder *(ICD-10)* or a reading disorder and a disorder of written expression *(DSM-IV)*.

2.1.2 The Prevalence of Dyslexia

Depending on the language, the estimations on the prevalence of dyslexia differ. The National Academy of Sciences [176] states that 10-17.5% of the population in the U.S.A. has dyslexia. The model of Shaywitz *et al.* [357] predicts that 10.8% of English speaking children have dyslexia while in Katusic *et al.* [189] the rates varied from 5.3% to 11.8% depending on the formula used. Brunswick [57] estimates 10% for English and 3.5% for Italian. Data on the prevalence of dyslexia in Spanish speakers are much more scarce: Galván Gómez [146] reports a 7.5% among school children in Perú; Carrillo *et al.* [71] found that 11.8% of the school children examined in Murcia (Spain) exhibited difficulties associated to dyslexia, and Jiménez *et al.* [183] report an 8.6% for a similar population in the Canary Islands (Spain).

We made an estimation of the presence of dyslexic texts in the Web to know their real impact and our results show that at least 0.63% of the English and 0.43% of the Spanish web pages found in the Web are specifically related to dyslexia. That is, for each 20 billion Web pages, there are at least one million pages containing dyslexic errors (see details of this study in Chapter 6).

2.1.3 Dyslexia among Languages

A considerable amount of studies confirm the biological foundations of dyslexia, with the exception of acquired dyslexia [398]. However, despite its universal neuro-cognitive basis, dyslexia manifestations are variable and culture-specific [155].

This variability is due to the different language orthographies[2] depending on their grade of consistency and regularity. English has an opaque –or deep– orthography in which the relationships between letters and sounds are inconsistent and many exceptions are permit-

[2]In character based languages, such as Mandarin Chinese, dyslexia is associated with multiple deficits, rather than with a core phonological deficit, since this writing system contains a large number of visual symbols or characters that represent units of meaning rather than phonemes as in an alphabet [169].

ted. English presents to the beginning reader a significantly greater challenge compared to other languages, such as Spanish. Spanish has a more regular alphabetic system containing consistent mappings between letters and sounds, that is, a transparent –or shallow– orthography. For instance, Italian dyslexics –shallow orthography– performed better on reading tasks than English and French dyslexics did –deep orthographies [271]. Dyslexia has been called a *hidden* disability due to the difficulty of its diagnosis in languages with shallow orthographies [398].

2.1.4 Types of Dyslexia

Dyslexia is more frequently *developmental*. However, *acquired* dyslexia also exists when specific disorders of reading or writing occur after a brain injury. Dyslexia occurs along a continuum and varies in severity [357], being a persistent and chronic condition [356].

Regarding developmental dyslexia, there is an extensive research literature on subtypes of dyslexia. Vellutino *et al.* [398] review four decades of studies of dyslexia –since the mid 70's to 2004– and decided not to address classifications of dyslexia because according to them, that "research has not been fruitful in enhancing our understanding of dyslexia subtypes at the cognitive level, with few studies finding evidence of relations between subtypes and biological or intervention findings [221]".

Please notice that the following classification of dyslexia comes from studies in English speaking population. English as such has such an opaque –or deep– orthography that having surface dyslexia affects to many processes of reading and writing, while for Spanish this type of dyslexia would not have a significant impact in daily life and would be hard to diagnose.

Researchers broadly agree on three different kinds of dyslexia:[3] phonological, surface and deep dyslexia. However, the delimitation of these three types is not clear and symptoms of different types of dyslexia overlap [141].

Phonological dyslexia[4] is a reading disorder characterized by im-

[3]These kinds are divided into acquired and developmental dyslexia, except for deep dyslexia, which is mostly acquired [86].

[4]Phonological dyslexia was first reported –and coined– by [29].

pairment in non–word[8] reading ability [87]. Surface dyslexia[5] is characterized by poor reading of low frequency irregular words, coupled with accurate reading of non–words. Errors made in reading irregular words tend to be regularizations[6] [268]. While phonological dyslexics use a visual reading route (read words at once), surface dyslexics use a phonological reading route (grapheme-to-phoneme conversion rules) [85]. While phonological dyslexics remember an orthographic and phonological vocabulary, surface dyslexics encounter problems when using the grapheme to phoneme conversion rules. People with deep dyslexia present semantic errors related to a lack of semantic representation retrieval, such as substitutions of entire words among others [90].

Although the classification of the types of dyslexia is based on reading models [85], any dyslexia involves spelling errors [177, 222, 337].

2.1.5 Difficulties of Dyslexia

People with dyslexia encounter problems, not only with some text presentation conditions, such as small font size [109, 228], but also with language-related conditions. Following we present the dyslexia-related difficulties according to their language level. We collected them from the cognitive neuroscience literature, with the exception of the discourse level, where there are recommendations from Web accessibility literature. We have included the poorer reading comprehension which characterized dyslexia in this level too because text comprehension depends on longer segments of texts, not only words.

(a) **Orthography**:

- Orthographically similar words, *addition* and *audition* [125, 376, 377];
- alternation of different typographical cases *e.g. ElefANte* (‘ellefant’) [229]
- letter recognition [35, 48];

[5]Surface dyslexia was fist modeled by [267].

[6]An example of a regularization would be saying /væs/ for the word $< vase >$ /vaz/ (decorative container).

- number recognition [80, 203]; and
- poor spelling, such as letter reversals, *trail* for *trial* [9, 56, 177, 417].

(b) **Phonology**:

- Irregular words, *vase*[7] [89, 125, 376]; and
- homophonic words or pseudo homophonic words, *weather* and *whether* [142, 266].

(c) **Morphology**:

- Derivational errors, *discomfortable* [243, 270, 376, 377, 394].

(d) **Lexicon**:

- New words, *fantabulous* [29, 100];
- pseudo–words and non–words,[8] *happisfaction* [123, 295, 376, 394];
- less frequent words, *pristine* [137, 147, 237, 338, 349];
- long words, *prestidigitation* [100, 269, 414];
- word additions, omissions and word recognition [54, 55, 177, 220];
- substitutions of functional words,[9] *of* by *for* [243, 270, 376, 414];
- confusions of small words, *in* by *is* [100, 404]; and
- misspellings recognition [171, 316].

(e) **Discourse**:

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

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

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

- Short sentences and short paragraphs benefit accessibility for people with dyslexia according to Web accessibility literature [44, 51, 261].
- Poor reading comprehension [398]. In dyslexia is related to decoding and not to problems in oral or listening comprehension [94], that is, in dyslexia, normally poor comprehension is caused by a decoding mistake, such as word recognition. Oral and reading comprehension need to be differentiated, since decoding and listening comprehension have been shown to have different implications in measuring comprehension [190]. Dyslexia affects decoding, but not listening comprehension [163, 244].

Additionally, there are visual difficulties associated with dyslexia [128] that could be alleviated by modifications of the visual display. The most studied in relationship with dyslexia is the visual stress syndrome (Meares-Irlen syndrome) [198]. The Meares-Irlen syndrome is characterized by symptoms of visual stress and visual perceptual distortions that are alleviated by using individually prescribed colored filters. Also patients susceptible to pattern glare, that is perceptual distortions and discomfort from patterns, are prone to Meares-Irlen syndrome and are also likely to find colored filters useful [129]. Kriss and Evans [198] compared colored overlays on a group of 32 children with dyslexia with a control group of 32 children. The difference in prevalence of the two groups did not reach statistical significance. The authors conclude that Meares-Irlen syndrome is prevalent in the general population and possibly a little more common in dyslexia. Children with dyslexia seemed to benefit more from colored overlays than non-dyslexic children. The authors stress that Meares-Irlen syndrome and dyslexia are separate entities and are detected and treated in different ways [198]. Also Jeanes *et al.* showed how color overlays helped the reading of children in school without taking into consideration dyslexia or other visual difficulties [182]. Gregor and Newell [158] and later Dickinson *et al.* [113] have shown that visual changes in the presentation of the text may alleviate some of the problems generated by dyslexia and the visual comorbidities related to dyslexia.

2.2 State of the Art

2.2.1 Eye Tracking

In this section we present a review of previous work about eye tracking and dyslexia.

Rayner [304] presents a review of the studies from the mid 70's to the 90's that have used eye movements to investigate cognitive processes. He argues that eye movement measures can be used to infer moment-to-moment cognitive processes in reading. For instance, shorter fixations are associated with better readability while longer fixations can indicate that processing loads are greater. As a matter of fact, non impaired readers present longer fixations at low-frequency words than at high-frequency words [175, 186, 300, 305, 311].

The eye movements of readers with dyslexia are different from regular readers. People with dyslexia as well as beginner readers, make longer fixations, more fixations, shorter saccades[10] and more regressions than normal readers [3, 122, 126, 208, 226].

During the 80's-90's it was discussed the extent to which eye movements are the cause of reading problems. If eye movements were a causative factor, then dyslexia could easily be diagnosed with a simple eye movement test. However, Tinker [378, 379] and Rayner [304, 307] argue quite strongly that eye movements are generally not a cause of reading disability but were a reflection of other underlying problems.

Following we discuss of three topics that suggest that eye movements might be related to the cause of dyslexia: (1) erratic eye movements, (2) instability during fixation, (3) and selective attentional deficit. After, we present three studies that confirm that eye movements are not *per se* the cause of dyslexia.

First, Pavlidis [272, 273, 274, 275] found that, opposite to regular readers, people with dyslexia made erratic eye movements in non-reading tasks. To Pavlidis, these eye movements are a contributing factor to a reading disability. He argued that faulty eye movements or some type of central temporal ordering problem could be a diagnostic of readers with dyslexia [273]. However, many other investigators could not replicate Pavlidis's findings; hence his hypothesis could not

[10] A rapid movement of the eye between fixation points.

be confirmed [42, 53, 258, 259, 369, 370]. Moreover, these results came from experiments composed of non-reading tasks (readers followed a dot that moved across a screen). For reading tasks, Pavlidis pointed out that the eye movements were inconclusive because one would not be able to tell differences between the eye movement characteristics of dyslexic and non-dyslexic readers. Rayner [304] points out that one possible explanation of this inconsistency in the results of experiments on dyslexia and eye movements is that the variability among readers with dyslexia is greater than in other groups. Pavlidis's results were found in a subgroup of participants with dyslexia [242, 291, 306, 307]. Indeed, case studies of readers with dyslexia [287, 288, 422] agree with Pavlidis in some of the eye movement characteristics of readers with dyslexia. Nevertheless, these differences are not necessarily the cause of dyslexia.

Second, Eden *et al.* [122], using non-reading tasks, found that children with dyslexia had worse eye movement stability during fixation of small targets than normal children. On the other hand, Raymond *et al.* [303] suggested that this instability during fixation did not differ from normal readers. Nonetheless, these results are also not extensible to reading tasks and are not necessarily related to the cause of dyslexia.

Third, Farmer and Klein [131] suggested that readers with dyslexia process less parafoveal[11] information on each fixation than regular readers. Rayner *et al.* [309] and Underwood and Zola [391] used the moving window technique to investigate this issue. They found that the perceptual span was smaller for readers with dyslexia than for regular readers, that is, they showed a selective attentional deficit in the letters belonging to words in parafoveal vision. This selective attentional deficit interferers with the processing of the currently fixated word. However, these findings do not necessarily mean that readers with dyslexia process parafoveal information less effectively than regular readers [310]. This also happens with regular readers and children when they are exposed to difficult texts. Regular readers obtain less parafoveal information when the fixated word is difficult to process

[11] Surrounding the fovea. Fovea is a small rodless area of the retina that affords acute vision.

[166]. Also, when children were given difficult text to read, their perceptual span gets smaller [308]. Later, Geiger and Lettvin [149], and Perry *et al.* [285] proposed that the cause of dyslexia was that readers with dyslexia processed parafoveal information more effectively than regular readers. Indeed, it was argued that people with dyslexia could markedly increase their reading ability by cutting a small window in an index card and reading the material inside the window as they moved it across the text [150]. But these findings could not be replicated by later studies [154, 193, 361]. Only Rayner *et al.* [309] could identify a dyslexic reader with the same characteristic described by Geiger and Lettvin [149]. This reader could identify parafoveal words and letters better than regular readers, and, when reading in a moving window experiment, he read better with a small window than with a larger one. However, this was an atypical reader with dyslexia and a selective attention deficit [309]. Hence, it cannot be argued that eye movements *per se* are the cause of dyslexia.

In addition, there are three studies that are consistent with the conclusion that eye movements reflect the difficulties that dyslexic individuals have reading and are not the cause of the reading problem.

First, Hyöna and Olson [174] found that readers with dyslexia show the typical word frequency effect in which low-frequency words are fixated longer (fixation duration, number of fixations, and regressions) than high-frequency words.

Second, Pirozzolo and Rayner [287] and Olson *et al.* [258] found that when people with dyslexia were given a text appropriate for their reading level, their eye movements (fixations, saccades, and regressions) were much like those of normal readers at that particular age level.

Third, Rayner [308] showed that regular children's eye movements (fixation durations, saccade lengths, and the size of the perceptual span) shared the characteristics of readers with dyslexia when they were given a text that was too difficult for them.

Taking into account all these studies the evidence suggests that for the vast majority of people with dyslexia they have a language processing deficit and that their eye movements simply reflect their difficulty processing language [304].

2.2.2 Text Accessibility

Following we present the user studies that have explored which **text presentation** conditions improve the reading performance of people with dyslexia. These studies come from accessibility literature.

Gregor and Newell [158] asked 12 students with dyslexia to test different colors, sizes, spacing, column widths, and similar letter highlighting to improve the subjective readability of MS Word documents. Later, the best parameters were found by testing seven people with dyslexia, which reported a subjective increase in readability. The results of this investigation were included in *SeeWord*[12] tool for *MS Word* [157]. After, Dickinson *et al.* [113] carried out a pilot study with six participants using the *SeeWord* software showing that reading accuracy improved as well as the subjectively rated reading comfort.

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, and multimedia elements such as subtitles and sign language interpretation for audio and video, and audio description of video material. 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 such as the need to control the speed of the highlighting of the text [286].

Kurniawan and Conroy [199] tested different color schemes for reading online with 27 users with dyslexia. The participants had to read five online articles and undertook comprehension questionnaires after reading. The comprehension of readers with dyslexia was poorer for the complex articles but the reading speed was not slower using the color scheme the users selected.

Santana *et al.* [347] 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. The customization settings included in *Firefixia* are based in previous user studies and recommendations and they include font type, font size, color, character spacing, line spacing, and column width.

[12] http://www.computing.dundee.ac.uk/projects/seeword/

2.2.3 Natural Language and String Processing

Text content in relationship with dyslexia has attracted the attention of many studies. Even though, the use of complicated language has been extensively pointed out as one of the key problems for people with dyslexia [228], all the existing NLP research and NLP based tools that we know are focused on the writing output of people with dyslexia such as spellchekers [160, 196, 206, 213, 276, 278, 277, 368]. For reading, all the existing applications at the moment only alter the design of the text (not the content), such as *SeeWord* [157],[13] *Claro Screen Ruler Suite* [79],[14] or *Colour Explorer*[15] or *Penfriend XL*,[16] or use text-to-speech technology such as *Claro Read*,[17] *ReadingPenTS Oxford*[18] or *DiTres*.[19]

In the existing research for NLP methods for people with special needs, dyslexia is sometimes mentioned as one of the disabilities to be addressed [7, 8, 412] and sometimes it is not [64, 72, 112]. We consider that dyslexia needs to be approached specifically rather than as part of all cognitive disabilities at large. In this section we select a set of NLP techniques that are related –or could be related with some modifications– to each of the language difficulties and the strengths of people with dyslexia.

Orthography: To address the problem that people with dyslexia find with spelling and visual word recognition, it could be used different orthographic similarity measures to created a complexity measure tailored for this target group. For intense, as Coltheart's orthographic neighborhood size metric (*ON*) [88] or the orthographic *Levenshtein 20* –based on Levenshtein distance– which uses a standard metric of string similarity [420]. Also an analysis of errors written by people with dyslexia could show dyslexic related orthographic features.

Phonology: Metrics such as *Phonix* [145], *Soundex* [340] or *Meta-*

[13] `http://www.computing.dundee.ac.uk/projects/seeword/`
[14] `http://www.clarosoftware.com/index.php?cPath=348.`
[15] `http://colour-explorer.software.informer.com/9.0/`
[16] `http://www.penfriend.biz/pf-xl.html.`
[17] `http://claroread.es/`
[18] `http://www.specialneedscomputers.ca/index.php?l=product_detail&p=929`
[19] `http://www.rehasoft.com/dislexia/ditres/`

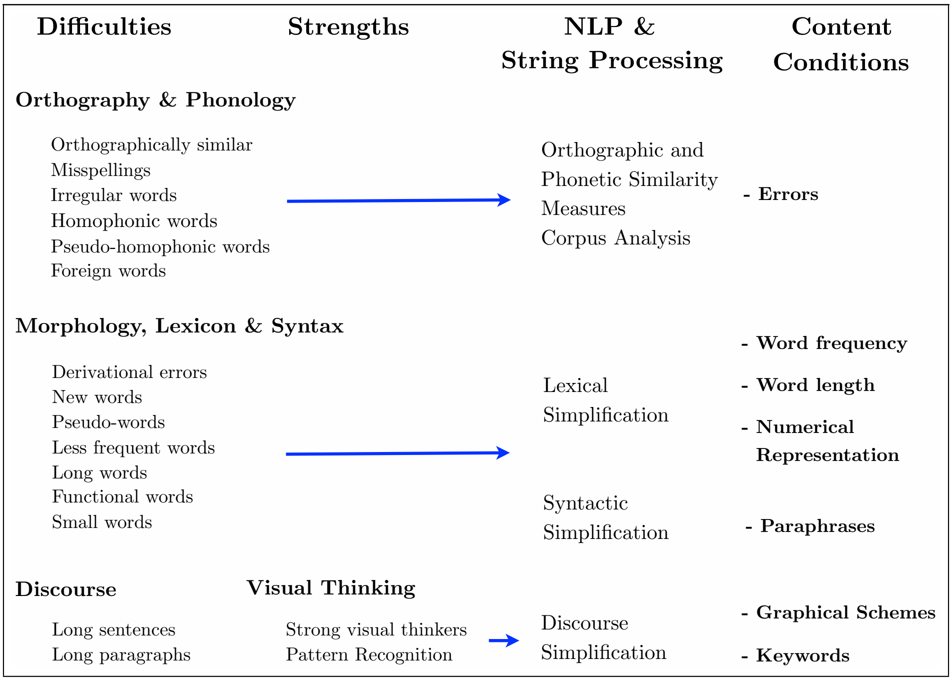


Figure 2.1: Difficulties and strengths of dyslexia in relationship with the NLP and string processing methods and the language conditions involved.

phone [201] compare words with regard to their phonetic similarity. This could be used for creating a complexity measure adapted to people with dyslexia. For instance, *Soundex* indexes words by sound, so each letter is encoded considering its pronunciation. Similar letter pronunciations are grouped together, *e.g.* <b>, <p>, <v>, <f> is one group and <m>, <n> is another group.[20] *Soundex*, *Phonix*, and *Metaphone* algorithms are all designed for English names [117]. We only found one commercial tool to detect Spanish phonetic similarity called *Signum* but we found no documentation. Also a phonetic analysis of errors written by people with dyslexia could show the phonetic characteristics regarding dyslexia.

Morphology and Lexicon: The morphological and lexical diffi-

[20] We use the linguistic conventions: '<>' for graphemes, '/ /' for phonemes and '[ ]' for allophones.

culties can be addressed with NLP methods such as paraphrasing and lexical simplification. Paraphrasing NLP methods recognize and generate phrases, sentences, or longer natural language expressions that convey almost the same information [13]. There are broadly three kinds of paraphrases [225]: (1) lexical paraphrases when individual lexical items have the same or similar meaning (synonyms such as *hot* and *warm* or hyperonyms such as *reply* and *say*); (2) phrasal paraphrases when there are phrasal fragments sharing the same semantic content (*take over, assume control of*); and (3) sentential paraphrases when two sentences represent the same semantic content (*I finished my work, I completed my assignment*). The use of automatic methods for generating paraphrases has been successfully applied for text simplification among other NLP tasks. Actually, paraphrasing is the main strategy used for text simplification [180]. Through different kind of paraphrases the target text is transformed into an equivalent text that is more understandable for a given user. Text simplification is beneficial for many groups of readers, such as non–native language learners [98], low literacy people [148], aphasic readers [72, 227] or deaf people, among others [180]. For instance, text simplification methods were applied to simplify newspaper texts for people with aphasia [72, 74] and Down syndrome [342] as well as to simplify online information for people with aphasia [112].

Discourse: Accessibility literature points out that short sentences and short paragraphs benefits accessibility for people with dyslexia [44, 51, 261]. NLP literature that tackles text simplification for severe cognitive disabilities adapt text by creating a simple and direct style using smaller vocabulary. Usually, one main idea is expressed by a single sentence and unnecessary details are omitted [46]. However, in the case of people with dyslexia details shall not be omitted and semantic information should remain, because normally in people with dyslexia the poor comprehension is caused by decoding mistakes, such as word recognition [163, 244]. Therefore, to make the discourse of the text more accessible we choose keyword highlighting and graphical schemes generation for people with dyslexia.

Previous work has shown that the readability of dyslexic students could be improved by using semantic maps [359] and the creation of graphical schemes is extensively recommended in education literature

[219, 299, 405]. However we did not find any NLP efforts to created mind maps automatically.

Regarding keyphrase[21] extraction, there is a vast amount of NLP literature on the topic [139, 191, 416]. The semantic data provided by keyphrase extraction can be used as metadata for refining NLP applications, such as summarization [105, 205] or text ranking [235]. Also, highlighting keyphrases in the text has been also used facilitate its skimming [390].

To measure the complexity of the text NLP researchers made use of different complexity measures such as the Automated Readability Index (ARI) [351], Coleman-Liau Index (CLI) [83] or Flesch-Kincaid Grade Level Readability Test (FK) [135]. Most of the indexes take into account the number of words per sentence and the number of letters per word. The few metrics that consider linguistic or cognitive knowledge [97, 106, 153] are not as established in literature as the ones mentioned before. Various factors have been applied to measure readability in dyslexics. Classic readability measures are useful to find appropriate reading material for dyslexics [197] and to measure comprehension. For instance, the Flesch-Kincaid readability degree was applied to access comprehension speeds and accuracy in dyslexic readers [199]. Other specific readability measures for dyslexic readers have been proposed in other domains such as information retrieval [360]. In this PhD thesis we do not make use of any of the text complexity measures but we take into consideration the parameters that different complexity measures use [115] to meet comparability requirements among the texts belonging to the same experiment.

In Figure 2.1 we show difficulties and strengths of dyslexia in relationship with the NLP methods ad the language conditions involved.

2.2.4 Recommendations

According to a survey by McCarthy and Swierenga [228], studies about dyslexia and accessibility are scarce compared to other groups of users with special needs. In the *Web Content Accessibility Guidelines* (WCAG) [63], dyslexia is treated as part of a diverse group of

[21] In the works mentioned ahead keyphrase and keywords are used interchangeably.

cognitive disabilities and they do not propose any specific guidelines for people with dyslexia. In some studies [138, 195], dyslexia is part of these cognitive disabilities to be addressed, while in some others dyslexia is not mentioned [10, 43]. Santana *et al.* [109] explain that this lack of explicit consideration of dyslexia specificities in the guidelines make the needs of users with dyslexia unfulfilled. McCarthy and Swierenga [228] point out that the key problems experienced by users with dyslexia when reading on the screen are poor color selections, text too small, and complicated language. According to Paciello [261] and Smythe [362] research about dyslexia and text accessibility is scarce but strong enough to guide more accessible design practices for dyslexia [261, 362].

Since dyslexic-accessible practices benefits also the readability of users without dyslexia [114, 199, 228, 290] and other special needs [130, 250] (see Section 1.1), the guidelines and recommendations for people with dyslexia overlap with many others. At the same time, customization by people with dyslexia is highly recommended [113, 158, 250], mainly because of the high variability of dyslexia, there is no universal profile of a user with dyslexia [157, 290].

The recommendations come from associations of dyslexia, educational studies, and user studies about reading on screen. Studies that take into account dyslexia from the Web accessibility approach are mainly focused on the design characteristics of a web page [228, 261, 362] such as special text formats for dyslexic users [130]. However, guidelines referring to content recommendations are vague [228]. While educational studies focus more in the content of the text, recommendations from associations and accessibility studies focus more on the presentation of the text.

Text Presentation

Following we present the recommendations for text presentation in relationship with the conditions studied in our experiments. These references are extended in their corresponding chapters together with work about the effect of the conditions below on people without dyslexia.

- **Font type (font face)**: There are recommendations from the British Dyslexia Association [52], Evett and Brown [130], Lockley [217], Hornsby [172] and specialized web pages [1, 120]. There are user studies by De Leeuw [108] and Sykes [373] as well as fonts specifically designed for people with dyslexia: *Sylexiad* [168], *Dyslexie* [108], *Read Regular*,[22] and *OpenDyslexic*[23] (See Chapter 9).

– **Text and background colors**: Bradford [51], the British Dyslexia Association [52] and Rainger [296] recommend different colors for people with dyslexia. Gregor and Newell [158] explored in a user study different color combinations. Regarding **text and background grey scale** we found Tseng's recommendations [388] (See Chapter 7).

– **Font size**: The effect of font size was studied with participants with dyslexia by O'Brien *et al.* [256] and Dickinson *et al.* [113]. There are also recommendations by the British Dyslexia Association [52], Rainger [296], Bradford [51], and Zarach [423] (See Chapters 7 and 8).

– **Spacing**: Regarding **character spacing** there is a user study by Zorzi *et al.* [424] and the recommendations of Rainger [296] and Pedley [279]. Bradford [51] tackles **paragraph spacing**, while Rainger [296], Pedley [279] and the British Dyslexia Association [52] give recommendations for **line spacing** (See Chapters 7 and 8).

– **Column width**: There are the recommendations of the British Dyslexia Association [52] and Bradford [51] and the user study by Schneps *et al.* [348] (See Chapter 7).

Text Content

Following we present the recommendations for text content in relationship with dyslexia. These references are extended in their corresponding chapters together with work about the effect of these conditions on people without dyslexia. In Figure 2.1 we put in relationship the text content conditions with the dyslexia related difficulties and the natural language processing methods.

[22] `http://www.readregular.com/`
[23] `http://opendyslexic.org/`

- **Graphical schemes**: since people with dyslexia are strong visual thinkers [409], the use of graphical schemes (mind maps) is extensively recommended in educational studies [75, 219, 299, 405] as they were found to be beneficial for disabled readers [359] (See Chapter 13).

- **Keywords**: in the education literature, highlighting keywords is a broadly recommended learning strategy [407]. Students with dyslexia and teachers are encouraged to highlight keywords to make texts more accessible [164, 280] (See Chapter A.9).

- **Verbal paraphrases**: people with dyslexia have special difficulties with short [100, 404] and functional words [243, 270, 376, 414]. Hence, the use of verbal paraphrases could reduce the use of such words in the sentence (See Chapter 12).

- **Word frequency and word length**: both parameters are strongly related to the difficulties that people with dyslexia find with very long and infrequent words [100, 174, 339, 358, 414] (See Chapter 10).

- **Numerical expressions**: people with dyslexia are more likely to have Mathematical learning difficulties [203] because dyslexia is comorbid with dyscalculia.[24] In fact, people with dyslexia find problems to recognize and recollect not only letters but also numbers [80] (See Chapter 11).

[24]A specific learning disability involving innate difficulty in learning or comprehending arithmetic. It is akin to dyslexia and includes difficulty in understanding numbers, learning how to manipulate numbers, learning mathematical facts, and a number of other related symptoms [61].

Chapter 3

Methodology

In this chapter we explain the methodology of the eye tracking experiments as well as the other methodologies used in this thesis coming from data analysis, linguistics, natural language processing, and usability.

3.1 Eye Tracking User Studies

To study the effect of the textual conditions on objective and subjective readability as well as comprehensibility, we conducted different experiments using eye tracking, comprehension tests, questionnaires, and interviews.

3.1.1 Design

On the experiments we used within-subjects design so each participant contributed to all the conditions. This way we know that the effects found are not due to the group. We always used a within-subjects design when possible except from one experiment that used hybrid-measures design (see Chapter 8). In all the experiments we counterbalanced the conditions and the texts to avoid sequence effects.

Independent Variables

The *independent variables* and their levels changed in each experiment. We performed twelve experiments (and eleven sub experiments). Three of them measure the effect of the presentation of the text, eight measure the effect of modifying the content of the text,

Experiment	**Independent Variables**
Errors I and *II* Chapter 4	[±Errors] *Error Awareness*
Text Presentation Chapter 7	*Color (text/background)* *Grey Scale (text)* *Grey Scale (background)* *Font Size* *Character Spacing* *Line Spacing* *Paragraph Spacing* *Column Width*
Font Chapter 9	*Font Type*
Wikipedia Chapter 8	*Font Size* *Line Spacing*
Word Frequency and Length Chapter 10	[±Frequent] [±Long]
Numerical Representations *Digits vs. Words* *Rounding vs. Decimals* *Percentages vs. Fractions* Chapter 11	 [±Digit] [±Round] [±Percentage]
Verbal Paraphrases Chapter 12	[±Simple]
Graphical Schemes and Keywords Chapter 13	[±Scheme] [±Keywords]
Lexical Simplification Chapter 14	*Lexical Simplification Strategy*
CASSA Chapter 15	*Lexical Simplification Strategy* *Word Frequency*

Table 3.1: Experiments and independent variables.

and two measure the effect of the presence of errors in the text. In Table 3.1 we show all the experiments and the conditions tested.

Dependent Variables		
Type	**Measure**	**Variable Name**
Objetive Variables	Readability	*Reading Time* *Fixation Duration* *Number of Fixations*
	Comprehensibility	*Comprehension Score*
Subjective Variables	Readability	*Readability Rating*
	Comprehensibility	*Comprehensibility Rating*
	Memorability	*Memorability Rating*
	Preferences	*Preference Rating*

Table 3.2: Dependent variables summary.

Dependent Variables

We quantified objective and subjective *readability* and *comprehensibility*. *Readability* refers to the ease a text can be read; and *comprehensibility* refers to the ease a text can be understood. Since readability strongly affects text comprehension [25], sometimes both terms have been used interchangeably [180]. However, when measuring the reading performance of people with dyslexia we need to separate them because they are not necessarily related. In the case of dyslexia, texts that might seen not readable for the general population, such as texts with errors, can be better understood by people with dyslexia, and *vice versa*, people with dyslexia find difficulties with standard texts (Chapter 4). In Table 3.2 we present a summary of the dependent measures used in the eye tracking experiments.

We used the following *dependent variables* to measure objective and subjective *readability* and *comprehensibility*.

Objective Variables

For quantifying objective *readability*, we used the following objective measures: *Reading Time*, *Fixation Duration*, and *Number of Fixations*. These data is extracted directly from the eye tracker software. For quantifying objective *comprehensibility*, we used a *Comprehension Score* as a dependent variable.

Reading Time: Shorter reading durations are preferred to longer ones since faster reading is related to more readable texts [413]. Therefore, we use *Reading Time*, that is, the time it takes a participant to completely read one text, as a measure of readability.

Fixation Duration: When reading a text, the eye does not move contiguously over text, but alternates saccades and visual fixations, that is, jumps in short steps and rests over pieces of text. *Fixation Duration* denotes how long the eye rests still on a single spot of the text. We use the average of the fixation durations as a metric for readability.

Number of Fixations: We use the total number of fixations while reading a text.

Fixation Duration and *Number of Fixations* have been shown to be a valid indicator of readability, eye movement measures can be used to infer moment-to-moment cognitive processes in reading. Shorter fixations are associated with better readability while a greater number of fixations and longer fixations can indicate that processing loads are greater [304]. For instance, people without dyslexia present more and longer fixations on low frequency words than on high frequency words [175, 186, 300, 305, 311].

There are three studies that show why fixation duration and number of fixations are valid indicators for people with dyslexia. These are the studies by Hyöna and Olson [174], Pirozzolo and Rayner [287] and Olson *et al.* [258] and Rayner [308] See Section 2.2.1 for the details of these studies.

Comprehension Score: Moving the eyes over a text does not guarantee its comprehension. To check that the text was not only read, but also understood, we measure text comprehension using questionnaires. We compute the *Comprehension Score* as the average of the answers (see their weights in Section 3.1.3). In some experiments we use the *Comprehension Score* as a *control variable* to guarantee that the recordings analyzed in this study were valid.

Subjective Variables

Readability Rating: We use this rating for quantifying subjective *readability*. The participants rated on a five-point Likert scale, how easy was to read the text.

Comprehensibility Rating: We use this rating for quantifying subjective *comprehensibility*. The participants rated on a five-point Likert scale, how easy was to understand the text.

Preference Rating: In addition, in some experiments we asked the participants to provide their personal preferences using questionnaires with Likert scales.

3.1.2 Participants

For each experiment we had between 23 and 48 participants with dyslexia, and between 23 to 104 participants without dyslexia that served as control group. Their ages ranged from 11 to 54 years old, and the age average is different in each experiment.

To guarantee that dyslexia was diagnosed in an authorized centre or hospital, the participants with dyslexia were asked to bring their diagnoses to the experiment. The Catalonian protocol of dyslexia diagnosis [81] does not consider the different kinds of dyslexia, extensively found in the literature that verses about dyslexia in English. Therefore, we can only guarantee that the participant was diagnosed in a authorized center or hospital but not the exact type of dyslexia. All the participants had normal or corrected to normal vision, and none of them were diagnosed with visual stress (Meares-Irlen) syndrome. The participants were asked about the languages they speak, their level of studies and about their reading habits.

On the following group D refers to the participants with dyslexia and group C for the non-dyslexic participants. We provide the details of the participants in each experiment.

3.1.3 Materials

To isolate the effects of the studied condition, the texts in which the conditions are inserted must be comparable in complexity. In this section we describe how we designed the base texts we used.

Text Content

All the texts used in the experiments meet the comparability requirements because they all share the parameters commonly used to compute text complexity [115]. Within each experiment the texts shared the following characteristics.

(a) They have the same genre.

(b) They have the same style.

(c) They are about similar topics.

(d) They have a similar discourse structure.

(e) They contain the same number of words.

(f) They contain the same number of sentences

(g) They have a similar word length average.

(h) Absence (or same number) of numerical expressions, acronyms, and foreign words.

(i) They contain the same number of unique named entities.

(j) They have the same number of alterations per text depending on the experiment. For instance the same number of highlighted words in the *Keywords* experiment.

Text Presentation

Since the presentation of the text has an effect on the reading speed of people with dyslexia [158], we used the same layout for all the texts within each experiment. The text was unjustified text since justified text alignment produces irregular spacing between words that make reading harder [51, 279, 296]. The font size used ranged from 14 points to 20 points and the column width did not exceeded 70 characters/column, as recommended by the British Dyslexia Association [52]. The color used was either black font on creme background[1] or black text on white background, one of the most frequently used color combinations.

[1]The CYMK are creme (FAFAC8) and black (000000). Color difference: 700, Brightness difference: 244.

El texto habla de: *'The text is about:'*

- ☐ Un sueño. *'A dream.'*
- ☐ Un parque de atracciones. *'An amusement park.'*
- ☐ Un helado de chocolate. *'A chocolate ice cream.'*

Figure 3.1: Comprehension literal item.

For the experiments where font type was not a condition, we choose to present the texts in *Arial* because of three reasons. First, *Arial* is the most common font used on screen for the Web [76]. Second, *Arial* is highly recommended in literature. For instance, Evett and Brown [130] put in comparison recommendations for readers with low vision and dyslexia, and both groups agree in using *Arial* and *Comic Sans*. Also, the British Dyslexia Association also recommends using *Arial*. Third, in Lockley's [217] study, *Arial* was the preferred font. Also in our *Font* experiment using eye tracking with 48 participants with dyslexia (Chapter 9), *Arial* was the font that lead to significantly shorter reading time [319].

Comprehension Questionnaires

Depending on the purpose of the experiment, the questionnaires contained different types of items: inferential and literal items. Inferential questions require a deep understanding of the text. Literal questions can be answered directly from the text (see Figure 3.1). In some test we included literal items related to details involving memory. We included these items in multiple-choice questions with three or four possible choices. The order of the correct answer was counterbalanced. The difficulty of the questions chosen was similar. One of the choices was always correct. The rest could be either wrong or partially correct. To compute the *Comprehension Score*, the choices counted 100%, 50%, and 0%, for the correct, partially correct and wrong, respectively. In some experiments we added the choice "I don't know" (0%).

Subjective Readability and Comprehension Questionnaires

To quantify the *Readability Rating* and the *Comprehensibility Rating* we used questionnaires. For each of the conditions, the participants

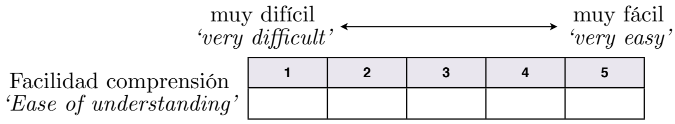


Figure 3.2: *Comprehensibility Rating* item.

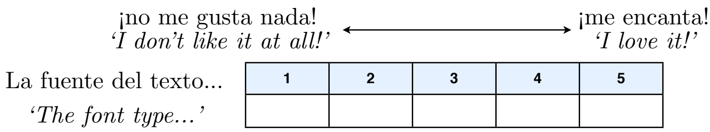


Figure 3.3: *Preference Rating* item.

rated on a five-point Likert scale, to which extent the text was easy to read and understand. Figure 3.2 gives one example scale.

Preference Questionnaires

To gather the participants' preferences, for each of the conditions, the participants rated on a five-point Likert scale, how much they liked certain text presentation parameter or certain text content modification. An example of the items is given in Figure 3.3. When we used different Likert scales we specify in each chapter.

3.1.4 Equipment and Software

The eye tracker we used was the Tobii 1750 [381], which has a 17-inch TFT monitor with a resolution of 1024×768 pixels. The time measurements of the eye tracker have a precision of 0.02 seconds. Hence, all time values are given with an accuracy of two decimals. The eye tracker was calibrated individually 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. For analyzing the eye tracking data we used Tobii Studio 3.0.

3.1.5 Procedure

The experiments were conducted at the Universitat Pompeu Fabra and lasted from 20 to 30 minutes each. They took place in a quiet room, where the participant was alone with the interviewer (author of this thesis), so that the participants could concentrate. Each participant performed the following three steps. First, we began with a questionnaire that was designed to collect demographic information. Second, the participants were given specific instructions. They were asked to read the texts in silence and complete the comprehension questionnaires. The eye tracker recorded the reading. They could not look back on the text when they had to answer the questions. Finally, each participant was asked to provide his/her preference ratings. Depending on the experiment, a semi-structured interview was carried out to collect feedback about the experiment and about the usefulness of the conditions tested.

3.1.6 Experiments

The twelve experiments (and eleven sub experiments, see Table 3.1) were conducted in two studies that lasted around 3 months each. In the first study we carried out the experiments of *Text Presentation*, *Word Frequency*, *Word Length*, *Graphical Schemes*, *Keywords*, *Verbal Paraphrases* and *Errors I*. After we analyzed the data of the first study we took the best results and carried out a second study where we conducted the following experiments: *Font Type*, *Wikipedia*, *Numerical Representations*, *Lexical Simplification* and *Errors II*. In the first study we had a total of 39 participants with dyslexia and 47 participants without dyslexia. In the second study we had a total of 56 participants with dyslexia and 104 participants without dyslexia. Depending on the age of the participant they performed a set of experiments, 2-4 experiments each. Some participants who were willing to help with the study came more than once to the lab and performed more experiments.

3.1.7 Statistical Analysis

For choosing the correct statistical tests to analyze the data, we first checked if each data set presented a normal and homogenous distri-

Design	Number of conditions	Parametric	Non-parametric
With-in subjects	2	Matched-pair t-test	Wilcoxon test
	> 2	One-way, repeated measures ANOVA	Friedman's test
	>2, 2 groups	Two-way, repeated measures ANOVA	Two-way Friedman's test
Between subjects	2	Independent measures t-test	Mann-Whitney test
	> 2	One-way, independent measures ANOVA	Kruskal-Wallis test
	>2, 2 groups	Two-way, independent measures ANOVA	

Table 3.3: Statistical tests.

bution. Whether the data was normally distributed was determined through Shapiro-Wilk tests. To check the homogeneity of variances we used either Barlett's or Levene's tests. Each data set corresponds with the data collected for one variable level of one group. That is, we checked the homogeneity of variance and the variance distribution for each of the levels of the conditions, groups, and dependent variables. Depending on the homogeneity and the variance distribution we chose either parametric or non-parametric tests for the statistical analyses. The debate of what analyses are admissible for Likert scales –parametric or non-parametric tests– is pretty contentious [67], and in this case we used non-parametric tests. Then, if the experiment followed a with-in subject design, we used paired tests for repeated measures, and if the experiment followed a between-subject design, we used independent-measures tests.

The parametric tests for repeated measures used were the matched-pair t-test (two conditions), the one-way repeated measures ANOVA (more than two conditions) and the two-way repeated measures ANOVA (more than two conditions, to compare groups). The parametric tests for repeated measures used were the independent measures t-test (two conditions), the one-way, independent-measures ANOVA (more than two conditions), and the two-way, independent-measures ANOVA (more than two conditions, to compare groups). However, most of the data sets were not homogeneous or normally distributed. Hence, the non-parametric tests for repeated measures used were the Wilcoxon test (two conditions), the Friedman's test (more than two conditions) and the two-way Friedman's test (more than two conditions, to compare groups). In the case that the measures were independent, we used the Mann-Whitney test (two conditions) and the Kruskal-Wallis test (more than two conditions). For post-hoc comparisons, we used the following complete pairwise post-hoc tests: paired t-tests with a Bonferroni adjustment or Holm correction (when we had a repeated measures ANOVA); pairwise Wilcoxon rank sum post-hoc comparison tests with a Bonferroni or Holm correction (when we had a Friedman's test), and Mann-Whitney tests (when we had a Kruskal-Wallis test). To test correlations we used either the Pearson (parametric) or the Spearman (non-parametric) correlation. We explain the tests used in each on the experiments (see Table 3.3).

Regarding the statistical notation, we used $\bar{x}$ for the mean, $\tilde{x}$ for the median, and s for the standard deviation.

The software used was the R 2.14.1 statistical tool [294].

3.2 Other Methodologies

In this thesis we used other methodologies that are further explained in detail in their corresponding chapter.

3.2.1 Web Data Analysis

In order to measure the presence of dyslexic errors in the Web we used a web mining approach, designing a measure to estimate a lower bound of the presence of dyslexia. We made use of the word frequencies in the Web using the hit counts of search engines together with real document frequencies computed from Yahoo! web search index (see Chapter 6).

3.2.2 Linguistics and Language Processing

For the analysis of the writing errors of people with dyslexia we used corpus creation, annotation and linguistic error analysis (see Chapter 5). For the creation of the lexical simplification algorithms LexSiS (Section 14.3) and CASSA (Section 15.3) we used NLP techniques such as lemmatization, statistical analysis, word sense disambiguation using the context and the exploitation of resources such as the Google Books Ngram Corpus [234]. To evaluate the quality of the NLP resources generated we used online questionnaires with rankings.

Throughout the thesis, examples of words with errors are preceded by an asterisk '*', followed by the correct word. Also, when giving examples with words in Spanish, we give right after the translation to English (between parenthesis and using single quotes).

3.2.3 Usability

We performed two usability studies, one for the *IDEAL eBOOK reader* for Android (14 participants with dyslexia) and one for the *DysWebxia Reader* for iOS (12 participants with dyslexia), see Sections 16.4.4 and 16, respectively. The usability evaluations followed a within-subject design, where all the participants had to perform some tasks using the think aloud protocol [211]. They also undertook a questionnaire and semi-structured interviews. A semi-structured interview is open, allowing new ideas to be brought up during the interview as a result of what the interviewee says.

Part II

UNDERSTANDING DYSLEXIA

In this part we present the first studies that we conducted to understand dyslexia. We decided to use the errors of people with dyslexia as the starting point because they have been used as a source of knowledge for various purposes such as studying dyslexia, diagnosing dyslexia, or for accessibility related purposes. First in Chapter 4 we present two eye tracking experiments that address how people with dyslexia read words with errors or words with jumbled letters, that is, when the internal letters of the word are randomly ordered. Second, in Chapter 5 we explain the design of the first corpus and a list of errors written by people with dyslexia. Then, we present the analyses of these written errors from a phonetic and a visual point of view. As the results show that dyslexic errors are visually and linguistically motivated, they motivate our subsequent work on the presentation and the content of the text in parts III and IV, respectively. Finally, in Chapter 6 we first classify of errors found in the Web and devise a methodology to estimate the prevalence of dyslexic errors in the Web using the written errors of people with dyslexia already mentioned. We obtain a lower bound of the presence of dyslexia in the Web, extended to major Internet domains, social media sites and throughout English and Spanish speaking countries.

Chapter 4

How People with Dyslexia Read Errors?

4.1 Introduction

Lexical quality broadly refers to the degree of quality of words in a text (spelling errors, typos, etc.) [283]. The presence of errors in the text is broadly used to rank its quality. For instance, in schools, the spelling error rate is one of the indicators of the quality of a text. In the context of the Web, lexical quality is a good estimator of the quality of the text of a website [151]. The main reason why lexical quality is used as an indicator is because the quality of words impacts the readers' readability and understanding. For example, the presence of errors in the text deteriorates the readability and the comprehension of regular readers.

However, the impact of the presence of errors in the readability and the comprehension of people with dyslexia has not yet been studied. In this chapter we explore the effect of lexical quality in people with dyslexia. The main contributions of this chapter are the following:

– The presence of errors in the text (lexical quality) is not a good indicator for the readability and the comprehensibility for people with dyslexia.

– In general, the presence of errors in the text does not impact the reading performance (readability and comprehension) of people with dyslexia, while it does for people without dyslexia. In the few cases that there is an effect of the presence of errors on people with dyslexia, those effects are significantly smaller than the ones observed for people without dyslexia under the same conditions.

– People with dyslexia are less aware of written errors than people without dyslexia.

– Lexical quality can be used as an accessibility metric for Web text comprehensibility, except for people with dyslexia.

Part of the results of this chapter are presented in Rello and Baeza-Yates [315, 316].

4.2 Related Work

There are two related work areas: neuropsychology literature about dyslexia and word recognition, and visual word recognition research.

4.2.1 Non-word Reading Deficit

In neuropsychology literature, one of the characteristics of people with dyslexia is that they present a non-word reading deficit. This deficit is not observed in people without dyslexia and it has been taken as evidence to support a phonological impairment underlying dyslexia [295]. The non-word reading deficit was first described for English by Rack, Snowling and Olson [295] and later for other languages with more straightforward grapheme-phone correspondences as German [415] or Spanish [353]. However, in languages with transparent orthographies it seems that speed problems (slower) are more evident and relevant than accuracy problems [353, 387]. In these they use studies non-words, that is, non existing but possible words in language for instance *molsmi* or *brigbert* would be examples of difficult non-words, and *blem* or *tig* examples of easy non-words [295]. The non-word reading deficit has been studied and discussed in relationship with other parameters such as the word length effect [367] or the effect of syllable frequency. Carreiras and Perea found no effects regarding syllable frequency [69], while Levelt and Wheeldon did find effects [209].

On the other hand there are a number of studies which explore the parameters involved in word recognition (not non–words), such as word and syllable frequency [68] as well as the effect of orthographic neighborhood in visual word recognition [70, 395]. Regarding people with dyslexia, they show inaccurate and slower word-recognition skills. According to Bruck [55], people with dyslexia strongly relay on the use of spelling-sound information, syllabic information, and context for word recognition. These difficulties in word-recognition are similar in both adults and dyslexic children [55].

4.2.2 Visual Word Recognition

In relationship with word recognition there are a number of studies that have explored to which extent visual abilities are related to word identification [398]. For instance, Vellutino [399], and Vellutino and

Scanlon [400] presented to two groups (good and poor readers) letters and words that were visually similar (,b> and <d>, *was* and *saw*). They found no differences among groups in memory for the letters and words that were visually similar. It was found that for people with and without dyslexia visual abilities were relatively poor predictors of word identification, spelling, pseudo–word decoding, and reading comprehension. This provided strong evidence that reading is primarily a linguistic skill [136, 143, 398].

In his PhD thesis, Rawlinson [302] explored the impact of letter position in word recognition. In 16 experiments he tested different kinds of letter randomizing, letter order reversal, word order reversal, letter reversal, letter image reversal, randomized vowels, substituted vowels, substituted letters (for vowels), and randomized replacement taking into consideration different grades of confusion in the replacements (for instance <b> with a <d> is more 'confusable' than replacing <b> with a <u>). The experiments show the flexibility of the reading process. One of the most relevant results for our study was that middle letter identification proceeds independently from their position. In other words, letter randomization in the middle of the word had little effect on the ability of skilled readers to understand the printed text [301].

Later in 2003 a text started to circulate in the Internet coining the term *The Cambridge University effect*,[1] even if the text does not refer to any true research project that was ever conducted at the University of Cambridge.

Velan and Frost [396] used rapid serial visual presentation[2] to tests the effects of letter transpositions between English and Hebrew with 28 students –bilingual in English and Hebrew– who read 40 sentences (20 in English and 20 in Hebrew). For Hebrew, they found that transpo-

[1] "Aoccdrnig to rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a total mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe."

[2] Rapid serial visual presentation is an experimental model to examine the temporal characteristics of attention. In [396] sentences were presented rapidly on the screen word by word, and participants had to reproduce the sequence of words perceived.

sitions deteriorated dramatically the performance of the participants, while transpositions in English had little effect on performance.

Grainger in 2008 [156] affirmed that the encoding of a letter within a word has become a key issue for any model of visual word recognition. There is substantial empirical evidence, obtained from different paradigms [2, 119, 312], that shows that transposed-letter pseudo–words are perceptually very similar to their base words, and they tend to be (initially) misperceived as their corresponding base words. We now focus in the literature that has studied the letter transposition effect on eye movements and reading cost.

Rayner *et al.* [312] performed an experiment with 30 college students who read 80 sentences in which letters were transposed, 40% of the words, that is, all content words longer than four letters. The readings were recorded via eye tracking and the comprehension was addressed via questionnaires. There was a 11% decrement in reading speed when the internal letters of the words were transposed. They report higher decrements for letter transpositions at the ending of the words –26% decrement–, and the beginning of the words –36% decrement–. Also, even if the participants answered the comprehension questions with high accuracy, 50% of them did not understand all the words in the sentences. This experiment was replicated in Japanese by Mori and Komatsu [241] and found the same pattern of reading delays but in smaller magnitude.

Later Johnson *et al.* [184] performed four experiments using eye tracking that support that exterior letters in the word play important roles in visual word recognition in transposed-letter silent reading. Consistently, in the experiments by White *et al.* [410] where transposed-letters were tested in different parts of the word (external, internal, beginning and end), show that word-initial transpositions cause the greatest interferences. Similarly, in Japanese, character transpositions at the beginning of words cause the largest delay in reading [241].

Further from letter-transposition, some studies have explored the effect of including numbers and symbols among the letters in the word (*e.g. YESTERDAY I SAW THE SECRE74RY*) [118], or the effect of using short message service (SMS) language *e.g. my hols wr gr8 (my holidays were great)* [281]. Results of both studies showed that there is a reading cost.

4.2.3 What is Missing?

Exploraing the effect of orthographic dyslexic errors and letter-transposition on the readability and comprehension of people with dyslexia.

4.3 Methodology

To study the effect of written errors on text readability and comprehensibility on the screen, we conducted two experiments with Spanish native speakers: *Errors I* and *Errors II*, with 44 participants (22 with dyslexia) and 78 participants (39 with dyslexia), respectively. All the participants had to read a set of texts with varying rates of written errors. Readability and comprehensibility were analyzed via eye tracking and comprehension tests, respectively. Via questionnaires we gathered the participants' subjective ratings of readability and comprehensibility, as well as their error awareness.

4.3.1 Design

In this section we explain the methodology which is specific of the experiments *Errors I* and *Errors II*. The rest of the methodological details are found in Chapter 3 where we explain the methodology shared by all the experiments. Refer to Table 4.1 for a summary of the *Errors I* experiment and Table 4.2 for a summary of the *Errors II* experiment.

Independent Variables

- **Errors I:** In our experimental design, [±Errors] served as an *independent variable* with 3 levels: [No errors], [16% errors−Explicit] and [16% errors+Explicit]. The condition [No errors] denotes that the text was presented without any errors, while [16% errors−Explicit] and [16% errors+Explicit] denote the conditions when the texts were presented with 12 errors per text (each text had 77 words, 16% error rate) and the participants knew that the text had errors. [16% errors−Explicit] denotes the condition where the text was presented with errors and the participant was not informed about the presence of errors. The condition [16% errors+Explicit] denotes when when the participant knew in advance that the text to read had errors.

The word length average in the texts with conditions [16% errors+Explicit] and [16% errors−Explicit] were $\bar{x} = 4.83 \pm 3.07$.

- **Errors II:** In our experimental design, [±Errors] served as an *independent variable* with 4 levels: [8% errors−Explicit], [8% errors+Explicit], [50% errors short+Explicit] and [50% errors long+Explicit]. The condition [8% errors−Explicit] denotes when the text was presented with 6 errors (in a text of 75 words, 8% error rate) and the participant was not informed about the presence of errors in the text. The condition [8% errors+Explicit] denotes when the text was presented with 6 errors and the participant was explicitly informed about the presence of errors in the text. The conditions [50% errors short+Explicit] and [50% errors long+Explicit] denote when 49.3% of the words in the text (37 of 75 words per text) were presented with their letters in random other, except the first and last letter of the word. For instance, **pitraa (pirata, 'pirate')* or **mdnuo (mundo, 'world')*. These two last conditions are visually explicit as almost all words have errors.

  Since the base texts used in *Errors I* for the [16% errors±Explicit] conditions were the same used in *Errors II* for the [8% errors±Explicit], the word length average is the same, $\bar{x} = 4.83 \pm 3.07$. The words in the text [50% errors short+Explicit] were short, with a word length average of $\bar{x} = 4.27 \pm 2.39$. The words in the text [50% errors long+Explicit] were longer, with a word length average of $\bar{x} = 5.12 \pm 3.65$.

We used a within-subject design, that is, each participant read all the texts, contributing to each of the conditions. To avoid sequence effects, we counter-balanced texts as far as possible. We needed to maintain that [−Explicit] condition was always before that its equivalent in errors [+Explicit] condition, that is, [8% errors−Explicit] was always before [8% errors+Explicit] and [16% errors−Explicit] was always before [16% errors+Explicit].

Dependent Variables

For quantifying readability and comprehension as well as error awareness we used the *dependent measures* presented in Tables 4.1 and 4.2. For quantifying objective readability we used *Reading Time*, *Fixation Duration*, and *Number of Fixations*, extracted from the eye tracking data. For quantifying the text comprehension of the texts we used *Comprehension Score*. To measure the readers subjective perception towards the text we used *Readability Rating* and the *Comprehensibility Rating*.

The dependent measures shared in most of the experiments are extensively explained in the Section 3.1.1. The only dependent variable specific of this experiment is *Error Awareness*. We define *Error Awareness* as how the participant perceived the amount of errors in the texts. These data were collected via questionnaires (see Section 4.3.3).

4.3.2 Participants

The details of the participants groups for both experiments are given in Table 4.1 and Table 4.2. In *Errors II* three of the participants with dyslexia were also diagnosed with attention deficit disorder. For more details about the participants please refer to Section 3.1.2.

4.3.3 Materials

Base Texts

To isolate the effects of the presence of errors in the text, the base texts were comparable in complexity. The texts had the same genre, style and similar topics: news about culture. They contain the same number of words: 77 in *Errors I* and 75 in *Errors II*.

Errors

The errors used in conditions [8% errors−Explicit], [8% errors+Explicit], [16% errors−Explicit] and [16% errors+Explicit] are dyslexic errors according to our classification of different kind of errors (Section 6.3.3). To generate the errors we dictated the texts to a person with dyslexia who copied the text. The written errors she wrote were used in the experiments. The rest of the errors were extracted from *DysList*, a list of 1,171 different word-error

¿Cuántos errores tenía el texto? *'How many errors had the text?:'*

☐ 100% de palabras. *'100% of misspelled words.'*

☐ 75% de palabras. *'75% of misspelled words.'*

☐ 50% de palabras. *'50% of misspelled words.'*

☐ 25% de palabras. *'25% of misspelled words.'*

☐ 0% de palabras. *'0% of misspelled words.'*

Figure 4.1: Error survey item.

pairs extracted from 83 texts written by children with dyslexia (Section 5.4). We used a total of 24 dyslexic errors *(Errors I* used the same texts and errors –half less– than *Errors II)*, that are found in Appendix A.1.

Text Presentation

We used the parameters detailed in Section 3.1.3 with black font on creme background and 20 points font size.

Comprehension Questionnaires

We used two comprehension items (one inferential and one literal) for each of the texts (Tables 4.1 and 4.2). See Section 3.1.3 for details about the creation of the comprehension questionnaires.

Error Survey

After reading each of the texts, except from the one with [No errors], the participants were asked to estimate how many errors the texts had. Below in Figure 4.1 we show the survey items for condition [50%±explicit]. All the possible answers for the error survey are given in Table 4.7.

4.4 Results

In this section, we present the analyses of the data of both groups, D and N. Shapiro-Wilk tests showed that only some of the data sets were normally distributed (only some from group C). Also, Levene tests showed that the data sets were not homogeneous.

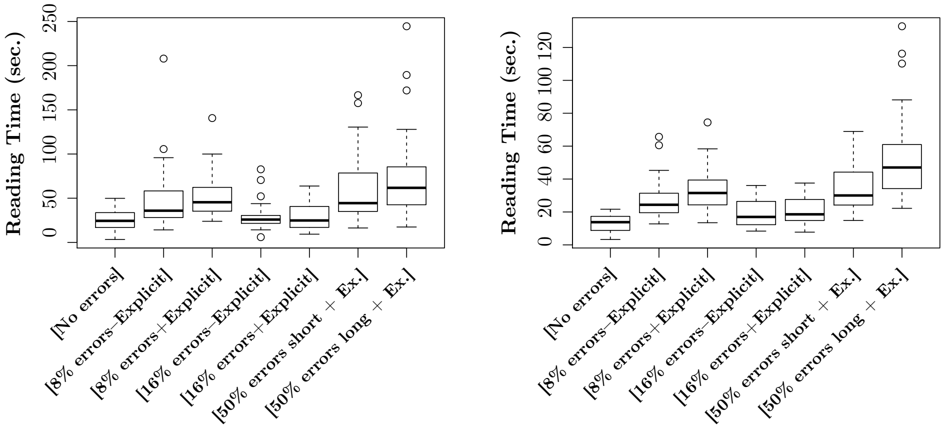


Figure 4.2: *Reading Time* means for *Errors I* and *Errors II* experiments for group D (left) and group C (right).

Hence, to study the effects of the conditions or readability and comprehensibility we used the two-way Friedman's non-parametric test for repeated measures plus a complete pairwise Wilcoxon rank sum post-hoc comparison test with a Bonferroni correction that includes the adjustment of the significance level. Then, to show effects of the conditions within groups, we divided the data for each group and used Friedman's non-parametric test for repeated measures plus a complete pairwise Wilcoxon rank sum post-hoc comparison test with a Bonferroni adjustment.

4.4.1 Objective Readability

Table 4.3 shows the main statistical measures for the *Reading Time*, *Fixation Duration* and *Number of Fixations* for each of the conditions of both experiments.

Errors I

Reading Time. There was a significant effect of [±Errors] on *Reading Time* ($\chi^2(2) = 20.61,\ p = 0.001$) (See Figure 4.2). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly longer reading times ($\bar{x} = 28.52$, $s = 15.51$ seconds) than the participants without dyslexia ($\bar{x} = 17.69$, $s = 8.04$ seconds, $p < 0.001$).

- **Group D:** We did not find a significant effect of the conditions on *Reading Time* in group D.

- **Group C:** There was a significant effect of [±Errors] on *Reading Time* in group C. Text with [No errors] lead to shorter reading times than the text with [16% errors−Explicit] ($p = 0.001$) and with [16% errors+Explicit] ($p < 0.001$).

Fixation Duration. There was a significant effect of [±Errors] on *Fixation Duration* ($\chi^2(2) = 20.61$, $p < 0.001$) (See Figure 4.3). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly longer fixation duration ($\bar{x} = 0.23$, $s = 0.07$ seconds) than the participants without dyslexia ($\bar{x} = 0.21$, $s = 0.04$ seconds, $p < 0.001$).

- **Group D:** We did not find a significant effect of the conditions on *Fixation Duration* in group D.

- **Group C:** In group C, there was a significant effect of [±Errors] on *Fixation Duration*. Text with [No errors] lead to shorter fixation durations than the text with [16% errors−Explicit] ($p = 0.003$) and with [16% errors+Explicit] ($p < 0.001$).

Number of Fixations. There was a significant effect of [±Errors] on the *Number of Fixations* ($\chi^2(2) = 8.72$, $p = 0.013$) (See Figure 4.4). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly more fixations ($\bar{x} = 117.48$, $s = 48.16$) than the participants without dyslexia ($\bar{x} = 82.86$, $s = 25.51$, $p < 0.001$).

- **Group D:** We did not find a significant effect of the conditions on *Number of Fixations* in group D.

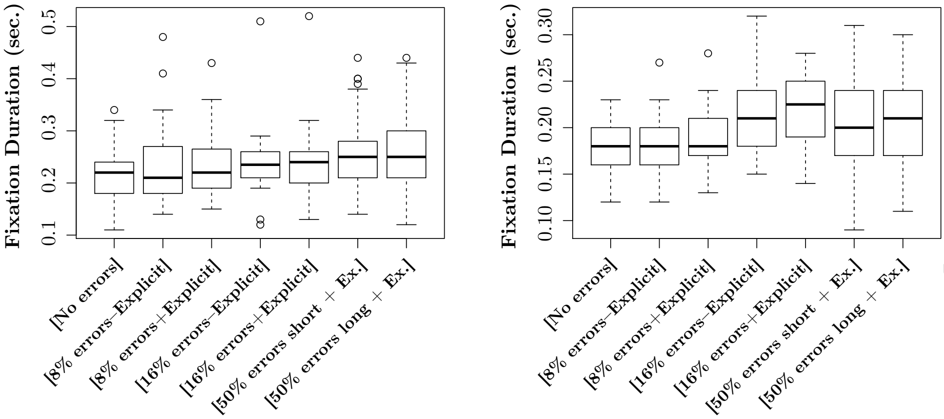


Figure 4.3: *Fixation Duration* means for *Errors I* and *Errors II* experiments for group D (left) and group C (right).

- **Group C:** In group C, there was a significant effect of [±Errors] on *Number of Fixations*. Text with [No errors] presented fewer fixations than the text with [16% errors−Explicit] ($p = 0.011$) and with [16% errors+Explicit] ($p < 0.001$).

Errors II

Reading Time. There was a significant effect of [±Errors] on *Reading Time* ($\chi^2(3) = 102.33,\ p < 0.001$) (See Figure 4.2). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly longer reading times ($\bar{x} = 56.98$, $s = 34.47$ seconds) than the participants without dyslexia ($\bar{x} = 36.50$, $s = 18.39$ seconds, $p < 0.001$).

- **Group D:** There was a significant effect of [±Errors] on *Reading Time* in group D. Texts with [8% errors−Explicit] lead to shorter reading times than texts with [50% errors long+Explicit] condition ($p < 0.001$).

- **Group C:** There was a significant effect of [±Errors] on *Reading Time* in group C. Texts with [8% errors−Explicit] lead to shorter fixation durations than texts with [8% errors+Explicit]

($p = 0.027$), [50% errors short+Explicit] ($p = 0.010$), and [50% errors long+Explicit] ($p < 0.001$). Also, texts with [50% errors long+Explicit] condition lead to longer reading times than [50% errors short+Explicit] ($p < 0.001$), [8% errors+Explicit] ($p < 0.001$), and [8% errors−Explicit] ($p < 0.001$).

Fixation Duration. There was a significant effect of [±Errors] on *Fixation Duration* ($\chi^2(3) = 54.87,\ p < 0.001$) (See Figure 4.3). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly longer fixation duration ($\bar{x} = 0.24$, $s = 0.07$ seconds) than the participants without dyslexia ($\bar{x} = 0.19$, $s = 0.04$ seconds, $p < 0.001$).

- **Group D:** We did not find a significant effect of the conditions on *Fixation Duration* for group D.

- **Group C:** For group C, there was a significant effect of [±Errors] on *Fixation Duration.* Texts with [8% errors−Explicit] lead to shorter fixation durations than texts with condition [50% errors short+Explicit] ($p = 0.036$), and [50% errors long+Explicit] ($p = 0.007$).

Number of Fixations. There was a significant effect of [±Errors] on *Number of Fixations* ($\chi^2(3) = 72.00,\ p < 0.001$) (See Figure 4.4). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly more fixations ($\bar{x} = 171.05$, $s = 82.53$) than the participants without dyslexia ($\bar{x} = 128.22$, $s = 63.70$, $p < 0.001$).

- **Group D:** There was a significant effect of [±Errors] on *Number of Fixations* in group D. Texts with [8% errors−Explicit] lead to shorter fixation durations than texts with conditions [8% errors+Explicit] ($p = 0.052$), and [8% errors+Explicit] condition ($p = 0.005$).

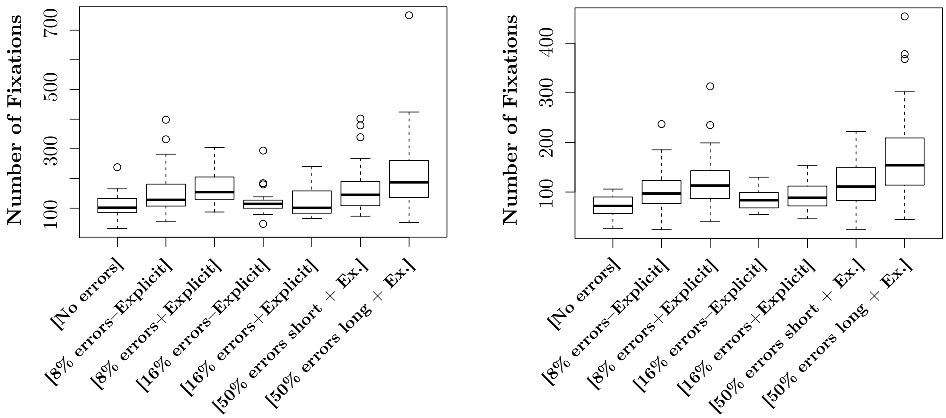


Figure 4.4: *Number of Fixations* means for *Errors I* and *Errors II* experiments for group D (left) and group C (right).

- **Group C:** For group C, there was a significant effect of [±Errors] on *Number of Fixations*. Texts with [50% errors long+Explicit] lead to more fixations than text with conditions [8% errors−Explicit] ($p < 0.001$), [8% errors+Explicit] ($p < 0.001$), and [50% errors short+Explicit] ($p < 0.001$).

4.4.2 Objective Comprehensibility

Table 4.4 shows the main statistical measures for the *Comprehension Score* for each of the conditions of the experiments *Errors I* and *II*.

Errors I

Comprehension Score. We did not find a significant effect of [±Errors] on the *Comprehension Score* ($\chi^2(2) = 2.60$, $p = 0.273$). However, the results of the post-hoc tests show that:

- **Between Groups:** We found no effects of [±Errors] on the *Comprehension Score* between groups ($p = 0.54$).

- **Group D:** We did not find any effects of [±Errors] on *Comprehension Score* in group D.

- **Group C:** For group C, there was a significant effect of [±Errors] on *Number of Fixations*. Texts with [16% errors+Explicit] presented a significant lower *Comprehension Score* than text with [No errors] ($p = 0.080$).

Errors II

Comprehension Score. We did not find a significant effect of [±Errors] on the *Comprehension Score* ($\chi^2(3) = 74.37,\ p < 0.001$) (Figure 4.5). The results of the post-hoc tests show that:

- **Between Groups:** We found no effects of [±Errors] on the *Comprehension Score* between the groups ($p = 0.170$).

- **Group D:** For group D, there was a significant effect of [±Errors] on the *Comprehension Score*. Texts with [50% errors short+Explicit] presented a significantly higher *Comprehension Score* than texts with [8% errors−Explicit] ($p = 0.004$), [8% errors+Explicit] ($p < 0.001$), and [50% errors long+Explicit] ($p < 0.001$).

- **Group C:** For group C, there was a significant effect of [±Errors] on the *Comprehension Score*. Texts with [8% errors−Explicit] presented a significantly lower *Comprehension Score* than texts with [50% errors short+Explicit] ($p = 0.002$). Texts with [8% errors+Explicit] presented a significantly lower *Comprehension Score* than texts with [50% errors short+Explicit] ($p < 0.001$) and [50% errors long+Explicit] ($p = 0.034$). Texts with [50% errors short+Explicit] presented significantly higher *Comprehension Score* than texts with [50% errors long+Explicit] ($p = 0.010$).

4.4.3 Reading Efficiency

For each participant we also calculated his/her *Reading Efficiency*, defined as the *Comprehension Score* divided by the *Reading Time*. In Table 4.5 we show the means of the *Reading Efficiency* per condition and group. Last column represents the ratio of the *Reading Efficiency* of group D divided by the one of group C. This ratio allows us to compare both experiments, because even if the complexity of the texts read might differ, the relative performance among groups should not.

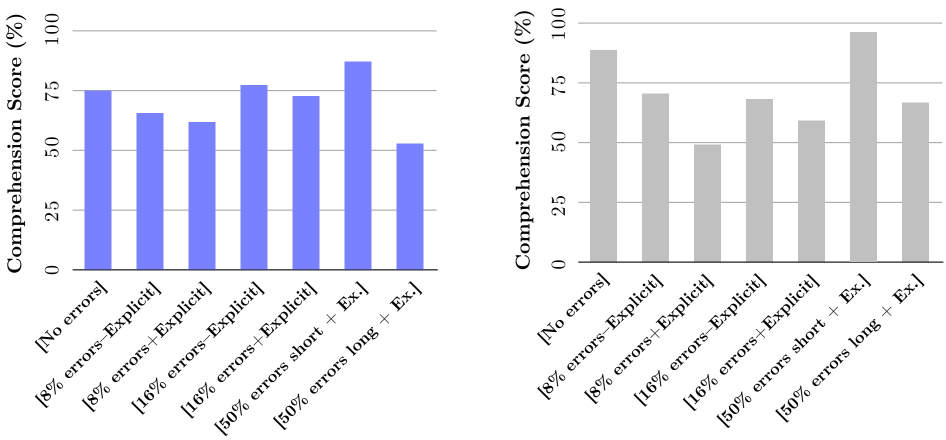


Figure 4.5: *Comprehension Score* means for *Errors I* and *Errors II* experiments for group D (left) and group C (right).

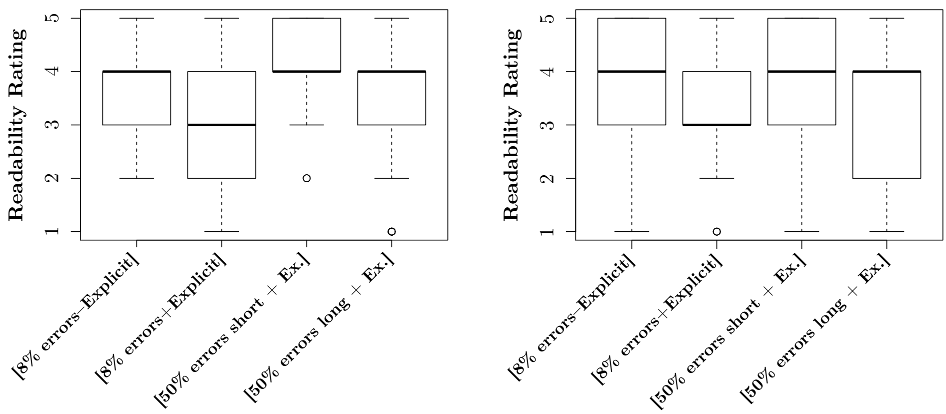


Figure 4.6: *Readability Rating* means of *Errors II* experiments for group D (left) and group C (right).

There was a significant effect of the *Reading Efficiency* between groups. Participants without dyslexia had a significantly higher *Reading Efficiency* ($\bar{x} = 3.15$, $s = 3.55$) than the participants with dyslexia ($\bar{x} = 2.08$, $s = 2.28$, $p < 0.001$).

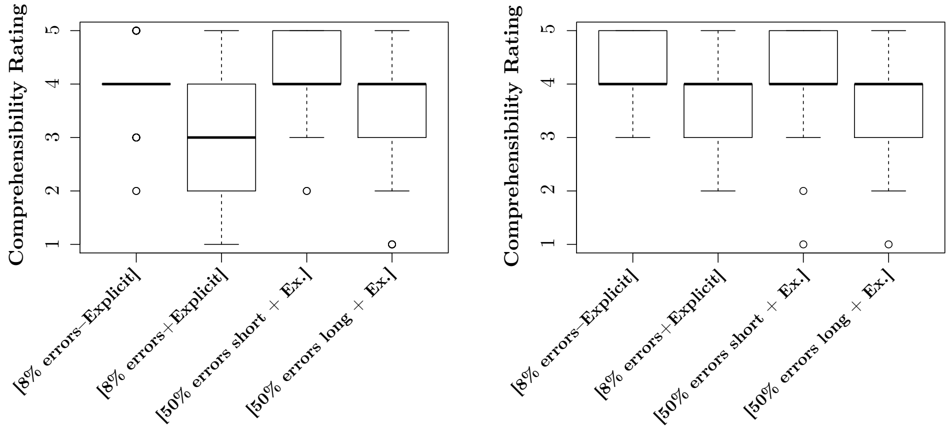


Figure 4.7: *Comprehensibility Rating* means of *Errors II* experiments for group D (left) and group C (right).

4.4.4 Subjective Readability

For *Errors II* we collected data for the *Readability Rating* and the *Comprehensibility Rating* of the participants. Table 4.6 shows the main statistical measures obtained.

Readability Rating. There was a significant effect of [±Errors] on *Number of Fixations* ($\chi^2(3) = 47.93$, $p < 0.001$), see Figure 4.6. The results of the post-hoc tests show that:

- **Between Groups:** We found no effects of [±Errors] on the *Readability Rating* between the groups ($p = 0.830$).

- **Group D:** There were significant effects of [±Errors] on the *Readability Rating* in group D. Texts with [8% errors+Explicit] we found to be less readable than texts with [8% errors−Explicit] ($p = 0.003$), and texts with [50% errors short+Explicit] ($p < 0.001$). Texts with [50% errors short+Explicit] were found to be more readable than texts with [50% errors long+Explicit] ($p = 0.005$).

- **Group C:** For group C, there was a significant effect of [±Errors] on the *Readability Rating*. Texts with [50% errors long+Explicit] were found less readable than text with [8% errors−Explicit] ($p = 0.025$).

4.4.5 Subjective Comprehensibility

Comprehensibility Rating. There was a significant effect of [±Errors] on *Number of Fixations* ($\chi^2(3) = 77.15$, $p < 0.001$), see Figure 4.7. The results of the post-hoc tests show that:

- **Between Groups:** There was a significant effect of [±Errors] on the *Comprehensibility Rating*. Participants without dyslexia found texts significantly more comprehensible ($\bar{x} = 3.94$, $s = 0.93$) than the participants with dyslexia ($\bar{x} = 3.67$, $s = 1.08$, $p = 0.011$).

- **Group D:** There were significant effects of [±Errors] on the *Comprehensibility Rating* in group D. Texts with [8% errors+Explicit] were found to be less comprehensible than texts with [8% errors−Explicit] ($p < 0.001$), and text with [50% errors short+Explicit] ($p < 0.001$). Texts with [50% errors long+Explicit] were found to be less comprehensible than texts with [8% errors−Explicit] ($p = 0.047$), and [50% errors short+Explicit] ($p = 0.003$).

- **Group C:** For group C, there was a significant effect of [±Errors] on the *Comprehensibility Rating*. Texts with [8% errors+Explicit] were found to be less comprehensible than texts with [8% errors−Explicit] ($p < 0.001$), and [50% errors short+Explicit] ($p = 0.001$). Texts with [50% errors long+Explicit] were found to be less comprehensible than texts with [8% errors−Explicit] ($p < 0.001$), and [50% errors short+Explicit] ($p = 0.004$).

4.4.6 Error Awareness

In Table 4.7 we show the distribution of the *Error Awareness Rate* by group and conditions for both experiments. In Figure 4.8 we show the relative percentages of the error awareness rate for each group and condition in comparison with the real error rate.

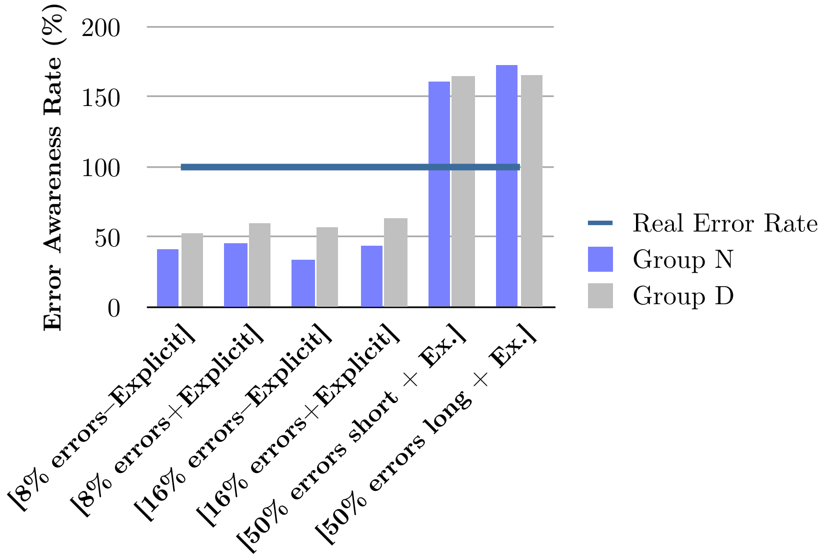


Figure 4.8: Relative percentages of the *Error Awareness Rate* per condition and group in comparison with the real error rate.

Errors I

Error Awareness Rate. In Figure 4.7 we present the percentages of the distribution of the *Error Awareness Rate*.

- **Between Groups:** There was no effect of the group on the *Error Awareness Rate* ($\chi^2(4) = 21.65,\ p < 0.001$). Hence there is a significant difference between groups on the awareness of errors.

- **Explicit:** There was an effect of the [±Explicit] condition on the *Error Awareness Rate* ($\chi^2(4) = 12.84,\ p = 0.012$). So the distribution of the *Error Awareness Rate* is not affected by the condition [±Explicit].

In *Errors I*, the Spearman's correlation coefficient between group D and group C on *Error Awareness Rate* is $\rho = 0.799$ ($p < 0.001$). In Table 4.8 we present the Spearman's correlation coefficients between conditions and groups.

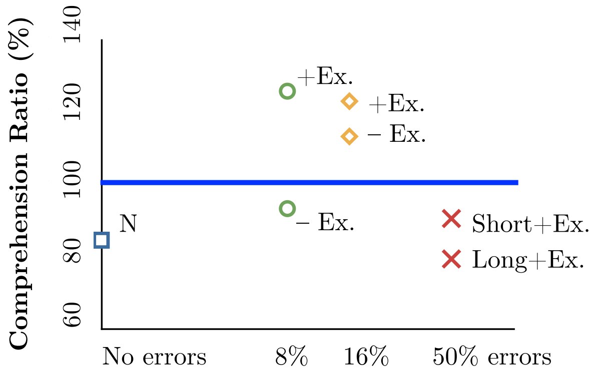


Figure 4.9: *Comprehension Score* percentage ratio per condition.

Errors II

Error Awareness Rate. In Figure 4.7 we present the percentages of the distribution of the *Error Awareness Rate*.

- **Between Groups:** There was no effect of the group on the *Error Awareness Rate* ($\chi^2(3) = 17.67,\ p = 0.001$). Hence the distribution of the *Error Awareness Rate* is not different among groups.

- **Explicit:** There was no effect of the [±Explicit] condition on the *Error Awareness Rate* ($\chi^2(3) = 9.16,\ p = 0.027$). Hence the distribution of the *Error Awareness Rate* is not affected by the condition [±Explicit].

In *Errors II*, the Spearman's correlation coefficient between group D and group C on *Error Awareness Rate* is $\rho = 0.470$ ($p < 0.001$). See Table 4.8 for all the the Spearman's correlation coefficients between conditions and groups for *Errors II*.

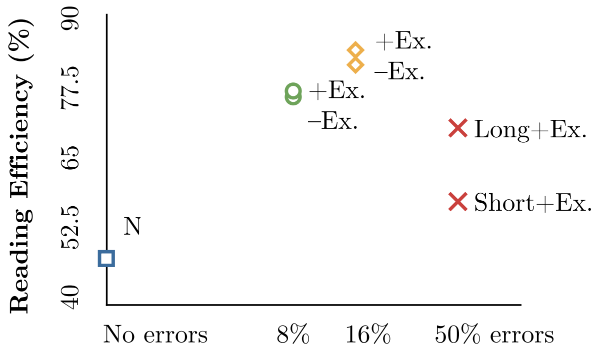


Figure 4.10: *Reading Efficiency* percentage ratio per condition.

4.5 Discussion

The main result of this chapter is that people with dyslexia is not consciously aware of the errors in the text. This fact is shown in our experiments in several ways. First, the reading time when the text has errors increases less in people with dyslexia. In fact, there are significant differences in *Errors I* for people without dyslexia while for people with dyslexia there is not. On the other hand, in both experiments there were significant differences on the number of fixations for the people with dyslexia, implying that they do see the errors but process them in a different way.

Moreover, the comprehension score does not seem to be affected by errors for people with dyslexia. In fact, in three of the experiments people with dyslexia had better average comprehension performance, a fact that we did not find in any of our other experiments (Chapters 8, 10, 11, 12, and 13). The texts of the experiments were designed to have an equivalent complexity, however we computed the relative percentage comprehension ratio to make the experiments more comparable. This ratio is plotted in Figure 4.9 where the improvements for the people with dyslexia are clearly seen.

Another experimental proof of the main result is on the awareness rate. There were significant differences between both populations, showing again that people with dyslexia were less aware of the errors.

Finally, to put everything together we computed reading efficiency, that is, the comprehension score divided by the reading time. Here again we compute the percentage ratio that is shown in Figure 4.10. Clearly the reading efficiency with respect to the people without dyslexia increases reaching almost 90% for the case of 16% errors. Also, the reading efficiency is not affected by the knowledge of knowing that there are errors (both cases are almost in the same place for 8% and 16% errors), a reaffirmation of our main finding. Moreover, for the case of letter transposition the people with dyslexia improves even more for the case of long words, which is exactly the case where they usually have more problems, as we will see later in this thesis (Chapter 10).

In summary, errors equalized the reading field between people with and without dyslexia. This implies that students with dyslexia should not be penalized for spelling errors, as they do not see them in the same way as people without dyslexia do.

Errors I Experiment		
Design	Within-subjects	
Independent Variables	[±Errors]	[No errors]
		[16% errors−Explicit]
		[16% errors+Explicit]
	Error Awareness	[+Explicit]
		[−Explicit]
Dependent Variables (Sec. 3.1.1)	*Reading Time*	(objective readability)
	Fixation Duration	
	Number of Fixations	
	Comprehension Score	(objective comprehensibility)
	Error Awareness Rate	(subjective error awareness)
Participants (Sec. 3.1.2)	Group D (22 participants)	12 female, 10 male **Age:** range from 13 to 37 ($\bar{x}$ = 20.59, s = 8.32) **Bilingual:** Catalan (12), English (1), Italian (1) **Education:** high school (11), university (8), no higher education (3) **Reading:** more 8 than hours (3), 4-8 hours (8), less than 4 hours/day (11)
	Group C (22 participants)	13 female, 9 male **Age:** range from 13 to 35 ($\bar{x}$ = 21.27, s = 8.89) **Bilingual:** Catalan (14), English (2), French (1) **Education:** high school (9), university (12), no higher education (1) **Reading:** more than 8 hours (0), 4-8 hours (9), less than 4 hours/day (13)
Materials (Sec. 3.1.3)	Base Texts	3 texts
	Errors	24 errors (12 errors per text)
	Text Presentation	
	Comprehension Quest.	3 inferential and 3 literal items (2 items of each type per text)
	Error Survey	2 items (1 item/text with errors)
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750	
Procedure (Sec. 3.1.5)	**Steps:** Instructions, demographic questionnaire, reading task (×3) comprehension questionnaires (×3) and error survey (×2)	

Table 4.1: Methodological summary of the experiment *Errors I*.

Errors II Experiment		
Design	Within-subjects	
Independent Variables	[±Errors]	[8% errors−Explicit]
		[8% errors+Explicit]
		[50% errors short+Explicit]
		[50% errors long+Explicit]
	Error Awareness	[+Explicit]
		[−Explicit]
Dependent Variables (Sec. 3.1.1)	*Reading Time*	(objective readability)
	Fixation Duration	
	Number of Fixations	
	Comprehension Score	(objective comprehensibility)
	Readability Rating	(subjective readability)
	Comprehensibility Rating	(subjective comprehensibility)
	Error Awareness Rate	(subjective error awareness)
Participants (Sec. 3.1.2)	Group D (39 participants)	20 female, 19 male **Age:** range from 11 to 45 ($\bar{x}$ = 21.15, s = 9.39) **Bilingual:** Catalan (13), English (1), French (1) **Education:** high school (20), university (16), no higher education (3) **Reading:** more than 5 hours (1), 3-5 hours (11), less than 2 hours/day (27)
	Group C (39 participants)	23 female, 16 male **Age:** range from 11 to 43 ($\bar{x}$ = 26.56, s = 8,79) **Bilingual:** Catalan (11), English (3) **Education:** high school (12), university (25), no higher education (2) **Reading:** more than 5 hours (4), 3-5 hours (17), less than 2 hours/day (18)
Materials (Sec. 3.1.3)	Base Texts	4 texts
	Errors	12 errors (6 errors per text) 50% words with random letters/text
	Text Presentation	
	Comprehension Quest.	4 inferential and 4 literal items (2 items of each type/text)
	Sub. Readability Quest.	4 Likert scales
	Sub. Comprehension Quest.	4 Likert scales
	Error Survey	4 items (1 item/text with errors)
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750	
Procedure (Sec. 3.1.5)	**Steps:** Instructions, demographic questionnaire, reading task (×4) comprehension questionnaires (×4), subjective readability questionnaire (×4) subjective comprehension questionnaire (×4) and error survey (×4)	

Table 4.2: Methodological summary of the experiment *Errors II.*

	Group D		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
Errors I	*Reading Time*			
[No errors]	24.38	**24.21** ± 12.16	13.80	**12.99** ± 8.03
[16% errors−Explicit]	25.84	30.29 ± 17.92	16.97	18.97 ± 8.44
[16% errors+Explicit]	24.81	31.06 ± 15.68	18.57	21.11 ± 5.23
Errors II	*Reading Time*			
[8% errors−Explicit]	35.89	**46.53 ± 31.05**	24.37	**26.94 ± 10.59**
[8% errors+Explicit]	45.45	51.85 ± 22.19	31.29	32.69 ± 11.78
[50% errors short+Ex.]	44.44	59.18 ± 33.94	30.36	34.22 ± 13.75
[50% errors long+Ex.]	61.73	70.68 ± 43.23	47.03	52.13 ± 23.87
Errors I	*Fixation Duration*			
[No errors]	0.22	0.22 ± 0.05	0.18	0.18 ± 0.03
[16% errors−Explicit]	0.24	0.24 ± 0.07	0.21	0.22 ± 0.05
[16% errors+Explicit]	0.24	0.25 ± 0.08	0.23	0.22 ± 0.04
Errors II	*Fixation Duration*			
[8% errors−Explicit]	0.21	0.23 ± 0.07	0.18	0.18 ± 0.03
[8% errors+Explicit]	0.22	0.23 ± 0.06	0.18	0.19 ± 0.03
[50% errors short+Ex.]	0.25	0.26 ± 0.07	0.20	0.21 ± 0.05
[50% errors long+Ex.]	0.25	0.25 ± 0.08	0.21	0.21 ± 0.04
Errors I	*Number of Fixations*			
[No errors]	101.5	108.09 ± 45.61	72.0	69.68 ± 21.35
[16% errors−Explicit]	114.5	121.55 ± 49.17	83.5	85.46 ± 22.11
[16% errors+Explicit]	101.0	122.82 ± 50.43	88.5	93.46 ± 27.64
Errors II	*Number of Fixations*			
[8% errors−Explicit]	128	146.68 ± 66.12	97	100.98 ± 39.71
[8% errors+Explicit]	154	169.96 ± 56.17	113	121.24 ± 49.31
[50% errors short+Ex.]	145	164.51 ± 74.07	111	115.29 ± 48.48
[50% errors long+Ex.]	187	203.51 ± 113.53	154	175.39 ± 82.94

Table 4.3: Median, mean and standard deviation of the *Reading Time*, *Fixation Duration*, and *Number of Fixations* for the *Errors* experiments.

	Group D		Group C		Ratio
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$	%
Errors I	*Comprehension Score*				
[No errors]	100	75.00 ± 42.96	100	**88.64 ± 30.60**	84.61
[16% errors−Explicit]	100	**77.27 ± 42.89**	100	68.18 ± 47.67	113.33
[16% errors+Explicit]	100	**72.73 ± 45.58**	100	59.09 ± 50.32	123.08
Errors II	*Comprehension Score*				
[8% errors−Explicit]	50	65.69 ± 35.34	100	**70.65 ± 37.00**	93.35
[8% errors+Explicit]	50	**61.76 ± 30.96**	50	49.20 ± 35.68	125.86
[50% errors short+Ex.]	100	87.25 ± 24.17	100	**96.08 ± 13.22**	90.61
[50% errors long+Ex.]	50	52.94 ± 35.23	50	**65.67 ± 35.02**	79.42

Table 4.4: Median, mean and standard deviation of the *Comprehension Score* for the *Errors* experiments.

	Group D	Group C	Ratio
	$\bar{x} \pm s$	$\bar{x} \pm s$	%
Errors I	*Reading Efficiency*		
[No errors]	4.30±4.27	**8.96**±7.40	48.00
[16% errors−Explicit]	3.71±3.64	**4.56**±3.74	81.36
[16% errors+Explicit]	3.06±2.82	**3.65**±3.89	83.90
Errors II	*Reading Efficiency*		
[8% errors−Explicit]	2.02±1.61	**2.66**±1.86	75.84
[8% errors+Explicit]	1.32±0.76	**1.72**±1.54	76.85
[50% errors short+Ex.]	1.85±1.10	**3.20**±1.30	57.81
[50% errors long+Ex.]	0.94±0.86	**1.34**±0.97	70.50

Table 4.5: Mean and standard deviation of the *Reading Efficiency* for the *Errors* experiments.

Errors II	**Group D**		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Readability Rating			
[8% errors−Explicit]	4	3.78 ± 0.78	4	**3.91 ± 1.00**
[8% errors+Explicit]	3	3.06 ± 1.07	3	**3.39 ± 1.16**
[50% errors short+Ex.]	4	**4.10 ± 0.81**	4	3.83 ± 0.99
[50% errors long+Ex.]	4	**3.37 ± 1.15**	4	3.25 ± 1.17
	Comprehensibility Rating			
[8% errors−Explicit]	4	4.02 ± 0.73	4	**4.39 ± 0.68**
[8% errors+Explicit]	4	3.06 ± 1.08	3	**3.59 ± 0.86**
[50% errors short+Ex.]	4	**4.20 ± 0.80**	4	**4.20 ± 0.81**
[50% errors long+Ex.]	4	3.39 ± 1.22	3.5	**3.57 ± 1.04**

Table 4.6: Median, mean and standard deviation of the *Readability Rating* and the *Comprehensibility Rating* for *Errors II*.

	Error survey	**Group D**	**Group C**	**Group D**	**Group C**
Errors I	*There are...*	*Error Awareness Rate* (%)		Number of Errors ($\bar{x}$, rel. %)	
[16% errors −Explicit]	*0-2 errors*	72.73	9.09	4.05 (33.71)	6.86 (57.20)
	3-5 errors	4.55	45.45		
	6-8 errors	4.55	22.73		
	9-11 errors	18.18	13.64		
	12 or more	0.00	9.09		
[16% errors +Explicit]	*0-2 errors*	22.73	0.00	5.27 (43.94)	7.63 (63.64)
	3-5 errors	40.91	36.36		
	6-8 errors	27.27	40.91		
	9-11 errors	4.55	18.18		
	12 or more	4.55	4.55		
Errors II	*There are...*	*Error Awareness Rate* (%)		Number of Errors ($\bar{x}$, rel. %)	
[8% errors −Explicit]	*0-2 errors*	80.39	57.41	2.47 (41.18)	3.16 (52.61)
	2-3 errors	5.88	11.11		
	3-4 errors	3.92	1.85		
	4-5 errors	5.88	7.41		
	5-6 errors	3.92	22.22		
[8% errors +Explicit]	*0-2 errors*	68.63	33.33	2.75 (45.75)	3.59 (59.80)
	2-3 errors	13.73	18.52		
	3-4 errors	1.96	18.52		
	4-5 errors	5.88	7.41		
	5-6 errors	9.80	22.22		
[50% errors short+Explicit]	*0% errors*	0	0	80.39 (160.78)	82.35 (164.71)
	25% errors	1.96	1.85		
	50% errors	9.80	7.41		
	75% errors	52.94	51.85		
	100% error	35.29	38.89		
[50% errors long+Explicit]	*0% errors*	0	0	86.27 (172.55)	82.84 (165.69)
	25% errors	0	0		
	50% errors	3.92	5.56		
	75% errors	47.06	55.56		
	100% error	49.02	38.89		

Table 4.7: Distribution of the *Error Awareness Rate* for the *Errors* experiments.

Error Awareness Rate				
Conditions	**Group D** and **Group C**			
	ρ	p	ρ	p
[16% errors−Ex.] and [16% errors+Ex.]	0.641	0.001	0.234	0.295
[8% errors−Ex.] and [8% errors+Ex.]	0.231	0.102	0.480	< 0.001
[50% errors short+Ex.] and [50% errors long+Ex.]	0.309	0.023	0.111	0.438

Conditions	**Group D** and **Group C**	
	ρ	p
[16% errors−Explicit]	0.188	0.402
[16% errors+Explicit]	0.108	0.631
All conditions, *Errors I*	0.799	< 0.001
[8% errors−Explicit]	0.088	0.541
[8% errors+Explicit]	0.027	0.851
[50% errors short+Explicit]	0.164	0.249
[50% errors long+Explicit]	0.187	0.190
All conditions, *Errors II*	0.470	< 0.001

Table 4.8: Spearman's correlation coefficients between conditions and groups for *Errors I* and *Errors II*.

Chapter 5

Writing Errors of Dyslexia

5.1 Introduction

The errors that people with dyslexia write are very valuable and have been used for various purposes, ranging from diagnosing dyslexia to software applications targeted to people with dyslexia. However, resources such as corpora or lists of dyslexic errors are scarce. In this chapter we present the creation of a corpus and a list of errors written by people with dyslexia in Spanish. We compare the different types of dyslexic errors in Spanish and English and present the linguistic, phonetic and visual analyses of the errors in Spanish. The main contributions of this chapter follow:

– The first approach to create a corpus of texts written by people with dyslexia in Spanish *(DysCorpus)*, guidelines for the annotation with linguistic information of dyslexic errors and, a comparison of our corpus with a similar corpus in English.

– The resource *DysList*, composed of a list of unique errors extracted from *DysCorpus*. The errors are annotated with linguistic, phonetic and visual information. To the best of our knowledge this is the largest resource of this kind.

– The different types of dyslexic errors in English and in Spanish present similar distributions frequencies and the differences in the distributions is expected due to the different orthographies of the languages.

– In Spanish substitution errors are the most frequent errors (58.84%), while transpositions the less frequent type (1.45 %). The most frequent error are: inserting or deleting a consonant (37.9%), deleting or inserting a vowel (37.5%), and substituting two letters of similar sound or deleting/inserting an <h> (15.4%).

– Dyslexic errors are visually motivated. More concretely, taking into account handwriting we found that:

  – 46.91% of the error letters had a mirror feature.
  – 38.23% of the error letters had a rotation feature.
  – 67.11% of the error letters are fuzzy letters.

– Dyslexic errors are phonetically motivated. More concretely:

  – The most frequent errors involve letters which the one-to-one correspondence between graphemes and phones is not maintained <b, v>, <g, j>, <c, z>, <c, s>, <r>, and <h> which corresponds with no sound in Spanish.
  – Vowels substitutions occurs more frequently in unrounded vowels, ([i], [e], [a]). Most of the substitutions of vowels occur in diphthong (0.94%), being [ea] and [eo] the two most frequent errors.
  – Consonant substitutions occurs more frequently in single consonants (46.37%). Most of the substitutions errors between consonants occur within the same class of consonants, only 5.52% of the errors between consonants do not have any phonetic feature in common.

The resources and the analyses of this chapter are presented in Rello *et al.* [323], Rello and Llisterri [332] and Rello *et al.* [328].

5.2 Related Work

In this section we present the previous approaches in relationship with corpora of texts written by people with dyslexia, lists of dyslexic errors, and the use of dyslexic errors as a source of knowledge.

5.2.1 Language Resources of Dyslexia

Corpora of Dyslexic Texts

To the best of our knowledge, there is only one corpus of dyslexic texts, the corpus used by Pedler [277] for the creation of a spell checker of real-word errors (see definition in Section 5.3.2) made by people with dyslexia. This corpus in English has 3,134 words and 363 errors [277]. It is composed of: word-processed homework (saved before it was spellchecked) produced by a third year secondary school student; two error samples used for a comparative test of spellcheckers [239]; and short passages of creative writing produced by secondary school children of low academic ability in the 1960s [170]. To develop a program designed to correct errors made by people with dyslexia, that initial corpus was enlarged to 21,524 words containing 2,654 errors, with over 800 real-word errors. The additional sources for that corpus were: texts from a student with dyslexia, texts from an online typing experiment [368], samples from dyslexic bulletin boards and mailing lists, and stories written by children with dyslexia.

Lists of Dyslexic Errors

Regarding lists of dyslexic errors, the only similar resource is the list of English confusion sets compiled by Pedler [277],[1] extracted from the corpus of texts written by people with dyslexia mentioned before. This list is composed of 833 confusion sets. A confusion set is a small group of words that are likely to be confused with one another, such as *weather* and *whether*.

5.2.2 The Knowledge of Dyslexic Errors

In general terms, errors could be used as a source of knowledge. For instance, the presence of errors in the textual Web has been used for detecting spam [289] or measuring the quality of web content [151].

Since the kinds of errors that people with dyslexia make are related to the types of difficulties that they have [371], their written errors have been used for various purposes such as (1) studying dyslexia, (2) diagnosing dyslexia, or (3) for accessibility related purposes.

[1] `http://www.dcs.bbk.ac.uk/~jenny/resources.html`

First, the analyses of writing errors made by people with dyslexia were used in previous literature to study different aspects of dyslexia [91, 14]. For instance, the specific types of dyslexic errors highlight different aspects of dyslexia [386], such as the phonological processing deficit [215, 240]. The dyslexic error rates vary depending on the language writing system [215]. However, compared to non-dyslexics, people with dyslexia present more errors attributable to phonological impairment, spelling knowledge, and lexical mistakes [371].

Second, since people with dyslexia exhibit higher spelling error rates than non-dyslexic people [82], there are diagnoses of dyslexia based on the spelling score [350, 385]. Also, the spelling error rate is being used as a diagnosing factor in the current official Catalonian protocols [81].

Third, the exploration of corpora of dyslexic errors [277, 328], was used for various accessibility related purposes such as the development of specialized tools like spellcheckers [196, 213, 277], text prediction software,[2] games [330], or word processors which perform text customization taking into account frequent writing errors [157].

5.2.3 What is Missing?

Spanish language resources –corpora and lists– composed of texts and errors, written by people with dyslexia.

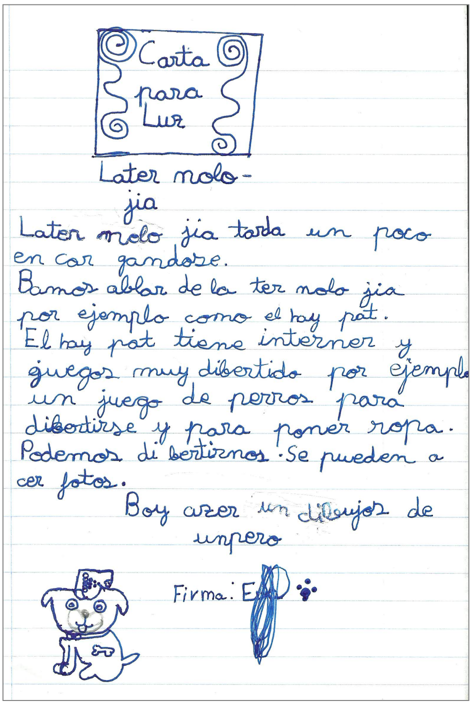
5.3 DysCorpus

In this section we explain how we collected texts written by children with dyslexia, extracted the errors, and classified them. We compare the types of errors that appear in our corpus with an English one.

5.3.1 Extracting Errors from Dyslexic Texts

Manifestations of dyslexia varies among languages [155] but also among subjects, and ages [398]. For instance, misspelling rate in dyslexic children is higher than in adults [371]. However, experiments evidence that adult with dyslexia have a continuing problem in the lexical domain, manifested in a poor spelling ability [371]. Therefore, we collected texts written by a similar population in terms of age,

[2] *Penfriend XL* (`http://www.penfriend.biz/`).


Carta para Luz

Later nolo-
jia

Later nolo jia tarda un poco
en car gandoze.
Bamos ablar de la ter nolo jia
por ejemplo como el hay pat.
El hay pat tiene interner y
guegos muy dibertido por ejemplo
un juego de perros para
dibertirse y para poner ropa.
Podemos di bertirnos. Se pueden a
cer fotos.
Boy azer un dibujos de
unpero

Firma: E

Carta para Luz
'Letter for Luz'

***Later nolojia**
La tecnología
'The technology'

***Later *nolo *jia tarda un poco en *car *gandose.**
La tecnología tarda un poco en cargarse.
'The technology takes a little while to load'

***Bamos *ablar de la *ter *molo *jia por ejemplo como el *hay *pot.**
Vamos a hablar de la tecnología, por ejemplo como el iPod.
'Let's talk about technology such as the iPod'

El *hay *pot tiene *interner y *guejos muy *dibertido
El iPod tiene Internet y juegos muy divertidos
'The iPod has internet and very fun games'

por ejemplo un juego de *perror para *dibertirse y para poner ropa.
por ejemplo un juego de perros para divertirse y para poner ropa.
'for example a game of dogs for fun and to put clothes.'

Podemos *di *bertirnos. Se pueden *a *cer fotos.
Podemos divertirnos. Se pueden hacer fotos.
We can have fun. You can take pictures.

***Boy *azer un *dibujos de un *pero**
Voy a hacer un dibujo de un perro.
'I'll draw a picture of a dog.'

Firma: Eva
'Signature: Eva'

Figure 5.1: Example of a handwritten text of a 9 year old girl with dyslexia (left) and its transcription in Spanish and English (right).

education, native language (Spanish), and that have been diagnosed with dyslexia. These texts were all handwritten and we transcribed them manually. The words that we were not able to transcript due to the illegibility of the handwriting were marked.

We used a total of 83 texts composed of 54 school essays and homework exercises provided by teachers from children and teenagers with dyslexia between 6 and 15 years old, and 29 texts provided by parents with children with dyslexia. Some of these texts were school essays and some other were written specifically for this study (Figure 5.1). Many of the texts came from schools in Catalonia, in which Catalan is taught alongside with Spanish. Since Spanish spelling might be influenced by the acquisition of Catalan orthographic rules, we included the feature "Language transfer" in the annotation of the errors.

From our text collection we manually extracted the misspelled words, without taking into account illegible handwritten words. We did not extracted capitalization or accentuation errors since most children among that age are still learning how to capitalize and accentuate in Spanish. From this set of words we extracted 894 different correct-misspelled pairs with a total of 1,171 errors. For instance, the words *accesibilidad (‘accessibility’)* and *sigilosamente (‘stealthily’)* are the ones that have more different errors (12). That is, there is more than one way to correct the mistake.

5.3.2 Types of Dyslexic Errors

The type of errors we found are consistent with previous studies in Spanish [14] and in English [277].[3] We classify errors as follows:

(a) Errors based on the degree of difference with the intended or target word:

> **Simple errors**. They differ from the intended word by only a single letter or two adjacent letters. They can be due to (i) substitution, **bonde (donde, ‘where’)*, (ii) insertion, **cerreza (cereza, ‘cherry’)*, (iii) omission, **mometo (momento, ‘moment’)* or (iv) transposition, **porceso (proceso, ‘process’)*. In Damerau’s corpus (non-dyslexic errors) [102], 80% of the misspellings were simple errors.[4]
>
> **Multi-errors**. They differ in more than one letter from the target word such as **pallazo (payaso, ‘clown’)*.
>
> **Word boundary errors**. They are run-ons and split words. A run-on is the result of omitting a space, such as **talvez (tal vez, ‘ maybe’)*. A split word occurs when a space is inserted in the middle of a word, such as **a drede (adrede, ‘intentionally’)*.

(b) Errors based on their correspondence with existing words:

[3] Examples with errors are preceded by an asterisk ‘*’. We use the standard linguistic conventions: ‘<>’ for graphemes, ‘/ /’ for phonemes and ‘[ ]’ for phones.

[4] The standard definition of edit distance [210] considers transpositions as two errors, while Damerau defined them as a single error.

Real-word errors. These are misspellings that result in another valid word. For instance, **pala ('shovel')* being the intended word *palabra ('word')*.

Non-word errors. Misspellings that do not result in another correct word.

(c) Errors based on their position:

First letter errors: **ace (hace, 'does')*.

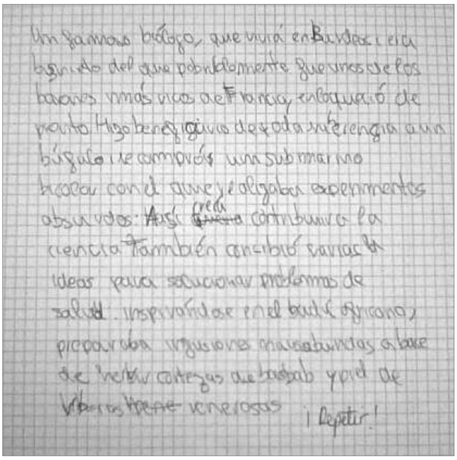
Middle letter errors: **cárzel (cárcel, 'prison')*.

Last letter errors: **interios (interior, 'interior')*.

One example of a fragment of our texts is given in Figure 5.2. An approximated literal translation for this example is:

> *'A famous biologist, who lived in Bordeaux and was great-grandson of who probably was one of the wealthiest barons in France, suddenly went mad. He chose a buffalo as the beneficiary of his inheritance and bought a bicolored submarine in which he made absurd experiments. He believed that with this he contributed to science. He also conceived various ideas to solve health problems inspired by African voodoo, preparing nauseating infusions based on boiled baobab barks and the skin of poisonous snakes.'*

Here we have the following simple errors: (i) substitution: **i (y, 'and'), *budú (vudú, 'voodoo'), *venerosas (venenosas, 'poisonous')*, and **baubab (baobab, 'baobab')*; (ii) insertion: **comprós (compró, 'bought')*; and (iii) omission: **expermentos (experimentos, 'experiments'), *unos (uno, 'some'), *beneficirio (beneficiario, 'beneficiary')*, and **nausabundas (nauseabundas, 'nauseating')*. There is also one multi-error word with one omission and one substitution, **pobrblemente (probablemente, 'probably')*. All of them are non-word errors with the exception of **unos*, which seems to be a concordance error. Most errors are in the middle (7) while we have three errors in the last position and two in the first position of the word. Notice that **i*, probably an error due to transfer from Catalan, counts for both positions.



Un famoso biólogo, que *viviá en Burdeos, *i era biznieto de que *pobrblemente fue *unos de los barones más ricos de Francia, enloqueció de pronto. Hizo *beneficirio de toda su herencia a un búfalo y se *comprós un submarino bicolor con el que realizaba *expermentos absurdos. Así creía contribuir a la ciencia. También concibió varias ideas para solucionar problemas de salud inspirándose en el *budú africano, preparaba infusiones *nausabundas a base de hervir cortezas de *baubab y piel de víboras *venerosas.

Figure 5.2: Story written in Spanish by a 14 years old boy with dyslexia.

We manually annotated the type(s) of error for each of the errors found in the corpus an compared our Spanish corpus with Pedler's corpus in English. In the case that there were two kind of errors we annotated as a multi-error, for instance, in **devidreo (de vidrio, 'of glass')* a boundary error is combined with a simple substitution error.

5.3.3 Comparing English and Spanish Errors

English and Spanish languages are archetypes of deep and shallow orthographies, respectively. Along an orthographic transparency scale for European languages, English appears as the language with the deepest orthography and Spanish as the second most shallow after Finnish [354].

In Tables 5.1 and 5.2 we compare the data of the English corpus described in [277] with our Spanish corpus. We compute the error ratio as the fraction of errors over the correctly spelt words we observe. As expected, Spanish dyslexics make less spelling errors (15%) than English dyslexics (20%), due to their different orthographies. However, the percentage of unique errors is almost the same.

Table 5.2 presents the distribution the different types of dyslexic errors for both corpus. To determine if an error was a real world error

Category	**English**	**Spanish**
Total words	3,134	1,075
Total errors	636	157
Error ratio	0.20	0.15
Distinct errors	577	144
Percentage	90.7	91.7

Table 5.1: Error ratio and percentage in English and Spanish corpora of dyslexic errors.

Category	**English**		**Spanish**	
	Number	%	Number	%
Simple errors	307	53	96	67
Multi errors	227	39	33	23
Word boundary errors	47	8	15	10
Real-word errors	100	17	30	21
Non-word errors	477	83	114	79
First letter errors	30	5	16	11
Total	577	100	144	100

Table 5.2: Distribution of errors in English and Spanish corpora.

we checked its existence in the *Royal Spanish Academy Dictionary* [313].

As expected, there is a greater percentage of multi-errors in a language with deep orthography –English– than in Spanish, *e.g.* **qría (creía, 'thought')*. However, first letter errors are almost two times more frequent in Spanish, *e.g.* **tula (ruta, 'way')*. This may look surprising according to Yannakoudakis and Fawthrop [419], whose findings report that the first letter of a misspelling is correct in the majority of cases, but in Spanish the letter *h* at the beginning of a word is not pronounced and this generates many more errors in that position (see Table 5.2).

The rest of the dyslexic error types are similar in both languages. There are slightly more real-word errors in Spanish, **dijo (digo, 'said')* or **llegada (llegaba, 'arrived')*. Simple errors are the most frequent ones in both languages. However, each error type has a different frequency. For instance, in our texts substitution errors, **dertro (den-*

tro, 'in'), are the most frequent ones (65% of the simple errors) while Ramírez and López [298] state that simple omissions are the most frequent kind for non dyslexic errors in Spanish.

5.4 DysList

In this section we explain how we make a new resource out of the corpus of text written by people with dyslexia: *DysList*, An Annotated Resource of Dyslexic Error. First, we explain the annotation of the dyslexic errors with linguistic, phonetic, and visual information and then we present the results of the analysis of the errors.

5.4.1 Annotation of Dyslexic Errors

We annotated each of the word-error pairs to create *DysList* with the following information:

- **Target word**: the intended word the person aimed to write.
- **Misspelled word**: the wrongly written word or tokens.
- **Damerau-Levenshtein distance**: the minimum number of edits (insertion, deletion, substitution, transposition) required to change the misspelled error into the (target) correct word [102, 210].[5]
- **Target and misspelled word frequencies**: defined as the number of hit counts in a major search engine for Web pages written in Spanish.
- **Target and misspelled length**: number of characters.
- **Error position**: the position in the target word where the error occurs.
- **Target word syllables**: number of syllables.

[5] The edit or Levenshtein distance [210] is the minimum number of substitutions, insertions and deletions to transform one string into another. The Damerau version [102] counts a transposition as a single error instead of two errors. Notice that might be more than one solution for the transformation associated to the edit distance.

- **Target syllable**: the structure of the syllable where the error occurs, such as CV, CVC, or CCV, among others.

- **Type of error**: a detail analyses of the different kind of dyslexic errors is given in [317].

  S **Substitution**: change one letter for another, for example **reelly (really)*.

  I **Insertion**: insert one letter, like in **situartion (situation)*. A word that has been split in two different tokens is counted as an insertion, like in **sub marine (submarine)*.

  D **Deletion**: omit one letter, as in **approch (approach)*. Run-on word boundary errors, like in **alot (a lot)*, are counted as one deletion.[6]

  T **Transposition**: reversing the order of two adjacent letters, for example **artcile (article)*.

- **Real word**: this Boolean attribute records if the error produced another real word. For instance, *witch* being *which* the intended word.

- **Visual information**: for each of the target and the error graphemes we annotate the letters involved in the error with the following visual information, considering both, handwritten and typewritten (*sans serif*) text. See Table 5.3.

  Mirror letter (handwriting/typewriting) such as <d> and <b> or <m> and <w>, with three possible values: vertical, horizontal, and none.

  Height (handwriting/typewriting): letters with descenders (*e.g.* <p, q>, or <g>), letters with ascenders (*e.g.* <t>, or <b>), both (*e.g.* <f>), and none (*e.g.* <n, m>, or <s>).

  Line (handwriting/typewriting): vertical (*e.g.* <m>), horizontal (*e.g.* <e>), and none (*e.g.* <o>).

[6]Notice that a deletion in the target word is an insertion in the misspelled word and *vice versa*.

Visual Feature	**Values**	**Letter(s)**
Mirror (type)	**V** = vertical, **H** = horizontal, **B** = both, **N** = none	**H** = <n, u>, **B** = <b, d, p,
Mirror (hand)	**V** = vertical, **H** = horizontal, **B** = both, **N** = none	**Y** = <g, h, m, n, u, v, w, y> **B** = <b, d, p, q>
Height (type)	**A** = ascender, **D** = descender, **B** = both, **N** = none	**A** = <b, d, f, h, k, l, t>, **D**= <g, j, p, q, y>
Height (hand)	**A** = ascender, **D** = descender, **B** = both, **N** = none	**A** = <b, d, h, k, l, t>, **D**= <g, j, p, q, y, z>, **B**= <
Line (type)	**V** = vertical, **H** = horizontal, **B** = both, **N** = none	**H** = <a, e, f, s>, **V** = <m, **B** = <k>
Line (hand)	**V** = vertical, **H** = horizontal, **B** = both, **N** = none	**H** = <k, z>, **V** = <m, w>
Rotation (type)	**Y** = yes, **N** = no **Y** = yes, **N** = no	**Y** = <a, e, d, b, p, q, n, u>
Rotation (hand)	**Y** = yes, **N** = no **Y** = yes, **N** = no	**Y** = <a, b, d, e, h, m, n, p, w, y>
Fuzzy (type)	**Y** = yes, **N** = no **Y** = yes, **N** = no	**Y** = <b, c, d, f, g, i, j, l, n, ñ p, q, t, u, v>
Fuzzy (hand)	**Y** = yes, **N** = no **Y** = yes, **N** = no	**Y** = <b, d, g, h, m, n, ñ, p, r, u, v, w, y, z>

Table 5.3: Visual features of the annotated target and error letters.

Rotation (handwriting/typewriting): Boolean attribute that indicates if the rotation of a letter produces another letter, such as <d> and <p>.

Fuzzy letters (handwriting/typewriting): Boolean attribute that indicates if the letter has similar visual letters (not due to rotate or mirror) such as <s> and <z>.

– **Phonetic information**: each of the target and the error phones associated to the graphemes in the text are annotated using traditional articulatory phonetic features [178]:

Phone type: vowel (*e.g.* [a]) or consonant (*e.g.* [p]); combinations of vowels forming a diphthong (*e.g.* [i̯a]) and consonant clusters in syllabic onsets (*e.g.* [pl]) have also been annotated as specific phone types.

For consonants:

Voicing: voiced (*e.g.* [b]) or voiceless (*e.g.* [p]).

Manner of articulation: plosive (*e.g.* [p]), nasal (*e.g.* [m]), trill (*e.g.* [r]), tap or flap (*e.g.* [ɾ]), fricative (*e.g.* [f]), lateral (*e.g.* [l]), approximant (*e.g.* [β̞]), and affricate (*e.g.* [tʃ]).

Place of articulation: bilabial (*e.g.* [p]), labiodental (*e.g.* [f]), interdental (*e.g.* [θ̟]), dental (*e.g.* d̪]), alveolar (*e.g.* [s]), palatal (*e.g.* [tʃ]), and velar (*e.g.* [k]).

For vowels:

Height: open (*e.g.* [a]), mid (*e.g.* [e]), and close (*e.g.* [i]).

Place of articulation: front (*e.g.* [i]), central (*e.g.* [a]), and back (*e.g.* [u]).

Lip rounding: rounded (*e.g.* [u]) or unrounded (*e.g.* [i]).

– **Language transfer**: some of the errors in the list were due to transference from Catalan to Spanish.[7] Hence we tagged the error caused by transference from Catalan. For instance, **accessiblidad (accesiblilidad, 'accessibility')* may be due to the existence of the word *accessibilitat* in Catalan.

5.4.2 Criteria for the Visual Characteristics

Since there are many handwriting alphabets we took into consideration a cursive alphabet frequently used to teach in Spanish schools (see Figure 5.3). In any case, visual features do not change much with a different cursive alphabet.

[7] Most of the texts come from Catalan schools where the rate of bilingual students (Catalan-Spanish) is high.

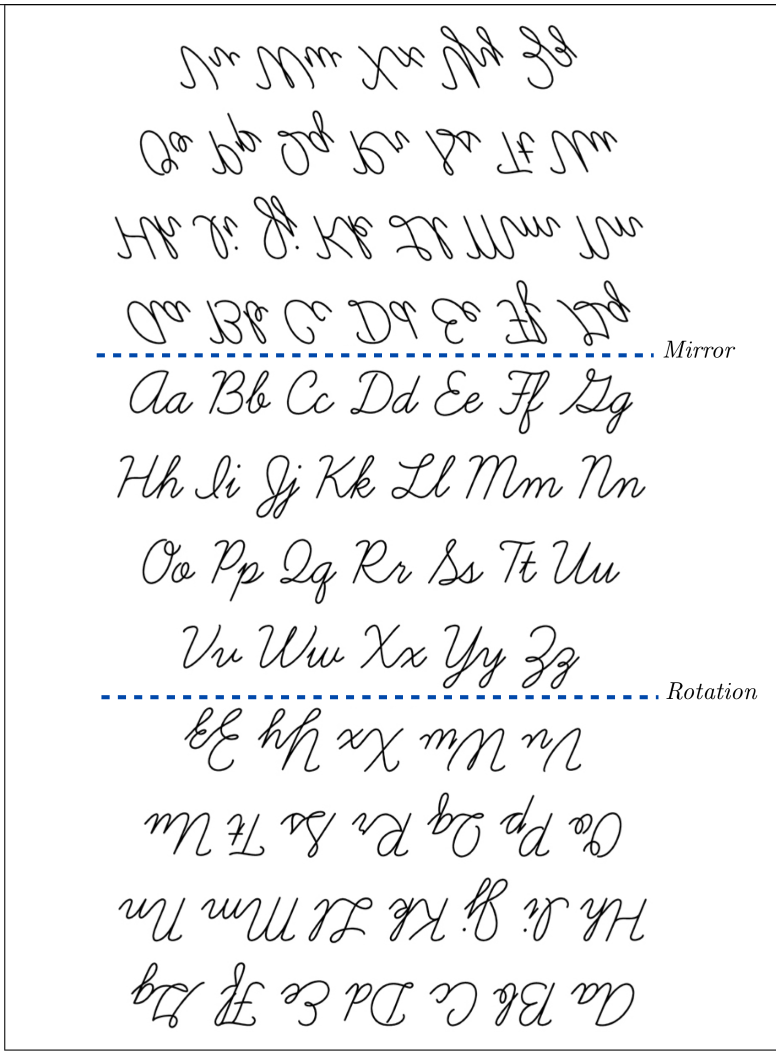


Figure 5.3: Handwritten cursive letters with visual transformations.

5.4.3 Criteria for the Phonetic Transcription

SAMPA symbols and phonetic features

The set of phones used for the transcription of *DysList* (Table 5.4) is based on the inventory presented in Llisterri and Mariño [216]. In this proposal, the phones required for the transcription of Spanish were selected after a study of the frequency of occurrence of more than 100,000 segments in a phonetic transcription of semi-spontaneous interviews. The final inventory was established by eliminating all the phones with a frequency of occurrence below 0.10% in the corpus analyzed.

A modification of the original proposal has been made to reflect the nature of the weak vowels in diphthongs; instead of using the traditional distinction between semi-vowels and semi-consonants, the realizations of /i/ and /u/ in diphthongs are considered non syllabic-vowels irrespectively of their position in the sequence of vowels [152]. They are represented as [i_^] and [u_^] using the X-SAMPA conventions [408] shown in Table 5.4.

Grapheme to phone correspondences for Spanish

Grapheme to phone correspondences for Spanish (Table 5.5) are based on those presented in Llisterri and Mariño [216], with some modifications:

- instead of using the traditional distinction between semi-vowels and semi-consonants, the realizations of /i/ and /u/ in diphthongs are considered non syllabic-vowels irrespective of their position in the sequence of vowels [152];

- the velar allophone of /n/ ([N]) has been introduced, since it is considered in the proposed inventory of phones (Table 5.4);

- since <q> is always followed by <u>, <qu> has been considered a single unit;

- the voiced allophone of /s/ ([z]) has been introduced, since it is considered in the proposed inventory of phones (Table 5.4);

- the criteria for the transcription of <w> follow the recommendations found in the *Pan-Hispanic Dictionary of Doubts* [314].

- the criteria for the transcription of <x> are based in Machuca [224]; although the realization of [s] would be possible when <x> is followed by a consonant such as in *éxtasis ('ecstasy')*, in a formal speaking style <x> would be realized as [ks] in this position. Hence, *México* and *Texas* are treated as exceptions [224];

– the criteria for the transcription of <y> are adapted from those presented in [152], although the plosive realization [178] is not considered.

The digraphs <ll>, <rr>, and <qu> are considered as a single unit, since these combination of letters represent a single sound: [L], [rr], and [k]. The combination <gu> is considered a single unit when it is used to represent the sound [g] or [G] (as in *guerra ('war')* [gerra] or in *seguir ('follow')* [seGir]).

The potential alternation between affricate ([dZ]) and fricative ([jj]) realizations of <y> in words such as *yo ('I')* or *cónyuge ('partner')* or of <hi> in words such as *hierba* has not been considered, since it depends on the speaking style; in these cases, the fricative realization is proposed.

It should be reminded that the transcription does not take into account phonetic phenomena that might take place between words; thus, *bata ('dressing gown')* is transcribed as [bata], although [b] would be realized a [B] in a sequence such as *mi bata ('my dressing gown')*.

5.5 Analysis of Dyslexic Errors

5.5.1 General Characteristics

Frequency: The target word web frequency ranged from 190, *arbolazo ('big tree')*, to 1,389,717,667 *en ('in')*. The errors words frequency ranged from 0, *aczecibilidad (accesibilidad, 'accessibility')*, to 1,178,165,310 in the real word error **ha (a, 'to')*. On average correct words were 4.63 more frequent than words with errors.

Length and error position: The lengths of the target words range from 1 to 20, with the mode at length 6 and an average length of 7.47 letters. Figure 5.4 gives the percentage distribution of target word lengths, the percentage distribution of the word positions where the errors appear, and the relative percentage of errors in the position (that is, 100 times the number of errors in that position divided by the total number of words that have that position).

Syllables: The number of syllables in the words containing spelling errors ranges from one to seven (Table 5.6 right). In these cases, we observed eleven types of syllables, with the distributions shown in

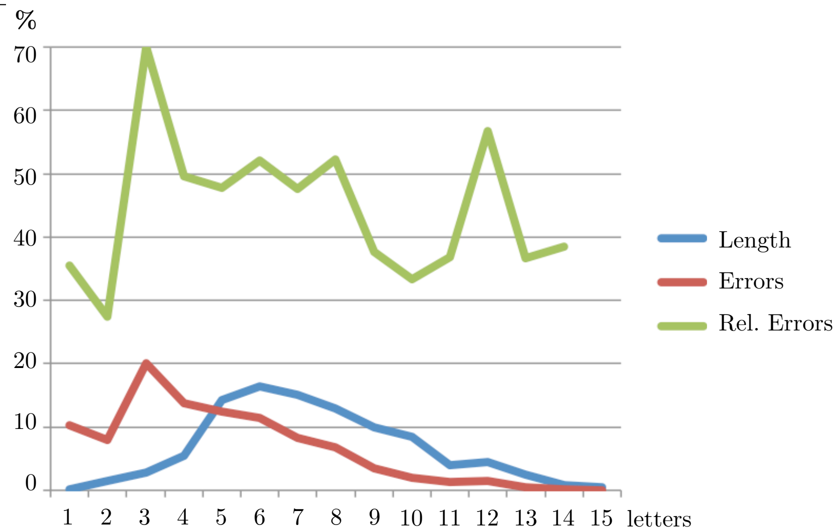


Figure 5.4: Percentage distribution of *DysList* word lengths, error positions and relative percentage of errors in each position.

Table 5.6 (left). 'None' refers to the boundary errors such as **a drede (adrede, 'in purpose')*.

Damerau-Levenshtein distance: In most cases the distance is just 1 (73.3%), with 21.6% of the cases at distance 2 and only 5.1% at distance 3 or greater.

Type of error: In Table 5.7 we give the percentages of every error type. As we can notice, substitution errors are the most frequent ones (near 60%), while Ramírez and López [298] state that simple omissions (deletions) are the most frequent kind of error for Spanish. Although dyslexia is popularly known for the transposition errors, less than 1% of the errors where of this type. This is consistent with [233] which states that only 30% of people with dyslexia have trouble with reversing letters and numbers.

In our analysis we consider some specific phonetic errors coming from digraphs that represent a single sound in Spanish (such as <ll> and <rr>). We found 229 different errors and the most frequent errors (down to 2%) are shown in Table 5.8. From this table we can notice that nine errors represent more than 40% of all errors found, showing the extreme bias of them (*i.e.* less than 4% of the unique errors cover more than 40% of the cases). The most frequent case produces more than 11% of the errors and involve two graphemes that in Spanish have the same phonetic realization, <b> and <v> (which is not the case in English). Analyzing this and other frequent cases, we found three large groups of errors:

– Inserting or deleting a consonant represent 37.9% of the errors, excluding <h> and <y>, which are included in the next cases.

– Deleting or inserting a vowel, including <y> that can have the same phonetic values as <i> in certain contexts, represent 37.5% of the errors.

– Substituting two letters that might have the same phonetic realization (*e.g.* <g> and <j>) or deleting/inserting an <h>, a letter that in Spanish most of the time does not correspond to any sound, represent 15.4% of the errors.

Notice that these three groups cover more than 80% of the errors.

We also studied the position of the errors without finding any important preference, although most errors occur inside the target word. The four most frequent cases were inserting an <h> at the beginning of the word (3.7%), substituting <b> by <v> at the first (2.8%) or third (2.1%) positions, and inserting an <e> in the second position (2.8%). Finally, only 8.97% of the errors were real word errors.

5.5.2 Visual Features Analysis

To analyze the visual features we used Chi-Square goodness of fit to establish whether or not an observed frequency distribution (in the error letters) differs from a theoretical distribution (the one of the correct letters). The percentages of error letters differ from the correct letters by typewritten visual features ($\chi^2(9) = 97.67,\ p < 0.001$) as well as handwriting visual features ($\chi^2(9) = 377.59,\ p < 0.001$). See Table 5.9 for the distribution of the visual features among the error and correct letters.

The distributions of the percentages of correct letters and errors letters in relationship with their mirror visual characteristics were significantly different for the typewritten case ($\chi^2(4) = 55.58,\ p < 0.001$) as well as for the handwriting format ($\chi^2(5) = 137.83,\ p < 0.001$). See the contingency table (Table 5.10) for the percentages. For handwriting almost half of the error letter (46.91 %) had at least of mirror feature.

The percentages of error letters differ from the correct letters by height visual features ($\chi^2(9) = 324.56,\ p < 0.001$) as well as handwriting visual features ($\chi^2(12) = 244.13,\ p < 0.001$). See the contingency table (Table 5.11) for the percentages. Most of the errors occurred in letters with no ascenders neither descenders, 62.93% for typewriting and 57.10% for handwriting.

The percentages of error letters differ from the correct letters by line visual features ($\chi^2(9) = 73.29,\ p < 0.001$) as well as handwriting visual features ($\chi^2(9) = 34.21,\ p < 0.001$). See the contingency table (Table 5.12) for the percentages.

The percentages of error letters differ from the correct letters by rotation visual features ($\chi^2(4) = 23.13,\ p < 0.001$) as well as handwriting visual features ($\chi^2(4) = 32.59,\ p < 0.001$). See the contingency table (Table 5.13) for the percentages. If the target letter had a rotation feature this lead to 31.55% of errors in typewriting and 38.23% for handwriting.

The percentages of correct fuzzy letters differ from the percentage of error fuzzy letters taking into account both, typewriting ($\chi^2(4) = 76.36,\ p < 0.001$) and handwriting typographies ($\chi^2(4) = 41.10,\ p < 0.001$). See the contingency table (Table 5.14) for the percentages. Most of the errors occur with fuzzy target letters, 68.95% for typewriting and 67.11% in handwriting.

5.5.3 Phonetic Analysis

Vowel substitutions account for 5.38% ($N = 63$) of the total number of errors in the corpus. The percentage distributions of the phonetic features of the error vowels differ from the ones of the correct vowels ($\chi^2(36) = 114.9,\ p < 0.001$).

After the transcription of the vowel graphemes according to their phonetic realization in Spanish, the percentage of substitutions errors concerning single vowels has been computed, as shown in Table 5.15.

In terms of shared features, the most frequent types of substitution errors involve one phonetic feature, lip rounding being the most frequent one. It is interesting to note that only 15.87% ($N = 10$) of the vowel substitution errors correspond to phones that do not have any feature in common.

Errors occur most frequently in unrounded vowels ([i], [e], [a]) as far as lip rounding is concerned, mid vowels ([e], [o]) if the degree of opening is considered and front vowels ([i], [e]) when place of articulation is taken into account.

The pattern arising from the study of the phonetic features involved in substitution errors is consistent with the most frequent substitutions found in the corpus (Table 5.15):

– [a] ([unrounded]) → [e] ([unrounded] [mid] [front])

– [e] ([unrounded] [mid] [front]) → [a] ([unrounded])

– [i] ([unrounded] [front]) → [e] ([unrounded] [mid] [front])

– [o] ([mid]) → [e] ([unrounded] [mid] [front])

Substitutions in vowel combinations forming a diphthong account for the 0.94% ($N = 11$) of the errors found in the corpus. The most frequent errors in this category –2 cases of each in the corpus– are found in the substitution of [i̯a] by [ea] and of [i̯o] by [eo]. The highest proportion of errors is observed in target [i̯a] and [o̯e] combinations. In terms of the result of the substitutions, [ea] and [eo] are the two most frequent errors. Given the small size of the sample, no further analysis has been performed, but the trend is coherent with the prevalence of substitutions involving [e] and [a] described for vowels.

Substitution errors in single consonants correspond to the 46.37% ($N = 543$) of the total number of errors in the corpus. The percentage distributions of the phonetic features of the error consonants differ from the ones of the correct consonants ($\chi^2(483) = 3133.96$, $p < 0.001$). They represent, then, the largest category of errors present in *DysList* and are summarized in Table 5.16.

It can be observed that the most frequent errors in consonants are related to the cases in which a one-to-one correspondence between graphemes and phones is not maintained. This results in two different graphemes having the same phonetic value:

– <b> and <v>: both realized as a bilabial plosive [b] or a bilabial approximant [β̞] according to the phonetic context.

– <j> followed by <a>, <o> or <u> and <g> followed by <e>, <i>: both realized as a velar fricative [x].

– <z> followed by <a>, <o> or <u> and <c> followed by <e> or <i>: both realized as an interdental fricative [θ].

– <c> followed by <a>, <o> or <u> and <qu> followed by <i> or <e>: both are realized as a velar plosive [k].

– <r> in word-initial position and after nasals or lateral consonants or <s> and <rr> between vowels: both are realized as an alveolar trill [r].

This is the reason of the high percentage of errors in target consonants [β̞] (18.23%), <x> (14.36%), <θ> (12.15%), and [k] (7.18%) and also in the consonants resulting from a substitution error: [β̞] (18.42%), [θ] (13.08%) [x] (10.50%), and [k] (6.45%) (Table 5.16). The lack of bi-univocal correspondence between phones and graphemes is also patent in the most frequent confusions in manner of articulation within the class of fricative consonants (24.68%) –to which [x] and [θ] belong–, within the group of approximant consonants (20.07%) –[β̞]– and within plosive consonants (14.55%) –[k]–. Taps and trills are also involved as target phones or as errors, although to a lesser extent. The same trend is observed when place of articulation is considered: the largest number of confusions occur within the class of bilabials (26.70%) –which includes [β̞]– and inside the group of velars (19.15%) –which includes [x] and [k]. The interdental consonant [θ] appears as the result of substitution errors in 13.08% of cases and as target phones in confusions in 11.97% of cases.

Confusions between [s] and [θ] (4.42%) and between [θ] and [s] (3.31%) observed in Table 5.16 might be in part explained by the geolectal phenomenon known as *seseo*, which consists in the systematic substitution of [θ] (interdental fricative) by [s] (alveolar fricative) so that [θ] is absent from the phonetic inventory of the speakers of the geographic areas in which this phenomenon occurs. The analysis of features of manner and place also point out in this direction if the confusions in the class of fricatives and in alveolar and interdental consonants are considered.

The presence of a 3.13% of cases in which [λ] appears as the result of a confusion error and the confusions between [j] and [λ] (2.03%) shown in Table 5.16 might be partially accounted for by the presence of *yeísmo*, *i.e.* a neutralization of the contrast between [j] (palatal approximant) and [λ] (palatal lateral) in favor of [j] which is common in most geographical varieties of Spanish. When substitutions in manner of articulation are considered, 2.58% of cases of confusions between laterals and approximants are found; part of the substitutions within the class of palatals (6.63%) may be also accounted for by the presence of *yeísmo*.

The 2.58% of confusions in [ɲ] (palatal nasal) that appear in Table 5.16 may be explained by the decision taken for the phonetic transcription of the corpus concerning a potential transfer from Catalan spelling rules. Since [ɲ] is spelled as <ñ> in Spanish and as <ny> in Catalan, it was considered that both <ñ> and <ny> were intended to represent the palatal nasal consonant.

Almost half of the substitutions found in consonants occur between phones that share their three features (48.43%), while confusions between consonants sharing one (19.52%) or two (26.15%) features are less commonly encountered. It is worth noting that confusions between consonants that do no have any phonetic feature in common take place in 5.52% of cases.

Finally, half of the consonant confusions in the corpus affect simultaneously voicing, manner and place features, a fact to be explained by the spelling irregularities mentioned earlier. When two features are involved in confusions, manner and place are simultaneously affected in 16.99% of cases, and voicing and place in 9.77% of cases. If the confusion involves only one feature, it can be either place of articulation (9.96%) or voicing (9.57%).

In summary, the analysis of consonant substitutions reveals that the spelling mistakes in cases of lack of one-to-one correspondence between phones and graphemes are an important source of confusions within the same class of consonants and are phonetically motivated.

Substitutions affecting combinations of consonants represent a 0.60% ($N = 7$) of the total number of errors in the corpus. More than half of the errors within this category –four cases– correspond to the target sound [ks], spelled as <x> in Spanish. The rest of the errors are found in heterosyllabic clusters formed by a plosive (or their approximant realizations) plus a liquid (*i.e.* a lateral or a rhotic consonant). No further phonetic analysis has been carried out due to the small size of the sample.

5.6 Discussion

The comparisons among different kind of dyslexic errors shed light on how dyslexia manifest. The dyslexic error types are similar in English and Spanish and at the same time their prevalence vary among languages and suggest that dyslexic accessible practices and tools are partially language dependent. Our Spanish list of dyslexic errors is still small but large enough to find some insights about dyslexic errors and to settle the annotation criteria.

As expected, English and in Spanish present similar distributions frequencies. In Spanish substitution errors are the most frequent errors while transpositions the less frequent type. Based on the visual analyses, we suggest that errors might be visually motivated, from the error handwriting-letter 46.91% of the errors had a mirror feature, 38.23% had a rotation feature, and 67.11% are fuzzy letter. A further analyses taking into account each letter frequency in general language would be needed for a stronger assumption. Dyslexic errors are phonetically motivated. The most frequent errors involve letters that do not have the one-to-one correspondence between graphemes and phones. Regarding vowels and consonants most of the errors occurs when they shared phonetic features. For instance, errors occur most frequently in unrounded vowels.

In fact, we believe that this collection is valuable if it allows the creation of more tools targeted to people with dyslexia, such as games to support children's spelling [330]. With respect to Pedler's confusion sets mentioned in Section 5.2, we believe our resource is of similar size and possibly more diverse as it includes a larger sample of the population. *DysList* resource is freely available in the Web.[8]

[8] `www.luzrello.com/resources.html` and `http://grupoweb.upf.edu/WRG/DysList.csv`

Symbol	**Type**	**Voicing**	**Manner** (cons.) **Height** (vowel)	**Place**	**Rounding**
p	consonant	voiceless	plosive	bilabial	
b	consonant	voiced	plosive	bilabial	
t	consonant	voiceless	plosive	dental	
d	consonant	voiced	plosive	dental	
k	consonant	voiceless	plosive	velar	
g	consonant	voiced	plosive	velar	
m	consonant	voiced	nasal	bilabial	
n	consonant	voiced	nasal	alveolar	
J	consonant	voiced	nasal	palatal	
N	consonant	voiced	nasal	velar	
tS	consonant	voiceless	affricate	palatal	
dZ	consonant	voiced	affricate	palatal	
B	consonant	voiced	approximant	bilabial	
f	consonant	voiceless	fricative	labiodental	
T	consonant	voiceless	fricative	interdental	
D	consonant	voiced	approximant	dental	
s	consonant	voiceless	fricative	alveolar	
z	consonant	voiced	fricative	alveolar	
jj	consonant	voiced	fricative	palatal	
j	consonant	voiced	approximant	palatal	
x	consonant	voiceless	fricative	velar	
G	consonant	voiced	approximant	velar	
l	consonant	voiced	lateral	alveolar	
L	consonant	voiced	lateral	palatal	
rr	consonant	voiced	trill	alveolar	
r	consonant	voiced	tap	alveolar	
i	vowel	voiced	close	front	unrounded
i_^	non-syllabic vowel	voiced	close	front	unrounded
e	vowel	voiced	mid	front	unrounded
a	vowel	voiced	open	central	unrounded
o	vowel	voiced	mid	back	rounded
u	vowel	voiced	close	back	rounded
u_^	non-syllabic vowel	voiced	close	back	rounded

Table 5.4: SAMPA symbols and phonetic features.

Grapheme	Context	Symbol
<b>	in word-initial position; after <m> or <n>	b
	all other cases	B
<c>	followed by <e> or <i>	T
	followed by <b, d, g> (preceding <a, o, u>), <m, n, ñ> or <v>	G
	all other cases	k
<ch>	all contexts	tS
<d>	in word-initial position; after <l>, <m> or <n>	d
	all other cases	D
<e>	all contexts	e
<f>	all contexts	f
<g>	in word-initial position followed by <r>, <l>, <a>, <o> or <u> after <m> or <n> followed by <a>, <o> or <u>	g
	followed by <i> or <e>	x
	all other cases	G
<h>	in word-initial position followed by <ie>	jj
	all other cases	no sound
<i>	in nuclear position in the syllable	i
	in non nuclear position in the syllable	i_^
<j>	all contexts	x
<k>	all contexts	k
<l>	all contexts	l
<ll>	all contexts	L
<m>	all contexts	m
<n>	followed by <p>, <b>, <v>, <m> or <f>	m
	followed by <c>, <q> and <a>, <o> or <u>	N
	all other cases	n
<ñ>	all contexts	J
<o>	all contexts	o
<p>	all contexts	p
<qu>	all contexts	k
<r>	in word-initial position; preceded by <l>, <n> or <s>	rr
	all other cases	r
<rr>	all contexts	rr
<s>	followed by <b>, <d>, <g>, <l>, <m>, <n> or <r>	z
	all other cases	s
<t>	in syllable-final position	D
	all other cases	t
<u>	without diaeresis preceded by <g> or <q>	no sound
	in nuclear position in the syllable	u
	in non nuclear position in the syllable	u_^
<v>	in word-initial position; after <m> or <n>	b
	all other cases	B
<w>	followed by a vowel in words of English origin	gu_^
	followed by a consonant in words of English origin	u_^
	in initial-word position in words of German origin	b
<x>	all contexts	ks
	in the words *México* and *Texas*	x
<y>	in word-initial position, after <n> or <l>	jj
	after a consonant different from <n> or <l>	j
	preceded and followed by a vowel after a syllable boundary	j
	preceded or followed by a vowel within the same syllable	i_^

Table 5.5: Grapheme to phone correspondences for Spanish.

Syllable Type	**Percentage**	**No. Syllables**	**Percentage**
CV	37.40	3	33.30
CVC	21.35	2	26.30
none	13.15	4	17.68
CCV	8.20	1	11.87
CVV	7.77	5	7.51
CVVC	6.06	6	3.25
VC	3.67	7	0.09
CCVC	1.54		
V	0.60		
VV	0.09		
CCVV	0.09		
CCVCC	0.09		

Table 5.6: Syllable types (left) and number of syllables (right) in words where errors occur.

Error Type	**Percentage**
Substitution	58.84
Insertion	13.40
Deletion	26.30
Transposition	1.45

Table 5.7: Percentages of dyslexic error types.

Error Type	**Letter(s)**	**Percentage**
S	<b, v>	11.36
D	*space*	6.75
S	<g, j>	5.46
D	<h>	4.53
I	*space*	3.07
S	<c, z>	2.82
S	<c, s>	2.22
D	<r>	2.22
I	<r>	2.13

Table 5.8: Percentages of frequent specific errors.

Visual Feature	**Letters**	**Percentage**
Mirror (type)	Correct	**none** = 26.81, **N** = 57.90, **H** = 3.93, **B** = 11.36
	Error	**none** = 33.39, **N** = 54.74, **H** = 4.01, **B** = 7.86
Mirror (hand)	Correct	**none** = 26.81, **N** = 47.65, **H** = 14.18, **B** = 11.36
	Error	**none** = 33.39, **N** = 39.28, **H** = 19.47, **B** = 7.86
Height (type)	Correct	**none** = 26.81, **A** = 19.04, **N** = 43.81, **D** = 10.33
	Error	**none** = 33.39, **A** = 11.44, **N** = 44.41, **D** = 10.76
Height (hand)	Correct	**none** = 26.81, **A** = 18.53, **N** = 42.70, **D** = 11.44,
	Error	**none** = 33.39, **A** = 11.44, **N** = 41.33, **D** = 13.83
Line (type)	Correct	**none** = 33.39, **V** = 0.85, **N** = 58.67, **H** = 13.66
	Error	**none** = 26.81, **V** = 1.11, **N** = 54.48, **H** = 11.02
Line (hand)	Correct	**none** = 26.81, **V** = 0.85, **N** = 71.22, **H** = 1.11
	Error	**none** = 33.39, **V** = 1.11, **N** = 62.43, **H** = 3.07
Rotation (type)	Correct	**none** = 26.81, **Y** = 22.63, **N** = 50.56
	Error	**none** = 33.39, **Y** = 18.19, **N** = 48.42
Rotation (hand)	Correct	**none** = 26.81, **Y** = 30.57, **N** = 42.61
	Error	**none** = 33.39, **Y** = 20.58, **N** = 46.03
Fuzzy (type)	Correct	**none** = 26.81, **Y** = 44.41, **N** = 28.78
	Error	**none** = 33.39, **Y** = 43.47, **N** = 23.14
Fuzzy (hand)	Correct	**none** = 26.81, **Y** = 44.66, **N** = 28.52
	Error	**none** = 33.39, **Y** = 41.59, **N** = 25.02

Table 5.9: Visual features of the annotated target and error letters.

Feature	Mirror (type)				Mirror (hand)			
	Error B	H	N	**Total** (%)	**Error** B	H	N	**Total** (%)
Correct								
Both	2.17	0	17.70	19.87	2.17	15.53	2.17	19.87
Horizontal	0	0	3.67	3.67	7.51	4.17	11.35	23.03
None	12.35	3.84	60.27	76.46	4.84	12.69	39.57	57.10
Total (%)	14.52	3.84	81.64		14.52	32.39	53.09	

Table 5.10: Percentages of errors with mirror visual features.

Feature	Height (type)				Height (hand)			
	Error D	A	N	**Total** (%)	**Error** D	A	N	**Total** (%)
Correct								
Both	0	0	0	0	0	0	1.00	1.00
Descender	11.35	3.51	3.67	18.53	11.52	3.84	5.18	20.54
Ascender	1.00	3.67	18.86	23.53	1.00	3.67	17.86	22.53
None	6.68	10.85	40.40	57.93	12.35	10.52	33.06	55.93
Total (%)	19.03	18.03	62.93		24.87	18.03	57.10	

Table 5.11: Percentages of errors with height visual features.

Feature	Line (type)				Line (hand)			
	Error H	V	N	**Total** (%)	**Error** H	V	N	**Total** (%)
Correct								
Horizontal	5.84	0.17	11.52	17.53	0	0	2.00	2.00
Vertical	0	0	1.67	1.67	0	0	1.67	1.67
None	7.85	1.67	71.29	80.81	5.84	1.84	88.65	96.33
Total (%)	13.69	1.84	84.48		5.84	1.84	92.32	

Table 5.12: Percentages of errors with line visual features.

Feature	Rotation (type)			Rotation (hand)		
	Error Yes	No	**Total** (%)	**Error** Yes	No	**Total** (%)
Correct						
Yes	7.01	24.54	31.55	10.18	28.05	38.23
No	18.03	50.42	68.45	18.20	43.57	61.77
Total (%)	25.04	74.96		28.38	71.62	

Table 5.13: Percentages of errors with rotation visual features.

Feature	Fuzzy (type)			Fuzzy (hand)		
	Error Yes	No	**Total** (%)	**Error** Yes	No	**Total** (%)
Correct						
Yes	49.75	16.86	66.61	43.07	18.53	61.60
No	19.20	14.19	33.39	24.04	14.36	38.40
Total (%)	68.95	31.05		67.11	32.89	

Table 5.14: Percentages of errors with fuzzy visual features.

	Error							**Total**
	a	e	i	i̯	o	u	u̯	(%)
Correct								
a	0	20.63	3.17	0	9.52	0	0	33.33
e	15.87	0	4.76	1.59	6.35	0	0	28.57
i	0	7.94	6.35	0	0	0	0	14.29
i̯	0	0	0	6.35	0	0	0	6.35
o	3.17	4.76	0	0	0	1.59	1.59	11.11
u	0	0	1.59	0	3.17	0	0	4.76
u̯	0	0	0	1.59	0	0	0	1.59
Total (%)	19.05	33.33	15.87	9.52	19.05	1.59	1.59	

Table 5.15: Percentage of vowel substitutions.

	Error																						**Total**
	b	β̞	d	ð̞	f	g	ɣ̞	j	ɲ	k	ks	l	λ	m	n	p	ɾ	rr	s	t	θ	x	(%)
Correct																							
b	7.73	0	0.37	0	0	0	0	0	0	0	0	0	0	0.18	0	0.18	0	0	0	0	0	0	8.47
β̞	0.18	16.57	0	0.74	0	0	0	0	0	0	0	0.37	0	0	0.18	0	0.18	0	0	0	0	0	18.23
d	0.37	0	0	0	0	0.18	0	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0	0	0.74
ð̞	0	0.74	0	0	0	0	0	0	0	0.18	0	0	0	0	0	0	0.74	0	0	0.37	0	0	2.03
f	0	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0	0	0	0	0	0	0.92	0	1.10
g	0	0	0	0	0	0	0	0	0	0.55	0	0	0	0	0	0	0	0	0	0.18	0	0.37	1.10
ɣ̞	0	0.55	0	0	0	0	0.37	0	0	0.18	0	0.18	0	0	0.18	0	0	0	0	0	0	0.55	2.03
j̞	0	0	0	0	0	0	0.18	0.92	0	0	0	0	2.03	0	0	0	0	0	0	0	0	0	3.13
ɲ	0	0	0	0	0	0	0.37	0.18	2.58	0	0	0	0.18	0	0.37	0	0	0	0	0	0	0.18	3.87
ʝ	0	0	0	0	0	0	0	0	0	0	0	0	0.55	0	0	0	0	0	0	0	0	0.37	0.92
k	0	0	0	0	0	0.18	0.18	0	0	3.13	0	0	0	0	0	0.55	0.18	0	0.18	0.18	2.39	0.18	7.18
l	0	0.18	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0.18	0.37	0	0.18	0.37	0	0	1.47
λ	0	0	0	0	0	0	0	0.37	0	0	0	0	0	0	0	0	0.18	0	0	0	0	0	0.55
m	0	0	0	0	0	0	0	0	0	0	0	0	0	1.66	0.37	0	0.18	0	0	0	0	0	2.21
n	0	0	0	0	0	0	0	0	0	0	0	0	0	1.10	0	0	1.10	0	0.55	0	0	0	2.76
ŋ̞	0	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0	0	0	0	0	0	0	0	0.18
p	0.18	0	0	0	0	0	0	0	0	0.18	0	0.18	0	0	0	0	0.18	0	0	0	0	0	0.74
ɾ	0	0.37	0	0	0	0	0	0	0	0	0	0.37	0	0	0.37	0.37	0	0.74	0.55	0.18	0	0	2.95
r	0	0	0	0	0	0	0	0	0	0	0	0	0.37	0	0	0	1.66	1.10	0	0.18	0.18	0	3.50
s	0	0	0	0	0	0	0	0	0	0.74	0	0.37	0	0	0.37	0	0.37	0.18	1.47	0	4.42	0.37	8.29
t	0.18	0	0.18	0.18	0	0.18	0	0	0	0	0	0.37	0	0	0	0	0	0	0	0	0	0	1.10
θ	0	0	0.37	0.55	0.37	0	0.18	0	0	1.29	0	0	0	0	0	0.18	0.55	0.18	3.31	0.18	4.60	0.37	12.15
tʃ	0	0	0	0.18	0	0.18	0.18	0	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0.18	0.92
x	0	0	0	0	0	1.84	3.50	0	0	0.18	0.18	0	0	0	0	0	0.18	0	0.18	0	0.37	7.92	14.36
Total (%)	8.66	18.42	0.92	1.66	0.37	2.58	4.97	1.47	2.58	6.45	0.18	1.84	3.13	3.31	2.03	1.47	5.89	2.21	6.45	1.84	13.08	10.50	

Table 5.16: Percentages of consonant substitutions.

Chapter 6

Prevalence of Dyslexia in the Web

6.1 Introduction

In this chapter we present a lower bound of the prevalence of dyslexia in the Web for English and Spanish, the most frequent languages in the world after Mandarin Chinese. Based on the analysis of corpora written by people with dyslexia, we use a representative data set of dyslexic words to calculate this lower bound of web pages containing English and Spanish dyslexic errors. We also present an analysis of dyslexic errors in major Internet domains, social media sites, and throughout English and Spanish speaking countries. To show the independence of our estimations from the presence of other kinds of errors, we compare them with the overall lexical quality of the Web, and with the error rate of non-corrected corpora. The presence of dyslexic errors in the Web motivates work in web accessibility for users with dyslexia and helps us to know the real impact of dyslexia in the Web. The contributions of this chapter are:

- A classification of errors found in the Web.
- A methodology to estimate the impact of dyslexic errors in the Web.
- A lower bound in the presence of dyslexia in the Web, extended to the major Internet domains, social media sites and throughout

English and Spanish speaking countries.

– A validation of our results by comparing them with large corpora and the overall lexical quality of the Web.

The results of this chapter are presented in Baeza-Yates and Rello [16], and Rello and Baeza-Yates [317].

6.2 Related Work

Web quality can be related to its contents (highly current, accuracy, source reputation, objectivity, etc.) or to its representation (spelling errors, various typos, sentences with low readability, grammatical errors, etc.). Most efforts are focused on assessing content quality, *e.g.* spam detection or source credibility. Ringlstetter *et al.* [336] propose filtering methods to retrieve cleaner corpora from the Web after investigating the distribution of orthographic errors of various types of web pages while Piskorski *et al.* [289] explore certain linguistic features for detecting spam.

Lexical quality refers to the degree of excellence of words in a text, including all kinds of spelling errors [283]. Our approach is mainly inspired by the work of Gelman and Barletta [151] that applies a spelling error rate as a metric to indicate the degree of quality of websites. They use a set of ten frequently misspelled words and hit counts of a search engine for this set, showing that web content quality and lexical quality are related.

6.3 Methodology

In this section we present our measure for sampling dyslexic errors in the Web and a description of the data sets we created, their sources and the criteria used for the selection of the content.

6.3.1 A Measure for Estimating Dyslexic Errors

A measure for estimating the amount of dyslexic errors in the Web should be independent of the size of the text or the number of pages in a website, to be able to compare this measure across websites or different web segments. One alternative could be to compute the rate of dyslexic errors, that is, the number of misspellings divided by the total numbers of words. However, that is hard to compute in the context of the Web due to its size. A solution is to use a sample of words and use the rate of spelling errors of those individual words to maintain independence of the text size. However, it is not trivial to find in the Web which are all possible misspells of a word for two reasons: the number of possible variations increases exponentially with the number of errors, and there might be more than one correct word at the same edit distance[1] for a given misspelled word.

A possible solution is then to find words that are frequent and that also have a frequent dyslexic misspell, using that occurrence ratio as a lower bound of the exact dyslexic misspell rate. As the frequency of the most frequent misspell is much lower than the correct version,[2] we can approximate the word rate of spelling errors just by dividing the most frequent misspell by the number of correct occurrences instead of using the total number of all possible misspells of the word (which as we said earlier is harder to determine).

To estimate a lower bound of the presence of dyslexia in the Web, we define a measure of the *Dyslexic Error Rate* (DER) as the average rate of the dyslexic misspells. That is, given a set of words W, we compute the relative ratio of the dyslexic misspell to the correct spelling averaged over this word sample scaled by 100 to obtain values that can be interpreted as a percentage. That is,

$$DER = 100 \cdot \text{mean}_{w_i \in W} \left(\frac{df_{\text{ misspelled } w_i}}{df_{\text{ correct } w_i}} \right) ,$$

where *df* is the document frequency[3] of each word as we will measure lexical quality across web pages and not number of words, since web pages have different number of words. Using the term frequency would be better, but that would imply that computing DER cannot be done using a standard search engine such as we propose here.

Since there is no reasonable way to know the overall frequency of words in the Web outside a search engine, for the frequencies we use

[1]See footnote 5 in Chapter 5.

[2]In fact, the distribution many times follows a power law, as the famous Britney Spears example: `http://www.netpaths.net/blog/britney-spears-spelling-variations/`.

[3]Document frequency is the number of documents where a term appears [18].

the hit counts of the Google search engine using the Advanced Search option to search only in English or Spanish websites depending on each case. Then, we compare our results using more than one search engine (Bing and Yahoo!) and validate them with real document frequencies computed from the Yahoo! web search index.[4] The similarity of DER among search engines varied depending on the moment when the queries were submitted and on data set used. While most of the differences were insignificant, we observed the greatest difference using English dyslexic simple errors data set between Google ($DER = 0.1023$) and Bing ($DER = 0.1559$) in November 2011.

For W we need to find words that have the following properties: they are frequent, they have a frequent dyslexic misspelling, and they are non ambiguous, that is, the word or the misspelled word cannot represent another word with the same spelling (*e.g.* a proper name, acronym or a foreign word).

Using this measure we can compute the impact of different kinds of errors depending on the data sets used for W that fulfill the conditions stated before. To pursue our goal, we created two new data sets. First, we created W_D (for English, W_Den, and Spanish, W_Dsp) composed of only dyslexic misspellings to compute the impact of different kinds of dyslexic errors. Second, we expanded W_D to W_E (for English, W_Een, and Spanish, W_Esp) including all types of errors to estimate the impact of dyslexic errors in comparison with other kinds of errors.

6.3.2 Selection Criteria for Dyslexic Errors

Sample W_D (see Appendix A.2) is composed of English and Spanish dyslexic errors extracted from texts produced by people with dyslexia.

First, we extracted all the misspellings from the sources described in the previous section. Second, we selected the errors that are related to a phonological and orthographical processing deficit. For instance, errors due the the similarity of sounds, **vidreo (vidrio, 'glass')*, or the graphemes, **pabre (padre, 'father')*.

Sample W_D includes non-word dyslexic errors from both simple –**childern (children)*– and multi-errors –**embueve (envuelve, 'wraps')*. Sample W_D includes all kinds of simple errors.

[4]These frequencies were obtained before Yahoo! started to use Bing.

We do not take into account first letter errors because in [277] a quarter of them are capitalization errors, another quarter are real word errors, and overall they present low frequency. We also discard word boundary errors because they sometimes involve more that one lexical unit.

To reduce the overlap of the errors produced by people with dyslexia in our data set with other kinds of errors, we selected them according to this principle: the errors related to the target word need to be unique and not ambiguous. For example, errors which coincide with other existing words in other languages are omitted, *i.e.* **couver (cover).* Similarly, errors which give as a result a proper name are also filtered **klene (clean)* [49]. Also, we only consider the cases which include letters with similar pronunciation that produce more confusion among dyslexic individuals than non-dyslexic individuals, such as <*m*|*n*>, <*m*|*p*>, <*b*|*p*> or <*b*|*m*>. Second, during the selection process, we pay special attention to examples with similar looking or symmetric letters, such as <*d*|*b*>, <*p*|*q*> and <*d*|*p*>, among others. These criteria are consistent with literature. For instance, it is specially frequent to find substitutions of orthographically similar letters, such as <*b*> and <*d*> as well as substitutions in letters with similar sounds in Spanish, such as /g/ by /d/ in **piegra (piedra, 'stone')* or /t/ and /k/ (written <*qu*> in the example) **pateque (paquete, 'packet')* [111].

Within the criteria we obtained a set composed of 40 words, 20 for English and 20 for Spanish covering simple and multiple dyslexic errors. The reduced size of W_D is explained by: (a) the difficulty of finding texts written by people diagnosed with dyslexia; (b) the lack of corpora of dyslexic errors, with the exception of Pedler's (2007) corpus for English, and ours for Spanish texts produced by dyslexics; and (c) the strict criteria that we establish for selecting the misspelled words. These criteria aim to guarantee, as much as possible, the uniqueness and unambiguity of the dyslexic word, constraining their selection.

Our estimations of dyslexia would not vary much using a larger sample of words. In Figure 6.1 we show the convergence of DER using the average of k words (k from 1 to 10) for five different random orderings of the simple dyslexic errors in Spanish W_Dsp, that is, ten words. We can see that already with seven of the words we get values

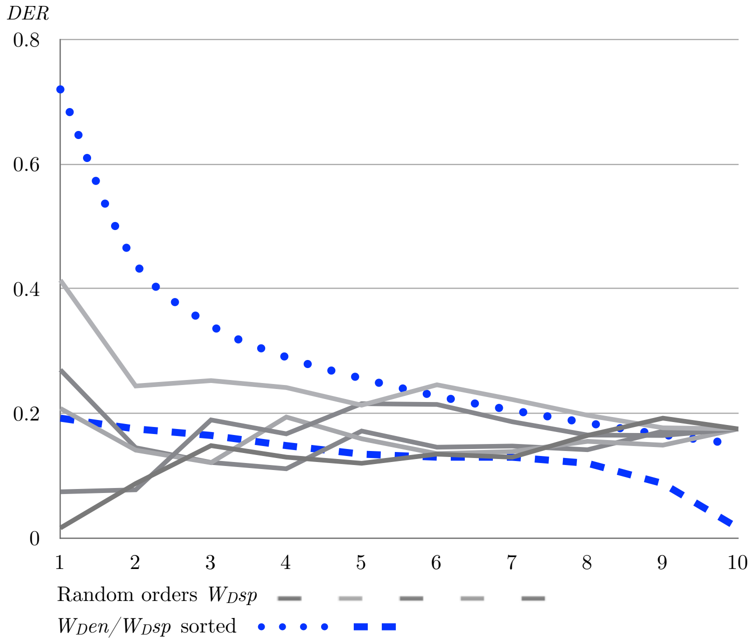


Figure 6.1: Partial DER for five random orderings of W_Dsp for the Web in Spanish, and the sorted individual misspell ratios for English and Spanish.

similar to ten words, so the results are already reliable above that. This shows that the accuracy of the measure improves as the size of W_D grows. We also give the sorted order of the ratios for W_Den and W_Dsp where we can see that the maximum and the minimum misspelling ratios differ by a factor of 4 for English and 20 for Spanish, being the maximum in Spanish for the pair **necestio (necesito, 'need')*. Both curves are similar and although the dyslexic error rate is not comparable across languages, this means that in our case the results will differ within one order of magnitude.

6.3.3 Error Classification

We classify the different kind of errors in the Web into five groups.

(a) **Dyslexic errors:** Among the different kinds of errors commonly made by dyslexics, that is, unfinished words or letters, omitted words or inconsistent spaces between words and letters [397], we only consider lexical errors, that is errors inside words such as multiple additions, transpositions, omissions or substitutions of letters. For instance, **unforchanely* instead of *unfortunately* [277].

(b) **Spelling errors:** Regular spelling errors produced by non-impaired native speakers in English or Spanish, such as the transposition error **recieve* instead of *receive.*

(c) **Typo errors:** Regular typos caused by the adjacency of letters in the keyboard, *i.e.* **dituation (situation).*

(d) **OCR errors:** Optical character recognition (OCR) errors, due to letters of similar shape, such as **tornorrow (tomorrow).*

(e) **Foreign errors:** Errors made by non-native speakers who use English or Spanish as a foreign language. For example, **receibe (receive)* is a typical error made by Spanish learners of English.

The other possibilities of errors related to the target with negligible frequency were discarded according to the conditions that W should fulfill (Section 6.3.1). Note that typos are behavioral errors, native and non-native misspellings are phonetic errors, OCR mistakes are visual errors, while dyslexic errors could be phonetic or/and visual.

Sample W_E (see Appendix A.2) is composed of 479 words in English ($W_E en$) and Spanish ($W_E sp$). They are divided in 20 subgroups composed of the target word and the different type of errors related to the intended word. There are no frequent words, and the words are relatively long. The average word length is 9.3 letters per word. This is longer than the average of letters per lemma (8.78) for the *Royal Spanish Academy Dictionary* [313], and the average number of letters per word (4.5) occurring in English texts [20].

Sample W_E contains all types of errors. In order to detect lexical errors produced by dyslexic individuals, it is required to distinguish pure dyslexic errors within all lexical errors. Therefore, we establish five classes of errors, taking into consideration the user disability, the user mother tongue and the source of the text.

6.3.4 Selection Criteria for Other Errors

Since dyslexic errors are the most difficult to find, our starting point was W_D. After identifying the dyslexic errors and their corresponding target word, we examined manually each of the different error types related to the target word. Then, we decided to include them or not in our sample according to this principle: the errors related to the target word need to be unique and not ambiguous. For instance, the real word *worried* could also be a typo from the intended word *worries* since *s* and *d* are adjacent in the keyboard. Similarly, the typo **dxplain (explain)* is also a proper name. Hence, named entities and real word errors were dismissed, as well as target words with more than three ambiguous errors. The great majority of the candidates was ambiguous and did not match the criteria.

The dyslexic errors contained in sample W_E are the multi-errors from W_D, a subgroup of all the possible errors made by dyslexic people. However, in the set of dyslexic multi-errors, not all the kinds of possible errors are taken into account. We also avoid taking into account errors which produce a syntactic anomaly, *i.e.* words that have no part of speech tags in common, such as the error **from (form)* or inflection errors, *i.e.* **storys (stories)*.

Regular spelling errors were created taking into account their high frequency in query logs and also general spelling error patterns taken from literature [298, 419].

Regular typos caused by the adjacency of letters in the keyboard, were generated by substituting each letter of the target word with the letter situated immediately left and right from the intended letter. Other cases have much smaller frequency (keys above or below). We discarded the cases in which the adjacent key was not a letter, *i.e.* **co,parison (comparison)*.

For generating the OCR errors we substituted the typical letters which are usually mistaken, for instance, $c \to e$, $rn \to m$ or $cl \to d$ [374].

To find the typical errors made by non-native speakers who use English or Spanish as foreign languages, we have taken into account errors caused by phonological transference from English or Spanish. For instance, **gobernment* is a typical error made by Spanish learners

English Error Class	Google Range (%) 2010	Google Range (%) 2011	Google Percentage 2010	Google Percentage 2011	Bing, Google and Yahoo! Range (%) 2010	Bing, Google and Yahoo! Percent. 2010
Spelling	1.10 – 91.57	2.58 – 92.99	51.99	55.77	47.39 – 91.57	63.73
Foreign	1.03 – 92.96	0.71 – 92.58	24.53	27.02	1.57 – 10.21	28.18
Typo	4.25 – 39.16	2.46 – 40.76	21.55	15.56	11.31 – 47.75	28.18
OCR	0.01 – 3.65	0.01 – 2.46	1.51	1.02	0.00 – 3.65	0.80
Dyslexia	0.00 – 3.40	0.01 – 3.56	0.42	0.63	0.01 – 3.10	**0.67**

Table 6.1: Range and percentages for the error classes in English according to Google, and taking into account the averages of the three search engines Bing, Google and Yahoo!.

of English, since the graphemes <*b*> and <*v*> are pronounced as /b/, and the phoneme /v/ does not exist in the standard Spanish phonemic system. Besides its translation in Spanish is written with *<b> ('gobierno')*.

6.4 Results

In this section we present the results of using sample W_E to estimate the percentages of different kinds of errors in the Web (Section 6.4.1). We used the sample W_D for estimating the lower bound of dyslexic errors in the Web and in diverse domains and websites (Sections 6.4.2 and 6.4.3). Finally we present a validation of our measure (Section 6.4.4).

6.4.1 All Errors in the Web

To compute DER we use Google to estimate the document frequency of each word in data set W_D. For the English data set in 2010 we have also used the average of the document frequency of three search engines Bing, Google and Yahoo!. Using W_E we computed the percentages of the different kinds of errors in the Web for English and Spanish (see Tables 6.1 and 6.2). Compared to other kinds of errors, the percentage of dyslexic errors is very low with an average of approximately 0.63% for English and 0.43% for Spanish. This percentage is very conservative because DER is a lower bound for dyslexia by defi-

Spanish Error Class	Google Range (%) 2011	Percentage
Spelling	21.85 – 93.98	54.11
Foreign	0.19 – 47.62	26.51
Typo	2.87 – 43.77	17.17
OCR	0.24 – 7.67	1.79
Dyslexia	0.02 – 1.83	**0.43**

Table 6.2: Range, percentages and average for the different error classes in Spanish.

Dyslexic Errors	Multi-Errors		Simple Errors		W_D
	Range	DER	*Range*	DER	DER
English	0.000*– 0.007	0.0019	0.004 – 0.720	0.1374	0.0697
Spanish	0.000*– 0.008	0.0018	0.016 – 0.414	0.1753	0.0886

Table 6.3: Range and DER of different dyslexic errors in the Web. A number larger than 0 but less than 0.0005 is represented by 0.000*.

nition and W_E only has dyslexic errors of the multi-error type, which are less frequent than simple errors. Simple dyslexic errors were not taken into consideration for W_E to avoid overlapping with other kind of errors.

We observe that the order of magnitude for all kinds of errors are the same in both languages, but dyslexic errors more frequent in English than in Spanish compared to the other misspelling types.

6.4.2 Dyslexia in the Web

To validate our results we use exact frequencies from Yahoo!'s web search index. In Table 6.3 we present the prevalence of dyslexia of the Web for English and Spanish.

To compare them with the rest of the Web, we chose six social media sites belonging to five different classes: blogs (Blogger) including micro-blogs (Twitter), social networks (Facebook), collaboration sites (Wikipedia), multimedia sites (YouTube) and opinions, including community question-answering systems (Y! Answers). To be able to estimate the overall impact of each site, we need to estimate the relative size of each of them. For this we use the total number of pages in the public content of each website according to Google's search engine. A search engine identifies this number by restricting the search to the pages on that site (this option is given in the advanced search page).

In Table 6.4 we compare each site and social media as a whole with other important web domains and the Web. For each site we also give the relative size of their (public) content.

Social media written in English has lower DER than the overall Web. However, compared to high quality sites *(.org* and *.edu)*, the presence of dyslexia in social media is higher in English. This should not be a surprise considering the diversity and sheer volume of social media content. It seems that the higher DER of *.com* may be due to the fact that most Web spam and social media content is part of *.com*.

Wikipedia values vary depending on the language and the kind of errors. We believe that the main contributor to this variability is the community section, since many examples were found in user, discussion and project pages of Wikipedia. Websites with *.edu* domain have the lowest presence of dyslexic errors for both languages and, among the social media sites; Blogger had the lowest occurrence of dyslexic errors.

6.4.3 Geographical Distribution

There are around 387 and 365 millions of Spanish and English speakers as first language, respectively. As a second language, English has the highest number of speakers reaching 1.4 billion, while Spanish is spoken as a second language by around 500 million people [212].

To compute the geographical distribution of dyslexic spelling errors among the countries where English and Spanish is spoken, we have taken into account the countries which have the highest populations of native English and Spanish speakers.

Since it is not possible to distinguish countries among the *.com, .edu, .net* and *.org* domains, the websites were geographically identified by the country domain (see Tables 6.5 and 6.6). For instance, we consider USA websites the ones with *.us* domain. According to the

English	Size	Multi-Errors		Simple Errors		W_D
Domain/Site	*(%)*	*Range*	*DER*	*Range*	*DER*	*DER*
.edu	1.66	0.000* – 0.000*	0.000*	0.000* – 0.074	0.0333	0.0167
.org	27.60	0.000* – 0.001	0.0002	0.002 – 0.310	0.0614	0.0308
.com	70.74	0.000* – 0.006	0.0011	0.010 – 0.793	0.1705	**0.0858**
Overall		0 – 0.005	0.0004	0.000* – 0.793	0.0884	0.0444
Blogger	12.23	0 – 0.001	0.0001	0.001 – 0.142	0.0438	0.0220
Youtube	17.60	0.000* – 0.000*	0.0001	0.001 – 0.277	0.0724	0.0363
Facebook	57.64	0.000* – 0.001	0.0003	0.003 – 0.267	0.0742	0.0373
Twitter	9.98	0 – 0.000*	0.0001	0.000* – 0.619	0.0804	0.0403
Wikipedia	0.40	0.000* – 0.006	0.0008	0.008 – 0.502	0.1244	0.0626
Y! Answers	2.21	0.000* – 0.002	0.0004	0.002 – 1.619	**0.2814**	**0.1409**
Overall		0.000* – 0.006	0.0003	0.000* – 1.619	0.1128	0.0566
English Web		0.000* – 0.007	0.0012	0.006 – 0.749	0.1290	0.0651

Spanish	Size	Multi-Errors		Simple Errors		W_D
Domain/Site	*(%)*	*Range*	*DER*	*Range*	*DER*	*DER*
.edu	0.90	0 – 0.000*	0.000*	0 – 0.036	0.0041	0.0021
.org	10.19	0 – 0.000*	0.0001	0.005 – 0.071	0.0348	0.0175
.com	88.91	0.000* – 0.006	0.0002	0.032 – 0.525	**0.1927**	**0.0965**
Overall		0 – 0.006	0.0001	0 – 0.528	0.0772	0.0387
Blogger	15.05	0.000* – 0.001	0.0002	0.007 – 0.075	0.0202	0.0102
Facebook	71.06	0 – 0.002	0.0004	0.001 – 0.068	0.0231	0.0118
Twitter	5.24	0 – 0.008	**0.0012**	0.004 – 0.068	0.0253	0.0133
Wikipedia	0.18	0.000* – 0.002	**0.0078**	0 – 0.095	0.0315	0.0197
Y! Answers	2.81	0.000* – 0.014	**0.0023**	0.036 – 0.153	0.0744	0.0384
Youtube	5.70	0.000* – 0.005	0.0005	0.005 – 0.661	0.0846	0.0426
Overall		0.000* – 0.014	**0.0011**	0.000* – 0.661	0.0432	0.0222
Spanish Web		0.000* – 0.008	0.0010	0.010 – 0.607	0.0996	0.0503

Table 6.4: Relative size, range, and *DER* for English and Spanish dyslexic errors in web domains and social media sites. The values over the average *DER* are highlighted and 0.000* represents a number larger than 0 but less than 0.0005.

Internet Systems Consortium Domain Survey[5] there are 2.1 millions of computers in *.us*[6] there are more than 463 million web pages with that domain. In fact, many websites have both, the *.com* and the country domains.

For English, we consider countries where it is an official or *de facto* official language, or national language. These are, in descending order of native speakers (in parenthesis): United States (215 M), United Kingdom (61 M), Canada (18.2 M), Australia (15.5 M), Nigeria (4 M), Ireland (3.8 M), South Africa (3.7 M), New Zealand (3.6 M) and Guyana (<1 M) [411].

Crystal [99] estimates that non-native speakers now outnumber native speakers by a ratio of three to one. Despite this fact, this estimation depends on how literacy or mastery of a language is defined and measured. So, we have added India (86.1 M) and Philippines (44 M), where English as a second language is widespread [411]. However, in India and Philippines, only 0.2 and 3.4 millions of speakers have English as a first language, respectively. The fact that these two countries mainly use English as a second language may have an influence on a higher rate of spelling errors in general, India having the highest rate of errors. Philippines is the fourth place after United Kingdom and Ireland. The results are given in Table 6.5.

For Spanish we considered the countries where it is an official language. These countries also present the highest percentage of Spanish native speakers of their populations. They are, in descending order of native speakers (in parenthesis): Mexico (104.1 M), Colombia (45.7 M), Spain (42 M), Argentina (36.3 M), Venezuela (28.3 M), Peru (25.0 M), Chile (17.0 M), Ecuador (11.9 M), Cuba (11.2 M), Dominican Republic (10.0 M), Guatemala (8.6 M), Honduras (8.0 M), Bolivia (6.0 M), El Salvador (6.2 M), Nicaragua (5.3 M), Costa Rica (4.5 M), Puerto Rico (3.8 M), Paraguay (3.7 M), Uruguay (3.2 M), Panama (3.0 M) and Equatorial Guinea (1.7 M) [411]. The results are given in Table 6.6.

Since our percentages are relative, the size of the country domain shall not have a great influence in the error rate. In countries with

[5] `www.isc.org/ds/` Consulted in April, 2012.

[6] `http://www.google.ca/advanced_search` Accessed in April, 2012.

Country, *Domain*	*Size*	*Range*	*DER*
India, *.in*	13.83	0.000*–0.004	**0.066**
U.K., *.uk*	39.15	0.000*–0.004	**0.050**
Ireland, *.ie*	2.51	0.000*–0.002	**0.040**
Philippines, *.ph*	3.20	0.000*–0.001	0.034
Canada, *.ca*	10.08	0.000*–0.001	0.034
Overall		0 –0.367	0.034
New Zealand, *.nz*	5.82	0.000*–0.001	0.032
Australia, *.au*	11.97	0.000*–0.001	0.028
U.S.A., *.us*	5.13	0.000*–0.001	0.023
South Africa, *.za*	7.90	0.000*–0.001	0.022
Nigeria, *.ng*	0.29	0.000*–0.001	0.008
Guyana, *.gy*	0.12	0.000*–0.000*	0.006

Table 6.5: Relative size (%), range and DER for a sample of frequent dyslexic errors in several English speaking countries' domains. The values over the DER average are highlighted and 0.000* represents a number larger than 0 but less than 0.0005.

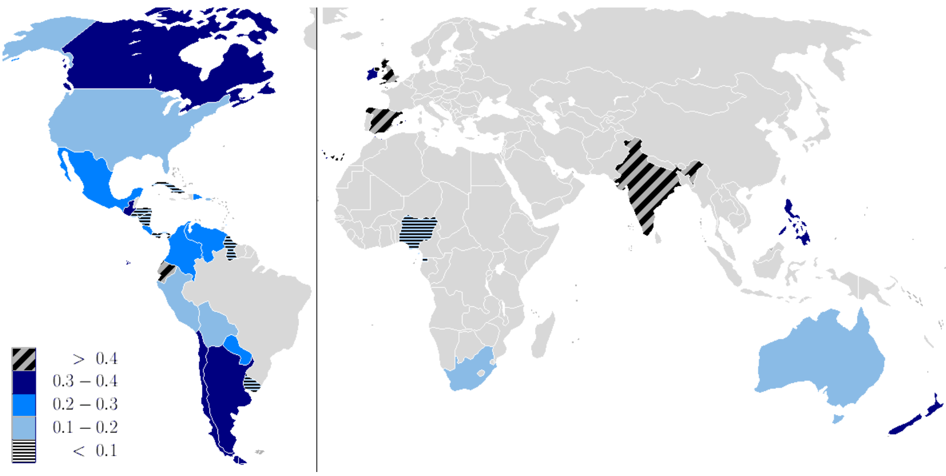


Figure 6.2: Geographical distribution of dyslexic errors in English and Spanish speaking countries' domains.

Country, *Domain*	*Size*	*Range*	*DER*
Ecuador, *.ec*	4.85	0.000*–0.350	**0.046**
Spain, *.es*	38.56	0.000*–0.121	**0.041**
Chile, *.cl*	7.90	0.000*–0.138	**0.039**
Guatemala, *.gt*	0.24	0 –0.316	**0.039**
Argentina, *.ar*	14.70	0.000*–0.128	**0.034**
Peru, *.pe*	4.66	0 –0.127	**0.028**
Bolivia, *.bo*	0.46	0 –0.120	**0.026**
Puerto Rico, *.pr*	0.69	0 –0.183	**0.021**
Costa Rica, *.cr*	0.84	0 –0.067	0.019
Overall		0 –0.350	0.019
Paraguay, *.py*	0.27	0 –0.076	0.018
Venezuela, *.ve*	2.61	0.000*–0.058	0.018
Mexico, *.mx*	6.61	0.000*–0.039	0.015
Colombia, *.co*	5.88	0.000*–0.056	0.015
Dominic. Rep., *.do*	1.55	0.000*–0.055	0.014
Honduras, *.hn*	0.12	0 –0.035	0.007
Nicaragua, *.ni*	0.88	0 –0.031	0.007
Panama, *.pa*	0.44	0 –0.020	0.006
Uruguay, *.uy*	5.50	0 –0.010	0.005
Cuba, *.cu*	1.47	0 –0.018	0.005
El Salvador, *.sv*	1.75	0 –0.016	0.004
Eq. Guinea, *.gq*	0.01	0 –0	0

Table 6.6: Relative size (%), range and DER for a sample of frequent dyslexic errors in several Spanish speaking countries' domains. The values over the DER average are highlighted and 0.000* represents a number larger than 0 but less than 0.0005.

small sizes such as Guyana and Equatorial Guinea, the low rate could be due to other reasons. For instance, even though English or Spanish are official languages in those countries, other native languages are spoken by the population as well; however official websites used English and Spanish. Surprisingly, *.us* has a relatively low rate of dyslexic errors maybe because of the fact that in the USA, the domain *.us* is less frequent than *.com* or *.net*, but USA has the highest number of

Internet users [179]. Notably, India, United Kingdom, Ecuador, and Spain have the highest rate of dyslexic errors. Figure 6.2 compares the results for all the countries studied.

6.4.4 Validating the Measure

To validate DER we have: (1) corroborated that DER rates have different order of magnitude in the Web than in non-corrected corpora written by native and non-native speakers, and (2) checked that DER is not correlated with the general lexical quality of the Web, that is, web pages with a high DER do not have a higher spelling error rate due to a greater presence of misspells, not necessarily dyslexic errors.

For the first validation we took in consideration the largest corpora available for English and Spanish: the Collins Word banks Online[7] with 550 million words, the British National Corpus (BNC)[8] with 100 million words and the Royal Spanish Academy Corpus or Current Spanish (CREA)[9] with 3.5 million words. These corpora are made of written and spoken language (non-aphasic) from various sources and with no corrected errors, although as the sources are of high quality we would expect to have a much lower DER. We also took into account the only available corpus we found composed of English essays written by students who use English as a foreign language, the Janus Pannonius University Corpus (JPC).[10] In these corpora we only found examples of simple errors such as **poeple (people)* but no examples of multi-errors. We computed DER for these corpora (see Table 6.7) and for both English and Spanish the DER was negligible (DER=0.001 for Spanish and DER=0.002 for English). However note that as expected, DER in the corpus of less quality, JPC, was higher.

For the second validation we considered the general spelling errors rate (not only dyslexic errors) for web pages. To compute the spelling error rate we use the lexical quality (LQ) of web pages. LQ is a

[7] `http://wordbanks.harpercollins.co.uk/auth/`

[8] `http://www.natcorp.ox.ac.uk/`

[9] `http://corpus.rae.es/creanet.html`

[10] `http://joeandco.blogspot.com.es/` Unfortunately, the rest of the corpora we found based on written essays of students are not available, such as, ICLE (International Corpus of Learner English), WriCLE (Written Corpus of Learner English), CEDEL2 (Corpus Escrito del Español L2), USE (Uppsala Student English Corpus), the Catalan-English Barcelona Corpus and Spencer Corpus.

	English			**Spanish**
Corpus	BNC	Collins	JPC	CREA
DER	0.000*	0.001	0.005	0.001

Table 6.7: *DER* in non-corrected corpora. The value 0.000* represents a number larger than 0 but less than 0.0005.

Pages in	2011		2012	
English	Google	Bing	Google	Bing
.org	0.038	0.075	0.066	0.044
.net	0.080	0.096	0.157	0.099
.com	0.051	0.081	0.149	0.219
Web	0.047	0.099	0.107	0.220

Table 6.8: *LQ* for English, two different search engines, two different years, and three major Internet domains.

measure similar to *DER* [17] which takes into consideration all kind of possible errors in the Web and gives as a result an estimation of the web page error rate. In Table 6.8 we present the *LQ* of two years for the overall Web and three major Internet domains and English pages. Specific *LQ* results in major and social media websites[11] are given in [318]. We found that many Web 2.0 sites have quite good *LQ* in spite of their collaborative nature, like Wikipedia or Flickr. We can notice that *LQ* increased almost in all cases from 2011 to 2012 (we include more values in the next section that shows the same). That is, the lexical quality is getting worse. There are a few factors that can explain this trend. First, the expansion of the Web 2.0, which has lower quality. In fact, correct spelling does not seem to be a goal since there are deliberate misspells. Second, most new users are young and they usually do not care much about spelling.

To show the value of *LQ* as an independent measure, we computed the Pearson correlation for the following measures in the top 13 common websites[12] of ComScore unique visitors in USA (December 2011)

[11] Such as Facebook, Flickr, Y! Answers, Twitter, Youtube, Blogger, etc.

[12] `amazon.com,aol.com,craigslist.org,ebay.com,espn.go.com, facebook.com,google.com,linkedin.com,msn.com,netflix.com,twitter.`

Measure	Alexa	Pages	Links	ComScore
LQ	0.803	0.775	0.678	0.779
Alexa		0.897	0.794	0.790
Pages			0.850	0.632
Links				**0.437**

Table 6.9: Pearson correlation for top English websites in early 2012.

Pearson correlation	**English**		**Spanish**	
	Countries	Social Media	Countries	Social Media
LQ/DER	0.097	-0.363	-0.200	-0.205

Table 6.10: Pearson correlation for DER and LQ in different web domains.

and the Alexa.com reach (February 2012): LQ, Alexa reach, number of pages in websites (as given by Google), number of in-links (as given by Alexa), and ComScore unique visitors. The results are given in Table 6.9, where we can observe that LQ is partially correlated to all these measures, but at the same time gives additional information. This shows that more content implies a higher misspelling rate and that web traffic does not imply better lexical quality. Therefore, due to this comparison of LQ with Web popularity, we believe that LQ is a good estimator of the lexical quality of a website.

Intuitively, DER could be correlated to the overall LQ of a website, because when the general error rate grows, DER should grow too. But our results show that this is not the case. We took the results related to the general lexical quality of the Web we presented in [17] and computed the Pearson correlation between measures (DER and LQ) for the English and Spanish speaking countries and the major social media sites. The data used is normally distributed (Shapiro–Wilk test). In Table 6.10, we can observe that DER is not correlated with the LQ measure. This shows that a higher misspelling rate does not imply a higher DER.

Note that in all the corpora used to validate DER, we assume that

`com`,`wikipedia.org`, and `yahoo.com`.

the majority of the authors involved are not dyslexics because: (a) the corpora are composed by written and spoken language made by non-aphasic from various sources, without mention to dyslexia or other aphasias in the descriptions of the corpora (BNC, Collins, CREA); plus the description of the corpus of essays written by students who use English as a foreign language (JPC) nothing is mentioned about the possible learning difficulties of the students; and (b) the measure of lexical quality uses the most frequent errors found in the Web, for instance, **becuase (because)*, and most people sometimes make this kind of errors, not only dyslexics.

Hence, although there is no conclusive evidence that these errors found in the Web were produced by dyslexic people, it is highly probable because of: the strict criteria in the selection of the dyslexic errors (see Section 6.3.4); the validation performed using non corrected and non native speaker corpora for both languages; and the fact that DER is not correlated with general lexical quality of the Web. Therefore, we believe that DER is a good estimation of the lower bound of the impact of dyslexic errors in the Web.

6.5 Discussion

Our lower bound shows that at least 0.07% and 0.09% of the web pages contain dyslexic errors for the English and Spanish Web, respectively. Although this is a small percentage, for each 20 billion Web pages, there are at least one million pages containing dyslexic errors. If we consider all spelling errors as dyslexic errors, the lower bound would increase to close to 0.2% and for each 10 billion pages, 20 million Web pages would contain dyslexic errors. These results could be surprising considering that the estimations of dyslexia among population are higher for English –from 5.3% [189] to 17.5% [176]– than for Spanish –from 7.5% [146] to 11.8% [71]–. However, these estimations are based on reading tests but not on writing misspells. If we take into consideration the error rates found in the corpora written by dyslexic people, then our results might be expected since there is a higher rate of multi-errors in English (39%) than in Spanish (23%) and a higher rate of simple errors in Spanish (67%) than in English (53%). Therefore, the high presence of simple errors in Spanish has an effect on the final estimation of DER being higher for Spanish than for English.

These results are conservative due to two reasons: the fact that DER was designed as a lower bound for making the estimation in the Web feasible (see Section 6.3.1), and the strict conditions that the words for the data sets must meet to assure as far as possible their dyslexic origin (see Section 6.3.2). For example, the errors in our data sets are long words and previous research on dyslexia reveals that error frequency is related to word length [277, 371], errors in shorter words being more frequent than in longer ones.

Our results should be taken with care, since DER is a lower bound and there is no consensus on the definition of dyslexia and previous user studies with dyslexics pointed out that dyslexia is highly variable and there is no typical profile of a dyslexic Internet user [290, 157]. However, our estimations are useful to consider the prevalence of dyslexia in the Web as well as to motivate dyslexic-accessible practices.

Our main findings are that:

– The amount of dyslexic texts in the Web is not as large as it could be. This suggests that the widespread use of spell checkers ameliorates the presence dyslexia in the Web so the prevalence of content with dyslexic errors is a function of both people and technology.

– The rate of dyslexic errors is independent from the rate of spelling errors in web pages. A comparison with our previous estimations made one year ago shows that the error percentages are growing. Hence, the increase of the Web does not correspond with the improvement of its lexical quality, which can be explained by the fact that each year we have a large number of new users.

– Spanish has a higher DER for simple dyslexic errors than English. However, if we compare the percentages of the different error kinds, there are more dyslexic errors in the English Web than in the Spanish Web.

– Even though Spanish orthography is shallower than English, the difference between these two languages in terms of dyslexic error rates in the Web is not as substantial as expected.

– Particular words can be used to detect dyslexic texts, and hence users with dyslexia. This can be used to improve Web accessibility as well as future spell checkers targeted to dyslexic users.

Part III

TEXT PRESENTATION

This third part of the thesis addresses how the presentation of the text leads to a better readability and comprehensibility of texts for people with dyslexia. First, we performed an exploratory experiment called *Text Presentation* experiment composed of eight sub experiments, where we investigated which of the text presentation parameters have an impact on the reading performance of readers with and without dyslexia. In that first experiment we tested color combinations for the font and the screen background, grey scales in the font, grey scales in the background, font size, column width as well as character, line and paragraph spacing (Chapter 7). The levels tested each of the conditions were based in previous research and recommendations. This experiment took part in the first round of the studies. Then, based on these results we conducted two more experiments: the *Wikipedia* experiment and the *Font* experiment.

In the *Wikipedia* experiment (Chapter 8) we tested the effect of font size and line spacing –and their interactions– in the context of the Web. We chose line spacing because it presented a strong correlation with reading performance in the previous experiment, and we chose font size because it was the only condition that leads to significant effects for each of its levels in the previous experiment. We decided to use Wikipedia because it is the most popular text heavy website as well as a world wide educational resource.

Since font size presented the greatest effects and font size and font type are related, we conducted the *Font* experiment (Chapter 9) selecting the most frequently used fonts in the Web as well as in printed texts together with fonts designed specifically for people with dyslexia. We found significant effects in both experiments, *Wikipedia* and *Font*. The combination of the results of the three experiments brings us a set of recommendations for displaying dyslexic-friendly text on screens.

Chapter 7

Colors, Sizes and Spacing

7.1 Introduction

The presentation of a text has a significant effect on the reading speed of people with dyslexia. To know which are good text presentation parameters we performed an eye tracking study with 92 people, 46 with dyslexia and 46 as a control group, where we measured the reading performance of the participants. We studied the following parameters: color combinations for the font and the screen background, font size, column width as well as character, line and paragraph spacing. We found that larger text and larger character spacing lead the participants to read significantly faster. We complemented our study with questionnaires to obtain the participants preferences for each of these parameters, finding other significant effects. These results provided evidence that people with dyslexia may benefit from specific text presentation parameters that make text on a screen more readable. To the best of our knowledge, this was the first time that eye tracking is applied with such an extensive group of people to define dyslexic-friendly text presentation recommendations. The main contributions are the following:

– Larger text lead people with and without dyslexia to read significantly faster.

– Larger character spacing lead people with and without dyslexia to read significantly faster.

The results of this chapter are presented in Rello *et al.* [331] and [335].

7.2 Related Work

According to a survey by McCarthy and Swierenga [228], studies about dyslexia and accessibility are scarce compared to other groups of users with special needs. However, as Al-Wabil *et al.* [4] claim, there are considerable barriers for people with dyslexia.

In the guidelines for accessible Web, that is, the *Web Content Accessibility Guidelines (WCAG)* [63], dyslexia is only one more disability within a diverse group of cognitive disabilities. According to Santana *et al.* [109] this lack of explicit consideration of dyslexia specificities in the guidelines make the needs of users with dyslexia unfulfilled.

We divide previous work related to dyslexic-friendly proposals in user studies regarding accessibility in general and each of the presentation parameters studied in this chapter.

7.2.1 Accessibility

There is a common agreement in specific studies about dyslexia and accessibility that the application of dyslexic-accessible practices benefits also the readability for non-dyslexic users [114, 199, 228]. Consequently, the guidelines for developing friendly websites to users with dyslexia [51, 296, 423] usually overlap with other disabilities, such as low vision [130], or with guidelines for low-literacy users [250]. For example, according to Zarach [423] their guidelines for enhance readability for people with dyslexia, also benefit people without dyslexia. However, there is no universal profile of a user with dyslexia and therefore some authors recommend using a customizable environment to enhance such users [158, 250].

The accessibility user studies related to text presentation were presented extensively in Section 2.2.2. We recall the main user studies on the accessibility field: *SeeWord* by Gregor *et al.* [157], work by Kurniawan and Conroy [199], and *Firefixia* by Santana *et al.* [347].

7.2.2 Gray Scales and Colors

Bradford [51], the British Dyslexia Association [52], Rainger [296] and Tseng [388] recommend different colors for people with dyslexia. Bradford [51] does not recommend using a pure black text on a pure white background due to its high contrast, as many people with dyslexia are sensitive to the brightness and this can cause the words to swirl or blur together. Tseng [388] suggests using light grey as background, such as the color with the following hexadecimal code: FFFFE5. Gregor and Newell [158] explored in a user study different color combinations. In their studied (text/background): brown/mucky green and blue/yellow pairs were chosen by people with dyslexia, as in the experiments carried out by Gregor and Newell [157, 158].

7.2.3 Font Size

Finding small font size is recalled as one of the main problems of people with dyslexia [228]. The recommendations by the British Dyslexia Association [52], Rainger [296], Bradford [51] and Zarach [423] for font size for this target group is 12 or 14 points. However, some readers with dyslexia may prefer a larger font [52, 113].

O'Brien *et al.* [256] compared the reading speeds using twelve different font sizes between two groups: children with dyslexia (aged 7 to 10 years) and without dyslexia (aged 6 to 8 years). They showed how dyslexic reading follows the same curve shape as skilled reading, with constant reading rates across large font sizes and a sharp decline in reading rates below a critical font size. Readers with dyslexia presented higher critical font sizes. In the following we focus on studies about font size and non-dyslexic readers.

Regarding people without dyslexia, in his article on the *Top 10 Mistakes in Web Design* [252] Jakob Nielsen argues that providing text in the right font size is crucial for the usability of any web page. Preferably, users should be allowed to adjust the font size to their individual needs. Yet, Nielsen also points out that users are typically unwilling to change fonts when viewing websites. Consequently, to ensure good readability, it is essential for websites to provide appropriate defaults. Nielsen recommends using font sizes of at least 10 points or 12 points for elderly readers. However, previous studies come to different conclusions about the ideal font size.

In the 1920's – 40's, Paterson and Tinker [264] studied the effect of typographical parameters on printed text. They measured the reading speed of 320 college students and found out that 10-point text yields the fastest reading compared to 6, 8, 12, or 14 points.

Bernard *et al.* [31] performed a study with 60 participants measuring reading time, preference, and errors while reading the text out loud using eight different font types and 10, 12, and 14 points. Fonts of 10 points were read significantly slower than fonts with 12 points. In a subsequent experiment, Bernard *et al.* [32] compared two fonts –*Arial* and *Times*– and two font sizes –10 and 12 points– with 35 participants. The experiment used the same dependent measures, 10-point *Arial* typeface again was read slower than the other conditions and the 12-point *Arial* typeface was preferred to the other typefaces.

In order to understand the impact of age on reading, Bernard *et al.* [30] studied the effects of font type and size on the legibility and reading time of online text by older adults. They compared legibility, reading time, and the participants' preferences of texts displayed with sans serif and serif fonts, and font sizes of 12 and 14 points. The 12-point serif fonts were read significantly slower than 14 serif and sans serif fonts, and participants preferred larger font sizes.

Darroch *et al.* [104] investigated the effect of font sizes ranging from 6 to 16 points, measuring the reading speed, reading accuracy, and subjective views among two groups, 12 old and 12 young readers. They did not find any significant differences, neither between the age groups, nor for the font sizes ranging from 6 to 16 points.

Banerjee *et al.* [19] performed a study with 40 participants, which had to read texts using the font sizes of 10, 12, and 14 points. The 14-point font leads to a significant faster reading (read aloud) and was preferred over 10 and 12 points.

Bhatia *et al.* [39] studied the effect of font size, italics, and number of colors on readability. A group of 180 undergraduate students had to take part in a text-reading experiment and indicate their preferences in a survey. The font sizes that Bhatia *et al.* tested were of 10, 12, and 14 points. Unfortunately, the survey responses did not reveal any significant effects.

The related work described until now measured preferences, reading time, and errors made while reading the text aloud to approximate readability. However, these measures have drawbacks. Subjective readability may not match objective readability. Participants may simply skim texts which are difficult to read, and hence result in misleadingly fast reading times. Reading aloud may introduce unsystematic variance through the extra reading activity, and is not valid for web reading.

As a remedy, Beymer and Russel [37] developed WebGazeAnalyzer, a system to monitor reading performance with an eye-tracker. This system allows, amongst other things, to record the eyes' fixations durations, which are an objective indicator of text readability [186, 305, 352]. The shorter the eyes fixate text parts at a time, the better the text's readability. Using this system in a between-subjects design with 82 participants, Beymer *et al.* [36] studied the effect of the font sizes 10, 12, and 14 points on readability and comprehension scores. When using 10 points font size, fixation durations resulted significantly longer as compared to 14 points. They also found significant differences related to the mother language of the participants: non-native English subjects had significantly longer fixations.

The findings from previous work unanimously indicate that the font sizes of 10 to 12 points, as suggested by Nielsen and other sources, might be too small and that bigger font sizes might be needed to achieve optimal readability of web pages.

7.2.4 Character Spacing

Regarding character spacing there is a user study by Zorzi *et al.* [424] and the recommendations of Rainger [296], and Pedley [279]. Zorzi *et al.* [424] conducted an experiment with 74 children with dyslexia (34 Italian and 40 French, aged between 8 and 14). The children read on paper texts with regular character spacing and extra large character spacing (an increase of 2.5pt in the standard letter spacing using 14 point Times font). The texts with larger character spacing lead to a better reading accuracy (number of errors) and speed (number of syllables per second). Consistently [279] recommends to create a slightly larger distance between individual words and reduce letter-spacing slightly, so that the letters within a word lie closer together. On the other hand, Rainger [296] suggests to have large spacing between letter combinations.

7.2.5 Line Spacing

Another key factor of legibility for people with dyslexia is line spacing [168]. Line spacing can be given in various units. In the web context, we often find values without units, such as 1.0. These values are factors that describe the line spacing relative to the default spacing. For example, for a default line spacing of 16 pixels, the factor 1.5 produces a line spacing of 24 pixels. Recommendations in previous work comprise line spacing of 1.3 in Pedley [279], 1.5 by British Dyslexia Association [52], and 1.5 to 2 lines in Rainger [296].

For people without dyslexia, we found no specific guidelines for line spacing of web texts. By default, browsers compute the line spacing relative to the font size. A spacing of 1.0 equates to 120% of the font size.[1] In best-practice recommendations, this spacing of 1.0 is often named as "generally the most readable and doesn't require that you do anything special".[2] However, no studies are cited.

According to a review by Bix [41], the vast amount of literature indicates that the optimal amount of spacing highly depends on other factors. Except for the general recommendation to avoid too little and too much spacing, no rules are given.

Paterson and Tinker [265] studied the effect of line spacing in printed text when performing a reading test (Chapman-Cook Speed of Reading Test) with 400 college students. They found that bigger line spacing (1.2 and 1.4 compared to 1.1) lead to faster readings. However the authors point out that such results may depend on other factors such as font type, column width, or font size, as already mentioned.

7.2.6 Paragraph Spacing

Bradford [51] tackles paragraph spacing suggesting that paragraphs –even when they have a single line– should always be spaced out with an empty line between each paragraph.

[1] `http://stackoverflow.com/questions/2262543/css-line-height-guide`

[2] `http://webdesign.about.com/od/styleproperties/qt/css_line_spacing.htm`

7.2.7 Column Width

There are the recommendations of the British Dyslexia Association [52] and Bradford [51] which suggest to avoid long lines –60 to 70 characters–, and to avoid narrow columns (long lines).

In a user study by Schneps *et al.* [348], they performed an experiment with 27 high school students with reading struggles. They tested line length and extra large spacing. Regarding line length they compare two screen dimensions: iPod Touch in portrait mode (5 cm $\times$ 7.5 cm) and the Apple iPad in landscape mode (19.7 cm $\times$ 14.8 cm). They found that using a small device improved readability (faster reading speeds, less number of fixations, and less regressive saccades). Regarding extra large spacing, they conclude that it improves comprehension in those most impaired.

7.2.8 What is Missing?

This study differs from the rest of the related approaches in the application of eye tracking to measure objective readability of a text –with the exception of [348]– with a greater number of participants 92 (46 with dyslexia).

7.3 Methodology

We conducted one experiment *Text Presentation* (with eight sub experiments) to study the effect of eight text presentation parameters on readability. In the experiments, 92 participants (46 with dyslexia) had to read a set of texts, which were altered to include the different values of the parameters.

7.3.1 Design

In Table 7.1 we show a summary of the experiment *Text Presentation* following the methodology explained in Section 3.

Independent Variables

In our experimental design, there were eight conditions that served as *independent variables.* The levels of the variables were chosen taking into account the difficulties that people with dyslexia find (see Section 2.1.5); previous user studies (see Section 2.2.2); and literature

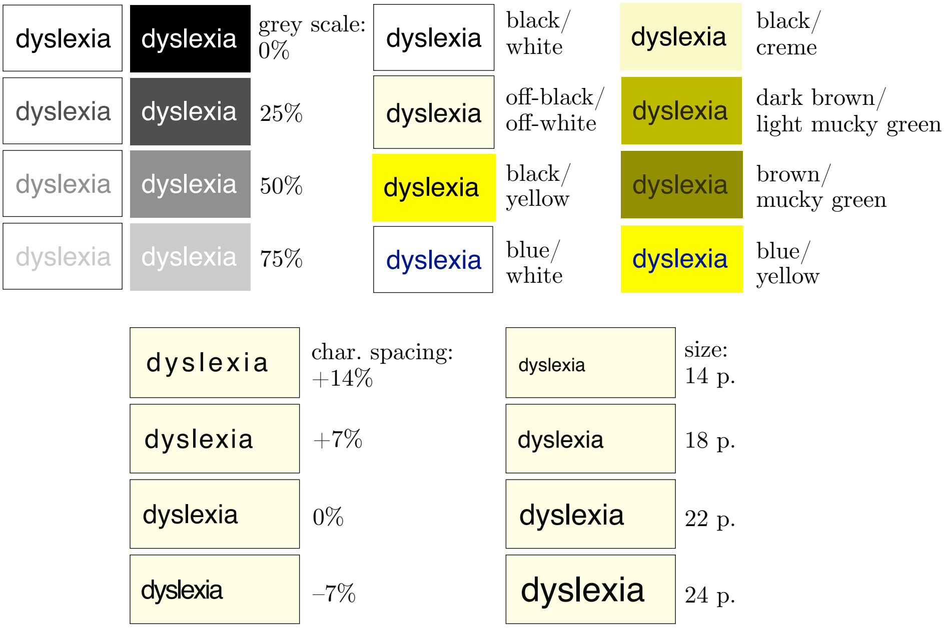


Figure 7.1: Text and background grey scales, colors, font size and character spacing of the *Text Presentation* experiment.

about recommendations for readers with dyslexia (see Section 2.2.4). More details of our parameters in comparison with the literature are discussed in Section 7.5. Next, we present the independent variables and their levels.

Text grey scale: This independent variable has four levels corresponding with four grey scale values for the text with white background: 0% (black font), 25%, 50% and 75% (See Figure 7.1).[3]

Background grey scale: This independent variable has four levels corresponding with four grey scale values for the background with white text: 0% (black background), 25%, 50% and 75% (See Figure 7.1).[3]

[3]The CYMK code for the colors and their contrast are shown in the Appendix A.4.

Text and background colors: We tried eight color pairs (text/background) from previous studies or recommendations (see Section 7.2.2): black/white, off-black/off-white, black/yellow, blue/white, black/creme, dark brown/light mucky green, brown/mucky green and blue/yellow (See Figure 7.1).[3]

Font size: This independent variable has four levels corresponding with four font sizes: 14, 18, 22, and 26 points (See Figure 7.1).

Character spacing: This independent variable has four levels corresponding with four distances between characters: -7%, 0%, +7%, and 14% (See Figure 7.1). The base level (0%) corresponds to the default spacing of the font.[4]

Line spacing: This independent variable has four levels corresponding with four values for spacing: 0.8, 1, 1.2, and 1.4 lines. The base level (1.0) corresponds to 120% of the font size (default spacing).

Paragraph spacing: This independent variable has four levels corresponding with four values for the spacing between paragraphs: 0.5, 1, 2, and 3 lines.

Column width: This independent variable has four levels corresponding with four values for column width tested: 22, 44, 66, and 88 characters per line (as the fonts have letters with variable width, this is the average number of characters per line).

The experiments followed a within-subjects design, so every participant contributed to each of the conditions in the experiments. The order of conditions was counter-balanced to cancel out sequence effects. No combinations of conditions were studied.

[4]Although there are others units that can be used, the simplest is to use a percentage of the current font size.

Dependent Variables

For quantifying objective readability and subjective preferences we used *Fixation Duration* and *Preference Rating* as *dependent variables*. To control comprehension we used *Comprehension Score* as a *control variable*. These dependent measures are explained in detail in Section 3.1.1.

In this experiment the *Comprehension Score* was computed as the percentage of correct answers, where the correct choice scored 100% and the others 0%. To guarantee that recordings analyzed in this study were valid, we used the comprehension score as a control variable, such that if the recording of a complete session did not have an overall 100% comprehension score, it was discarded from the analysis.

In the *Text Presentation* experiment the *Preference Rating* was computed differently as in the rest of the experiments that use Likert scales. We asked them to select the texts that they found the easiest to read, that is, the most readable one. For each condition, they wrote their answers in a paper questionnaire while they saw the options on the screen. The participant could see the options as much time and as many times as desired. Whenever the participant selected one, two, three or four options as most readable, we gave the weights 1, 0.5, 0.33 and 0.25, respectively, to those options. To calculate the average preference score we added the weights divided by the number of participants.

7.3.2 Participants

The details of the participants groups for both experiments are given in Table 7.1. Two of the participants with dyslexia were diagnosed with attention deficit disorder. For more details about the participants please refer to Section 3.1.2

7.3.3 Materials

Texts

For the reading tests we used two stories. The first story[5] was written in verse and contains 733 words, while the second story is a fragment

[5] *Los Encuentros del Caracol Aventurero (The Encounters of the Adventurous Snail)* by Federico García Lorca.

in prose[6] with 204 words (see the texts in the Appendix A.3).

We divided the overall text in 36 fragments and each of them was presented to the participants with a different condition. To maintain the independence of the variables, there were no combinations among the conditions levels. The texts fragments belonging to different levels of the same condition were comparable to each other. They have the same number of words and the same number of syllables for the shorter passages (texts containing less than 22 words). The shorter texts were extracted from the story written in verse so they were very similar to each other having the same rhythm and meter.

Depending on the length of the text fragment, some of them were presented in a single slide while others were presented in groups in the same slide. There were a total of 20 slides. As we already mentioned, the order of conditions was counter-balanced to cancel out sequence effects.

Text Presentation

We used the parameters detailed in Section 3.1.3 with black font on creme background (except from the slides with the condition *Color*) and 20 points font size (except from the slides with the condition *Font Size*).

Comprehension Questionnaires

We used multiple-choice questions with three possible choices, one correct choice and two wrong (Table 7.1). See Section 3.1.3 for details about the comprehension questionnaires. An example of an item is given in Figure 7.2 (top).

Preferences Questionnaire

We replay the slides that the participant read and through a paper questionnaire, the participant chose what s/he thought was the best reading alternative between the options given for each of the parameters. The questionnaire had eight items, one for each experiment, and four to eight possible choices depending on the number of levels of the variable. Each item was composed of one statement and the options.

[6] From the book *¿Soy dix-leso? (Am I dyx-leso?)* of the *Papelucho* series by Marcela Paz.

¿Por qué querían matar las hormigas a su compañera?
'Why did the ants wanted to kill their colleague?'

☐ Porque era egoísta. *'Because she was selfish.'*

☐ Porque veía las estrellas. *'Because she saw the stars.'*

☐ Porque no rezaba. *'Because she did not pray.'*

■ Leo mejor el texto... *'I read best the text...'*

1 ☐ 2 ☐ 3 ☐ 4 ☐

Figure 7.2: Comprehension and preference items examples of the experiment *Text Presentation*.

The statement was always the same. See an example in Figure 7.2 (bottom).

7.4 Results

In this section, we present the analyses of the data of both groups, group D and group C. The comprehension score was used to filter the reading recordings. Recall that if the reading recording did not had an overall 100% comprehension score, we discarded it. In this case, only one recording from the group D and two from the group C were discarded.

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

- Two-way ANOVA test for repeated measures plus a complete pairwise post-hoc comparison using paired t-tests with a Bonferroni adjustment, to show effects of the conditions on fixation duration among groups D and N.

- Then, to show effects of the conditions on fixation duration within groups, we divided the data and used a one-way ANOVA test for repeated measures plus a complete pairwise post-hoc comparison using paired t-tests with a Bonferroni adjustment.

- At the end, we used Pearson's Chi-squared test to show effects of the participants' choices.

Please, refer to Table 7.2 for all the means and standard deviations. The percentage shows their fixation extra time in comparison with the lowest value and the best results are shown in boldface.

7.4.1 Text Grey Scale

Objective Readability

- **Between Groups:** With two-way ANOVA, we found significant effects for the groups ($F(1,366) = 14.52,\ p < 0.001$) on fixation duration. But we did not find any effects of text grey scale ($F(3,366) = 0.09,\ p = 0.968$) on fixation duration. We also did not find interaction effects of text grey scale and the group ($F(3,366) = 1.40,\ p = 0.242$).

  The results of the post-hoc tests show that:

  - Participants with dyslexia had significantly longer fixation times ($\bar{x} = 0.23,\ s = 0.08$ seconds) than the participants without dyslexia ($\bar{x} = 0.21,\ s = 0.06$ seconds, $p < 0.001$) (Table 7.2).

- **Within Groups:** We did not find a significant effect of text grey scale on fixation duration in group D ($F(3,183) = 0.46,\ p = 0.711$) nor in group C ($F(3,183) = 1.31,\ p = 0.274$). See Figure 7.3 for the means of the fixation durations.

Preferences

- **Within Groups:** Participants with and without dyslexia found black text significantly more readable than text presented with different grey scales ($\chi^2(3) = 15.13,\ p = 0.002$ for group D and $\chi^2(3) = 39.87,\ p < 0.001$ for group C). See Figure 7.3.

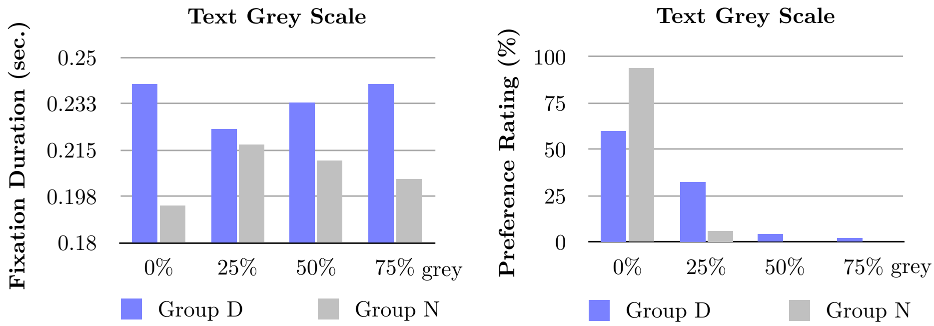


Figure 7.3: *Fixation Duration* means and *Preference Rating* (score in %) for text grey scales.

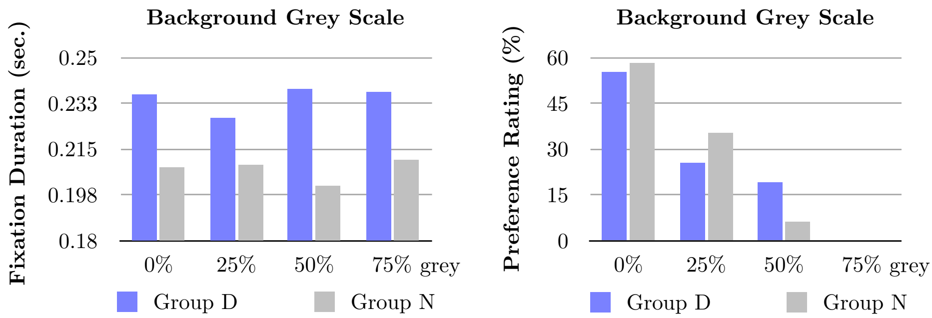


Figure 7.4: *Fixation Duration* means and *Preference Rating* (score in %) for background grey scales.

7.4.2 Background Grey Scale

Objective Readability

- **Between Groups:** With a two-way ANOVA, we found significant effects for the groups ($F(1, 376) = 14.63,\ p < 0.001$) on fixation duration. But we did not find any effects of background grey scale ($F(3, 376) = 0.14,\ p = 0.938$) on fixation duration. We also did not find interaction effects of background grey scale and groups ($F(3, 376) = 0.25,\ p = 0.860$).

  The results of the post-hoc tests show that:

  - Participants with dyslexia had significantly longer fixation times ($\bar{x} = 0.24$, $s = 0.08$ seconds) than the participants without dyslexia ($\bar{x} = 0.21$, $s = 0.06$ seconds, $p < 0.001$) (Table 7.2).

- **Within Groups:** We did not find a significant effect of background grey scale on fixation duration in group D ($F(3, 188) = 0.19$, $p = 0.904$) or group C ($F(3, 188) = 0.20$, $p = 0.893$). See Figure 7.4 for the means of the fixation durations.

Preferences

- **Within Groups:** Participants with and without dyslexia found pure black background significantly more readable than text presented with different grey scales ($\chi^2(3) = 11.10$, $p = 0.011$ for group D and $\chi^2(3) = 21.79$, $p < 0.001$ for group C). See Figure 7.4.

7.4.3 Text and Background Colors

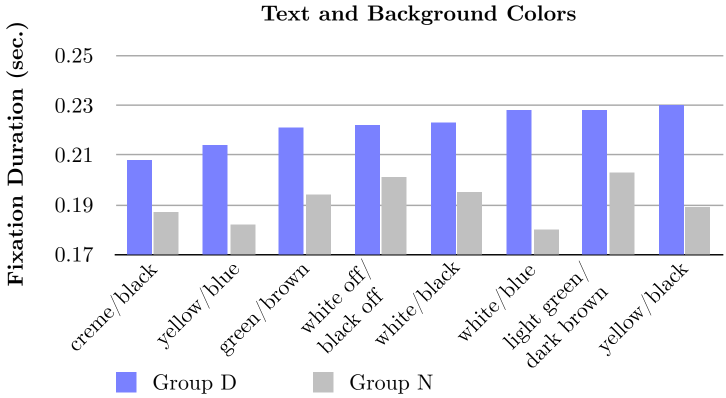
Objective Readability

- **Between Groups:** With a two-way ANOVA, we found significant effects for the groups ($F(1, 732) = 40.03$, $p < 0.001$) on fixation duration. But we did not find any effects of text and background colors ($F(7, 732) = 0.85$, $p = 0.548$) on fixation duration. We also did not find interaction effects of text and background colors and groups ($F(7, 732) = 0.53$, $p = 0.814$).

  The results of the post-hoc tests show that:

  - Participants with dyslexia had significantly longer fixation times ($\bar{x} = 0.22$, $s = 0.08$ seconds) than the participants without dyslexia ($\bar{x} = 0.19$, $s = 0.06$ seconds, $p < 0.001$) (Table 7.2).

- **Within Groups:** We did not find a significant effect of text and background colors on fixation duration in group D ($F(7, 366) = 0.47$, $p = 0.858$) or group C ($F(7, 366) = 1.15$, $p = 0.332$). See Figure 7.5 for the means of the fixation durations.



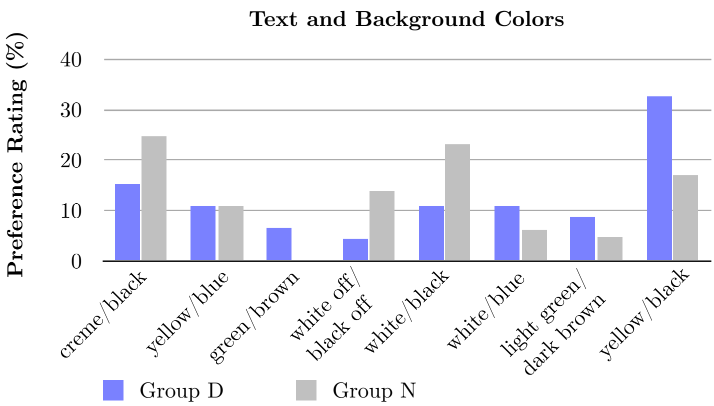


Figure 7.5: *Fixation Duration* means and *Preference Rating* (score in %) for text and background colors.

Preferences

- **Within Groups:** Participants with and without dyslexia did not find any of the text and background colors significantly more readable ($\chi^2(7) = 11.82$, $p = 0.107$ for group D and $\chi^2(7) = 10.29$, $p = 0.172$ for group C). See Figure 7.5.

7.4.4 Font Size

Objective Readability

- **Between Groups:** With a two-way ANOVA, we found signif-

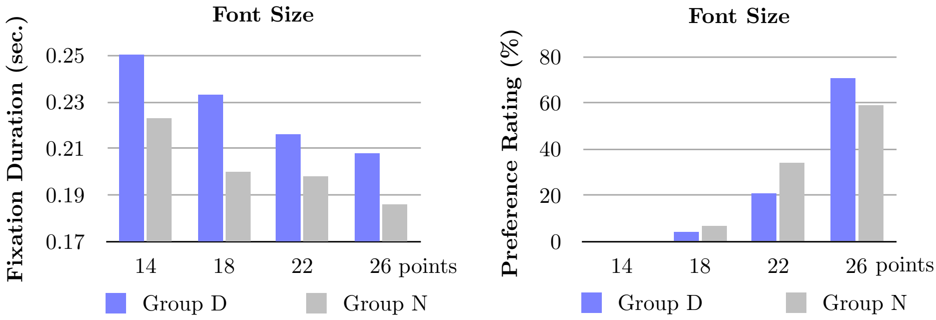


Figure 7.6: *Fixation Duration* means and *Preference Rating* (score in %) for font size.

icant effects for the groups ($F(1, 372) = 19.71,\ p < 0.001$) and for font size ($F(3, 372) = 8.80,\ p < 0.001$) of fixation duration. We did not find any interaction effect for font size and groups ($F(3, 372) = 0.46,\ p = 0.712$).

The results of the post-hoc tests show that:

- Group D had significantly longer fixations durations ($\bar{x} = 0.23$, $s = 0.07$ seconds) than group C ($\bar{x} = 0.20$, $s = 0.05$ seconds, $p < 0.001$) (Table 7.2).
- Participants present significant longer fixation durations with 14 points font size than with 18 points ($p = 0.054$), 22 points ($p = 0.002$) and 26 points ($p < 0.001$) (Table 7.2).

- **Within Groups:** We found a significant effect of font size spacing on fixation duration in group D ($F(3, 186) = 4.97,\ p = 0.002$) and in group C ($F(3, 186) = 4.04,\ p = 0.008$). See Figure 7.6 for the means of the fixation durations.

  The results of the post-hoc tests show that:

  - For group D, the font size of 26 points lead to significant shorter fixation durations ($\bar{x} = 0.21$, $s = 0.05$ seconds) than texts with 14 points ($\bar{x} = 0.26$, $s = 0.09$ seconds) ($p = 0.003$); and font size of 22 points lead to significant shorter fixation durations ($\bar{x} = 0.22$, $s = 0.06$ seconds) than texts with 14 points ($p = 0.003$) (Table 7.2).

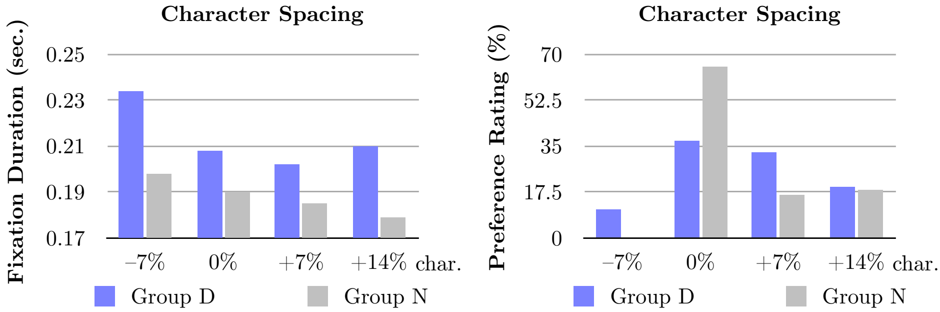


Figure 7.7: *Fixation Duration* means and *Preference Rating* (score in %) for character spacing.

- Participants in group C had significant shorter fixation durations with 26 points font size ($\bar{x} = 0.19$, $s = 0.04$ seconds) than with 14 points ($\bar{x} = 0.22$, $s = 0.07$ seconds, $p = 0.005$) (Table 7.2).

Preferences

- **Within Groups:** Participants with and without dyslexia found texts sizes of 26 points significantly easier to read than the rest of the sizes ($\chi^2(3) = 9.05$, $p = 0.03$ for group D and $\chi^2(3) = 20.79$, $p < 0.01$ for group C). See Figure 7.6.

7.4.5 Character Spacing

Objective Readability

- **Between Groups:** With a two-way ANOVA, we found significant effects for the groups ($F(1, 368) = 16.35$, $p < 0.001$) and for character spacing ($F(3, 368) = 2.86$, $p = 0.037$) of fixation duration. We also found a significant interaction of character spacing and groups ($F(3, 368) = 0.52$, $p = 0.665$).

  The results of the post-hoc tests show that:

  - There is a significant difference of fixation duration between group D ($\bar{x} = 0.21$, $s = 0.07$ seconds) and group C ($\bar{x} = 0.19$, $s = 0.05$ seconds, $p < 0.001$) (Table 7.2).

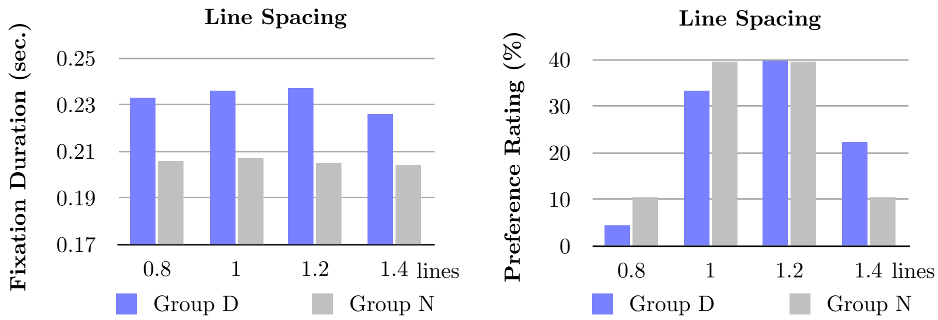


Figure 7.8: *Fixation Duration* means and *Preference Rating* (score in %) for line spacing.

  - Participants present significant longer fixation durations with character spacing -7% than with spacing +7% ($p = 0.035$) and spacing +14% ($p = 0.013$) (Table 7.2).

- **Within Groups:** We did not find a significant effect of character spacing on fixation duration in group D ($F(3, 184) = 1.90,\ p = 0.132$) or group C ($F(3, 184) = 1.28,\ p = 0.282$). See Figure 7.7 for the means of the fixation durations.

Preferences

- **Within Groups:** Participants with dyslexia did not find any of the options significantly easier to read ($\chi^2(3) = 2.03,\ p = 0.567$), while participants without dyslexia found text with 0% character spacing significantly more readable ($\chi^2(3) = 21.54,\ p < 0.001$). See Figure 7.7.

7.4.6 Line Spacing

Objective Readability

- **Between Groups:** With a two-way ANOVA, we found significant effects for the groups ($F(1, 372) = 17.79,\ p < 0.001$) on fixation duration. But we did not find any effects of line spacing ($F(3, 372) = 0.24,\ p = 0.870$) on fixation duration. We also did not find interaction effects of line spacing and groups ($F(3, 372) = 0.11,\ p = 0.955$).

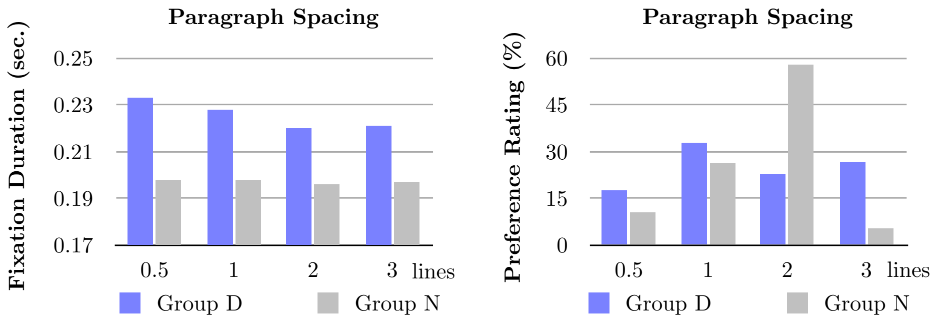


Figure 7.9: *Fixation Duration* means and *Preference Rating* (score in %) for paragraph spacing.

The results of the post-hoc tests show that:

- Group D had significantly longer fixations durations ($\bar{x} = 0.23$, $s = 0.07$ seconds) than group C ($\bar{x} = 0.21$, $s = 0.06$ seconds, $p < 0.001$) (Table 7.2).

- **Within Groups:** We did not find a significant effect of line spacing on fixation duration in group D ($F(3, 186) = 0.27$, $p = 0.849$) or group C ($F(3, 186) = 0.03$, $p = 0.993$). See Figure 7.8 for the means of the fixation durations.

Preferences

- **Within Groups:** Participants did not find any of the options of line spacing significantly easier to read, $\chi^2(3) = 2.16$, $p = 0.539$ for group D and $\chi^2(3) = 3.18$, $p = 0.365$ for group C. See Figure 7.8.

7.4.7 Paragraph Spacing

Objective Readability

- **Between Groups:** With a two-way ANOVA, we found significant effects for the groups ($F(1, 374) = 28.55$, $p < 0.001$) on fixation duration. But we did not find any effects of paragraph spacing ($F(3, 374) = 0.45$, $p = 0.715$) on fixation duration. We

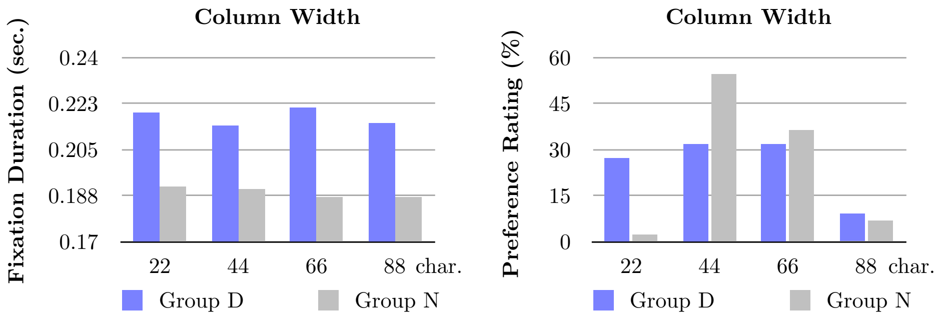


Figure 7.10: *Fixation Duration* means and *Preference Rating* (score in %) for column width.

also did not find interaction effects of paragraph spacing and groups ($F(3, 374) = 0.23,\ p = 0.873$).

The results of the post-hoc tests show that:

  - Group C had significantly shorter fixations durations ($\bar{x} = 0.20$, $s = 0.05$ seconds) than group D ($\bar{x} = 0.23$, $s = 0.06$ seconds, $p < 0.001$) (Table 7.2).

- **Within Groups:** We did not find a significant effect of paragraph spacing on fixation duration in group D ($F(3, 187) = 0.54,\ p = 0.652$) or group C ($F(3, 187) = 0.02,\ p = 0.995$). See Figure 7.9 for the means of the fixation durations.

Preferences

- **Within Groups:** Participants did not find any of the options of line paragraph significantly easier to read, $\chi^2(3) = 2.36,\ p = 0.502$ for group D and $\chi^2(3) = 2.81,\ p = 0.421$ for group C. See Figure 7.9.

7.4.8 Column Width

Objective Readability

- **Between Groups:** With a two-way ANOVA, we found significant effects for the groups ($F(1, 372) = 30.78,\ p < 0.001$) on fixation duration. But we did not find any effects of column width ($F(3, 372) = 0.16,\ p = 0.927$) on fixation duration. We also did not find interaction effects of column width and groups ($F(3, 372) = 0.21,\ p = 0.886$).

The results of the post-hoc tests show that:

- Group C had significantly shorter fixations durations ($\bar{x} = 0.19$, $s = 0.05$ seconds) than group D ($\bar{x} = 0.22$, $s = 0.05$ seconds, $p < 0.001$) (Table 7.2).

- **Within Groups:** We did not find a significant effect of column width on fixation duration in group D ($F(3, 186) = 0.17,\ p = 0.913$) or group C ($F(3, 186) = 0.20,\ p = 0.895$). See Figure 7.10 for the means of the fixation durations.

Preferences

- **Within Groups:** Participants with dyslexia did not find any of the options of column width significantly easier to read ($\chi^2(3) = 0.85,\ p = 0.839$), while participants without dyslexia found the option of 44 characters per line significantly more readable ($\chi^2(3) = 14.75,\ p = 0.002$). See Figure 7.10 for the means of the preferences.

7.5 Discussion

In general, participants without dyslexia read significantly faster and had shorter fixation durations than participants with dyslexia. For the font size variable, participants with dyslexia had significantly shorter fixation durations when using bigger fonts compared to smaller fonts.

Although no more significant effects were found, people with dyslexia were in general more sensitive to text-presentation changes, since they presented larger differences in fixation duration among the different conditions.

Regarding the differences between the groups, our results are consistent with other eye tracking studies that found significant differences among the two populations [3, 122, 126, 208, 226].

Regarding how sensitive was fixation duration with respect to the parameters, there are clearly two different groups for people with dyslexia, even when normalizing with respect to the range of values used in each parameter. The first group is the set of parameters that affect reading performance in a large percentage (10% or above), which are in order of importance: font size, character spacing, text and background color, and text grey scale. This group seems to suggest that the most important high-level characteristic that the text needs, is to let people to distinguish letters well. The second group, where impact is 5% or less, is formed by paragraph spacing, background gray scale, line spacing, and column width. This suggests that distinguishing words and lines is less important, and that most probably, word spacing also does not have a large impact in readability (unless it gets really small).

For group C the order above changes a bit, but groups are the same. In the first group, character spacing moves to the last position (fourth). In the second group paragraph spacing moves to the last position while column width and line spacing are swapped. However, as the readability impact of these three parameters mentioned is 2% or less, the order is not really relevant. Next, we discuss each parameter in detail.

Text Grey scale

Using a pure black text on a pure white background is not recommended for people with dyslexia [51]. However, we found no user studies about text grey scale and readability on people with dyslexia. Our readability results are consistent with recommendations in literature, people with dyslexia presented the shortest fixation mean with a 25% text grey scale, hence text written in dark grey instead of black might improve the readability of people with dyslexia. Most of participants (93,88%) chose black over white as the most readable option, maybe because it is one of the most frequent color combination used in computer screens.

Background Grey scale

We did not find recommendations about background grey scale for people with dyslexia apart from the suggestion of using light grey as background [388], such as the color with the following hexadecimal code: FFFFE5 [109]. Our results for readability are again consistent with the recommendations, since the shortest fixation mean appears with a 25% background grey scale for participants with dyslexia. However, most of our participants said that grey actually did not help them. Further experiments shall be done about the role of the background, because light on dark has different readability requirements than dark on light [84].

Text and Background Color

Poor color selections are one of the key problems encountered by people with dyslexia when reading [228]. Although the pair off-black/off-white is the one recommended for Web accessibility for dyslexics [51], it was the least selected one by the participants with dyslexia (only two selected it). The most preferred color pair chosen by our participants (black/yellow) has the highest contrast color combination, which is not consistent with [51], who recommends to avoid high contrast. Moreover, according to [284], such high contrast creates so much vibration that it diminishes readability. Our explanation is that this pair was chosen because it has the highest contrast. It seems more readable at first sight, although eye-tracker data showed that it was actually the hardest contrast to read. Consistently, mucky green/brown and blue/yellow pairs were chosen by people with dyslexia, as in the experiments carried out by Gregor and Newell [157, 158].

Surprisingly, the most selected pair (black/yellow) has the highest mean for the fixation durations (0.23 seconds). As comparison, the average of the color combinations is 0.22 seconds. On the other hand, the color pair that was the fastest to read was black/creme (mean of 0.21 for the fixation duration). This pair of colors is used by the British Dyslexia Association for their website.[7]

According to the W3C algorithm [418], brightness differences of less than 125 and color differences of less than 500 are not supposed

[7] http://www.bdadyslexia.org.uk/

to be good. All the pair colors selected by the participants match this guideline, except dark brown/light mucky green pair (brightness difference: 107, color difference: 240). However, the readability of colored background/text pairs is influenced by the size of the text [84] and the size used in this study was 20 points.

For group C, higher preferences scores were strongly correlated with shorter fixations. In this case blue/white and blue/yellow where the best pairs of colors, suggesting than in future experiments we should also try blue/creme.

Font Size

Another of the key problems experienced by people with dyslexia is finding the text too small [228]. Although the recommended font size for this target group is 12 or 14 points [4, 51, 52], some readers with dyslexia may prefer a larger font [52, 113].

For both groups, texts presented with 26 points size led to significant shorter fixations durations than texts with 14 points. For participants with dyslexia, texts with 22 points also led to significant shorter fixations durations than texts with 14 points. Hence, larger font sizes are more readable for people with and without dyslexia.

Unexpectedly, more than half of the participants with and without dyslexia, 24 and 26 respectively, chose our biggest option (26 points). None chose the smaller option, 14 points, which is already a recommended and relatively big font size. Since all the columns had the same width (a mean of 50 characters for 12 points), column width could not influence these decisions. Further investigations shall the done to find the font size preferred by people with and without dyslexia, as clearly there must be a turning point where a very large size starts making the reading more difficult.

Character Spacing

In [279], it is recommended to create a slightly larger distance between individual words and reduce letter-spacing slightly, so that the letters within a word lie closer together while [296] suggests to have large spacing between letter combinations. Our results are consistent with the experiment that Zorzi *et al.* [424] conducted on paper –not on screen– where larger character spacing improved the reading performance of children with dyslexia (reading speed and reading errors). Our results show that most of the participants without dyslexia significantly prefer the standard spacing among characters (32 users) and that participants with dyslexia prefer either the standard separation or more separated characters (17 users for 0% and 15 for +7%).

Line Spacing

Even though it is recommended to use a line spacing of 1.5 to 2 [296], our results show that line spacing does not have a significant effect on the participants' readability and preferences.

Paragraph Spacing

According to [51], paragraphs –even when they have a single line– should always be spaced out with an empty line between each paragraph. However, in our results we did not find an effect of paragraph spacing on readability and on the user preferences.

Column Width

Accordingly to [51, 52], which recommend to avoid long lines –60 to 70 characters– and to avoid narrow columns [52], most of the participants preferred the intermediate values: paragraphs with lines of 44 (14 users in group D and 24 in group C) or 66 characters (14 users in group D and 16 in group C). Our results are not comparable with Schneps *et al.* [348] since they used 34 point font and other devices in their experiment. Some of the participants said that they preferred the text with the widest column because they believed it was shorter than the others. Since the texts had the same number of words a side effect of having a wider column width was that the text had less lines and seemed to be shorter.

Limitations of the Study

One limitation of the study is that we only used fixation duration as a measure of reading performance. Other measures such as reading errors were not used because the reading was done in silence trying to emulate a natural online reading. Reading time was not also used because the texts lengths were not the same. Comprehension could also have been used as a measure, but we used it only as a control variable because the text representation was modified for different fragments between the two stories and it was not possible to discriminate if the comprehension was due to comprehending a single fragment or the whole story.

Also, the texts tested were small so our results are not extensible to longer passages of text such as emails or heavy-text web pages. For instance, we tested the effect of font size on Wikipedia web pages using eye tracking and a significant effect for people with dyslexia started at 18 points font size instead of 22 points as in the present study (see Chapter 8).

Another limitation is that this study does not take into consideration the interaction effects between variables. All the conditions were tested independently. While some studies found interactions [380] of the parameters others did not [36]. For instance, Tinker [380] compared font size and line width showing that long lines, very short lines, and small type size, and the combinations of these lead to significantly slower readings. Using eye tracking, Beymer *et al.* compared font size and font type and found no significant effects. Also, we could not find no interaction effects between font size and line width when reading Wikipedia texts (see Chapter 8).

Text Presentation Experiment		
Design	Within-subjects	
Independent Variables	*Color (text and background)*	black/white off-black/off-white black/yellow blue/white black/creme blue/yellow dark brown/light mucky green brown/mucky green
	Text Grey Scales	0% (black font), 25%, 50%, and 75% (white background)
	Background Grey Scales	0% (black background), 25%, 50%, and 75% (white font)
	Font Size	14, 18, 22, and 26 points
	Character Spacing	-7%, 0%, +7%, and 14% char.
	Line Spacing	0.8, 1, 1.2, and 1.4 lines
	Paragraph Spacing	0.5, 1, 2, and 3 lines
	Column Width	22, 44, 66, and 88 characters/line
Dependent Variables (Sec. 3.1.1)	*Fixation Duration* *Preference Rating*	(objective readability) (subjective preferences)
Control Variable	*Comprehension Score*	(objective comprehensibility)
Participants (Sec. 3.1.2)	Group D (46 participants)	26 female, 20 male **Age:** range from 11 to 45 ($\bar{x}$ = 20.70, s = 7.87) **Education:** high school (22), university (21), no higher education (3)
	Group C (46 participants)	27 female, 19 male **Age:** range from 13 to 37 ($\bar{x}$ = 23.50, s = 8.16) **Education:** high school (16), university (28), no higher education (2)
Materials (Sec. 3.1.3)	Base Texts Text Presentation Preferences Quest.	36 text fragments 8 items (1 item/condition)
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750	
Procedure (Sec. 3.1.5)	**Steps:** Instructions, demographic questionnaire, reading task comprehension questionnaire, preference questionnaires	

Table 7.1: Methodology for the *Text Presentation* experiment.

Text Presentation	**Group D**		**Group C**		**Group D**	**Group C**
	$\bar{x} \pm s$	%	$\bar{x} \pm s$	%	%	%
Text Grey Scale	*Fixation Duration*				*Preference Rating*	
0% (black)	0.24±0.07	7.62	**0.19**±0.05	–	**60.00**	**93.88**
25%	**0.22**±0.08	–	0.22±0.07	11.86	32.60	6.12
50%	0.23±0.09	4.48	0.21±0.06	8.76	4.35	–
75%	0.24±0.09	7.62	0.20±0.06	5.15	2.17	–
Background Grey Scale	*Fixation Duration*				*Preference Rating*	
0% (black)	0.24±0.09	3.96	0.21±0.06	3.48	**55.32**	**58.33**
25%	**0.23**±0.07	–	0.21±0.06	3.98	25.53	35.42
50%	0.24±0.07	4.85	**0.20**±0.06	–	19.15	6.25
75%	0.24±0.09	4.41	0.21±0.07	4.98	–	–
Text/Background Colors	*Fixation Duration*				*Preference Rating*	
black/creme	**0.21**±0.07	–	0.19±0.05	3.89	15.22	**24.62**
blue/yellow	0.21±0.08	2.88	0.18±0.05	1.11	10.87	10.78
green/brown	0.22±0.07	6.25	0.19±0.06	7.78	6.52	–
off-black/off-white	0.22±0.08	6.73	0.20±0.06	11.67	4.35	13.85
black/white	0.22±0.07	7.21	0.20±0.06	8.33	10.87	23.08
light green/dark brown	0.23±0.07	9.62	0.20±0.04	12.78	8.70	4.62
blue/white	0.23±0.07	9.62	**0.18**±0.05	–	10.87	6.15
black/yellow	0.23±0.09	10.58	0.19±0.05	5.00	**32.61**	16.92
Font Size	*Fixation Duration*				*Preference Rating*	
14 points	0.26±0.09	24.52	0.22±0.07	19.89	–	–
18 points	0.23±0.07	12.02	0.20±0.05	7.53	10.42	6.82
22 points	0.22±0.06	3.85	0.20±0.05	6.45	39.58	34.09
26 points	**0.21**±0.05	–	**0.19**±0.04	–	**50.00**	**59.09**
Character Spacing	*Fixation Duration*				*Preference Rating*	
-7%	0.23±0.09	15.84	0.20±0.06	10.61	10.87	–
0%	0.21±0.07	2.97	0.19±0.05	6.15	**36.96**	**65.31**
+7%	**0.20**±0.06	–	0.19±0.05	3.35	32.60	16.33
+14%	0.21 ±0.06	3.96	**0.18**±0.05	–	19.57	18.37
Line Spacing	*Fixation Duration*				*Preference Rating*	
0.8 lines	0.23±0.07	3.10	0.21±0.05	0.98	4.44	10.42
1 line	0.24±0.07	4.42	0.21±0.05	1.47	33.33	**39.58**
1.2 lines	0.24±0.07	4.87	0.21±0.06	0.49	**40.00**	**39.58**
1.4 lines	**0.23**±0.06	–	**0.20**±0.06	–	22.22	10.42
Paragraph Spacing	*Fixation Duration*				*Preference Rating*	
0.5 lines	0.23±0.06	5.90	0.20±0.05	1.02	17.56	10.53
1 line	0.23±0.06	3.64	0.20±0.05	1.02	**32.82**	26.32
2 lines	**0.22**±0.05	–	**0.20**±0.04	–	22.90	**57.89**
3 lines	0.22±0.05	0.45	0.20±0.05	0.51	26.72	5.26
Column Width	*Fixation Duration*				*Preference Rating*	
22 char./line	0.22±0.06	2.34	0.19±0.04	2.14	27.27	2.27
44 char./line	**0.21**±0.06	–	0.19±0.05	1.60	**31.81**	**54.55**
66 char./line	0.22±0.06	3.27	**0.19**±0.04	–	**31.81**	36.36
88 char./line	0.22±0.05	0.47	**0.19**±0.04	–	9.09	6.82

Table 7.2: *Fixation Duration* and *Preference Rating* results.

Chapter 8

A Dyslexic-friendly Wikipedia

8.1 Introduction

One of the most used websites in education is Wikipedia. According to *Alexa Internet* [5], in February 2013, Wikipedia was the sixth most popular website worldwide, that is, the most popular text-heavy website of the Web. Being one of the main repositories of knowledge, students from all over the world consult it to do their exercises. In fact, there is growing effort from Wikipedia to support education, such as the *Wikipedia Education Program*. But, around 10% of the people have dyslexia, which difficult their access to text-based websites. Fortunately, thanks to the fact that more and more of this reading involves online resources on the Web, we are able to alter and improve the presentation of educational resources for children with dyslexia. Presenting online text in more dyslexic-friendly ways may not only impact these children's reading performance but also their success in education. Hence, how could Wikipedia be presented to be more readable for this target group?

In the *Web Content Accessibility Guidelines* (WCAG) [63], dyslexia is treated as part of a diverse group of cognitive disabilities. They do not contain specific guidelines for text presentation for people with dyslexia. A common limitation of most previous studies about text presentation on screen for people with dyslexia is that they used isolated texts [158, 199]. Yet, most of the text we encounter in the Web is embedded into websites with navigation bars, images, and sidebars containing, such as advertisements or additional links. By ignoring these contextual factors, it is not clear how the results of these studies will generalize to real-world usage. What is missing, is studying the effect of the combination of text presentations parameter on reading and comprehension in context, that is, of a text that is embedded into a standard website.

One of the crucial factors for readability is font size [250, 264] together with line spacing [265]. For instance, the search phrase "text too small" is a frequent discussion topic in the Web.[1] Also, ensuring good readability is an easy way of making information more accessible to people with special needs, such as elderly people [104], or people with print disabilities [292], such as people with low vision [130].

In this chapter, we present the first study with 28 participants with dyslexia (and 104 without dyslexia) that experimentally compares the effect of 6 font sizes and 4 line spacing on objective and subjective readability and comprehensibility of texts in Wikipedia. Our results significantly differ from previous recommendations, presumably, because this is the first work to cover a wide range of values and to study them in the context of an actual website. Our main contributions are:

– For people with dyslexia, font size has a significant effect on objective and subjective readability and comprehensibility, while line spacing has not.

– For people with dyslexia, reading improves up to a font size of 18 points, and beyond that we do not see further improvements. For people without dyslexia font size has significant effects on readability and comprehension. Hence, both aspects improve significantly with increasing font size, until 18 points.

– For people without dyslexia, line spacing has a significant effect on comprehension, suggesting that too small or too large spacing may impair comprehension.

[1] The query "font too small" in Google gives 369,000 hits (March 14th, 2014).

  - On the basis of our findings, we recommend to use 18 points font size and line spacing ranging from 1.0 to 1.4 for good readability and comprehension of web text content for people with and without dyslexia.

The results of this chapter are presented in Rello *et al.* [333].

8.2 Related Work

We divide previous work into general guidelines, and previous studies related to dyslexic readers for font size and line spacing, in addition to the studies mentioned in Sections 7.2.3 and 7.2.5.

8.2.1 Font Size

A detailed explanation of the related work on font size in Section 7.2.3. The general finding that repeats throughout previous work is that people read and comprehend texts better with increasing font sizes. Recommendations typically suggest font sizes ranging from 10 to 16 points. However, it remains unclear from which point on increasing font size is no longer beneficial.

Dependencies of Font Size

Previous findings also indicate that font size is interdependent with font type [31, 36]. Most of the previous work applies to the two most common fonts used on screen and printed texts, *Arial* and *Times*, respectively [76]. One of the reasons is that font size can result into different letter sizes for different font types,[2] so parts of the observed effects might be due to the actual size of the letters. In consequence, research on the effect of font size needs to consider the font type, *e.g.*, by at least making clear for which font type the findings are valid, or consider letters of the same real size, even if they are different in point size. In addition, notice that if the column width is fixed, the number of character per column depends on the font size.

[2] For the interested reader, see Figure 1 in Boyarski *et al.* [50], who compared *Times* with *Georgia* and *Verdana*.

8.2.2 Line Spacing

Line spacing refers to the distance between the baselines of two text lines. The concept is also known as *leading* from the days of hand-typesetting and *line-height* in cascading style sheets (CSS). The bigger the line spacing, the further two sentences are apart vertically.

In our previous study (Chapter 7), line spacing was strongly correlated with reading performance: the narrower the space between the lines, the slower the participants read.

8.2.3 What is Missing?

In all presented previous studies, bigger fonts led to better results, either in terms of readability or in terms of preference, except from Beymer *et al.* [36] who also use comprehension scores. Previous work indicates that bigger font sizes will result into more readable websites. However, none of the studies increased font size beyond 16 points to study the limits of this improvement. About line spacing, no conclusive evidence has been reported. Also, in all previous studies, including the ones which use eye tracking, study the effect of text presentations parameters (a) independent from each other, *i.e.,* combinations of text parameters are not assessed, and (b) independent from the context, *i.e.,* the text is studied isolated without the other content that typically appears on web pages.

8.3 Methodology

To study the effect of font size and line spacing on readability and comprehensibility of websites, we conducted an experiment. 28 participants with dyslexia (104 participants without dyslexia) had to read six Wikipedia entries related to animals with varying font sizes and line spacing. We chose Wikipedia, since it is the most-visited text based website.[3] Readability and comprehensibility were analyzed via eye tracking, comprehension tests, and subjective feedback.

[3] Other, more visited websites, such as `google.com`, contain almost no text and are hence not useful for this study.

Wikipedia Experiment		
Design	Hybrid-measures: Within-subjects for *Font Size* Between-subjects for *Line Spacing*	
Independent Variables	*Font Size*	10, 12, 14, 18, 22, and 26 points
	Line Spacing	0.8, 1.0. 1.4, and 1.8 lines
Dependent Variables	*Fixation Duration*	(objective readability)
(Sec. 3.1.1)	*Comprehension Score*	(objective comprehensibility)
	Readability Rating	(subjective readability)
	Comprehensibility Rating	(subjective comprehensibility)
Participants (Sec. 3.1.2)	Group D (28 participants)	15 female, 13 male **Age:** range from 14 to 38 ($\bar{x}$ = 21.36, s = 7.49) **Education:** high school (14), university (11), no higher education (3) **Reading:** more 8 than hours (3.57%), 4-8 hours (57.14%), less than 4 hours/day (39.29%)
	Group C (104 participants)	61 female, 43 male **Age:** range from 14 to 54 ($\bar{x}$ = 30.24, s = 9.13) **Education:** high school (7), university (92), no higher education (4) **Reading:** more than 8 hours (13.04%), 4-8 hours (39,13%), less than 4 hours/day (47.83%)
Materials (Sec. 3.1.3)	Base Texts	24 Wikipedia articles
	Text Presentation	
	Compren. Quest.	24 literal and inferential items
	Sub. Readability Quest.	24 Likert scales
	Sub. Comprehension Quest.	24 Likert scales
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750	
Procedure (Sec. 3.1.5)	**Steps:** Instructions, demographic questionnaire, reading task (× 6) comprehension questionnaire (× 6), subjective readability quest. (× 6), subjective comprehension quest. (× 6)	

Table 8.1: Methodological summary for the *Wikipedia* experiment.

8.3.1 Design

In Table 8.1 we show a summary of the methodology used for the *Wikipedia* experiment.

Independent Variables

In our experimental design, *Line Spacing* and *Font Size* served as independent variables with 4 and 6 levels, respectively.

- For *Font Size*, we used the levels 10, 12, 14, 18, 22, and 26 points. We chose to study *font size* because it is the only text presentation parameter which had a significant reading texts with participants with dyslexia in Chapter 7. We chose 10 points because it is suggested as minimum font size in standard usability guidelines. The other font sizes were chosen because they are recommended in previous work: 12 points in [30], 14 points in [19, 30, 31, 38, 39], and 18, 22, 26 points in Chapter 7.

- For *Line Spacing* we tested the levels 0.8, 1.0., 1.4, and 1.8 lines. We chose *Line Spacing* because of its strong correlation with reading performance [331]. Recommendations for people with dyslexia are: 1.3 in [279], 1.4 [331], 1.5 [52], and 1.5 to 2 lines [296]. Since line spacing has never been studied in comparison with font size, we chose to study the browser's default line spacing (1.0). For Firefox, the browser we used, this equals to 120% points of the font size.[4] We chose 1.4 since many style guidelines suggest to using slightly increased line spacing and 1.0 because is the default in word processors.

We used a hybrid-measures design. Each participant read six texts with one *Line Spacing* and six different *font sizes*. Hence, for *Font Size*, we collect repeated measures, while for *Line Spacing*, we obtain between-group data. The order of conditions was counter-balanced to cancel out sequence effects.

Dependent Variables

For quantifying objective readability and comprehensibility, we used *Fixation Duration* and *Comprehension Score* as *dependent measures*, respectively. To measure the readers subjective perception towards the text we used *Readability Rating* and the *Comprehensibility Rating*

[4] That is, 0.8 equals to 96%, 1.0 equals to 120%, 1.4 equals to 168%, and 1.8 equals to 216% of the font size.

The definition and details of these *dependent measures* can be found in Section 3.1.1. The *Comprehension Score* was computed as the percentage of correct answers, where the correct choice scored 100% and the others 0%.

8.3.2 Participants

The details of the participants groups for both experiments are given in Table 8.1. For more details about the participants please refer to Section 3.1.2.

8.3.3 Materials

Wikipedia Entries

The Wikipedia entries need to be similar to isolate the effects of the text presentation. Since Wikipedia entries are heterogeneous, it is challenging to find sufficiently similar entries. We decided against modifying text, because otherwise the experiment does not show readability and comprehension in real context of the Web. Thus, we went through the articles of the Spanish Wikipedia related to animals and chose 24 articles which share the following comparable characteristics:

(a) All texts used in the experiment cover the same *genre* and the same *topic*, namely animals. We chose animals because they are a topic of general interest, not technical or academic.

(b) They all have a similar number of words in the first and the second paragraphs, ranging from 40 to 60 words for each of the paragraphs.

(c) They have a similar discourse structure: title, the first paragraph presents the animal and the place where the animal lives, the second and paragraph gives more information which differs depending on the entry, the third paragraph explains more details.

(d) All texts had low frequencies (ranging from two to five) of numerical expressions, acronyms, and foreign words, because these type of words are processed differently than regular words [110, 352]

Figure 8.1: Example of the beginning of a Wikipedia article.

and people with dyslexia specially encounter problems with such words (Section 2.1.5).

For each of the selected Wikipedia articles, we obtained the HTML source code. To alter the presentation, we used a browser plug-in (StyleBot) to modify the style sheet (CSS) to change font size and line spacing.

Text Presentation

The layout of the Wikipedia Articles was always the same: the paragraphs were located in roughly the same position of the screen. Each article contained one image on the top-right of the content page (see Figure 8.1). All the entries used the *sans-serif* font *Arial*, which Wikipedia uses as default on Firefox and other browsers on MS Windows.

Comprehension Questionnaires

Each of the questionnaires was composed of six multiple-choice questions, one for each of the Wikipedia articles. An example of each type of items is given in Figure 8.2. We used both literal and inferential questions and multiple-choice questions with four possible choices, one correct choice, two wrong choices and one "I don't know".

Segun lo que acabas de leer en la Wikipedia
'According to what I just read on Wikipedia:'

☐ El gorila tiene un ADN muy similar al de los humanos.
'The gorilla's DNA is similar to the human's.'

☐ El gorila vive en los bosques del sur de África.
'The gorilla lives in the forests of southern Africa.'

☐ El gorila es un primate carnívoro.
'The gorilla is a carnivorous primate.'

☐ No lo sé, creo que lo no ponía o al menos yo no lo recuerdo.
'I do not know, I think it was not in the text, or at least I do not remember it.'

Figure 8.2: Comprehension item example of *Wikipedia* experiment.

8.4 Results

In this section, we present the analysis of the data from the eye tracker (fixation duration), the comprehension tests, and the perception ratings.

We used a three-way ANOVA test for repeated measures to show effects of the conditions between groups D and N. To test for significant effects within groups, we used two-way ANOVA and pairwise, Holm-corrected, t-tests for parametric, normally distributed or homogeneously-distributed scores, and Friedman Test and pairwise, Holm-corrected Wilcoxon-Signed-Rank tests for non-parametric or non-normally-distributed scores. Whether the data was normally distributed was determined through Shapiro-Wilk tests. Barlett's tests were used to check the homogeneity of variances.

8.4.1 Differences between Groups

Objective Readability

- **Between Groups:** We found significant effects for the groups D and C ($F(1, 691) = 43.95$, $p < 0.001$) on *Fixation Duration*. Participants with dyslexia had significantly longer fixation times ($\bar{x} = 0.25$, $s = 0.09$ seconds) than the participants without dyslexia ($\bar{x} = 0.22$, $s = 0.05$, $p < 0.001$). We found significant effects of *Font Size* ($F(5, 691) = 20.44,\ p < 0.001$) on *Fixation Duration*. But we did not find any effects of *Line Spacing* ($F(3, 691) = 0.80,\ p = 0.497$) on *Fixation Duration*.

- **Interactions:** We also did not find interaction effects of *Font Size* and groups ($F(5, 691) = 1.11,\ p = 0.357$) on *Fixation Duration*; but there were interaction effects for *Line Spacing* and groups ($F(3, 691) = 3.37,\ p = 0.018$) on *Fixation Duration*. We did not find any interaction effects of *Line Spacing* and *Font Size* ($F(15, 691) = 1.11,\ p = 0.342$) on *Fixation Duration*. We did not find any interaction effects of *Line Spacing*, *Font Size* and groups ($F(15, 691) = 0.29,\ p = 0.996$) on *Fixation Duration*.

Objective Comprehensibility

- **Between Groups:** Dyslexia had significant smaller comprehension scores ($\bar{x} = 54.94$, $s = 49.91\%$) than the participants without dyslexia ($\bar{x} = 76.31$, $s =$37.53%, $p < 0.001$). We found significant effects of *Font Size* ($F(5, 1474) = 8.26,\ p < 0.001$) and of *Line Spacing* ($F(3, 1474) = 2.64,\ p = 0.049$) on the *Comprehension Score*.

- **Interactions:** There were interaction effects of *Font Size* and groups ($F(5, 1474) = 3.55,\ p = 0.004$) on *Comprehension Score*; and there were interaction effects for *Line Spacing* and groups ($F(3, 1474) = 5.78,\ p < 0.001$) on *Comprehension Score*. We also found interaction effects between *Line Spacing* and *Font Size* ($F(15, 1474) = 8.28,\ p < 0.001$) on *Comprehension Score*. We also found interaction effects between *Line Spacing*, *Font Size* and groups ($F(15, 1474) = 4.42,\ p < 0.001$) on *Comprehension Score*.

Subjective Readability

- **Between Groups:** We found significant effects for the groups D and C ($F(1, 1474) = 5.17,\ p = 0.023$) on the *Readability Rating*. Participants with dyslexia had significantly lower *Readability Rating* ($\bar{x} = 1.87$, $s = 1.28$) than the participants without dyslexia ($\bar{x} = 3.71$, $s = 1.29$, $p < 0.001$). We found significant effects of *Font Size* ($F(5, 1474) = 67.24$, $p < 0.001$) on the *Readability Rating*. But we did not find any significant effects of *Line Spacing* ($F(3, 1474) = 1.41$, $p = 0.238$) on the *Readability Rating*.

- **Interactions:** There were no interaction effects of *Font Size* and groups ($F(5, 1474) = 2.15$, $p = 0.059$) on the *Readability Rating*. But there were interaction effects for *Line Spacing* and groups ($F(3, 1474) = 2.66$, $p = 0.048$) on *Readability Rating*. We could not find interaction effects between *Line Spacing* and *Font Size* ($F(15, 1474) = 1.14$, $p = 0.317$) on *Readability Rating*. We did not find interaction effects between *Line Spacing*, *Font Size* and groups ($F(15, 1474) = 0.52$, $p = 0.930$) on the *Readability Rating*.

Subjective Comprehensibility

- **Between Groups:** We found significant effects for the groups D and C ($F(1, 1474) = 3.80$, $p = 0.052$) on the *Comprehensibility Rating*. Participants with dyslexia had significantly lower *Comprehensibility Rating* ($\bar{x} = 3.42$, $s = 1.01$) than the participants without dyslexia ($\bar{x} = 3.58$, $s = 1.04$, $p < 0.001$). We found significant effects of *Font Size* ($F(5, 1474) = 35.77$, $p < 0.001$) and *Line Spacing* ($F(3, 1474) = 6.02$, $p = 0.001$) on the *Comprehensibility Rating*.

- **Interactions:** There were not interaction effects of *Font Size* and groups ($F(5, 1474) = 1.76$, $p = 0.121$) on *Comprehensibility Rating*. Similarly, we could not find interaction effects for *Line Spacing* and groups ($F(3, 1474) = 0.89$, $p = 0.447$) on *Comprehensibility Rating*. We found interaction effects between *Line Spacing* and *Font Size* ($F(15, 1474) = 1.82$, $p = 0.030$) on *Comprehensibility Rating*. We also did not find interaction effects between *Line Spacing*, *Font Size* and groups ($F(15, 1474) = 0.49$, $p = 0.945$) on *Comprehensibility Rating*.

Wikipedia	**Group D**			**Group C**		
	$\tilde{x}$	$\bar{x} \pm s$	%	$\tilde{x}$	$\bar{x} \pm s$	%
Font Size	*Fixation Duration*					
10 points	0.30	0.30 ± 0.13	136.36	0.25	0.25 ± 0.06	125.00
12 points	0.29	0.28 ± 0.10	127.27	0.23	0.24 ± 0.05	120.00
14 points	0.27	0.27 ± 0.10	122.73	0.22	0.22 ± 0.05	110.00
18 points	0.23	0.23 ± 0.08	104.55	0.21	0.21 ± 0.04	105.00
22 points	0.21	$\mathbf{0.22 \pm 0.08}$	100.00	0.20	$\mathbf{0.20 \pm 0.04}$	100.00
26 points	0.21	0.23 ± 0.08	104.55	0.20	$\mathbf{0.20 \pm 0.04}$	100.00
Line Spacing	*Fixation Duration*					
0.8 lines	0.25	0.26 ± 0.12	113.04	0.21	0.22 ± 0.05	100.00
1 line	0.24	0.25 ± 0.05	108.70	0.21	0.22 ± 0.06	100.00
1.4 lines	0.26	0.28 ± 0.10	121.74	0.22	$\mathbf{0.22 \pm 0.05}$	100.00
1.8 lines	0.21	$\mathbf{0.23 \pm 0.10}$	100.00	0.21	$\mathbf{0.22 \pm 0.05}$	100.00

Table 8.2: Median, mean and standard deviation of *Fixation Duration* in seconds for groups C and D.

8.4.2 Font Size

Objective Readability

- **Group D:** Figure 8.3 (left) shows the average fixation duration for each of the *Font Size* conditions (Table 8.2). There was a significant main effect of *Font Size* on *Fixation Duration* ($F(1, 156) = 15.51$, $p < 0.001$). We found:

  - For 10 points font size, participants had significantly longer fixation durations than 14 points ($p = 0.018$), as well as 18, 22, and 26 points ($p < 0.001$, each).
  - For 12 points font size, participants had significantly longer fixation durations than 18, 22, and 26 points ($p < 0.001$, each).
  - For 14 points font size, participants had significantly longer fixation durations than 18, 22, and 26 points ($p < 0.001$, each).
  - For the font sizes 18, 22, and 26 points, we could not find

significant differences between the fixation durations.

This data indicates that until 18 points font size, the fixation duration decreases with increasing font size. Beyond 18 points, increasing the font size led to no significant improvement in our data. Hence, our results provide evidence that bigger font sizes lead to better readability. However, from 18 points font size on, no significant improvement of the readability was found.

- **Group C:** Figure 8.4 (left) shows the average fixation duration for each of the *Font Size* conditions (Table 8.2). There was a significant effect of *Font Size* on *Fixation Duration* ($F(1, 574) = 81.48,\ p < 0.001$). Indeed,

  - For 10 points font size, participants had significantly longer fixation durations than for 14, 18, 22, and 26 points ($p < 0.001$, each).
  - For 12 points font size, participants had significantly longer fixation durations than for 18, 22, and 26 points ($p < 0.001$, each).
  - For 14 points font size, participants had significantly longer fixation durations than for 18, 22, and 26 points ($p < 0.001$, $p < 0.001$, $p < 0.010$, respectively).
  - For the font sizes 18, 22, and 26 points, we could not find significant differences between the fixation durations.

Objective Comprehensibility

- **Group D:** Figure 8.5 (left) shows the comprehension scores for each of the *Font Size* conditions (Table 8.3). There was a significant effect of *Font Size* on the comprehension score ($F(1, 165) = 9.37,\ p = 0.003$). We found that:

  - For 10 points font size, participants had significantly lower comprehension scores than for 18, 22, and 26 points ($p < 0.001$, $p < 0.001$, $p = 0.005$, respectively).

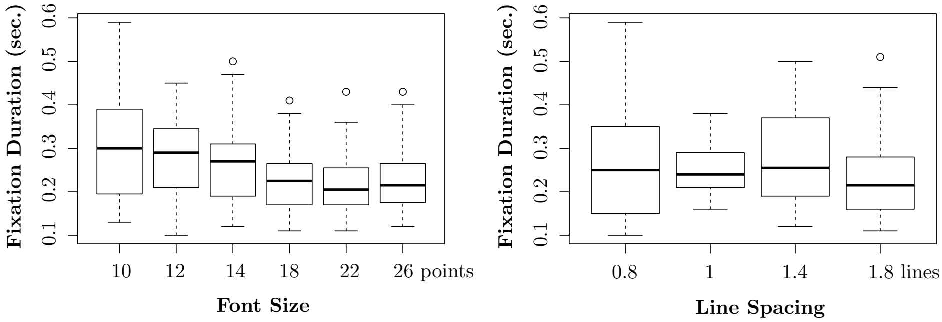


Figure 8.3: *Fixation Duration* means by font size (left) and line spacing (right) for group D.

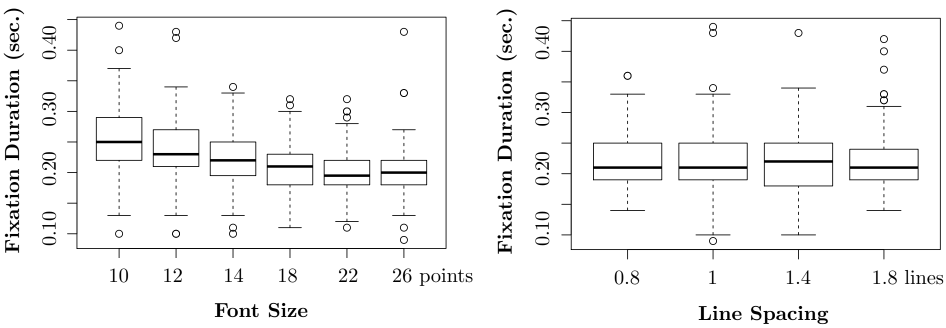


Figure 8.4: *Fixation Duration* means by font size (left) and by line spacing (right) for group C.

  – For 12 points font size, participants had significantly lower comprehension scores than 18 and 22 points ($p < 0.001$, each).
  – For the font sizes 14, 18, 22, and 26 points, we could not find significant differences between the comprehension scores.
  – Similarly, we could not find significant differences between the comprehension scores for the font sizes 10 and 12 points.

These results indicate that the comprehension is significantly better for the larger font sizes (18, 22, 26 points) than for the smaller font sizes (10, 12 points) that we tested.

– **Group C:** Figure 8.6 (right) shows the comprehension score

Wikipedia	Group D			Group C		
	$\tilde{x}$	$\bar{x} \pm s$	%	$\tilde{x}$	$\bar{x} \pm s$	%
Font Size	*Comprehension Score*					
10 points	0	33.33 ± 48.04	45.00	100	65.91 ± 47.67	73.23
12 points	0	40.74 ± 50.07	55.00	100	81.82 ± 38.79	90.91
14 points	100	51.85 ± 50.92	70.00	100	75.56 ± 43.22	83.96
18 points	100	74.07 ± 44.66	100.00	100	90.00 ± 30.17	100.00
22 points	100	70.37 ± 46.53	95.00	100	85.87 ± 35.02	95.41
26 points	100	59.26 ± 50.07	80.01	100	84.44 ± 36.45	93.82
Line Spacing	*Comprehension Score*					
0.8 lines	100	66.67 ± 47.71	100.00	100	79.83 ± 40.30	92.51
1 line	100	55.56 ± 50.40	83.34	100	78.88 ± 40.94	91.41
1.4 lines	0	47.61 ± 50.55	71.41	100	78.36 ± 41.33	90.81
1.8 lines	50	50.00 ± 50.61	75.00	100	86.29 ± 34.53	100.00

Table 8.3: Median, mean and standard deviation of the *Comprehension Score* in seconds for groups C and D.

distribution for each of the *Font Size* conditions (Table 8.3). There was a significant effect of *Font Size* on the comprehension score ($\chi^2(5) = 27.29,\ p < 0.001$).

In particular:

– For 10 points, participants had significantly lower comprehension scores than for 18 and 22 points ($p < 0.001$ and $p < 0.001$).

Subjective Readability

– **Group D:** There was a significant effect of *Font Size* on subjective readability rating ($F(1, 135) = 72.19,\ p < 0.001$). Figure 8.7 (left) shows the subjective readability ratings by *Font Size* (Table 8.4). Pairwise post-hoc comparisons showed significant differences between almost all conditions. Namely,

  – For the font sizes 10, 12, 14, and 18 points, readability ratings increase significantly with increasing font size. This

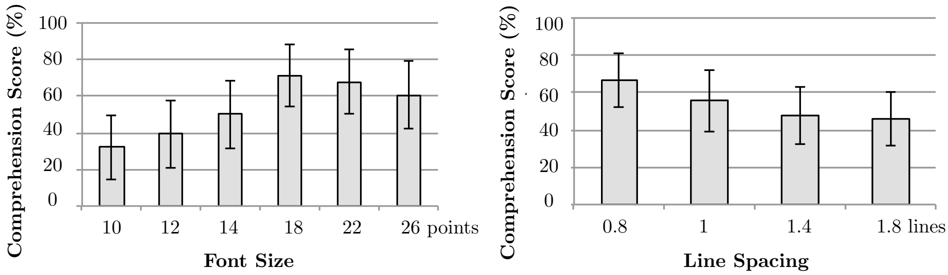


Figure 8.5: *Comprehension Score* means by font size (left) and by line spacing (right) for group D.

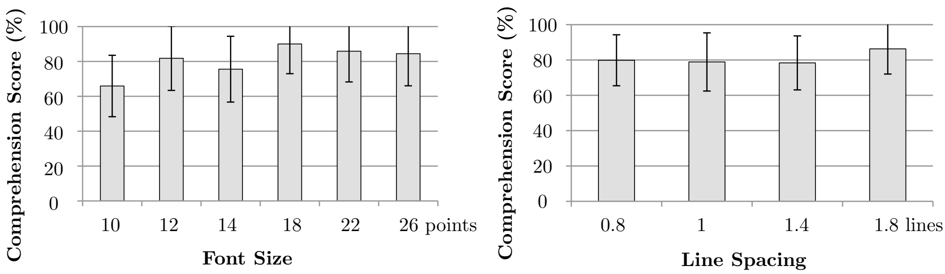


Figure 8.6: *Comprehension Score* means by font size (left) and by line spacing (right) for group C.

means that the readability ratings for the conditions are: $10 < 12 < 14 < 18$ points ($p < 0.001$, each).

– For the font sizes 18 and 22 points, we found no significant difference in the ratings ($p = 0.324$).

– For 26 points font size, the readability ratings are significantly lower than for 22 points ($p < 0.001$).

These results indicate that subjective readability increases with increasing font size, but that it achieves maximum for 18 points, and decreases beyond that point.

– **Group C:** Figure 8.8 (left) shows the distribution of the subjective readability ratings by *Font Size* (Table 8.4). There was a significant effect of *Font Size* on subjective readability ($\chi^2(5) = 135.85,\ p < 0.001$). In fact,

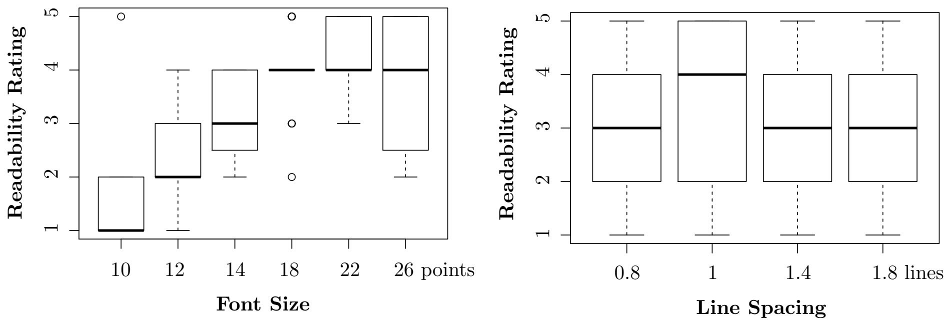


Figure 8.7: *Readability Rating* means for font size (left) and line spacing (right) for group D.

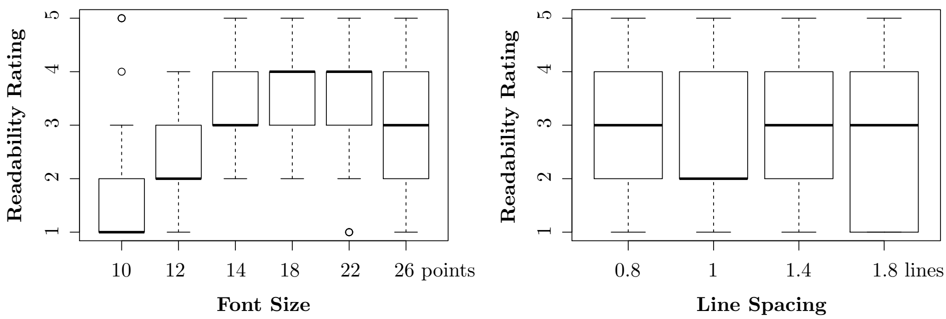


Figure 8.8: *Readability Rating* means for font size (left) and line spacing (right) for group C.

- For 10 points, readability ratings were significantly lower than for all other sizes ($p < 0.001$, each).
- For 12 points, readability ratings were significantly lower than for all larger sizes ($p < 0.001$, each).
- For 14 points, readability ratings were significantly lower than for 18 points ($p < 0.001$).
- For 26 points, readability ratings were significantly lower than for 18 points ($p < 0.001$).

Subjective Comprehensibility

- **Group D:** Analog to the effect on subjective readability, there was a significant effect of *Font Size* on subjective comprehension

Wikipedia	Group D			Group C		
	$\tilde{x}$	$\bar{x} \pm s$	%	$\tilde{x}$	$\bar{x} \pm s$	%
Font Size	*Readability Rating*					
10 points	1	1.61 ± 0.89	37.79	1	1.54 ± 0.94	40.42
12 points	2	2.30 ± 0.97	53.99	2	2.14 ± 0.83	56.17
14 points	3	3.04 ± 0.77	71.36	3	3.29 ± 0.91	86.35
18 points	4	4.00 ± 0.74	93.90	4	3.81 ± 0.82	100.00
22 points	4	4.26 ± 0.69	100.00	4	3.48 ± 1.16	91.34
26 points	4	3.57 ± 1.20	83.80	3	2.96 ± 1.38	77.69
Line Spacing	*Readability Rating*					
0.8 lines	3	3.03 ± 1.30	90.99	3	2.93 ± 1.32	97.34
1 line	4	3.33 ± 1.46	100.00	2	2.70 ± 1.23	89.70
1.4 lines	3	3.02 ± 1.12	90.69	3	3.01 ± 1.16	100.00
1.8 lines	3	3.22 ± 1.35	96.70	3	2.80 ± 1.39	93.02

Table 8.4: Median, mean and standard deviation of the *Comprehensibility Rating* for groups C and D.

rating ($F(1, 135) = 48.05,\ p < 0.001$). Figure 8.9 (right) shows the subjective comprehensibility ratings by *Font Size* (Table 8.5). Pairwise post-hoc comparisons showed a similar pattern, as we found on the readability ratings. There were significant differences between almost all conditions:

– For the font sizes 10, 12, and 14 points, comprehensibility ratings increase significantly with increasing font size. This means that the readability ratings for the conditions are: $10 < 12 < 14$ points ($p < 0.001$, each).

– No significant differences were found between 14 and 26 points ($p = 0.254$), 18 and 22 points ($p < 0.703$), 18 and 26 points ($p = 0.088$), and 22 and 26 points ($p = 0.184$).

– Nevertheless, the ratings for font size 14 points are significantly lower than for 18 points ($p < 0.001$) and 22 points ($p < 0.001$).

These results indicate that comprehension is highest for font sizes 18, 22, and 26 points.

Wikipedia	**Group D**			**Group C**		
	$\tilde{x}$	$\bar{x} \pm s$	%	$\tilde{x}$	$\bar{x} \pm s$	%
Font Size	*Comprehensibility Rating*					
10 points	2	2.26 ± 0.75	54.72	3	2.93 ± 1.05	70.94
12 points	3	2.87 ± 0.97	69.49	3	3.09 ± 0.99	74.82
14 points	3	3.48 ± 0.73	84.26	4	3.82 ± 0.77	92.49
18 points	4	4.13 ± 0.63	100.00	4	4.13 ± 0.77	100.00
22 points	4	4.04 ± 0.82	97.82	4	3.93 ± 0.95	95.16
26 points	4	3.74 ± 0.69	90.56	4	3.87 ± 0.90	93.70
Line Spacing	*Comprehensibility Rating*					
0.8 lines	3	3.28 ± 0.94	88.41	4	3.60 ± 0.99	92.31
1 line	4	3.71 ± 0.91	100.00	4	3.90 ± 0.87	100.00
1.4 lines	4	3.50 ± 0.99	94.34	4	3.54 ± 1.05	90.77
1.8 lines	3	3.28 ± 1.14	88.41	3.5	3.49 ± 1.07	89.49

Table 8.5: Median, mean and standard deviation of the *Comprehensibility Rating* for groups C and D.

- **Group C:** Figure 8.10 (left) shows the distribution of subjective comprehensibility ratings by *Font Size* (Table 8.5). There was a significant effect of *Font Size* on comprehension ratings ($F(1, 135) = 69.64,\ p < 0.001$). In addition:
  - For 10 points, comprehension ratings were significantly lower than for 14, 18, 22, and 26 points ($p < 0.001$, each).
  - For 12 points, comprehension ratings were significantly lower than for 14, 18, 22, and 26 points ($p < 0.001$, each), as well.
  - No significant differences were found between 10 and 12 points. Similarly, no significant differences were found between 14, 18, 22, and 26 points.

8.4.3 Line Spacing

Objective Readability

- **Group D:** Figure 8.3 (right) shows the average fixation duration for each of the *Line Spacing* conditions. We did not

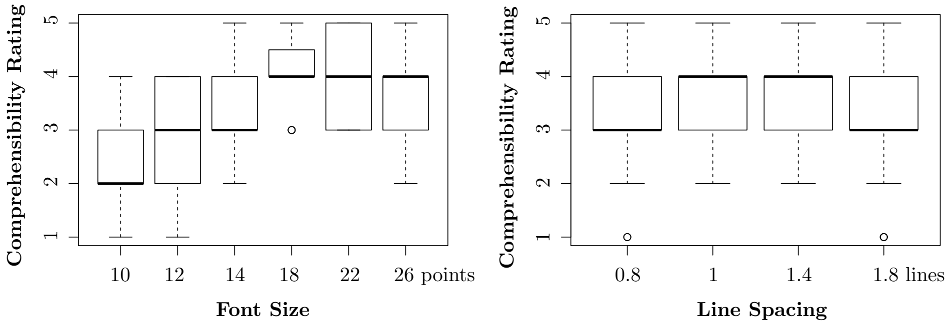


Figure 8.9: *Comprehensibility Rating* means for font size (left) and line spacing (right) for group D.

find a significant effect of *Line Spacing* on *Fixation Duration* ($F(1, 156) = 0.90,\ p = 0.345$). Hence, in contrast to font size, line spacing did not have an effect on readability.

- **Group C:** Figure 8.4 (right) shows the average fixation duration for each of the *Line Spacing* conditions. We did not find a significant effect of *Line Spacing* on *Fixation Duration* ($F(1, 574) = 0.06,\ p = 0.804$).

Objective Comprehensibility

- **Group D:** Figure 8.5 (right) shows the comprehension scores for each of the *Line Spacing* conditions. There was a significant effect of *Line Spacing* on the comprehension score ($F(1, 165) = 4.21,\ p = 0.042$). The comprehension score was significantly higher for 0.8 lines spacing than 1.8 lines spacing ($p < 0.001$).

- **Group C:** Figure 8.6 (right) shows the comprehension score distribution for each of the *Line Spacing* conditions. For *Line Spacing*, we did not find a significant effect on the comprehension scores ($\chi^2(3) = 3.35,\ p = 0.341$).

Subjective Readability

- **Group D:** For *Line Spacing*, we found not significant effects on the subjective readability rating ($F(1, 135) = 0.11,\ p = 0.737$) of the texts. See Figure 8.7 (right).

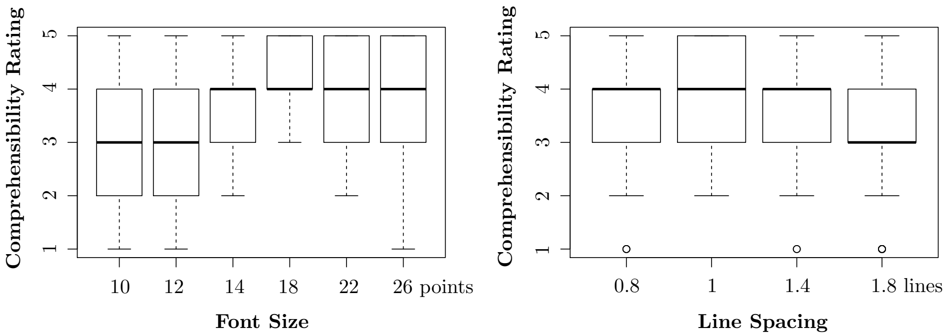


Figure 8.10: *Comprehensibility Rating* means for font size (left) and line spacing (right) for group C.

- **Group C:** Figure 8.8 (right) shows the distribution of the readability ratings for *Line Spacing*. We did not find a significant effect on the subjective readability rating ($\chi^2(3) = 6.19,\ p = 0.102$).

Subjective Comprehensibility

- **Group D:** For *Line Spacing*, we found no significant effects on the subjective comprehensibility ($F(1, 135) = 0.19,\ p = 0.661$) of the texts. See Figure 8.9 (right).

- **Group C:** Figure 8.10 (right) shows the distribution of comprehension ratings by *Line Spacing*. We found significant effects on the subjective comprehensibility ratings ($\chi^2(3) = 8.99,\ p < 0.001$):

  - For 1.0 line spacing comprehension ratings were significantly higher than for 1.8 ($p < 0.001$).

8.4.4 Interactions

Objective Readability

Within Groups: The interaction plot in Figures 8.11 and 8.12 show the interaction between *Font Size* and *Line Spacing* for the fixation duration. In group D (Figure 8.11 left), for increasing font size the fixation duration decreases similarly for any line spacing. Only

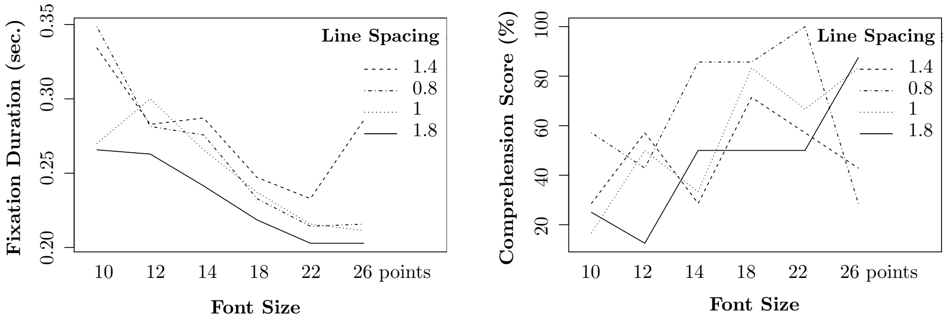


Figure 8.11: Interaction between font size and line spacing for *Fixation Duration* and *Comprehension Score* for group D.

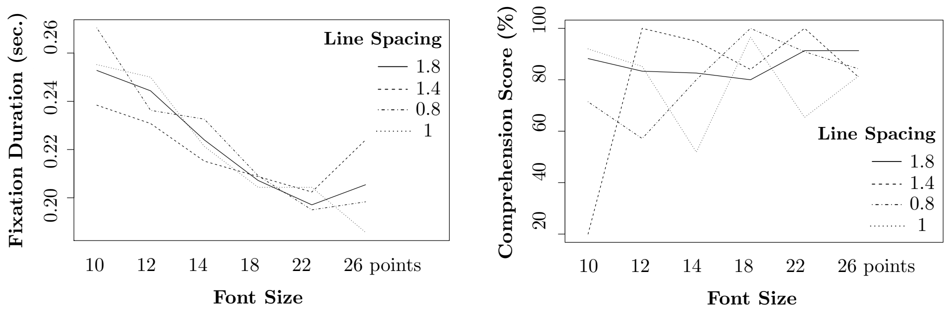


Figure 8.12: Interaction between font size and line spacing for *Fixation Duration* and *Comprehension Score* for group C.

for a line spacing of 1.4, we can see an increase at 26 points font size, which may indicate the these two values in combination decrease readability. However, no significant interaction effect on fixation duration was found ($p = 0.068$) for group D. For group C, (8.12, left) figure shows that the interaction for fixation duration appear to be stable from 10 points to 22 points, and only for 26 points slight variances can be observed. However, despite the large number of participants (N = 104), the effect is not statistically significant ($p = 0.074$).

Objective Comprehensibility

Within Groups: The interaction plot in Figures 8.11 and 8.12 show the interaction between *Font Size* and *Line Spacing* for the comprehension scores for both groups. Similarly, no significant interaction effect on the comprehension scores was found for group D ($p = 0.153$)

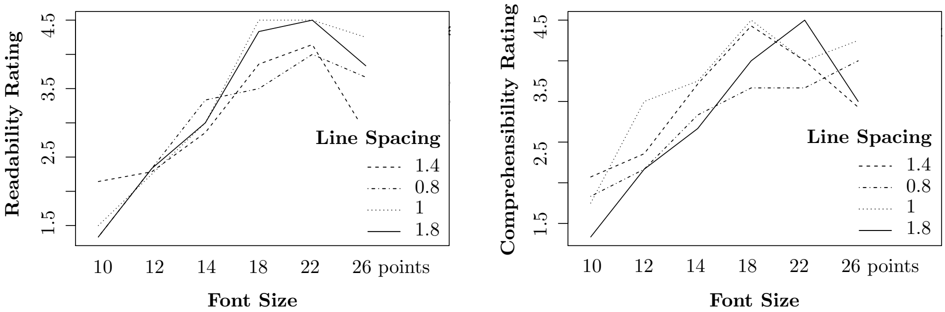


Figure 8.13: Interaction between font size and line spacing for *Readability Rating* and *Comprehensibility Rating* for group D.

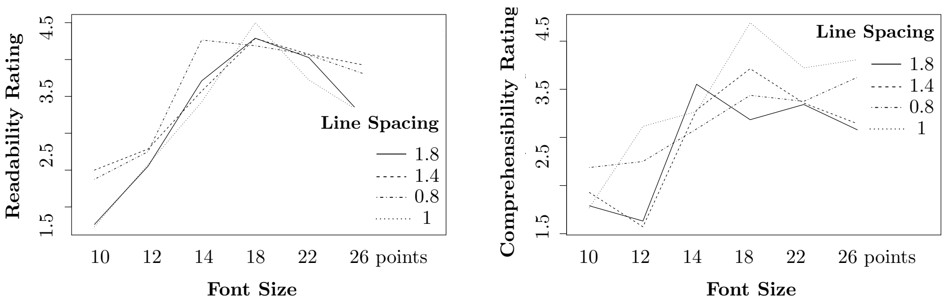


Figure 8.14: Interaction between font size and line spacing for *Readability Rating* and *Comprehensibility Rating* for group C.

and group C ($p = 0.810$), see Figures 8.11 and 8.12 (right), respectively.

Subjective Readability

Within Groups: The interaction plot in Figures 8.13 and 8.14 show the interaction between *Font Size* and *Line Spacing* for the readability ratings. For group D (Figure 8.13, left), there were no significant effects for subjective readability ratings ($p = 0.986$), nor for group C ($p = 0.780$), see Figure 8.14 (left).

Subjective Comprehensibility

Within Groups: The interaction plot in Figures 8.13 and 8.14 show the interaction between *Font Size* and *Line Spacing* for the comprehensibility ratings. For group D (Figure 8.13, right), there were no significant effects for subjective comprehensibility ratings ($p = 0.751$), nor for group C ($p = 0.575$), see Figure 8.14 (right).

8.5 Discussion

We found significant effects of font size on fixation duration, comprehension scores, and subjective ratings. For all people objective readability increased significantly until 18 points font size. Beyond this font size, no significant reduction of the fixation durations could be found. For participants with dyslexia, the comprehension scores were significantly higher for larger font sizes (18, 22, 26 points) than for smaller font sizes (10, 12 points). The participants without dyslexia gave more correct answers to comprehension questions for 18 and 22 points than for 10 points.

For line spacing, the only significant effect we found was in the comprehension score for participants with dyslexia: for the largest line spacing (1.8) the comprehension scores were significantly lower than for the lowest (0.8). Other than that, line spacing did not have any significant effect in our setup.

Similarly, for both groups, subjective readability increased with increasing font size, being highest at 18 and 22 points, and stabilizes with larger sizes. For participants with dyslexia, subjective comprehensibility increased too, with increasing font size, being highest for the larger font sizes (18, 22, 26 points). For participants without dyslexia, subjective comprehensibility ratings were higher from 14 to 26 points than for 10 and 12 points, and larger for 1.0 line spacing than for 1.8 line spacing.

Font Size

Font size had significant effects on all dependent measures. Subjective and objective readability increased steadily until 18 points for both groups.

Our results regarding font size are not consistent with previous studies and recommendations. Previous studies using eye tracking with regular readers [38] recommend 14 points (comparing the sizes of 10, 12, and 14 points), while 26 points are recommended with readers with dyslexia (comparing fonts of 14, 18, 22 and 26 points) (Chapter 7). Also, web design recommendations for readers with dyslexia recommend 12 or 14 points [4, 51, 52] or bigger [52, 113].

However, since no further significant improvements were observed beyond 18 points, but a drop in subjective readability for 26 points has been found, the results indicate that a local maximum might exist between 18 to 26 points.[5] A local maximum is to be expected, as increasing font size will required to have less and less text in a single line, which leads to more frequent eye jumps, scrolling, and the loss of overview [121]. On the basis of our results, we recommend to use font size of 18 points for text in the Web. Indeed, 18 points strikes the balance between having the best readability, comprehension, and subjective perception scores.

Subjective readability was higher for the larger font sizes (14 to 26 points) than for the very small sizes (10 and 12 points). This matches the findings from the comprehension questionnaires, where our subjects gave more wrong answers for 10 and 12 points than for 18 and 26 points. These findings indicate that small font sizes have a negative effect on how easy a text can be understood. This is a notable insight, as 10 and 12 points happen to be font sizes that are very commonly used and recommended in texts or websites.

Line Spacing

Existing recommendations regarding line spacing for readers with dyslexia are heterogeneous. Previous work has suggested 1.3 [279], 1.4 [331], 1.5 [52], and 1.5 to 2 lines [296]. Since in our setup, line spacing hardly had significant effects, we can neither confirm nor disprove these recommendations. The only significant effect of line spacing in our experiment was found on the comprehension scores, which were higher for 0.8 than for 1.8 spacing. This can be seen as indicator, that too much line spacing leads to decreased reading performance. However, from the data we cannot make assumptions about the intermediate line spacing 1.0 and 1.4. For people without dyslexia, subjective comprehension was higher for the standard spacing compared to the largest spacing (1.8). Hence, subjectively, wide line spacing is detrimental to comprehension. Therefore, while the data does not provide clear evidence about which line spacing to use, it indicates that deviating too much from the standard spacing may make texts

[5] For fixation duration, the minimal value was attained for 22 points.

more difficult to understand.

Our findings confirm general guidelines, which suggest to using the default (1.0) or slightly larger line spacing. Since our results are not highly conclusive, we support the assumption by Bix [41] that line spacing is not a major factor on readability and that the ideal line spacing depends on other factors. It could be subordinated to aesthetic considerations or user preferences.

Limitations of the Study

One of the limitations of our study is that we only considered the first three paragraphs of Wikipedia articles. When using eye tracking to study reading, it has been found that the initially measured fixation durations are longer, since users are still in a familiarization phase [251, 255]. The heat map in Figure 8.15 shows that this effect occurred in our setup, too. However, the heat map also shows that the fixation durations normalize when reading on. Then, since we assume that people often only read parts of web pages, we conclude that despite the short lengths of the texts, our findings have practical validity, that is, this familiarization also happens when people read web pages [60].

In comparison to previous work [19, 31, 32, 36, 104], we did not measure reading time. We did so for two reasons. First, we wanted to create a natural setting, in which reading as fast as possible is neither a goal nor an indicator for readability. Second, reading fast can, in our opinion, be misleading. For example, in case of bad readability, participants might become frustrated and start skimming the text instead of reading it with full attention. Our decision is backed up by findings from Beymer *et al.* [36], who found significantly longer fixations for smaller fonts but no significant effect of font size on reading speed.

Another limitation of the study is that we used a fixed line width as the browser window was maximized throughout the study. It could be possible that increasing the line width when increasing the font size would have eliminated some of the positive effects. However, previous research [355] actually predicts the opposite effect: in a reading study with 20 students, the highest line width led to fastest readings speeds. Therefore, if we had increased line width with font size, we might have even found more pronounced effects. Yet, the typical browser will not change its size when changing the font size. Hence, our design has

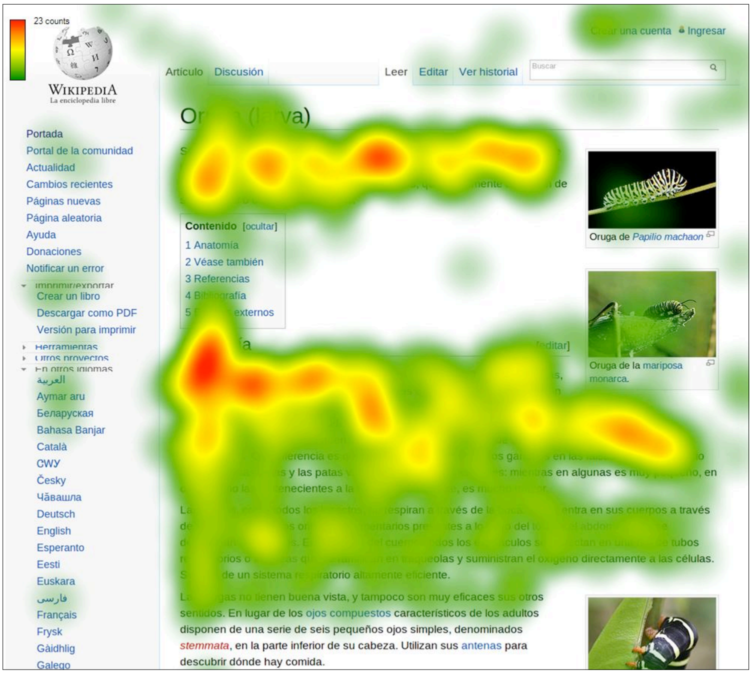


Figure 8.15: Heat map of a Wikipedia article used in the study (18 points and 1.0 line spacing).

practical and allows applying our findings to typical reading patterns.

Previous work [31, 50] has shown that readability of texts also depend on the font type. Therefore, our findings might not be generalizable to every font (Chapter 9). However, we use one of the most widely used fonts to present texts in the Web, as *Arial* is the default sans-serif font in most modern web browsers. More important, we believe that our work shows a clear indication that bigger font sizes lead to better reading and comprehension, encouraging designers to, regardless of the font type, think about and argue for bigger font sizes.

Chapter 9

Good Fonts for Dyslexia

9.1 Introduction

There is evidence that the presentation of the text has a significant effect on a text's accessibility for people with dyslexia [158]. At the same time, any typewritten text has to be written using one or several font types. Although the selection of font types is crucial in the text design process, empirical analyses of reading performance of people with dyslexia has focused more on font size [256] rather than on font type. To the best of our knowledge, there are no experiments that objectively measure the impact of the font type on reading performance.

In this chapter, we present the first study that measures the impact of the font type on the reading performance of people with dyslexia using eye tracking, as well as asking them their personal preferences. We used a within-subject design, 97 subjects (48 subjects dyslexia) read 12 texts with 12 different fonts. On the basis of our results, we present recommendations for font styles and a set of more accessible fonts for people with dyslexia. The contributions of this chapter are:

- Font types have a significant impact on readability of people with and without dyslexia.
- What is good for people with dyslexia regarding font types is also good for people without dyslexia.
- Good fonts for people with dyslexia are *Helvetica*, *Courier*, *Arial*, *Verdana* and *Computer Modern Unicode*, taking into consider-

ation reading performance and subjective preferences. On the contrary, *Arial It.* should be avoided since it decreases readability.

– *Sans serif*, *roman* and *monospaced* font types increased the reading performance of our participants, while *italic* fonts did the opposite.

This chapter was presented in Rello and Baeza-Yates [319].

9.2 Related Work

The relationship between fonts and dyslexia has drawn the attention of many fields, such as psychology, arts, and accessibility. We divide related work in: fonts recommended for people with dyslexia, fonts designed for this target group, and related user studies.

9.2.1 Recommendations

Most of the recommendations come from associations for people with dyslexia and they agree in using *sans-serif* fonts. The British Dyslexia Association recommends to use *Arial*, *Comic Sans* or, as alternatives to these, *Verdana*, *Tahoma*, *Century Gothic*, and *Trebuchet* [52]. However, the website does not disclose on the basis of which evidence these recommendations are made. Evett and Brown's [130] recommendations for readers with low vision as well as readers with dyslexia are put in comparison, giving as a result the recommendation of using also *Arial* and *Comic Sans.* Lockley [217] recommended to avoid italics and fancy fonts, which are particularly difficult for a reader with dyslexia, and also point to *Arial* as preferred font. Another font recommended in 2010 was *Sassoon Primary* but not anymore [120].

The only recommendation for *serif* fonts has been done by the International Dyslexia Centre [172] and that was for *Times New Roman.* According to AbilityNet [1], *Courier* is easier to read by people with dyslexia because it is monospaced.

In the *Web Content Accessibility Guidelines* (WCAG) [63], dyslexia is treated as part of a diverse group of cognitive disabilities and they do not propose any specific guidelines about font types for people with dyslexia.

Surprisingly, none of the typefaces recommended by the dyslexia organizations mentioned above were ever designed specifically for readers with dyslexia.

9.2.2 Fonts Designed for People with Dyslexia

We found four fonts designed for people with dyslexia: *Sylexiad* [168], *Dyslexie* [108], *Read Regular*,[1] and *OpenDyslexic.*[2] The four fonts have

[1] `http://www.readregular.com/`

[2] `http://opendyslexic.org/`

in common that the letters are more differentiated compared to regular fonts. For example, the shape of the letter 'b' is not a mirror image of 'd'. From these fonts, we choose to study *Open Dyslexic* (both roman and italic styles), because it is the only open sourced and hence free. This font has been already integrated in various tools.

9.2.3 User Studies

There are several studies on text presentation and people with dyslexia regarding font and background colors [331], font [256, 333] or letter spacing [424].

The closest work to ours is a study with people with dyslexia [108] that compared *Arial* and *Dyslexie*. They conducted a word-reading test with 21 students with dyslexia (Dutch One Minute Test). *Dyslexie* did not lead to faster reading, but could help with some dyslexic-related errors in Dutch. In [373], text design for people with dyslexia is explored with a qualitative study with eleven students using class observations, interviews, and questionnaires. In some tasks, the participants needed to choose the font they prefer, but no analyses of the chosen fonts is presented.

9.2.4 What is Missing?

What is missing is a sound investigation into the effect of the most frequent fonts on reading performance. Our experiment advances previous work by providing this evidence via quantitative data from eye tracking measurements. In addition, by testing 12 different fonts with 48 participants, we compare a greater number of font types with a larger number of participants than previous studies. We selected the fonts on the basis of their popularity and frequency of use in the Web (Section 9.3.1).

9.3 Methodology

To study the effect of font type on readability and comprehensibility of texts on the screen, we conducted an experiment where 97 participants (48 with dyslexia) had to read 12 comparable texts with varying font types. Readability and comprehensibility were analyzed via eye tracking and comprehension tests, respectively, using the latter as a

This is Arial	This is Myriad
This is Arial It.	This is OpenDyslexic
This is Computer Modern	*This is OpenDyslexic It.*
`This is Courier`	This is Times
This is Garamond	*This is Times It.*
This is Helvetica	This is Verdana

Figure 9.1: Font types used in the *Font* experiment.

control variable. The participants' preference was gathered via questionnaires.

9.3.1 Design

In Table 9.1 we show a summary of the methodology of *Font* experiment.

Independent Variables

In our experimental design, *Font Type* served as an *independent variable* with 12 levels: *Arial*, *Arial Italic*, *Computer Modern Unicode (CMU)*, *Courier*, *Garamond*, *Helvetica*, *Myriad*, *OpenDyslexic*, *OpenDyslexic Italic*, *Times*, *Times Italic*, and *Verdana* (See Figure 9.1). We use for brevity *OpenDys* for the corresponding fonts.

We chose to study *Arial* and *Times* because they are the most common fonts used on screen and printed texts, respectively [76]. *OpenDyslexic* was selected because is a free font type designed specifically for people with dyslexia and *Verdana* because is the recommended font for this target group [52]. We choose *Courier* because is the most common example of monospaced font [76]. *Helvetica* and *Myriad* were chosen for being broadly used in graphic design and for being the typeface of choice of Microsoft and Apple, respectively. We chose *Garamond* because is claimed to have strong legibility for printed materials [76] and we selected *CMU* because is widely used in scientific publishing, as is the default of the typesetting program TeX, as well as a free font supporting many languages [194].

We also made sure that the fonts cover variations of essential font characteristics:

– [±Italic] served as independent variable with two values: *italic* denotes the condition where the text was presented using an italic type, that is a cursive typeface, and *roman* denotes the condition where the text was presented in a roman type. We study the italic types of *Arial*, *OpenDyslexic*, and *Times*.

– [±Serif] served as independent variable with two values: *serif* denotes the condition where the text was presented with typefaces with serifs, small lines trailing from the edges of letters and symbols, and *sans serif* denotes the condition where the text used typefaces without serifs. In our set of fonts there are three *serif* fonts –*CMU*, *Garamond*, and *Times*– and four *sans serif* fonts –*Arial*, *Helvetica*, *Myriad*, and *Verdana*–.

– [±Monospace] served as independent variable with two values: *monospaced* denotes the condition where the text was presented using a monospaced type, that is, a font whose letters and characters each occupy the same amount of horizontal space, and *proportional*, where the text was presented using proportional fonts. We chose the most commonly used monospaced font, the *roman serif* font *Courier*, and we compare it with the rest of the *roman* and *serif* fonts that are *proportional*: *CMU*, *Garamond* and *Times*.

– [±Dyslexic] served as independent variable with two values: [+Dyslexic] denotes the condition where the text was presented using a font type which was specifically designed from people with dyslexia *OpenDyslexic*; and [−Dyslexic] where the text was presented using other fonts. To compare with *OpenDyslexic* we choose *sans serif*, *roman* fonts *Arial*, *Helvetica*, *Myriad*, and *Verdana*. To cover the italic variant *OpenDyslexic It.*, [±Dyslexic It.] served as independent variable with two values: [+Dyslexic It.] which correspond with *OpenDyslexic It.* and [−Dyslexic It.] which correspond with the only *sans serif*, *italic* font we had in our set, *Arial it.*.

We used a within-subject design, that is, each participant read 12 different texts with 12 different fonts, hence, contributing to each condition. We counter-balanced texts and fonts to avoid experimental sequence effects. Therefore, the data with respect to text-font combinations was evenly distributed.

Dependent Variables

For quantifying readability, we used two *dependent measures*: *Reading Time* and *Fixation Duration*, both extracted from the eye tracking data. To collect the participant preferences, we used the subjective *Preference Rating* through questionnaires. The definition and details of these *dependent measures* can be found in Section 3.1.1.

For the *Comprehension Score* we used multiple-choice questions. If the reader did not chose the correct answer, the corresponding text was discarded from the analysis. We used this comprehension question as a *control variable* to guarantee that the recordings analyzed in this study were valid.

9.3.2 Participants

The demographic data of the participants groups are given in Table 9.1. For more details about the participants please refer to Section 3.1.2.

9.3.3 Materials

Texts

All the texts used in the experiment meet the comparability requirements because they all share the parameters commonly used to compute readability [115]. All the texts were extracted from the same book, *Impostores* (Impostors), by Lucas Sánchez [346]. We chose this book because its structure (32 chapters) gave us the possibility of extracting similar texts. Each chapter of the book is an independent story and it starts always by an introductory paragraph. Thus, we went through the book and selected the introduction paragraphs sharing the following characteristics:

(a) Same genre and same style.

(b) Same number of words (60 words). If the paragraph did not have that number of words we slightly modified it to match the number of words.

(c) Similar word length, with an average length ranging from 4.92 to 5.87 letters.

(d) Absence of numerical expressions, acronyms, and foreign words, because people with dyslexia specially encounter problems with such words [100].

Text Presentation

We used the parameters detailed in Section 3.1.3 with black text on white background and 14 points font size.

Comprehension Questionnaires

Each of the literal items was composed of a three multiple-choice questions. The order of the correct answer was counterbalanced. An example of one of these questions was previously given in Figure 3.1. The difficulty of the questions chosen was similar.

Preference Questionnaires

For each of the twelve text-font pairs, the participants rated on a five-point Likert scale, how much did they like the font type used in the text presentation. An example of the items is given in Section 3.1.3, Figure 3.3.

9.4 Results

For group D, Shapiro-Wilk tests showed that nine and eight out of the twelve data sets were not normally distributed for the *Reading Time* and *Fixation Duration*, respectively. For group C, Shapiro-Wilk tests showed that three and eleven out of the twelve data sets were not normally distributed for the *Reading Time* and *Fixation Duration*, respectively. Shapiro-Wilk tests showed that no data sets were normally distributed for the *Preference Rating* of both groups. Also, Levene tests showed that none of the data sets had an homogeneous variance for all the measures and both groups.

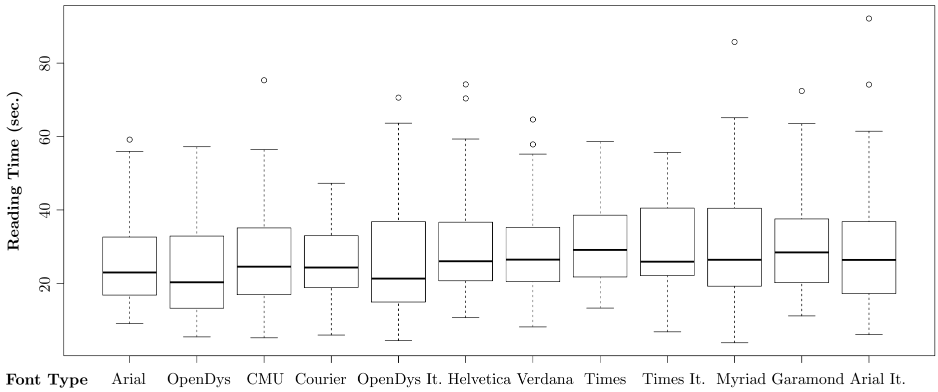


Figure 9.2: *Reading Time* box plots by *Font Type* for group D (ordered by average *Reading Time* for group D).

Hence, to study the effects of *Font Type* on readability, comprehensibility and preferences we used the two-way Friedman's non-parametric test for repeated measures plus a complete pairwise Wilcoxon rank sum post-hoc comparison test with a Bonferroni correction that includes the adjustment of the significance level. Then, to show effects of the conditions within groups, we divided the data for each group and used Friedman's non-parametric test for repeated measures plus a complete pairwise Wilcoxon rank sum post-hoc comparison test with a Bonferroni adjustment. To study the effect of the second level independent variables, *Italics*, *Serif*, and *Monospace*, we use Friedman's non-parametric test for repeated measures and the Wilcoxon rank sum post-hoc comparison test with a Bonferroni adjustment. For these reasons we later include the median and box plots for all our measures in addition to the average and the standard deviation.

9.4.1 Font Type

Reading Time. Table 9.2 shows the main statistical measures for the *Reading Time* for each of the *Font Type* conditions.

There was a significant effect of *Font Type* on *Reading Time* ($\chi^2(11) = 13.99,\ p < 0.001$).

For group D, *Reading Time* and *Fixation Duration* had a Spearman's correlation coefficient of $\rho = 0.987$ ($p < 0.001$). For group C the

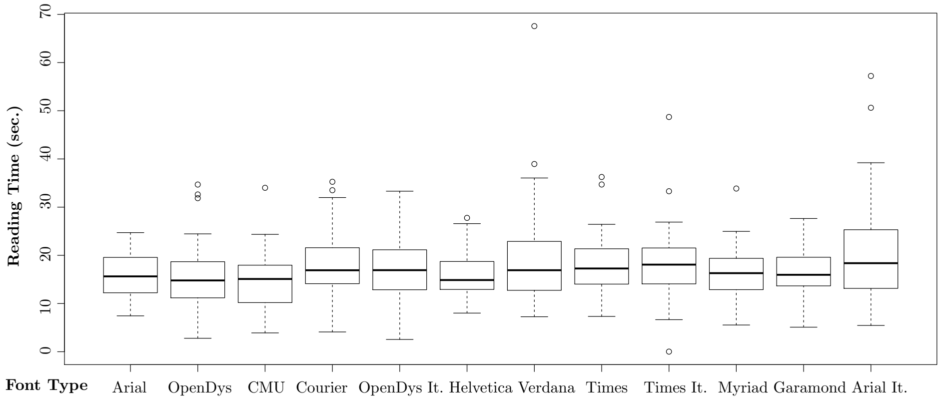


Figure 9.3: *Reading Time* box plots by *Font Type* for group C (ordered by average *Reading Time* for group D).

correlation of *Reading Time* and *Fixation Duration* had a Spearman's correlation coefficient of $\rho = 0.551$ ($p = 0.063$).

The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly longer reading times ($\bar{x} = 24.94$, $s = 12.15$ seconds) than the participants without dyslexia ($\bar{x} = 17.69$, $s = 5.73$ seconds, $p < 0.001$). For *Reading Time* between groups, the Spearman's correlation coefficient is $\rho = 0.657$, and it is statistically significant ($p = 0.024$).

- **Group D:** There was a significant effect of *Font Type* on *Reading Time* ($\chi^2(11) = 31.55,\ p < 0.001$) (Table 9.2, Figure 9.2).

  The results of the post-hoc tests show that:

  – *Arial It.* had the longest reading time mean. Participants had significantly longer reading times using *Arial It.* than *Arial* ($p = 0.011$), *CMU* ($p = 0.011$), and *Helvetica* ($p = 0.034$).

- **Group C:** There was a significant effect of *Font Type* on *Reading Time* ($\chi^2(11) = 85.07,\ p < 0.001$) (Table 9.2, Figure 9.3).

  The results of the post-hoc tests show that:

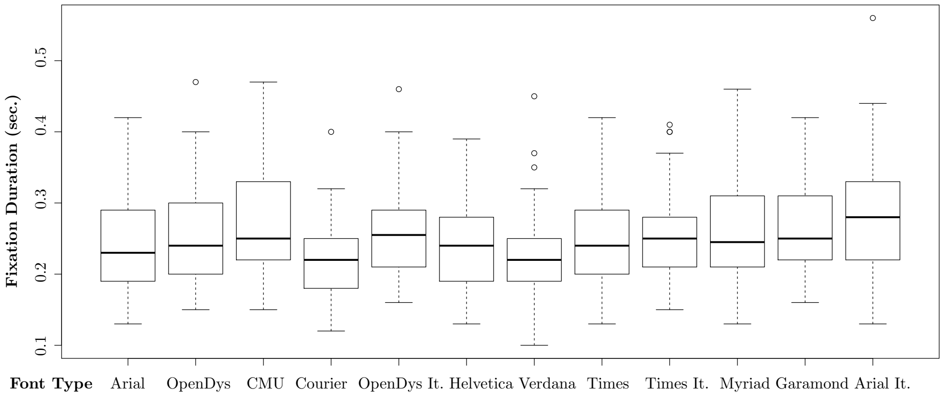


Figure 9.4: *Fixation Duration* box plots by *Font Type* for group D (ordered by average *Reading Time* for group D).

- *Arial It.* had the longest reading time means. Participants had significantly longer reading times using *Arial It.* than *Arial* ($p < 0.001$), *CMU* ($p = 0.001$), *Courier* ($p = 0.043$), *Garamond* ($p < 0.001$), *Helvetica* ($p = 0.013$), and *Times It.* ($p = 0.033$).
- *Arial* had the shortest reading time mean. Participants had significantly shorter reading times using *Arial* than *Courier* ($p = 0.022$), *OpenDys It.* ($p = 0.031$), *Times* ($p < 0.001$), *Times It.* ($p = 0.003$), *Arial It.* ($p < 0.001$), and *Verdana* ($p < 0.001$).
- *Verdana* has the second longest reading time mean. Participants had significantly longer reading times with *Verdana* than with *Arial* ($p < 0.001$), *CMU* ($p < 0.001$), *Courier* ($p = 0.088$), *Garamond* ($p = 0.003$), *Helvetica* ($p = 0.021$), and *Myriad* ($p = 0.054$).
- Participants had significantly shorter reading times reading with *CMU* than with *Times* ($p = 0.023$), and *Verdana* ($p < 0.001$).

Fixation Duration. Table 9.3 shows the main statistical measures for the *Fixation Duration* for each of the *Font Type* conditions. There was a significant effect of *Font Type* on *Fixation Duration* ($\chi^2(11) =$

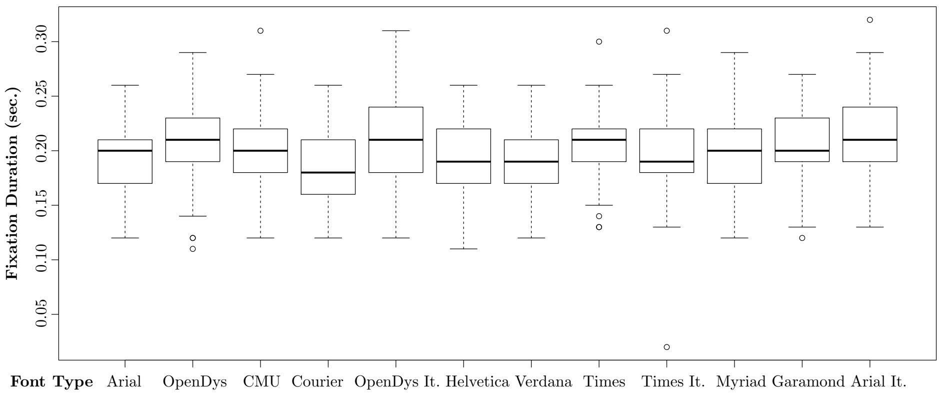


Figure 9.5: *Fixation Duration* box plots by *Font Type* for group C (ordered by average *Reading Time* for group D).

180.16, $p < 0.001$).

The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly longer fixation duration means ($\bar{x} = 0.25$, $s = 0.07$ seconds) than the participants without dyslexia ($\bar{x} = 0.20$, $s = 0.04$ seconds, $p < 0.001$). For *Fixation Duration* between groups, the Spearman's correlation coefficient is $\rho = 0.717$, and it is statistically significant ($p = 0.009$).

- **Group D:** There was a significant effect of *Font Type* on *Fixation Duration* ($\chi^2(11) = 93.63$, $p < 0.001$) (Table 9.3, Figure 9.4). The results of the post-hoc tests show that:

  – *Courier* had the lowest fixation duration mean. Participants had significantly shorter fixation durations reading with *Courier* than with *Arial It.* ($p < 0.001$), *CMU* ($p < 0.001$), *Garamond* ($p < 0.001$), *Times It.* ($p < 0.001$), *OpenDys It.* ($p = 0.001$), and *Arial* ($p = 0.046$).

  – *Helvetica* had the third lowest fixation duration mean. Participants had significantly shorter fixation durations reading with *Helvetica* than with *Arial It.* ($p < 0.001$), *CMU* ($p = 0.001$), and *Garamond* ($p = 0.006$).

  - Participants had significantly shorter fixation durations reading with *Arial* than with *CMU* ($p = 0.020$).
  - *Arial It.* had the highest fixation duration mean. Participants had significantly longer fixation durations reading with *Arial It.* than with *Courier* ($p < 0.001$), *Helvetica* ($p < 0.001$), *Arial* ($p < 0.001$), *Times It.* ($p < 0.001$), *Times* ($p = 0.003$), *Myriad* ($p = 0.004$), *Garamond* ($p = 0.011$), and *Verdana* ($p = 0.049$).

- **Group C:** There was a significant effect of *Font Type* on *Fixation Duration* ($\chi^2(11) = 95.99,\ p < 0.001$) (Table 9.3, Figure 9.5). The results of the post-hoc tests show that:

  - *Courier* had the lowest fixation duration mean. Participants had significantly shorter fixation durations reading with *Courier* than with *Arial It.* ($p < 0.001$), *CMU* ($p = 0.001$), *Garamond* ($p = 0.001$), *Myriad* ($p = 0.004$), *OpenDys It.* ($p = 0.001$), and *Times* ($p = 0.001$).
  - *Verdana* had the second lowest fixation duration mean. Participants had significantly shorter fixation durations reading with *Verdana* than with *Arial It.* ($p < 0.001$), *CMU* ($p = 0.030$), *Garamond* ($p = 0.029$), *Arial* ($p = 0.003$), and *Arial It.* ($p = 0.031$).
  - *Arial It.* had the highest fixation duration mean. Participants had significantly longer fixation durations reading with *Arial It.* than with *Arial* ($p < 0.001$), *Courier* ($p < 0.001$), *Times It.* ($p = 0.001$), *Times* ($p = 0.046$), and *Verdana* ($p < 0.001$).

Preference Rating. Table 9.4 shows the main statistical measures for the *Preference Rating* for each of the *Font Type* conditions.

There was a significant effect of *Font Type* on the *Preference Rating* ($\chi^2(11) = 120.92,\ p < 0.001$).

For group D, *Reading Time* and *Preference Rating* had a Spearman's correlation coefficient of $\rho = -0.998$, ($p < 0.001$). For group C the correlation of *Reading Time* and *Preference Rating* had a Spearman's correlation coefficient of $\rho = -0.767$, ($p = 0.004$). This is

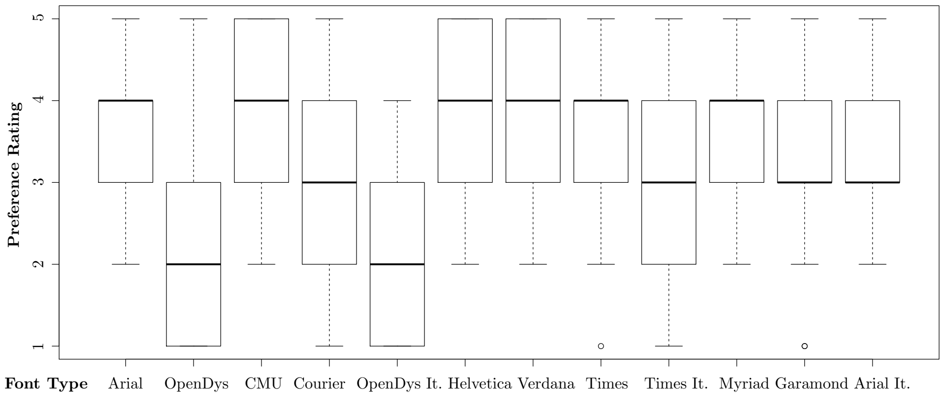


Figure 9.6: *Preference Rating* box plots by *Font Type* for group D (ordered by average *Reading Time* for group D).

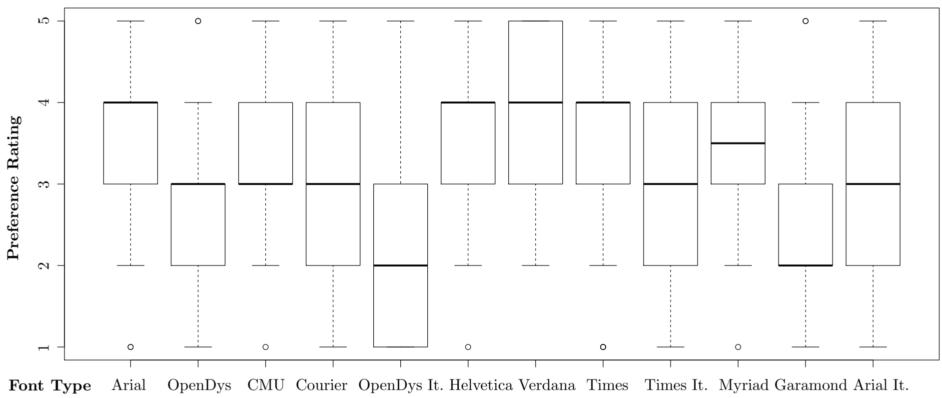


Figure 9.7: *Preference Rating* box plots by *Font Type* for group C (ordered by average *Reading Time* for group D).

expected because larger ratings are better while smaller times are better.

The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly lower preferences ratings ($\bar{x} = 3.14$, $s = 1.21$) than the participants without dyslexia ($\bar{x} = 3.32$, $s = 1.17$, $p = 0.024$). For *Fixation Duration* between groups, the Spearman's correlation coefficient is $\rho = 0.912$, and this is statistically significant

($p < 0.001$).

- **Group D:** There was a significant effect of *Font Type* on subjective preference ratings ($\chi^2(11) = 79.61,\ p < 0.001$) (Table 9.4, Figure 9.6). Pairwise post-hoc comparisons showed significant differences between the following conditions:

  – *Verdana* is significantly preferred over *Arial It* ($p < 0.001$), *OpenDys* ($p = 0.002$), *OpenDys It.* ($p = 0.004$), *Garamond* ($p = 0.008$), and *Times It.* ($p = 0.041$).

  – *Helvetica* is significantly preferred over *OpenDys It.* ($p = 0.010$), *OpenDys* ($p = 0.020$), and *Arial It.* ($p = 0.031$).

  – *Arial* was significantly more preferred than *Arial It.* ($p = 0.028$), and *OpenDys It.* ($p = 0.050$).

  – *Garamond* was significantly less preferred than *Verdana* ($p = 0.008$), *Times* ($p = 0.023$), *Arial* ($p = 0.023$), and *CMU* ($p = 0.030$).

  - Hence, participants significantly preferred *Verdana* and *Helvetica* to other fonts and significantly disliked *Garamond* in comparison with others.

- **Group C:** There was a significant effect of *Font Type* on subjective preference ratings ($\chi^2(11) = 50.65,\ p < 0.001$) (Table 9.4, Figure 9.7). Pairwise post-hoc comparisons showed significant differences between the following conditions:

  – *Verdana* is significantly preferred over *Courier* ($p = 0.005$), *OpenDys* ($p = 0.012$), and *OpenDys It.* ($p = 0.001$).

  – *Helvetica* is significantly preferred over *OpenDys* ($p = 0.011$), and *OpenDys It.* ($p = 0.004$).

  – *Arial* is significantly preferred over *Courier* ($p = 0.050$), and *OpenDys It.* ($p < 0.001$).

  – *OpenDys It.* was significantly less preferred than *Arial* ($p < 0.001$), *Arial It.* ($p = 0.005$), *CMU* ($p = 0.002$), *Garamond* ($p = 0.021$), *Helvetica* ($p = 0.004$), *Myriad* ($p = 0.006$), *Times* ($p = 0.005$), and *Verdana* ($p = 0.001$).

– Hence, participants without dyslexia significantly preferred *Verdana*, *Helvetica* and *Arial* to other fonts and significantly disliked *OpenDys* and *OpenDys It.* in comparison with others.

9.4.2 Italics

Reading Time. There was a significant effect of [±Italic] on *Reading Time* ($\chi^2(1) = 27.27,\ p < 0.001$). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly longer reading times than the participants without dyslexia ($p < 0.001$). In Table 9.5 we show the medians, means and standard deviations of each group.

- **Group D:** We did not find a significant effect of [±Italic] on *Reading Time* ($p = 0.120$) (Table 9.5).

- **Group C:** There was a significant effect of [±Italic] on *Reading Time* for participants without dyslexia ($p = 0.001$). The reading time mean of fonts in [+Italic] (*Arial It.*, *OpenDys. It.*, and *Times It.*), ($\bar{x} = 20.11,\ s = 7.55$), was significantly larger than the reading time mean of the fonts in [−Italic] or *roman* (*Arial*, *OpenDys* and *Times*), ($\bar{x} = 17.40,\ s = 5.62$) (Table 9.5).

Fixation Duration. There was a significant effect of [±Italic] on *Fixation Duration* ($\chi^2(1) = 8.07,\ p = 0.005$). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly longer reading times than the participants without dyslexia ($p < 0.001$). In Table 9.5 we show the medians, means and standard deviations of each group.

- **Group D:** There was a significant effect of [±Italic] on *Fixation Duration* for participants with dyslexia ($p = 0.040$). The fixation duration mean of the fonts [+Italic] (*Arial It.*, *OpenDys. It.*, and *Times It.*), ($\bar{x} = 0.27,\ s = 0.07$), was significantly larger than the fixation duration mean of the fonts [−Italic] (*Arial*, *OpenDys*, and *Times*), ($\bar{x} = 0.25,\ s = 0.07$) (Table 9.5).

- **Group C:** We did not find a significant effect of [±Italic] on *Fixation Duration* ($p = 0.280$) (Table 9.5).

Preference Rating. There was a significant effect of [±Italic] on *Preference Rating* ($\chi^2(1) = 40.11,\ p < 0.001$). The results of the post-hoc tests show that:

- **Between Groups:** We found no effect between group on *Preference Rating* ($p = 0.460$) (Table 9.5).

- **Group D:** There was a significant effect of [±Italic] on *Preference Rating* for participants with dyslexia ($p = 0.002$). The *Preference Rating* for the fonts [−Italic] (*Arial*, *OpenDys*, and *Times*), ($\bar{x} = 3.21,\ s = 1.22$), was significantly higher than for the fonts [+Italic] (*Arial It.*, *OpenDys. It.*, and *Times It.*), ($\bar{x} = 2.73,\ s = 1.22$) (Table 9.5).

- **Group C:** There was a significant effect of [±Italic] on *Preference Rating* for participants without dyslexia ($p = 0.018$). The *Preference Rating* for the fonts [−Italic] (*Arial*, *OpenDys*, and *Times*), ($\bar{x} = 3.24,\ s = 1.17$), was significantly higher than for the fonts [+Italic] (*Arial It.*, *OpenDys. It.*, and *Times It.*), ($\bar{x} = 2.86,\ s = 1.16$) (Table 9.5).

9.4.3 Serif

Reading Time. We did not find a significant effect of [±Serif] on *Reading Time* ($\chi^2(1) = 1.65,\ p = 0.199$). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly longer reading times than participants without dyslexia ($p < 0.001$). In Table 9.6 we show the medians, means and standard deviations of each group.

- **Group D:** We did not find a significant effect of [±Serif] on *Reading Time* for people with dyslexia ($p = 0.480$). The visit duration means were of [+Serif] fonts were not significantly longer that the ones from *sans serif* or [−Serif] font types (Table 9.6).

- **Group C:** Similarly, we did not find a significant effect of [±Serif] on *Reading Time* for people without dyslexia ($p = 0.41$) (Table 9.6).

Fixation Duration. There was a significant effect of [±Serif] on *Fixation Duration* ($\chi^2(1) = 9.31,\ p = 0.002$). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly longer fixation means than participants without dyslexia ($p < 0.001$). In Table 9.6 we show the medians, means and standard deviations of each group.

- **Group D:** There was a significant effect of [±Serif] on *Fixation Duration* for people with dyslexia ($p = 0.015$). Indeed, the fixation duration mean of the [+Serif] fonts, ($\bar{x} = 0.26,\ s = 0.07$), was significantly larger than the fixation duration mean of the *sans serif* fonts [−Serif], ($\bar{x} = 0.24,\ s = 0.07$) (Table 9.6).

- **Group C:** Similarly, there was a significant effect of [±Serif] on *Fixation Duration* for people without dyslexia ($p = 0.007$). The fixation duration mean of the [+Serif] fonts, ($\bar{x} = 0.20,\ s = 0.03$), was significantly larger than the fixation duration mean of the [−Serif] fonts, ($\bar{x} = 0.19,\ s = 0.03$) (Table 9.6).

Preference Rating. There was a significant effect of [±Serif] on *Preference Rating* ($\chi^2(1) = 13.88,\ p < 0.001$). The results of the post-hoc tests show that:

- **Between Groups:** Participants with dyslexia had significantly lower preference ratings than the participants without dyslexia ($p < 0.001$). In Table 9.6 we show the medians, means and standard deviations of each group.

- **Group D:** There was a significant effect of [±Serif] on *Preference Rating* for people with dyslexia ($p < 0.001$). They significantly preferred [−Serif] fonts, ($\tilde{x} = 3.60,\ s = 1.04$), than [+Serif] fonts, ($\tilde{x} = 3.11,\ s = 1.15$) (Table 9.6).

- **Group C:** We did not find a significant effect of [±Serif] on *Preference Rating* for people without dyslexia ($p = 0.091$) (Table 9.6).

9.4.4 Monospace

Reading Time. We did not find a significant effect of [±Monospace] on *Reading Time* ($\chi^2(1) = 3.40,\ p = 0.065$) (Table 9.7).

- **Between Groups:** There was a significant effect of [±Monospace] on *Reading Time* between groups ($p < 0.001$). Participants with dyslexia had significantly longer reading times than the participants without dyslexia. In Table 9.7 we show the medians, means and standard deviations of each group.

- **Group D:** We did not find a significant effect of [±Monospace] on *Reading Time* for people with dyslexia ($p = 0.63$). The visit duration means were of [−Monospace] or *proportional* fonts were not significantly longer that the ones from [+Monospace] font types. (Table 9.7).

- **Group C:** Similarly, we did not find a significant effect of [±Monospace] on *Reading Time* for people without dyslexia ($p = 0.97$) (Table 9.7).

Fixation Duration. There was a significant effect of [±Monospace] on *Preference Rating* ($\chi^2(1) = 25.28,\ p < 0.001$). The results of the post-hoc tests show that:

- **Between Groups:** There was a significant effect of [±Monospace] on *Fixation Duration* between groups ($p < 0.001$). Participants with dyslexia had significantly higher fixation durations than the participants without dyslexia. In Table 9.7 we show the medians, means and standard deviations of each group.

- **Group D:** There was a significant difference of [±Monospace] on *Fixation Duration* ($p < 0.001$). We found that the fixation duration mean of the [+Monospace] font, ($\bar{x} = 0.22,\ s = 0.05$), was significantly shorter than the fixation duration mean of the *proportional* or [−Monospace] fonts, ($\bar{x} = 0.26,\ s = 0.07$) (Table 9.7).

- **Group C:** There was a significant difference of [±Monospace] on *Fixation Duration* ($p = 0.002$). Similarly, the fixation duration mean of the [+Monospace] font, ($\bar{x} = 0.19,\ s = 0.03$), was significantly shorter than the fixation duration mean of the [−Monospace] fonts, ($\bar{x} = 0.20,\ s = 0.03$) (Table 9.7).

Preference Rating. There was a significant effect of [±Monospace] on *Preference Rating* ($\chi^2(1) = 6.45,\ p = 0.011$). The results of the post-hoc tests show that:

- **Between Groups:** We found no effect between group on *Preference Rating* ($p = 0.055$). In Table 9.7 we show the medians, means and standard deviations of each group.

- **Group D:** We did not find a significant effect of [±Monospace] on the participants with dyslexia preferences ($p = 0.79$) (Table 9.7).

- **Group C:** There was a significant effect of [±Monospace] on *Preference Rating* for participants without dyslexia ($p = 0.003$). They significantly preferred [−Monospace] fonts (*CMU*, *Garamond* and *Times*), ($\bar{x} = 3.59,\ s = 0.98$), than the [+Monospace] font (*Courier*), ($\bar{x} = 2.85,\ s = 1.30$) (Table 9.7).

9.4.5 Dyslexic Fonts

Reading Time. There was a significant effect of [±Dyslexic] and [±Dyslexic It.] on *Reading Time*, ($\chi^2(1) = 27.67,\ p < 0.001$) and ($\chi^2(1) = 25.27,\ p = 0.065$), respectively (Table 9.8).

- **Between Groups:** There was a significant effect of [±Dyslexic] and [±Dyslexic It.] on *Reading Time* between groups ($p < 0.001$

for both cases). Participants with dyslexia had significantly longer reading times than the participants without dyslexia. In Table 9.8 we show the medians, means and standard deviations of each group.

- **Group D:** We did not find a significant effect of [±Dyslexic] and [±Dyslexic It.] on *Reading Time* for people with dyslexia ($p = 0.25$) and ($p = 0.094$), respectively. The visit duration means of [+Dyslexic] and [+Dyslexic It.] fonts were not significantly longer than the ones from [−Dyslexic] and [−Dyslexic It.], respectively (Table 9.8).

- **Group C:** We did not find a significant effect of [±Dyslexic] on *Reading Time* for people without dyslexia ($p = 0.27$). To the contrary, we found a significant effect of [±Dyslexic It.] on *Reading Time* for people without dyslexia ($p = 0.035$). The visit duration means of [+Dyslexic It.] font *OpenDys It.*, ($\bar{x} = 18.73,\ s = 5.60$), were significantly longer than the ones from [−Dyslexic It.], ($\bar{x} = 22.38,\ s = 9.02$) (Table 9.8).

Fixation Duration. There was a significant effect of [±Dyslexic] on *Fixation Duration* ($\chi^2(1) = 7.45,\ p = 0.006$). In contrast, we did not find a significant effect of [±Dyslexic It.] on *Fixation Duration* ($\chi^2(1) = 2.91,\ p = 0.088$) (Table 9.8).

- **Between Groups:** There was a significant effect of [±Dyslexic] and [±Dyslexic It.] on *Fixation Duration* between groups ($p < 0.001$ for both cases). Participants with dyslexia had significantly longer fixation duration means than the participants without dyslexia. In Table 9.8 we show the medians, means and standard deviations of each group.

- **Group D:** We did not find a significant effect of [±Dyslexic] and [±Dyslexic It.] on *Fixation Duration* for people with dyslexia $p = 0.25$ and $p = 0.16$, respectively (Table 9.8).

- **Group C:** We found a significant effect of [±Dyslexic] on *Fixation Duration* for participants without dyslexia ($p = 0.027$). The fonts which were not designed for people with dyslexia

[−Dyslexic] (*Arial*, *Helvetica*, *Myriad*, and *Verdana*), ($\bar{x} = 0.19,\ s = 0.03$), lead to significantly shorter fixations durations than *OpenDyslexic* ([+Dyslexic]), ($\bar{x} = 0.2,\ s = 0.04$). To the contrary, we did not find a significant effect of [±Dyslexic It.] on *Fixation Duration* for people without dyslexia ($p = 0.37$) (Table 9.8).

Preference Rating. There was a significant effect of [±Dyslexic] on *Preference Rating* ($\chi^2(1) = 94.32,\ p < 0.001$) as well as there was an effect of [±Dyslexic It.] on *Preference Rating* ($\chi^2(1) = 8.1,\ p = 0.004$) (Table 9.8).

- **Between Groups:** We did not find a significant effect of [±Dyslexic] on *Preference Rating* between groups ($p = 0.21$). Similarly, we did not find a significant effect of [±Dyslexic It.] on *Preference Rating* between groups ($p = 0.83$). In Table 9.8 we show the medians, means and standard deviations of each group.

- **Group D:** We found a significant effect of [±Dyslexic] on *Preference Rating* for participants with dyslexia ($p < 0.001$). The fonts which were not designed for people with dyslexia (*Arial*, *Helvetica*, *Myriad*, and *Verdana*), ($\bar{x} = 3.60,\ s = 1.04$), were preferred to *OpenDyslexic*, ($\bar{x} = 2.57,\ s = 1.15$). In contrast, we did not find a significant effect of [±Dyslexic It.] on *Preference Rating* for people with dyslexia ($p = 0.06$) (Table 9.8).

- **Group C:** We found a significant effect of [±Dyslexic] on *Preference Rating* for participants without dyslexia ($p < 0.001$). The [−Dyslexic] fonts (*Arial*, *Helvetica*, *Myriad*, and *Verdana*), ($\bar{x} = 3.86,\ s = 0.91$), were preferred to *OpenDyslexic* [+Dyslexic], ($\bar{x} = 2.24,\ s = 1.09$). There was a significant effect of [±Dyslexic It.] on *Preference Rating* for people without dyslexia ($p < 0.001$). The [−Dyslexic It.] font type, ($\bar{x} = 3.36,\ s = 0.93$) was significantly more preferred than the [+Dyslexic It.] font *OpenDys It.*, ($\bar{x} = 2.03,\ s = 1.05$) (Table 9.8).

9.5 Discussion

The differences in reading performance between groups are consistent with the majority of eye tracking literature from experimental psychology [3, 122, 126, 208, 226]. The eye movements of readers with dyslexia are different from regular readers. People with dyslexia as well as beginning readers, make longer fixations, more fixations, shorter saccades and more regressions than normal readers. However, participants with dyslexia had significantly lower preferences ratings than the participants without dyslexia. This can be explained by their reading difficulties caused by dyslexia.

The correlations between groups were significant for all the measures. The fonts which are more readable for people with dyslexia are also beneficial for people without dyslexia. People without dyslexia also prefer the fonts that people with dyslexia prefer are also preferred. While there were strong positive correlations between the two measures for reading performance (*Reading Time* and *Fixation Duration*). The correlations between this two measures were negative and significant for both groups. What is beneficial for the readability for people with and without dyslexia is not necessarily what they prefer.

First, our results on reading performance provide evidence that font types have an impact on readability for both people with and without dyslexia. Second, these results are consistent with most of the current text design recommendations for people with dyslexia. Fonts *sans serif* and in *roman* style lead to shorter fixation durations in our participants with dyslexia, as recommended by Lockley [217]. However, these styles did not lead to significant shorter reading durations for people with dyslexia. For people without dyslexia *italic* fonts lead to longer reading times and fixation durations.

Overall, the reading time of the *italic* fonts was always worse than its *roman* counterpart, confirming the commonly established fact that cursive letters are harder to read for people with dyslexia. Both groups preferred *roman* fonts. Although *sans serif*, *monospaced* and *roman* fonts lead to significant shorter fixation durations for people with dyslexia, we did not find a significant difference in reading time. Hence, our conclusions towards these characteristics are weaker. People without dyslexia presented the same behavior regarding *sans serif* and *monospaced* fonts, they presented longer fixation durations for *serif* and *monospaced* fonts. But only people with dyslexia preferred *sans serif* fonts, and surprisingly people without dyslexia preferred *proportional* fonts.

The fonts designed specifically for dyslexia, *OpenDys* and *OpenDys It.*, did not lead to a better or worse readability. As De Leeuw [108] shows, *OpenDys* did not lead to faster reading. However, we did not performed a reading out loud test with words, which is what might improve with the use of specially designed fonts [108]. For participants without dyslexia, *OpenDys It.* lead to shorter reading ties than *Arial It* and the non-dyslexics fonts *Arial*, *Helvetica*, *Myriad*, and *Verdana* lead to shorter fixations durations compared to *OpenDys*. In addition, both groups significantly preferred *Verdana* or *Helvetica* over *OpenDys*, and *Verdana*, *Helvetica* and *Arial OpenDys It.* Participants without dyslexia were more extreme with their preferences and also preferred *Arial It.*, *CMU*, *Myriad*, *Times* and *Garamond* over *OpenDys It.*, even if they objectively read faster with *OpenDys It.* compared to *Arial It.*

Although *Arial* had the shortest reading time for both groups and is highly recommended in literature for dyslexia [52, 130, 217], we cannot conclude that this font type leads to better readability because we only found significant differences with respect to *OpenDys It.* and *Arial It* in participants with dyslexia. However, for people with and without dyslexia *Arial It.* did lead to significant longer reading times than *Helvetica*, *Arial*, and *CMU*. It also lead to significant longer fixation durations than most of the fonts. Hence, we recommend avoiding *Arial It.* Moreover, participants with dyslexia significantly preferred *Arial* to *Arial It.*

The two fonts that lead to shorter fixation durations were *Courier* and *Helvetica*. Hence the use of these fonts might help people with dyslexia to read faster. This is consistent with the recommendation of AbilityNet [1] to use *Courier* and with Lockley [217] to use sans serif fonts in the case of *Helvetica*. Also, *Helvetica* was the second most significantly preferred font by our participants after *Verdana*.

Regarding reading time, more significant differences were found for participants without dyslexia while regarding fixation duration more effects were found within the participant with dyslexia. Similar to participants with dyslexia, *Arial It.* had longer reading times, *Arial* and

CMU presented shorter reading times, and shorter fixation durations were found using *Courier*. The discordant font was *Verdana*. While *Verdana* did not lead to shorter reading times for people with dyslexia and even presented shorter fixation durations than *Arial It.*, it seems to have the opposite effect for people without dyslexia. For the control group, *Verdana* had the second longest reading time mean and significantly longer compared to *Arial*, *CMU*, *Courier*, *Garamond*, *Helvetica*, and *Myriad*. Surprisingly, *Verdana* was preferred over *Courier*, which objectively lead to lower fixations in people without dyslexia. However, *Verdana* also had the second lowest fixation duration mean for participants without dyslexia.

One way to understand these results is to build the partial order obtained by considering all the order relations that are valid for the average values in *Reading Time* and the *Preference Ratings*. The result is given in Figure 9.8 (a), where the fonts can be grouped in four different levels. However, not all of these order relations are significant. Hence, the partial orders, (b) and (c), show the significant relations for *Reading Time* and *Preference Ratings*, respectively. In the case of (b), thicker lines indicate that those relations are also significant for *Fixation Duration*. From these partial orders, the only three fonts that are not dominated in both partial orders, (b) and (c), are *Helvetica*, *CMU*, and *Arial*. These can be considered good fonts for dyslexia when we also consider the subjective preferences of the participants. The next two in importance are *Verdana* and *Times*.

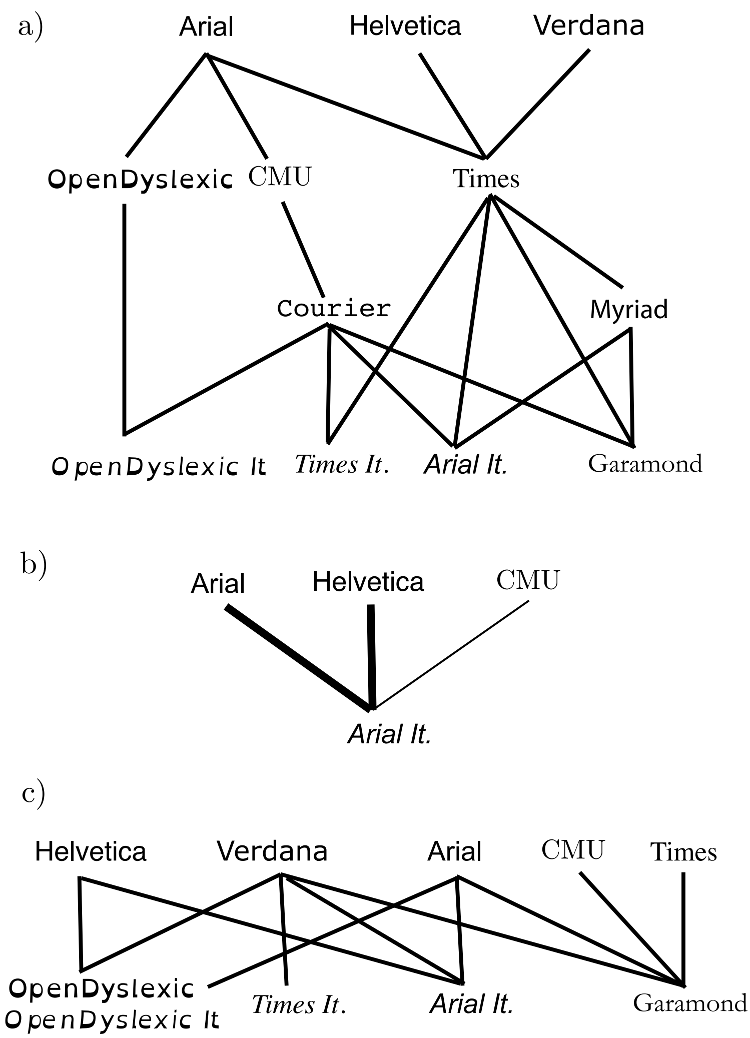


Figure 9.8: Partial order (group D) obtained from the means order of *Reading Time* and *Preference Rating* (a), and the partial order for significant differences in *Reading Time* (b) and *Preference Rating* (c).

Font Experiment		
Design	Within-subjects	
Independent Variables	*Font Type*	*Arial, Arial Italic* *Computer Modern Unicode (CMU)* *Courier, Garamond* *Helvetica, Myriad* *OpenDyslexic, OpenDyslexic Italic* *Times, Times Italic* *Verdana*
	[±Italic]	[−Italic] [+Italic]
	[±Serif]	[−Serif] [+Serif]
	[±Monospace]	[−Monospace] [+Monospace]
	[±Dyslexic]	[− Dyslexic] [+ Dyslexic]
	[±Dyslexic It.]	[− Dyslexic It.] [+ Dyslexic It.]
Dependent Variables (Sec. 3.1.1	*Reading Time* *Fixation Duration*	(objective readability)
	Preference Rating	(subjective preferences)
Control Variable	*Comprehension Score*	(objective comprehensibility)
Participants (Sec. 3.1.2)	Group D (48 participants)	22 female, 26 male **Age:** range from 11 to 50 ($\bar{x}$ = 20.96, s = 9.98) **Education:** high school (26), university (19), no higher education (3)
	Group C (49 participants)	28 female, 21 male **Age:** range from 11 to 54 ($\bar{x}$ = 29.20, s = 9.03) **Education:** high school (17), university (27), no higher education (5)
Materials (Sec. 3.1.3)	Texts Text Presentation Comprehension Quest. Preferences Quest.	12 story beginnings 12 literal items (1 item/text) 12 items (1 item/condition)
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750	
Procedure (Sec. 3.1.5)	**Steps:** Instructions, demographic questionnaire, reading task (× 12), comprehension questionnaire (× 12), preferences questionnaire (× 12)	

Table 9.1: Methodological summary for the *Font* experiment.

Font Type	**Group D**			*Font Type*	**Group C**		
	$\tilde{x}$	$\bar{x} \pm s$	%		$\tilde{x}$	$\bar{x} \pm s$	%
	Reading Time				*Reading Time*		
Arial	24.22	28.35 ± 12.39	100	Arial	15.63	11.83 ± 4.32	100
OpenDys	23.81	29.17 ± 15.79	103	Helvetica	16.78	12.41 ± 4.17	105
CMU	26.06	29.58 ± 12.05	104	CMU	17.21	13.03 ± 4.29	110
Courier	29.73	29.61 ± 10.87	104	OpenDys	16.34	13.09 ± 5.85	111
OpenDys It.	25.44	29.68 ± 14.44	105	Garamond	16.71	13.33 ± 4.38	113
Helvetica	27.18	31.05 ± 15.04	109	Courier	16.90	13.75 ± 5.50	116
Verdana	28.97	31.16 ± 13.03	110	Myriad	17.86	13.87 ± 5.32	117
Times	29.30	31.68 ± 11.81	112	OpenDys It.	17.63	14.34 ± 6.14	121
Times It.	28.55	32.38 ± 12.34	114	Times It.	18.50	14.55 ± 6.38	123
Myriad	26.95	32.66 ± 14.80	115	Times	18.35	14.75 ± 5.09	125
Garamond	30.53	33.30 ± 15.45	117	Verdana	18.86	15.69 ± 7.88	133
Arial It.	29.68	34.99 ± 16.60	123	Arial It.	21.25	17.24 ± 7.53	146

Table 9.2: Median, mean and standard deviation of *Reading Time* in seconds for groups C and D. We include the relative percentage for *Reading Time*, with respect to the smallest average value, Arial.

Font Type	**Group D**			*Font Type*	**Group C**		
	$\tilde{x}$	$\bar{x} \pm s$	%		$\tilde{x}$	$\bar{x} \pm s$	%
	Fixation Duration				*Fixation Duration*		
Courier	0.22	0.22 ± 0.05	100	Courier	0.18	0.19 ± 0.03	100
Verdana	0.22	0.23 ± 0.07	105	Verdana	0.19	0.19 ± 0.03	100
Helvetica	0.24	0.24 ± 0.06	109	Arial	0.20	0.19 ± 0.03	100
Arial	0.23	0.24 ± 0.07	109	Helvetica	0.19	0.19 ± 0.04	100
Times	0.24	0.25 ± 0.07	114	Times It.	0.19	0.19 ± 0.04	100
Myriad	0.25	0.25 ± 0.07	114	Myriad	0.20	0.20 ± 0.04	105
Times It.	0.25	0.26 ± 0.06	118	Garamond	0.20	0.20 ± 0.03	105
OpenDys	0.24	0.26 ± 0.07	118	Times	0.21	0.20 ± 0.03	105
OpenDys It.	0.26	0.26 ± 0.07	118	CMU	0.20	0.20 ± 0.04	105
Garamond	0.25	0.27 ± 0.07	123	OpenDys	0.21	0.21 ± 0.04	111
CMU	0.25	0.27 ± 0.08	123	OpenDys It.	0.21	0.21 ± 0.04	111
Arial It.	0.28	0.28 ± 0.08	127	Arial It.	0.21	0.21 ± 0.04	111

Table 9.3: Median, mean and standard deviation of *Fixation Duration* in seconds for groups C and D. We include the relative percentage for *Fixation Duration* with respect to the smallest average value.

Font Type	**Group D**			*Font Type*	**Group C**		
	$\tilde{x}$	$\bar{x} \pm s$	%		$\tilde{x}$	$\bar{x} \pm s$	%
	Preference Rating				*Preference Rating*		
Verdana	4	3.79 ± 0.98	100	Verdana	4	3.97 ± 0.95	100
Helvetica	4	3.62 ± 1.08	96	Helvetica	4	3.97 ± 0.92	100
Arial	4	3.60 ± 1.13	95	Arial	4	3.84 ± 0.80	97
Times	4	3.45 ± 1.15	91	CMU	4	3.79 ± 0.90	95
Myriad	3.5	3.40 ± 0.99	90	Myriad	4	3.66 ± 0.93	92
CMU	3	3.31 ± 0.98	87	Times	4	3.64 ± 0.89	92
Courier	3	3.14 ± 1.39	83	Arial It.	3	3.36 ± 1.30	85
Arial It.	3	2.90 ± 1.10	77	Garamond	3	3.33 ± 0.93	84
Times It.	3	2.86 ± 1.20	75	Times It.	3	3.18 ± 1.07	80
Garamond	2	2.57 ± 1.15	68	Courier	3	2.85 ± 1.08	72
OpenDys	3	2.57 ± 1.15	68	OpenDys	2	2.24 ± 1.09	56
OpenDys It.	2	2.43 ± 1.17	64	OpenDys It.	2	2.03 ± 1.04	51

Table 9.4: Median, mean and standard deviation of the *Preference Rating* for groups C and D. We include the relative percentage for *Preference Rating* with respect to the highest average value, Verdana.

Font Type	**Group D**			*Font Type*	**Group C**		
	$\tilde{x}$	$\bar{x} \pm s$	%		$\tilde{x}$	$\bar{x} \pm s$	%
	Reading Time				*Reading Time*		
[−Italic]	27.04	29.74 ± 13.40	100	[−Italic]	16.69	17.40 ± 5.62	100
[+Italic]	28.77	32.35 ± 14.62	109	[+Italic]	18.91	20.11 ± 7.55	116
	Fixation Duration				*Fixation Duration*		
[−Italic]	0.24	0.25 ± 0.07	100	[−Italic]	0.20	0.20 ± 5.62	100
[+Italic]	0.26	0.27 ± 0.07	108	[+Italic]	0.20	0.21 ± 7.55	105
	Preference Rating				*Preference Rating*		
[−Italic]	3	3.21 ± 1.22	100	[−Italic]	3	3.24 ± 1.17	100
[+Italic]	3	2.73 ± 1.20	85	[+Italic]	3	2.86 ± 1.16	88

Table 9.5: Median, mean and standard deviation of *Reading Time*, *Fixation Duration*, and *Preference Rating* for [±*Italic*].

Font Type	**Group D**			*Font Type*	**Group C**		
	$\tilde{x}$	$\bar{x} \pm s$	%		$\tilde{x}$	$\bar{x} \pm s$	%
	Reading Time				*Reading Time*		
[−Serif]	27.08	30.80 ± 13.84	100	[−Serif]	17.48	17.98 ± 6.58	100
[+Serif]	29.06	31.53 ± 13.21	102	[+Serif]	16.78	18.07 ± 5.09	101
	Fixation Duration				*Fixation Duration*		
[−Serif]	0.24	0.24 ± 0.07	100	[−Serif]	0.20	0.20 ± 0.03	100
[+Serif]	0.25	0.26 ± 0.07	108	[+Serif]	0.19	0.19 ± 0.03	95
	Preference Rating				*Preference Rating*		
[−Serif]	4	3.60 ± 1.04	100	[−Serif]	4	3.84 ± 0.94	100
[+Serif]	3	3.11 ± 1.15	86	[+Serif]	4	3.62 ± 0.95	94

Table 9.6: Median, mean and standard deviation of *Reading Time*, *Fixation Duration*, and *Preference Rating* for $[\pm Serif]$.

Font Type	**Group D**			*Font Type*	**Group C**		
	$\tilde{x}$	$\bar{x} \pm s$	%		$\tilde{x}$	$\bar{x} \pm s$	%
	Reading Time				*Reading Time*		
[+Monospace]	29.73	29.61 ± 10.87	100	[+Monospace]	16.90	18.57 ± 6.24	100
[−Monospace]	29.06	31.53 ± 13.21	106	[−Monospace]	17.46	18.04 ± 5.09	97
	Fixation Duration				*Fixation Duration*		
[+Monospace]	0.22	0.22 ± 0.05	100	[+Monospace]	0.18	0.19 ± 0.03	100
[−Monospace]	0.25	0.26 ± 0.07	118	[−Monospace]	0.20	0.20 ± 0.03	105
	Preference Rating				*Preference Rating*		
[+Monospace]	3	3.14 ± 1.39	100	[+Monospace]	3	2.85 ± 1.30	79
[−Monospace]	3	3.11 ± 1.15	99	[−Monospace]	4	3.59 ± 0.98	100

Table 9.7: Median, mean and standard deviation of *Reading Time*, *Fixation Duration*, and *Preference Rating* for $[\pm Monospace]$.

Font Type	**Group D**			*Font Type*	**Group C**		
	$\tilde{x}$	$\bar{x} \pm s$	%		$\tilde{x}$	$\bar{x} \pm s$	%
	Reading Time				*Reading Time*		
[+Dyslexic]	23.81	29.17 ± 15.79	100	[+Dyslexic]	16.34	17.04 ± 6.08	100
[−Dyslexic]	27.08	30.80 ± 13.84	106	[−Dyslexic]	16.85	18.01 ± 6.59	106
[+Dyslexic It.]	25.44	29.68 ± 14.44	100	[+Dyslexic It.]	17.63	18.73 ± 5.60	100
[−Dyslexic It.]	29.68	34.99 ± 16.60	118	[−Dyslexic It.]	21.25	22.38 ± 9.02	119
	Fixation Duration				*Fixation Duration*		
[+Dyslexic]	0.24	0.26 ± 0.07	108	[+Dyslexic]	0.21	0.21 ± 0.04	111
[−Dyslexic]	0.24	0.24 ± 0.07	100	[−Dyslexic]	0.19	0.19 ± 0.03	100
[+Dyslexic It.]	0.25	0.26 ± 0.07	100	[+Dyslexic It.]	0.21	0.21 ± 0.04	100
[−Dyslexic It.]	0.28	0.28 ± 0.08	108	[−Dyslexic It.]	0.21	0.21 ± 0.04	100
	Preference Rating				*Preference Rating*		
[+Dyslexic]	3	2.57 ± 1.15	71	[+Dyslexic]	2	2.24 ± 1.09	58
[−Dyslexic]	4	3.60 ± 1.04	100	[−Dyslexic]	4	3.86 ± 0.91	100
[+Dyslexic It.]	2	2.42 ± 1.27	83	[+Dyslexic It.]	2	2.03 ± 1.05	60
[−Dyslexic It.]	3	2.90 ± 1.10	100	[−Dyslexic It.]	3	3.36 ± 0.93	100

Table 9.8: Median, mean and standard deviation of *Reading Time*, *Fixation Duration*, and *Preference Rating* for the [±*Dyslexic*] and [±Dyslexic It.] conditions.

Part IV

TEXT CONTENT

In Part III we addressed how certain text presentations can benefit the reading performance. However, there are still problems found by persons with dyslexia that remain unsolved. Given that dyslexia is a learning disability that affects language, we can assume that accessibility can be approached not only from the presentation of the text point of view, but also from the text content. Actually, the use of complicated language has been extensively pointed out as one of the key problems for this target group.

In this part of the thesis we explore how different text modifications, ranging from the lexical to the discourse level, affect the reading performance of people with dyslexia. Starting with the lexical level, in Chapter 10 we explore how the frequency and the length of the words affect the reading performance of people with dyslexia carrying out two experiments: *Word Frequency* and *Word Length*. Later, since words and numbers are processed differently, and around 40% of the dyslexic population has also dyscalculia, in Chapter 11 we explore how different numerical representations affect the reading performance of people with dyslexia. For this case we performed three experiments addressing six types of numerical representations: *Digits vs. Words*, *Rounding vs. Decimals* and *Percentages vs. Fractions*.

Having found statistical effects in the lexical level (Chapters 10 and 11) we move to the next language level, syntax. In Chapter 12 we investigate how syntactic simplification via verbal paraphrases affects the reading performance of people with dyslexia and conducted the *Verbal Paraphrases* experiment, finding no statistical effects. This lead us to go further and explore the discourse level. Hence, in Chapter 13 we explore how two of the most recommended pedagogical strategies for people with dyslexia affect their reading performance: the use of mind maps (*Graphical Schemes* experiment) and the inclusion highlighted keywords in the text for the main ideas (*Keywords* experiment).

Chapter 10

Word Frequency and Length

10.1 Introduction

Previous work has studied how to optimize the presentation of the text, but adapting the text content has not received that much attention. Also, there are applications that modify the text layout for users with dyslexia, but these only modify its design but not its content. However, According to cognitive science literature, people with dyslexia specifically encounter problems with less frequent words and long words [100, 174, 414].

The goal of this chapter is to study to which extent word frequency and word length impacts text readability and comprehensibility for people with dyslexia. Likewise to lexical simplification, which often involves replacing difficult words by their simpler synonyms [45], in two experiments we replaced nouns by synonyms with diverse lengths and frequencies to study its effect on readability and comprehensibility. This would reveal whether people with dyslexia could benefit from lexical simplification tools, such as a browser plug-in that allows interactive substitution of complex words by simpler synonyms. Hence, we present an eye tracking study that addresses this goal. In an experiment with 46 people, 23 with dyslexia and 23 without as a control group, we compare texts where words were substituted by shorter/longer and more/less frequent synonyms.

Our hypotheses are:

H1.1 *High frequency words increase readability for people with* *dyslexia.*

H1.2 *High frequency words increase comprehensibility for people with dyslexia.*

H2.1 *Shorter words increase readability for people with dyslexia.*

H2.2 *Shorter words increase comprehensibility for people with dyslexia.*

The main contributions are:

– Frequent words improve readability for people with dyslexia.

– Shorter words improve comprehensibility for people with dyslexia.

– The effect of word length and frequency is not significant for people without dyslexia.

The results of this chapter are presented in Rello *et al.* [322].

10.2 Related Work

We chose to study word frequency and word length because they are related to the word's processing time [305, 352], and they are strongly related to the difficulties that people with dyslexia find. Previous work related to our research can be divided into: studies from experimental psychology about word frequency and length in dyslexia, and work about the effect of text content on the comprehension abilities of people with dyslexia.

10.2.1 Word Frequency and Length

Among the language difficulties that people with dyslexia find, the additional difficulty that less frequent words (*e.g. pristine*) and longer words (*e.g. prestidigitation*) has been specially stressed [100, 174, 414]. The role of word frequency is so crucial in dyslexia that there is even a diagnosis of dyslexia based on the performance while reading frequent words [200].

The closest work to ours comes from experimental psychology [174], which studies the effect of word length and word frequency in relation with eye fixation patterns among readers. Their results show that low frequency and long words present longer gaze durations and more reinspections. Their approach focuses on finding eye-movement patterns to study particular words with the aim to discard the oculomotor dysfunction hypothesis of dyslexia [174]. They analyze single words while we analyze the whole text.

Since word frequency and word length are naturally related in language, we studied them in two different experiments. The correlation originates from the fact that expressions which are frequently used, tend to become shorter over time [185]. As stressed by Rayner *et al.* [305], to unveil cause and effect relationships, these dimensions have to be studied individually.

10.2.2 Text Content and Comprehension

Previous research has studied the effect of long sentences with difficult structures [358], sentence context [246], and the incorporation of graphical schemes [327], among others, on text comprehension of readers with dyslexia. More related to our work are [339, 358] who have suggested that the text could be made more difficult by the inclusion of low frequency and long words.

Most research in this area focus on English, but our study considers Spanish language. Dyslexia manifestations vary depending on different language orthographies [57].

As explained in Section 2.1.3, English and Spanish have different orthographies. Therefore, findings from one language do not necessarily apply to the other language. In fact, English has a deep orthography where the relationships between letters and sounds are inconsistent and many exceptions are permitted. On the other hand, Spanish has a shallow orthography with a more regular alphabetic system that contains consistent mappings between letters and sounds, the second shallowest of European languages [354].

10.2.3 What is Missing?

What is missing in previous research is knowledge about two factors of lexical complexity –word frequency and word length– in the Spanish language, and their impact on the reading performance and the comprehension of people with dyslexia, as well as the integration of these findings in tools for people with dyslexia.

10.3 Methodology

We study word frequency and word length as two independent variables in two different experiments. Nonetheless, the inherent relationship between frequency and length has constrained the selection criteria of the target words, *i.e.* long words are inevitably less frequent. In the experiments, 46 participants (23 with dyslexia) had to read four texts, which were altered to include more/less frequent and longer/shorter words. In Table 10.1 we show a summary of the methodology of the *Word Frequency* and the *Word Length* experiments.

10.3.1 Design

Independent Variables

For each experiment there were two conditions. In the *Word Frequency* experiment, [±Frequent] served as independent variable with two levels: [+Frequent] denotes the condition where suitable words were replaced by more frequent synonyms and [−Frequent] denotes the condition where suitable words were replaced by less frequent synonyms. In the *Word Length* experiment, [±Long] served as independent variable with two levels: [−Long] denotes the condition where suitable words were replaced by shorter synonyms and [+Long] denotes the condition where suitable words were replaced by longer synonyms.

The experiments followed a within-subjects design, so every participant contributed to each of the conditions in both experiments. The order of conditions was counter-balanced to cancel out sequence effects.

Dependent Variables

To measure objective readability and comprehensibility, we consider *Reading Time*, *Fixation Duration*, and a *Comprehension Score* as dependent variables. We counted the choices as explained in Section 3.1.1.

Word Frequency and **Word Length** Experiments		
Design	within-subjects	
		Word Frequency
Independent Variables	[±Frequent]	[+Frequent] [−Frequent]
		Word Length
	[±Long]	[+Long] [−Long]
Dependent Variables (Sec. 3.1.1)	*Reading Time*	(objective readability)
	Fixation Duration	
	Comprehension Score	(objective comprehensibility)
Participants (Sec. 3.1.2)	Group D (23 participants)	12 female, 11 male **Age:** range from 13 to 37 ($\bar{x}$ = 20.74, s = 8.18) **Education:** high school (11), university (10), no higher education (2) **Reading:** more than 8 hours (13.0%), 4-8 hours (39.1%), less than 4 hours/day (47.8%)
	Group C (23 participants)	13 female, 10 male **Age:** range from 13 to 35 ($\bar{x}$ = 20.91, s = 7.33) **Education:** high school (6), university (16), no higher education (1) **Reading:** more than 8 hours (4.3%), 4-8 hours (52.2%), less than 4 hours/day (43.5%)
Materials (Sec. 3.1.3)	Texts	4 texts (2 texts/experiment)
	Synonym Pairs	15 for *Word Frequency* Exp. 6 for *Word Length* Exp.
	Text Presentation	
	Compren. Quest.	8 inferential items (2 items/text)
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750	
Procedure (Sec. 3.1.5)	**Steps:** (per experiment) Instructions, demographic questionnaire, reading task (× 2), comprehension questionnaire (× 2), and preferences questionnaire (× 2)	

Table 10.1: Methodology for *Word Frequency & Length* experiments.

10.3.2 Participants

The details of the participants groups for both experiments are given in Table 10.1. Three of the participants were also diagnosed with attention deficit disorder. Regarding their reading habits we think that the participants interpreted reading time as schooling time and that is why the numbers may seem exaggerated. For further explanation about the participants please refer to Section 3.1.2.

10.3.3 Materials

To study the effects of word length and frequency, we need to study target words in context, that is, as part of a text. The rationale behind this is that readability and comprehensibility pertain to longer segments of texts [173]. To isolate the effects of these variables, the texts need to be comparable in complexity. In this section, we describe how we designed the texts, the target words, and the questionnaires that were used in this study.

Base Texts

As basis for our tests, we picked four short texts with an average length of 60.63 words. For *Word Frequency* experiment, we used two texts about the consequences of wars in Pakistan and Somalia. These were the most similar pairs of texts we could find in the Spanish Simplex corpus [47]. For the *Word Length* experiment, we created two mystery stories, one about a wizard and one about a car. In the following, we denote these texts as *Pakistan*, *Somalia*, *Wizard*, and *Car*.

To meet the comparability requirements among the texts belonging to the same experiment, we adapted the texts maintaining as much as possible the original text. We matched the readability of the texts by making sure that the parameters commonly used to compute readability [115], had the same or similar values:

(a) They have the same number of target words: Fifteen [±Frequent] word pairs for each of the texts in the *Word Frequency* experiment and six [±Long] word pairs for each of the texts in the *Word Length* experiment. Only in *Somalia* (*Word Frequency*) two target words are repeated in the text.

(b) Within each experiment, the texts use the same genre, international news.

(c) They are about similar topics.

(d) They contain the same number of sentences: four sentences for each of the texts in the *Word Frequency* experiment and three in the *Word Length* experiment.

(e) The texts of each experiment have the same number of words per text (93 words for both texts in *Word Frequency* and 33 words in *Word Length*).

(f) All the texts have similar word length, with an average length ranging from 4.89 to 5.50 letters.

(g) They contain the same number of unique named entities and they do not contain foreign words, numerical expressions or acronyms.

Synonym Pairs

Word length and frequency were selected taking into consideration the problems that people with dyslexia encounter. To control that other dyslexic-related difficulties did not interfere in the selected target words, we also took into account linguistic criteria during the selection of the target word pairs of synonyms:[1]

(a) We did not include ambiguous names because they require more processing than unambiguous words [305].

(b) We did not change compound nouns or collocations in the texts. An instance like *secretario general ('secretary of state')*, which is composed by two tokens but one meaning, were discarded.

(c) We did not use irregular words, foreign words, non–words, and new words, like *fantabuloso ('fantabulous')* [87, 100].

(d) We took into consideration only common names. Uncommon names were discarded because they are more likely to be irregular, foreign, or new words.

(e) We did not take into consideration phonetically similar words, that is, homophonic words (e.g. *weather* and *whether*) or pseudo homophonic words (e.g. *addition* and *audition*) [266] and orthographically similar words [125].

[1] Although pure synonyms do not exist [223], we consider as synonyms all different words with almost identical or similar meaning including the different word senses, by taking into account the context of the word.

To apply our criteria (a) to (d) we used linguistic knowledge from the Royal Spanish Academy Dictionary [313]. To control for orthographic and phonetic similarity (e) of the target word, we consulted the database of indexes of frequency, length, and orthographic neighbors[2] in Spanish [282], which enables us to determine how many neighbors the target word has and how frequent are those neighbors. We can assume that orthographic and phonetic similarities are related in Spanish, because, as we mentioned before, it has a shallow orthography.

Word Frequency

For creating the pair of [±Frequent] synonyms we first extracted the nouns of the texts according to the previous selection criteria. Then, we checked their synonyms using a synonym dictionary[3] and created a list of synonyms for each target word. We discarded nouns with no synonyms such as *million* or *kilometer*. We subsequently computed the relative frequencies of each of the synonyms for each of the lists using the advanced search of a major search engine.[4] Then, we manually selected the pair of synonyms (most frequent and less frequent) for the context where the target noun occurred in the text. Each pair of [±Frequent] has a frequency difference of at least one order of magnitude, for instance *morada ('house')*, is 67 times less frequent than *casa ('house')*. In the Appendix A.5 we present the pairs of synonyms with their frequency ratio. To maintain both factors independent, short nouns with less than four letters were discarded. Still, we could only partially control for length differences: the average length per word in the high frequency synonyms is 7.62 letters while for the low frequency synonyms is 9.56 letters.

Word Length

Finding synonym pairs with a relatively large difference in word length is challenging for Spanish because 79% of the words in the dictionary have between 6 and 11 letters. Note that the average word length in

[2]Neighbors are all the words with the same length as the target word that differs in a single letter [88], that is, the Hamming distance between them is one. For instance, the word *casa* has many neighbors like *masa, cosa, cama* and *caso*.

[3]`http://www.wordreference.com/sinonimos/`

[4]`http://www.google.com/advanced_search/`

Spanish is 8.78 letters per lemma, being similar to English where is 8.99 letters.[5]

Therefore, for creating the pair of [±Long] synonyms we took all the Spanish lemmas from the Royal Spanish Academy Dictionary [313] and selected the longest words. Then, we looked up for the synonyms[6] of these words to create the lists of synonyms. However, the majority of the longest words in Spanish do not have synonyms, for example *electroencefalograma ('electroencephalogram')*. From the list of synonyms we selected synonym pairs where the [+Long] synonym at least doubled the length of its [−Long] counterpart. For instance *apartamento ('flat')* is 2.75 times longer than *piso ('flat')*. In Appendix A.5, we present the pairs of synonyms with their length ratio.

Text Presentation

We used the parameters detailed in Section 3.1.3 with black font and creme background, plus 20 points font size. Each line did not exceed 62 characters per line.

Comprehension Questionnaires

We used multiple-choice questions with three possible choices, one correct choice, one partially correct choice, and one wrong choice. See Section 3.1.3 for details about the creation of the comprehension questionnaires. An example question is given in Figure 10.1.

10.4 Results

In this section we present the analyses of the data from the eye tracker (reading time and fixation duration) and the comprehension tests. First, we analyzed the differences among groups (D and N) and then the effect of the conditions within each group. A Shapiro-Wilk test showed that the three data sets were not normally distributed. However, a Barlett's test showed that they were homogeneous. Hence, we use Student's matched-pair t-tests for repeated measures to show statistically significant effects.

[5] We used the Royal Spanish Academy Dictionary [313] and the Longman Dictionary of Contemporary English [218].

[6] `http://www.wordreference.com/sinonimos/`

¿De qué trata el texto?
'What is the text about?'

☐ Sobre la acción de la Cruz Roja en Somalia.
'About the Red Cross action in Somalia'.

☐ Sobre las consecuencias de la guerra en la población de Somalia.
'About the consequences of the war on the Somalia population'.

☐ Sobre el incremento del precio de los productos en Somalia.
'About the price increase of products in Somalia'.

Figure 10.1: Comprehension question example for *Word Frequency.*

10.4.1 Differences between Groups

Next we present the differences we found among the groups.

Reading Time. Tables 10.2 and 10.3 present the means, standard deviations and medians of *Reading Time* for both groups.

- **Between Groups:** We found a significant difference between the groups regarding *Reading Time* ($t(124.71) = 5.43$, $p < 0.001$). Group D had significantly longer reading times ($\bar{x} = 31.86$, $s = 20.28$ seconds) than the participants of group C ($\bar{x} = 18.32$, $s = 10.53$ seconds).

Fixation Duration. Tables 10.2 and 10.3 present the means, standard deviations and medians of *Fixation Duration* for both groups.

- **Between Groups:** There was a significant difference between the groups regarding *Fixation Duration* ($t(140.11) = 8.21$, $p < 0.001$). Participants with dyslexia had significantly longer fixation times ($\bar{x} = 0.23$, $s = 0.05$ seconds) than the participants without dyslexia ($\bar{x} = 0.17$, $s = 0.03$ seconds).

Comprehension Score Table 10.2 presents the means, standard deviations and medians of the *Comprehension Score* for both groups.

- **Between Groups:** Participants with dyslexia answered fewer questions correctly ($\bar{x} = 88.75\%$, $s = 21.15\%$) than participants without dyslexia ($\bar{x} = 91.25\%$, $s = 19.24\%$). However, the

Word Frequency	Group D		Group C	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Reading Time			
[−Frequent]	48.97	53.35 ± 19.10	30.05	29.25 ± 10.77
[+Frequent]	39.70	41.99 ± 13.00	22.50	24.17 ± 6.18
	Fixation Duration			
[−Frequent]	0.24	0.25 ± 0.05	0.18	0.18 ± 0.03
[+Frequent]	0.23	0.22 ± 0.04	0.17	0.17 ± 0.03
	Comprehension Score			
[−Frequent]	100	87.50 ± 22.21	100	87.50 ± 22.21
[+Frequent]	100	90.00 ± 20.52	100	95.00 ± 15.39

Table 10.2: Results of the *Word Frequency* experiment.

difference between the groups was not statistically significant ($t(149.59) = -1.87,\ p = 0.063$).

10.4.2 Word Frequency

Reading Time. Table 10.2 presents the means, standard deviations and medians of *Reading Time* for both groups.

- **Within Groups:** For the participants with dyslexia we found a significant effect of [±Frequent] on *Reading Time* ($t(33.49) = -2.120,\ p = 0.035$). Using more frequent words lead to significantly shorter reading times ($\bar{x} = 41.99,\ s = 13.00$ seconds) than using less frequent words ($\bar{x} = 53.35,\ s = 19.10$ seconds). For the control group we found no significant effect on reading time ($t(30.283) = -1.83,\ p = 0.077$).

Fixation Duration. Table 10.2 presents the means, standard deviations and medians of *Fixation Duration* for both groups.

- **Within Groups:** For the participants with dyslexia, we found a significant effect of [±Frequent] on *Fixation Duration* ($t(35.74) = -2.15,\ p = 0.038$). Using more frequent words lead to significantly shorter fixation times ($\bar{x} = 0.22,\ s = 0.04$ seconds) than

Word Length	**Group D**		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Reading Time			
[−Long]	11.53	13.74 ± 6.05	8.49	9.59 ± 3.11
[+Long]	18.42	21.23 ± 10.28	10.66	11.78 ± 4.24
	Fixation Duration			
[−Long]	0.22	0.22 ± 0.04	0.19	0.17 ± 0.03
[+Long]	0.25	0.23 ± 0.05	0.19	0.18 ± 0.03
	Comprehension Score			
[−Long]	100	88.64 ± 26.42	100	90.91 ± 19.73
[+Long]	50	65.91 ± 35.81	100	86.36 ± 22.79

Table 10.3: Results of the *Word Length* experiment.

using less frequent words ($\bar{x} = 0.25$, $s = 0.05$ seconds). For the control group, no significant effect on fixation duration was found ($t(37.40) = -1.04$, $p = 0.307$).

Comprehension Score See Table 10.2 for the means, standard deviations, and medians of the *Comprehension Score*.

- **Within Groups:** We did not found a significant effect on the *Comprehension Score* for group D ($t(37.76) = 0.370$, $p = 0.714$) nor for group C ($t(33.82) = 1.24$, $p = 0.223$).

10.4.3 Word Length

Reading Time. Table 10.3 presents the means, standard deviations, and medians of *Reading Time* for both groups.

- **Within Groups:** For group D, we found a significant effect on reading time ($t(33.97) = 2.94$, $p = 0.006$). Using shorter words lead to significantly shorter reading times ($\bar{x} = 13.74$, $s = 6.05$ seconds) than using longer words ($\bar{x} = 21.23$, $s = 10.28$ seconds). For the control group, there was no significant effect on reading time ($t(38.57) = 1.96$, $p = 0.058$).

Fixation Duration. Table 10.3 presents the means, standard deviations and medians of *Fixation Duration* for both groups.

- **Within Groups:** We found no significant effect on the fixation duration in group D ($t(40.002) = 0.763,\ p = 0.450$), nor in group C ($t(41.57) = 0.991,\ p = 0.327$).

Comprehension Score See Table 10.3 for the means, standard deviations, and medians of the *Comprehension Score*.

- **Within Groups:** For the participants with dyslexia, we found a significant effect on the comprehension score ($t(38.67) = -2.40,\ p = 0.022$). Shorter words led to significantly higher text comprehension ($\bar{x} = 88.64\%,\ s = 26.42\%$) than longer words ($\bar{x} = 65.91\%,\ s = 35.81\%$). For the control group, changing the word length had no significant effect on text comprehension ($t(41.16) = -0.707,\ p = 0.484$).

10.5 Discussion

In general, participants without dyslexia read significantly faster and had shorter fixation durations than participants with dyslexia. However, participants with dyslexia read significantly faster and have significantly shorter fixation durations using more frequent words. Using shorter words caused participants with dyslexia to read significantly faster and significantly increased their text comprehension. For the people without dyslexia, no differences in reading time, fixation duration, or text comprehension were found.

Regarding the differences between the groups, our results are consistent with other eye tracking studies to diagnose dyslexia that found statistical differences between the two populations [3, 122, 126, 208, 226].

Since shorter reading times and fixation durations are associated with better readability, our findings support **H1.1**: *High frequency words increase readability for people with dyslexia.* The effect of frequency is more pronounced in people with dyslexia than in people without dyslexia, where no significant results were found. Our results are consistent with previous results for both groups. Word frequency was found to have a powerful influence on word recognition tasks for people with [174] or without [305, 352] dyslexia and some reading experiments have demonstrated that readers spend more time looking at low-frequency words than at high-frequency words [174].

In Figure 10.2 we consider the two visual behaviors that we use as surrogate variables for readability (reading time and fixation duration), visualizing the distribution of the data in such 2D space. A more meaningful feature emerges that can be obtained from a linear combination of both of them.[7] The emerging feature can be interpreted as a readability variable that defines the readability axis. Thus, any readability improvement that moves along the direction determined by the arrow approaches the "ideal situation", which is characterized by target values reached by people without dyslexia in the most favorable conditions (*e.g.* the behavior on people without dyslexia in texts with higher frequency words). The use of more frequent words bring the average fixation time of group D closer to group C.

Since word frequency had no significant effect on text comprehension, we cannot confirm **H1.2**: *High frequency words increase comprehensibility for people with dyslexia.*

Since shorter reading times are associated with better readability, our findings support **H2.1**: *Shorter words increase readability for people with dyslexia.* Although no statistical significance has been found for length regarding fixation duration, we can observe a qualitative improvement in the readability through the use of shorter words since the mean fixation duration for group D also comes closer to the mean fixation duration for group C.

One possible explanation of the lower significance of word length than word frequency is that *Wizard* might have been more difficult to read than *Car*. According to some of the comments of the participants during the open questions, the fact that there was an additive set in front of the noun in the text *Wizard* made it a bit more complicated for further reading. In Spanish, adjectives that are postponed to nouns are frequent and natural (unmarked syntactic structure for most of the cases). Therefore this might be a new variable to take into

[7]Doing a least-squares linear regression we obtain the formula *Reading time* = 173.3 *Fixation time* + 2.2 seconds (Pearson correlation of 0.51 and $p < 0.001$).

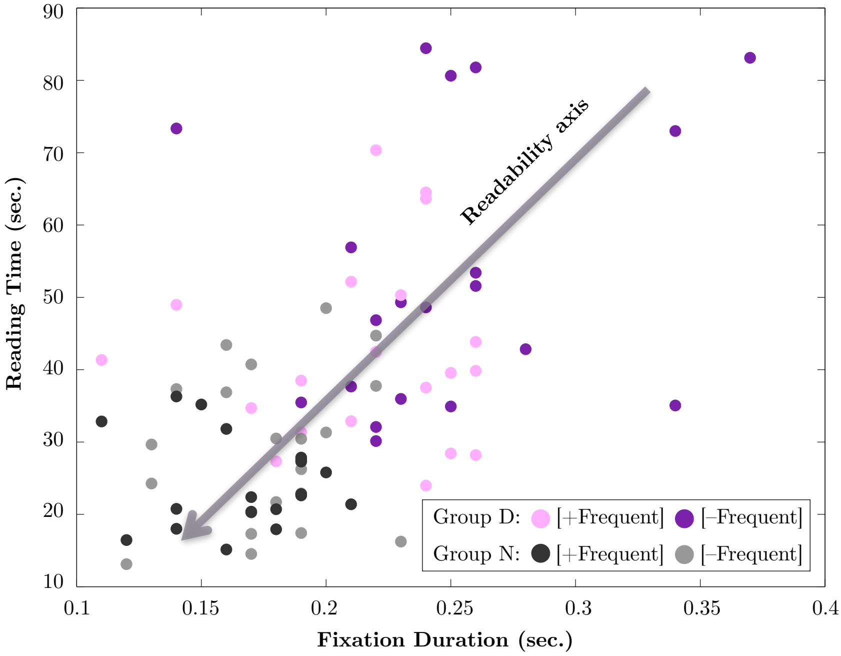


Figure 10.2: Readability as a function of reading time and fixation duration.

consideration in further work.

Because participants with dyslexia had a significantly increased text comprehension with texts having shorter words, our findings support **H2.2**: *Shorter words increase comprehensibility for people with dyslexia.*

Limitations of the Study

One of the limitations of our study is that the inferences made from **H2.1** and **H2.2** could be due to the low frequency of long words and not to their length. Through the experimental design we maintained as much as possible both factors –frequency and length– separated for studying both effects independently. However a total dissociation was not possible, as we could not find long words that were more frequent than their shorter synonyms. Although Rayner and Duffy [305] explain the necessity to study both effects separately, such separation between token frequency of linguistic expressions and their length does not exist in natural language [62]. Words [+Long] are in average 59.45 less frequent than their correspondent [−Long] synonyms. However this frequency ratio is not as high as in the experiment *word frequency* where [+Frequent] were 1249.17 times more frequent in average than their [−Frequent] synonyms (see the Appendix A.5 for the frequency ratios for each pair). Since using shorter words usually implies more frequent words, **H1.1** and **H2.1** reinforce each other.

However, the reasons why longer and less frequent words receive longer fixations are different [174], because while less frequent words require more processing, the word-length effect can be ascribed to acuity limitations of the visual system.[8] On the other hand, longer words may imply more fixations instead of longer fixations.

[8]Long words extend beyond the fovea where the acuity is greatest, thus increasing the need for making a fixation and even a re-fixation on a word [174].

Chapter 11

Numerical Expressions

11.1 Introduction

People with dyslexia find problems to recognize and recollect not only letters but also numbers [80, 248]. Although dyscalculia[1] and dyslexia are two different disabilities, they are comorbid [204][2] and people with dyslexia are more likely to have mathematical learning difficulties [203]. In addition, a large percentage of information expressed in daily news or reports contain numerical expressions such as economical statistics or demographic data. Numerical information can have different representations such as using digits or words, rounded numbers or decimals, fractions instead of percentages, etc. According to cognitive studies, numbers in a text are processed in a different way than words [110], and the presence of numbers in the text impacts the reading process [230]. For these reasons we are particularly interested in studying how the different kinds of numerical expressions affect the readability and the comprehension of a text.

The main goal of this chapter is to study the effect of different representations of numerical expressions with respect to readability and comprehensibility for people with dyslexia. To the best of our

[1] A specific learning disability involving innate difficulty in learning or comprehending arithmetic. It is akin to dyslexia and includes difficulty in understanding numbers, learning how to manipulate numbers, learning mathematical facts, and a number of other related symptoms [61].

[2] Comorbidity indicates a medical condition (in this case dyscalculia) existing simultaneously but independently with another condition (dyslexia).

knowledge, this is the first study that addresses the cognitive load of number representation using eye tracking in any language, even more for people with dyslexia. We conducted three experiments with 72 persons (36 with dyslexia) using eye tracking and comprehension questionnaires.

We tested the following hypotheses:

H1.1 *Readability increases if digits are used instead of words for representing numerical expressions.*

H1.2 *Comprehensibility increases if digits are used instead of words for representing numerical expressions.*

H2.1 *Readability increases if rounded numerical expressions are used instead of unrounded expressions (with decimals).*

H2.2 *Comprehensibility increases if rounded numerical expressions are used instead of unrounded expressions (with decimals).*

H3.1 *Readability increases if numerical expressions are expressed in percentages instead of fractions.*

H3.2 *Comprehensibility increases if numerical expressions are expressed in percentages instead of fractions.*

From our results we can quantify the impact of numerical expressions in the reading process for people with or without dyslexia, and it is possible to apply this information to the adaptation of numerical information so texts are more accessible to the widest number of readers. The three main contributions of this chapter are:

– Numerical information represented as digits improve readability for people with dyslexia but do not help comprehension.

– No significant results were found concerning the influence of rounding.

– Numerical information represented as percentages improve readability for people with dyslexia.

The results of this chapter are presented in Rello *et al.* [329].

11.2 Related Work

Related work is divided into two areas experimental psychology, and text simplification in natural language processing.

11.2.1 Numerical Expressions

Experimental psychology and cognitive neuropsychology have dealt with the study of number processing and calculation over the last two decades. Many researchers have studied the cognitive processes that are responsible for number processing and calculation, with the goal of contributing to the improvement of teaching and learning. For example, [167, 345] present findings about how the frequency of use of a word or number is an influential variable in the reading process. In addition, it seems that numerical expressions most frequently used require less time for recognition.

Researchers in psychophysics [304] have also studied various aspects of reading: impact of text context on readability and eye movements during reading, among others. When numerical expressions are expressed in digits there is a faster access to its semantic representation than when expressed in words [103, 133, 134, 181].

From experimental psychology there is evidence regarding the importance of the frequency of use of numerical expressions. For example, Brysbaert [58] investigated number processing by looking at reading time using eye tracking and showed effects for the frequency of the number. Overall, it is a generally accepted hypothesis that the probability of making a fixation on a particular linguistic unit is determined by the perceptual or informational relevance of the unit, and the degree of difficulty of processing required for its identification. Due to the quantitative nature of this measure, it is not easy to unravel what are the specific reasons justifying the preference for fixations on certain kinds of linguistic units. Thus, the visual processing of words has been shown to be significantly affected by factors such as length, frequency, type of vocabulary, predictability, or word ambiguity [305].

11.2.2 Text Simplification

Currently there is plenty of automatic text simplification research based on cognitive aspects. The main objectives are to identify simplification operations that can be applied to adapt a text using some kind of automatic means. Most of the text simplification approaches disregard the treatment of numerical expressions [45, 73], except [27] that is a numerical expressions simplification system designed on the basis of corpus analyses [26, 116].

11.2.3 What is Missing?

We found no previous work that investigates the effect of numerical representations in readers with dyslexia using eye tracking, neither for people in general. Also, to the best of our knowledge, there are no user evaluations regarding the impact of simplifying numerical expressions.

11.3 Methodology

We designed three different experiments to study the effect of different representations of numerical expressions with respect to readability and comprehensibility. In Table 11.1 we show a summary of the methodology of the numerical representation experiments: *Digits vs. Words*, *Rounding vs. Decimals*, and *Percentages vs. Fractions* experiments.

11.3.1 Design

Independent Variables

In the *Digits vs. Words* experiment, the independent variable [±Digit] had two levels: [+Digit] denotes the condition where numbers in the text were written in digits, *i.e. 22*, and [−Digit] denotes the condition where numbers were written using words, *i.e. veintidos ('twenty two')*.

In the *Rounding vs. Decimals* experiment, the independent variable [±Round] had again two levels: [+Round] denotes the condition where numbers were rounded, without decimals *i.e. 19*, and [−Round] denotes the condition where numbers in the text were written with decimals, *i.e. 19.45*.

In *Percentages vs. Fractions* experiment, the independent variable

[±Percentage] also had two levels: [+Percentage] denotes the condition where numbers in the text were written using percentages, *i.e. 25%*, and [−Percentage] denotes the condition where numbers in the text were written using fractions, *i.e. 1/4*.

The experiment followed a within-subjects design, so every participant contributed to each of the conditions in the experiments. The order of conditions was counter-balanced to cancel out sequence effects.

Dependent Variables

To measure objective comprehensibility and readability, we used *Fixation Duration* and *Comprehension Score*, respectively. The *Comprehension Score* was calculated as the percentage of correct answers, where the correct choice scored 100% and the others 0% using multiple-choice questions with three possible choices. To measure subjective comprehensibility and readability, we used a *Readability Rating* and *Comprehensibility Rating*, respectively. The dependent variables are explained in detail in Section 3.1.1.

11.3.2 Participants

The participants' demographic data is presented in Table 10.1. For further explanation about the participants please refer to Section 3.1.2.

11.3.3 Materials

Base Texts

To study the effects of numerical expressions we need to study target numerical expressions as part of a text. As basis for our manipulations, we created six texts, ranging from 39 to 63 words, with an average length of 55 words. To meet the comparability requirements among the texts belonging to the same experiment [115] the text had the following characteristics:

(a) They have the same number of target words: Seven [±Digit] numerical expressions pairs for each of the texts in *Digits vs. Words*, seven [±Round] numerical expressions pairs for each of the texts in *Rounding vs. Decimals*, and four [±Percentage] target numerical expressions in *Percentages vs. Fractions.*

(b) They share the same genre and are about similar topics: fast food ingredients.

(c) They contain the same number of sentences: three sentences in *Percentages vs. Fractions* and four sentences in the other two experiments.

(d) They have the same number of words per text (39 words in *Percentages vs. Fractions* and 63 words for the other two experiments).

(e) All the texts have a similar word length average ranging from 4.88 to 5.24 letters.

(f) They do not contain named entities, foreign words, or numerical acronyms.

Numerical Expressions

For the numerical expressions we used the following criteria:

(a) We chose numerical expressions denoting the same number because the frequency of the number expressed has an impact on its processing time [167, 345]. Therefore the same numerical expressions with different representation were chosen for the three experiments.

(b) We did not include ambiguous numerical expressions because they require more processing than unambiguous ones [305]. For instance, depending on the context *cien ('hundred')* could also mean "very fast" in Spanish, *i.e. Estoy trabajando a cien ('I am working very fast')* and *Las manzanas están a cien ('The apples cost one hundred').*

(c) We did not used orthographically similar numerical expressions, (*i.e.* mirror numbers '6' and '9') in the same text since such representations can be a major difficulty for people with dyslexia [125].

(d) In the experiment *Digits vs. Words*, we selected different numerical expressions, *i.e.* with decimals, rounded, two or three digits and percentages (see the Appendix A.6 for the data used).

(e) In the experiment *Rounding vs. Decimals*, we rounded the original numbers, always using modifiers when there was a loss of precision, *i.e. un poco ('a little'), casi ('almost').*

(f) In the experiment *Percentages vs. Fractions*, we selected frequently occurring percentages and their corresponding fractions. We only used exact fractions in the experiment.

Text Presentation

We used the parameters detailed in Section 3.1.3 with black font and creme background, plus 20 points font size with and average of 62 characters per line.

Comprehension Questionnaires

We used two comprehension items (one inferential and one literal) for each of the texts. The literal question asked about one details of the text that were expressed using a numerical expression. It was always phrased with the same number representation that was used in the text. Two examples, an inferential question (i) and a question devoted to details (ii) are given in Figure 11.1. See Section 3.1.3 for details about the creation of the comprehension questionnaires.

Subjective Readability and Comprehension Questionnaires

Each participant performed a questionnaire with 20 questions that were rated using a five-point Likert scale. For 10 of the statements the participant was asked about how easy was to read the text, that is, readability, while for the other 10 statements the participant was asked about how easy was to comprehend the text, that is, comprehensibility. Each of the statements contained a numerical expression, using one of the following representations: rounded, unrounded, percentage and fractions, where eight were written in words and twelve in digits. In the Appendix A.6 we present the set of numerical expressions used.

(i) El texto trata sobre:
'The text is about:'

- ☐ La descomposición de una hamburguesa.
  'The decomposition of a burger.'
- ☐ La creación de una hamburguesa.
  'The creation of a burger.'
- ☐ La composición de una hamburguesa.
  'The composition of a burger.'

(ii) Una porción de patatas fritas tiene un máximo de:
'A portion of french fries has a maximum of:'

- ☐ 200 kilocalorías.
  '200 kilocalories.'
- ☐ 300 kilocalorías.
  '300 kilocalories.'
- ☐ 400 kilocalorías.
  '400 kilocalories.'

Figure 11.1: Two questions of the comprehension tests.

11.4 Results

In this section we present the analysis of the results of the eye tracking and comprehension tests. A Shapiro-Wilk test showed that the three data sets were not normally distributed. However, a Barlett's test showed that they were homogeneous. Hence, in order to test our hypotheses, differences between groups and conditions were tested by means of (Bonferroni corrected) Student's t-tests. First, we study the differences between both groups. Then, we analyze the impact of the different numerical expressions in objective and subjective readability and comprehensibility.

11.4.1 Differences between Groups

Next we present the differences we found among the groups for the three experiments. The results between groups with-in each experiment lead to the same statistical significances.

Reading Time. In Tables 11.2, 11.3, and 11.4 we present the means, standard deviations, and medians of *Reading Time* for both groups. There we found a significant difference between the groups regarding *Reading Time* ($t(184.82) = 5.67$, $p < 0.001$). Group D had significantly longer reading times ($\bar{x} = 33.50$, $s = 15.72$ seconds) than the participants of group C ($\bar{x} = 22.81$, $s = 11.05$ seconds).

Fixation Duration. In Tables 11.2, 11.3, and 11.4 we present the means, standard deviations, and medians of *Fixation Duration* for both groups. There was a significant difference between the groups regarding *Fixation Duration* ($t(175.02) = 6.07$, $p < 0.001$). Participants with dyslexia had significantly longer fixation times ($\bar{x} = 0.22$, $s = 0.05$ seconds) than the participants without dyslexia ($\bar{x} = 0.19$, $s = 0.03$ seconds).

Comprehension Score. There was a significant difference between the groups regarding *Comprehension Score* ($t(186.29) = -2.00$, $p = 0.046$). Group D had less correct answers ($\bar{x} = 85.15$, $s = 28.77$) than for group C ($\bar{x} = 92.31$, $s = 21.78$) (see Table 10.2).

Readability Rating. We calculated the standard deviation for all the statements. For the readability answers the standard deviation was significantly higher in group D ($\bar{x} = 1.20$, $s = 0.46$) than in group C ($\bar{x} = 0.87$, $s = 0.41$), with $p = 0.007$. This means that participants with dyslexia had a higher variability in their ratings.

Comprehensibility Rating. For the comprehensibility answers, we found no significant difference between groups in their standard deviation ($p = 0.157$) for group D ($\bar{x} = 1.04$, $s = 0.44$) and for group C ($\bar{x} = 0.89$, $s = 0.37$).

Taking into account both, *Readability Rating* and *Comprehensibility Rating*. We found a high Pearson correlation of 0.95 between the answers of both groups. Therefore, groups C and D generally agreed in their answers with respect to readability and comprehensibility. In Figures 11.2 and 11.3 we show the histograms of the subjective ratings.

11.4.2 Digits vs. Words

Reading Time. In Table 11.2 we present the means, standard deviations, and medians of *Reading Time* for both groups. There was a significant effect of [±Digit] on *Reading Time* for group C ($p = 0.029$) and group D ($p = 0.051$). All the participants had shorter reading times when the numerical expressions of the text were presented with [+Digit] and not written with letters (see Table 11.2).

Fixation Duration. In Table 11.2 we present the means, standard deviations, and medians of *Fixation Duration* for both groups. We did

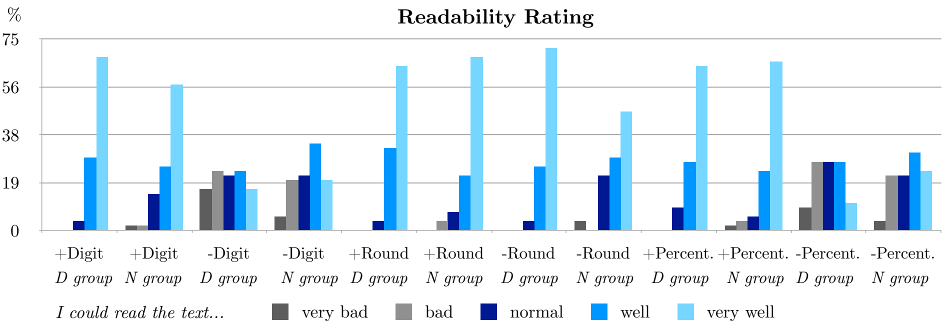


Figure 11.2: *Readability Rating* means for the *Numerical Representation* experiments.

not find statistical significance in readability ($p < 0.444$) for group C (see Table 11.2). However, we found statistical significance in readability for group D taking into account the mean of fixation time ($p < 0.054$). This result supports our **H1.1** hypothesis.

Comprehension Score. No statistical significance was found for both groups. Hence we reject hypothesis **H1.2** ($p < 0.241$ for group C and $p < 0.269$ for group D).

Readability Rating. We also found significance within groups for readability ($p < 0.001$ for both groups), Participants significantly found numbers written in digits more readable than in letters ($p < 0.001$) (Figure 11.2).

Comprehensibility Rating. We found significance within groups for comprehensibility ($p < 0.001$ in group D and $p = 0.014$ in group C) Participants significantly found numbers written in digits more understandable ($p < 0.001$) (Figure 11.3).

11.4.3 Rounding vs. Decimals

Reading Time. In Table 11.3 we present the means, standard deviations, and medians of *Reading Time* for both groups. We did not find statistical significance of [±Round] on *Reading Time* for group C ($p = 0.390$) and group D ($p = 0.759$) (see Table 11.3)

Fixation Duration. In Table 11.3 we present the means, standard deviations, and medians of *Fixation Duration* for both groups. We did

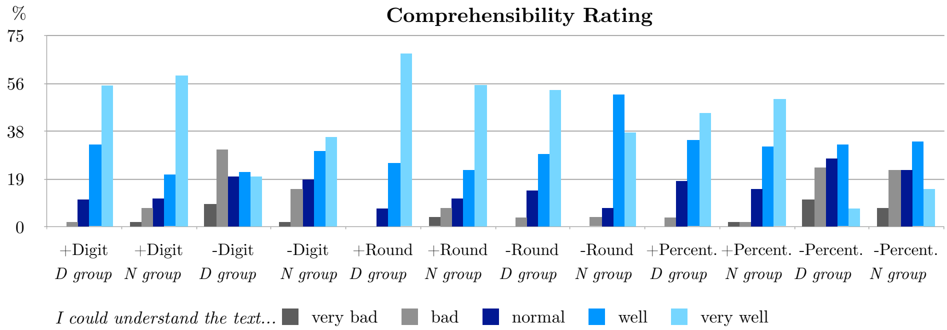


Figure 11.3: *Comprehensibility Rating* means for the *Numerical Representation* experiments.

not find statistical significance in readability in group C ($p < 0.867$) nor in group D ($p < 0.685$) when reading texts with rounded numerical expressions taking into account the mean of fixation time.

Comprehension Score. We also refute **H2.2** because we did not find statistical significance for comprehensibility in both groups ($p < 0.310$ in group C and $p < 0.695$ in group D).

Readability Rating. No significant differences were found within groups for readability ($t(53.94) = 0.48$, $p = 0.634$ in group D and $p = 0.111$ in group C) (Figure 11.2).

Comprehensibility Rating. No significant differences were found within groups for comprehensibility ($p = 0.163$ in group D and $p = 0.888$ in group C) (Figure 11.3).

11.4.4 Percentages vs. Fractions

Reading Time. In Table 11.4 we present the means, standard deviations, and medians of *Reading Time* for both groups. There was a significant effect of [±Percentage] on *Reading Time* for group D ($p = 0.054$), while we did not find any effects for group C ($p = 0.5415$). People with dyslexia had shorter reading times when the numerical expressions of the text were presented with [+Percentage] and not written with fractions (see Table 11.4).

Fixation Duration. In Table 11.4 we present the means, standard deviations, and medians of *Fixation Duration* for both groups. We did

not find statistical significance in readability for group C ($p < 0.462$) taking into account the mean of fixation time (see Table 11.4). However, our results confirm **H3.1** because we found statistical significance in readability for group D ($p < 0.046$) when reading texts with numerical expressions in percentages. This group reads faster texts with expressions in percentages than texts with numerical information in fractions.

Comprehension Score. On the other hand, we reject **H3.2** because we did not find statistical significance results for comprehensibility in both groups ($p < 0.170$ for group C and $p < 0.474$ for group D, see again Table 11.4).

Readability Rating. There was a significant effect of the condition on the *Readability Rating* ($p < 0.001$ for group D and group C). Participants significantly found percentages more readable than fractions ($p < 0.001$) (Figure 11.2).

Comprehensibility Rating. We found significance within groups for comprehensibility ($p < 0.001$ for group D and group C). Participants significantly found percentages more understandable than fractions ($p < 0.001$) (Figure 11.3).

11.5 Discussion

With respect to differences between the use of digits and the use of numerical expressions in words, results indicate a statistically significant improvement in performance for readability in people with dyslexia when digits are employed. In contrast, for group C, we found no significant differences regarding fixations duration. Group C had shorter reading times, perhaps due to the fact that the texts with numerical expressions in letter were longer (had more words). This agrees with the fact that numerical expressions described using words require a longer number of words and/or characters in comparison with the corresponding versions using digits. Overall length is an already known parameter that creates difficulties for people with dyslexia, so the reduction in length involved in phrasing a number in digits should make it easier to read for them (Chapter 10). Results for comprehensibility in experiment *digits vs. words* are not statistically significant for both groups.

With respect to differences between the use of rounded and unrounded numbers, none of the differences found are statistically significant. Rounding numbers even with modifiers such as "around" or "almost" did not have the expected effect in our experiments.

With respect to differences between the use of percentages and fractions, there is a statistically significant increase in readability for group D when percentages are used instead of fractions. In contrast, no differences were found for group C. Again, results for comprehensibility are not statistically significant in either case. There is an apparent contradiction in that for group D percentages seem to be easier to read but more difficult to understand. A possible explanation might be related to the nature of these expressions. From a conceptual point of view, both percentages and fractions convey the relative proportion between two quantities: the value of the percentage and 100 in the case of percentages, and the value of the numerator and the value of the denominator in fractions. However, the reference value in the case of percentages is implicit (or conveyed by the % sign). This implies that for fractions, two quantities have to be read, whereas only one needs to be read for percentages. This may account for the comparative ease for group D of reading percentages (only one quantity to read) vs. fractions (two different quantities to read). Participants that took this experiment did it because they were not tired and still willing to read more. In most cases participants with dyslexia were adults, and as such they have reading skills that are similar to adults without dyslexia. Note that fixation duration for these participants is shorter when reading percentages. More over, the percentages used in the texts were the most commonly used (see Appendix A.6) and the more frequent the word, the shorter the eye fixation [174].

The standard deviations of the subjective ratings reveal that people with dyslexia made a greater difference between readability and comprehensibility than people without dyslexia. For people without dyslexia, easier reading was correlated to text comprehension while participants with dyslexia dissociated these two elements, perhaps due do the nature of dyslexia, which affects reading but not comprehension of the language.

The higher variability of scores indicates that for people with dyslexia the representation of numbers has a much bigger impact on readability. However, the correlation of the answers of people with or without dyslexia is high. Hence, both groups generally agree in their rates with respect to readability and comprehensibility.

The significant difference in the number representations for *Digits vs. Words* and *Percentages vs. Fractions*, are consistent with the quantitative data from the eye tracker where we found significant variations. Hence, the performance and the preferences of our participants with respect to these number representations are consistent.

Numerical Representation Experiments		
Design	within-subjects	
Independent Variables		**Digits vs. Words**
	[±Digit]	[+Digit]
		[−Digit]
		Rounding vs. Decimals
	[±Round]	[+Round]
		[−Round]
		Percentages vs. Fractions
	[±Percentage]	[+Percentage]
		[−Percentage]
Dependent Variables (Sec. 3.1.1)	*Fixation Duration*	(objective readability)
	Comprehension Score	(objective comprehensibility)
	Readability Rating	(subjective readability)
	Comprehensibility Rating	(subjective comprehensibility)
Participants (Sec. 3.1.2)	Group D (36 participants)	21 female, 15 male **Age:** range from 16 to 50 ($\bar{x}$ = 23.38, s = 11.27) **Education:** high school (17), university (16), no higher education (3)
	Group C (36 participants)	22 female, 14 male **Age:** range from 16 to 50 ($\bar{x}$ = 26.94, s = 10.61) **Education:** high school (13), university (21), no higher education (2)
Materials (Sec. 3.1.3)	Texts	6 texts (2 texts/experiment)
	Numerical Expressions	14 in *Digits vs. Words Exp.* 14 in *Rounding vs. Decimals Exp.* 8 in *Percent. vs. Fractions Exp.*
	Text Presentation	
	Compren. Quest.	6 inferential, 6 literal items (2 items of each type/text)
	Sub. Readability Quest.	6 Likert scales (1/condition level)
	Sub. Comprehension Quest.	6 Likert scales (1/condition level)
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750	
Procedure (Sec. 3.1.5)	**Steps:** (per experiment) Instructions, demographic questionnaire, reading task (× 6), comprehension questionnaire (× 6), subjective readability questionnaire (×6), and subjective comprehension questionnaire (× 6)	

Table 11.1: Methodology of the *Numerical Representation* experiments.

Digits vs. Words	**Group D**		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Reading Time			
[−Digit]	37.66	42.73 ± 17.02	26.81	30.33 ± 13.12
[+Digit]	26.81	32.55 ± 14.83	19.30	22.46 ± 8.14
	Fixation Duration			
[−Digit]	0.22	0.24 ± 0.05	0.20	0.19 ± 0.02
[+Digit]	0.21	0.21 ± 0.04	0.19	0.19 ± 0.03
	Comprehension Score			
[−Digit]	100	76.51 ± 38.62	100	95.00 ± 15.39
[+Digit]	100	85.35 ± 23.50	100	87.50 ± 31.93

Table 11.2: Results for *Digits vs. Words* experiment.

Rounding vs. Decimals	**Group D**		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Reading Time			
[−Round]	29.50	38.55 ± 16.32	21.96	23.97 ± 10.64
[+Round]	36.17	40.28 ± 15.27	26.07	27.26 ± 10.65
	Fixation Duration			
[−Round]	0.21	0.22 ± 0.04	0.19	0.19 ± 0.03
[+Round]	0.21	0.23 ± 0.04	0.20	0.19 ± 0.03
	Comprehension Score			
[−Round]	100	91.67 ± 18.16	100	90.87 ± 27.19
[+Round]	100	90.63 ± 27.19	100	97.22 ± 12.50

Table 11.3: Results for *Rounding vs. Decimals* experiment.

Percentages vs. Fractions	**Group D**		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Reading Time			
[−Percentage]	21.92	24.76 ± 7.31	12.99	14.76 ± 5.46
[+Percentage]	19.22	20.07 ± 5.79	13.25	16.29 ± 8.20
	Fixation Duration			
[−Percentage]	0.22	0.23 ± 0.05	0.18	0.18 ± 0.03
[+Percentage]	0.20	0.20 ± 0.06	0.18	0.19 ± 0.03
	Comprehension Score			
[−Percentage]	100	88.89 ± 22.36	100	96.86 ± 12.50
[+Percentage]	100	80.55 ± 35.94	100	88.19 ± 22.36

Table 11.4: Results for *Percentages vs. Fractions* experiment.

Chapter 12

Verbal Paraphrases

12.1 Introduction

Lexical complexity such as word frequency, verb complexity and lexical ambiguity has an effect on the readability and comprehensibility for people without dyslexia [305]. According to cognitive neuroscience studies, people with dyslexia find difficulties with functional [266] and short words [371]. On the other hand, in natural language processing, text simplification is the process of transforming a text into an equivalent which is easier to read and to understand, preserving its meaning. One of the alternatives for text simplification is the use of verbal paraphrases. One of the more common verbal paraphrase pairs is the one composed by a lexical verb (*to hug*) and by a support verb plus a noun collocation (*to give a hug*). Functional and short words are present in those verbal paraphrases where *dar un paseo ('to go for a walk')*, support verb plus a noun collocation, could be simplified by a lexical verb *pasear ('to walk')*, in order to create simpler texts for people with dyslexia, by reducing the number of short and functional words form the text.

The goal of this chapter is to present the impact of lexical simplification through Spanish verbal paraphrases in readability and comprehensibility for people with and without dyslexia. This way we aim to find out whether lexical simplification systems targeted for people with dyslexia shall include verbal paraphrases. We conducted an eye tracking study with a group of 46 participants, 23 with confirmed dyslexia and 23 in the control group. To the best of our knowledge, this is the first time that the effect of verbal paraphrases is measured in terms of readability and comprehensibility for people with and without dyslexia using this methodology.

We tested the following hypotheses:

H1 *Syntactic simplification via lexical verbal paraphrases increase the readability of people with dyslexia.*

H2 *Syntactic simplification via lexical verbal paraphrases increase the comprehensibility of people with dyslexia.*

From our results, we conclude the following:

– We did not find significant effects, so tools that can perform this kind of paraphrases automatically might not have a large effect on people with dyslexia.

The analysis of this chapter was presented in Rello *et al.* [324],

12.2 Related Work

Related work to our study belong to different fields: (a) experimental psychology studies which takes into account the impact of language complexity in reading comprehension and performance of people with dyslexia, and (b) natural language processing (NLP) literature about paraphrases and their use in text simplification.

12.2.1 Text Complexity

Text complexity and dyslexia also has been studied in experimental psychology. Word frequency, verb complexity and lexical ambiguity are related to the processing time of words [305, 352]. Hyönä and Olson measure the effect of word length and word frequency in relation with eye fixation patterns and show that low frequency and long words present longer gaze durations and more re-inspections in both, readers with and without dyslexia [174]. In that work, the analysis is focused on target words [174] while we measure the whole text and the integration of target words in the overall text. The rationale behind this is that readability and comprehensibility pertain to longer segments of texts [173]. Comprehension in people with dyslexia was studied in correlation with syntax complexity including long sentences with complex structures [358], the sentence context [246], or the word fluency [101], among others.

12.2.2 Text Simplification

In NLP a paraphrase is an alternative surface form in the same language expressing the same semantic content as the original form [225]. The use of automatic methods for generating paraphrases has been successfully applied for text simplification among other NLP tasks. For instance, in [180] paraphrasing is used to remove difficult syntactic structures for deaf learners of written English and Japanese. Paraphrasing methods were applied to simplify for people with aphasia newspaper texts [72, 74] and online information [112].

12.2.3 What is Missing?

However, there are no studies for Spanish which approach readability and comprehension of people with dyslexia taking into consideration one common verbal paraphrasing pair [22] used for lexical simplification. That is, the pair composed of a lexical verb (*abrazar, 'to hug'*) and by a support verb plus a noun collocation (*dar un abrazo, 'to give a hug'*).

12.3 Methodology

To study the effect of verbal simplification in people with dyslexia we conducted the *Verbal Paraphrases* experiment with 46 participants (23 with dyslexia), as summarized in Table 12.1.

12.3.1 Design

Independent Variables

In the experiment, [±Simple] served as independent variable with two levels: [+Simple] denotes the condition where the verbal meanings are presented by lexical verbs (*to hug*), and [−Simple] denotes the condition where the verbal meanings are presented by a noun collocation and a lexical verb (*to give a hug*). More details about the verbal paraphrases selection and linguistic criteria are given in Section 12.3.3.

Verbal Paraphrases Experiment		
Design	Within-subjects	
Independent Variables	[±Simple]	[+Simple] [−Simple]
Dependent Variables (Sec. 3.1.1)	*Reading Time* *Fixation Duration* *Comprehension Score* *Preference Rating*	(objective readability) (objective comprehensibility) (subjective preferences)
Participants (Sec. 3.1.2)	Group D (23 participants)	12 female, 11 male **Age:** range from 13 to 37 ($\bar{x}$ = 20.74, s = 8.18) **Education:** high school (11), university (10), no higher education (2) **Reading:** more than 8 hours (13.0%), 4-8 hours (39.1%), less than 4 hours/day (47.8%)
	Group C (23 participants)	13 female, 10 male **Age:** range from 13 to 35 ($\bar{x}$ = 20.91, s = 7.33) **Education:** high school (6), university (16), no higher education (1) **Reading:** more than 8 hours (4.3%), 4-8 hours (52.2%), less than 4 hours/day (43.5%)
Materials (Sec. 3.1.3)	Texts Verbal Paraphrases Text Presentation Compren. Quest.	2 texts 18 (9/experiment) 4 inferential items (2 items/text)
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750	
Procedure (Sec. 3.1.5)	**Steps:** Instructions, demographic questionnaire, reading task (× 2), comprehension questionnaire (× 2), and preferences questionnaire (× 1)	

Table 12.1: Methodology for the *Verbal Paraphrases* experiment.

The experiment followed a within-subjects design, so every participant contributed to each of the conditions in both experiments.

The order of conditions was counter-balanced to cancel out sequence effects.

Dependent Variables

To measure objective readability and comprehensibility, we consider *Reading Time*, *Fixation Duration*, and a *Comprehension Score* as dependent variables. We explain them in detail in Section 3.1.1. For the *Comprehension Score* 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, we counted the choices 100%, 50%, and 0%, respectively.

12.3.2 Participants

Three of the participants were also diagnosed with attention deficit disorder. Regarding their reading habits we think that the participants interpreted reading time as schooling time and that is why the numbers may seem exaggerated. The rest of the details of the participants are given in Table 10.1. For further explanation about the participants please refer to Section 3.1.2.

12.3.3 Materials

Base Texts

We selected two very similar newspaper texts from the Spanish Simplex corpus [47]. To meet the comparability requirements among the texts [115], we slightly adapted the texts maintaining as much as possible the original text. Next, we present the characteristics shared by the texts:

(a) They are about similar topics: a literature award (Text *María*) and a cinema award (Text *Alex*). See the Appendix A.7 an example of texts use.

(b) They have the same number of substitutions: nine verbal paraphrase pairs [+Simple] and [−Simple]. See the Appendix A.7 for all the paraphrase pairs used in the experiment.

(c) They share the same genre: culture news.

(d) They have the same number of sentences per text, five sentences.

(e) They have the same number of words per text, 100 words.

(e) All the texts have a similar word length average ranging from 4.87 to 5.19 letters per word.

(f) They contain the same number of named entities mentioned for the first time.

(i) The texts do not contain numerical expressions, foreign words or acronyms.

Target Verbal Paraphrases

Under 18% of manual simplification operations made by experts in newspaper articles are lexical changes [45]. One of the most common simplification solutions done manually in Spanish is the substitution of the combination of the support verb and a deverbal noun by the corresponding verb alone [115]. That is, *dar un paseo ('to go for a walk')* by *pasear ('to walk')*, or *dar un abrazo ('to give a hug')* by *abrazar ('to hug')*. Although these kind of lexical simplifications are frequent in manual simplifications, their automatic computational process is still challenging [115]. Thus, there are specific linguistic resources developed for such tasks, such as the *Badele.3000* database [23].

Badele.3000 is a database that contains more than 3,600 high frequency Spanish nouns and 2,800 high frequency Spanish verbs, including 23,000 collocations made from the combinations of both kinds of words. The paraphrase pairs consisting of a verb and a verb-noun collocation were manually extracted [24]. As an expert created *Badele.3000* created manually, the linguistic validity of the paraphrases pairs used in our study is guaranteed.

The selected pairs of synonymic paraphrases are composed of a support verb plus a noun collocation and a lexical verb. According to the manual simplifications [115], the lexical verb alone is considered to be simpler; for instance:

[−Simple] Sus lectores **tenían confianza** en ella.
'Her readers **had trust** *in her.'*

[+Simple] Sus lectores **confiaban** en ella.
'Her readers **trusted** *her.'*

According to cognitive neuroscience studies, it would also be expected that people with dyslexia might find more difficult to read the [−Simple] option since they have more frequent errors with functional [266] and short words [371]. However, from a linguistic point of view it is not clear which option is simpler.

Linguists agree in differentiate lexical words and functional words [223]. Lexical words have a lexical meaning which is less ambiguous than the grammatical meanings expressed by functional words. Functional words are prepositions, pronouns, auxiliary verbs, and conjunctions, among others. Support verbs have been considered as functional words because they are semantically empty, for instance the verb *dar ('to give')* is a support verb in *dar un abrazo ('to give a hug')*.[1]

Since functional words do not have a lexical representation their processing is different than lexical words [65]. There are still many open questions about the different levels of word processing by the human brain. However, in the case of dyslexia a special emphasis has been made for errors in functional words [266]. To the best of our knowledge, there is no formal explanation behind errors in functional words. They could be due to their nature (*i.e.* lack of lexical content) or could be simply due to the fact that higher errors rates are observed for shorter words [371].

On the other hand, word processing depends on the complexity of the morphological components of the word [66]. For instance, *paseo ('walk')* is simpler than *pasear ('to walk')* because it is composed by one lexeme while *pasear* is made by one lexeme plus one derivative morpheme *pasear* = *paseo* + *ar*. Since it is not trivial to access the complexity of the paraphrase pairs from a linguistic point of view, we take as our criteria the empirical analysis observed in manual simplifications performed by experts [115].

[1] However, Barrios [22] analyzed extensively the meaning of support verbs concluding that some of them are not fully empty.

Verbal Paraphrases	**Group D**		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Reading Time			
[−Simple]	43.90	47.43 ± 14.61	28.13	27.83 ± 6.99
[+Simple]	41.47	44.40 ± 17.23	24.07	25.17 ± 5.48
	Fixation Duration			
[−Simple]	0.23	0.23 ± 0.06	0.19	0.18 ± 0.04
[+Simple]	0.23	0.23 ± 0.07	0.18	0.18 ± 0.04
	Comprehension Score			
[−Simple]	100	67.50 ± 43.75	100	77.50 ± 34.32
[+Simple]	100	67.50 ± 43.75	100	75.00 ± 38.04

Table 12.2: Results for the *Verbal Paraphrases* experiment.

Text Presentation

We used the parameters detailed in Section 3.1.3 with black font with creme background and 20 points font size. Each line had an average of 62 characters.

Comprehension Questionnaires

We used inferential items with multiple-choice questions with three possible choices, one correct choice, one partially correct choice, and one wrong choice. More details are shown in Section 3.1.3.

12.4 Results

The Shapiro-Wilk test showed that the *Fixation Duration* data sets were normally distributed while the *Reading Time* and the *Comprehension Score* data sets were not. Also, none of the data had an homogeneous variance (Bartlett test). Hence, to study the effect of *Verbal Paraphrases* on readability and comprehensibility we used the Wilcoxon non-parametric test for repeated measures.

Reading Time. In Table 12.2 we present the means and standard deviations.

- **Between Groups:** We found a significant difference between the groups regarding *Reading Time* ($V = 810,\ p < 0.001$).

- **Within Groups:** We did not find a significant effect of *Verbal Paraphrases* on *Reading Time* for the participants with dyslexia ($V = 134,\ p = 0.2873$), and for the participants without dyslexia ($V = 134,\ p = 0.294$).

Fixation Duration. The means and standard deviations are presented in Table 12.2.

- **Between Groups:** We found a significant difference between the groups regarding *Fixation Duration* ($V = 697,\ p < 0.001$). See Table 12.2.

- **Within Groups:** Similarly, there were found no significant effects of *Verbal Paraphrases* on *Fixation Duration* for the participants with ($V = 103,\ p = 0.9553$) or without dyslexia ($V = 87,\ p = 0.965$). To estimate the likelihood that we missed revealing an existing effect of verbal paraphrases on the mean of fixation durations, we calculated the achieved statistical power. Given a p-value of 0.873, an effect size of 0.052 (Cohen's d), and a sample size of 40, the achieved power is 0.880. Hence, the probability of not committing a Type II Error is 88%, that is, the likelihood that an unrevealed effect exists is only 12%.

Comprehension Score. See Table 12.2 for the means and standard deviations of the *Comprehension Score*.

- **Between Groups:** We did not find effects between the groups for the *Comprehension Score* ($V = 94,\ p = 0.283$)

- **Within Groups:** We found no significant effects of *Verbal Paraphrases* on *Comprehension Score* for the participants with ($V = 39,\ p = 1$) or without dyslexia ($V = 24.5,\ p = 0.851$).

12.5 Discussion

Regarding the differences between the groups, our results are consistent with other eye tracking studies to diagnose dyslexia that found statistical differences between the two populations [3, 122, 126, 208, 226].

Since *Verbal Paraphrases* had no significant effect on readability, we cannot confirm **H.1** *Syntactic simplification via verbal paraphrases increase the readability of people with dyslexia.*

One possible reason for this is that the texts were not very complex. If the texts contained very complicated sentences it could have been more probable to find effects of the simplification strategy.

Since *Verbal Paraphrases* had no significant effect on comprehension, we cannot confirm **H.2** *Syntactic simplification via lexical verbal paraphrases increase the comprehensibility of people with dyslexia..*

One possible reason for this is that the kind of simplifications that we performed were mild. If more extreme text simplification strategies were applied, the text would have changed more and, most probably, effects would have been found.

Chapter 13

Keywords and Graphical Schemes

13.1 Introduction

Generally, people with dyslexia are poor readers but strong visual thinkers. Actually, the role of visual thinking is crucial in dyslexics and its development may be helpful for a number of tasks such as visual analysis and pattern recognition [409]. Two of the most recommended pedagogical strategies in education literature for people with dyslexia are the use of mind maps (graphical schemes) [299, 75, 405] and highlighting keywords [164, 280]. These strategies are also recommended for students without dyslexia [407] (keywords) or reading disabilities [219] (graphical schemes). In fact, the inclusion of semantic maps was found to be beneficial for reading comprehension of general disabled readers in [359]. However, to the best of our knowledge, no estimation of the effect of graphical schemes or keywords on the objective readability and comprehensibly for people with dyslexia has been done.

In this chapter we present the first studies that explore graphical schemes and keywords using eye tracking with 46 (23 with dyslexia) and 62 Spanish native speakers (31 with dyslexia), respectively. First, we explored the relation between reading performance (readability and comprehensibility) and the visual conceptual schemes that aim to make the text more clear for these specific target readers. The second study explores the modification of the text presentation in relationship with its semantics, by highlighting the main ideas of the text (keywords) in boldface.

Our hypotheses are:

H1.1 *Graphical schemes in the text increases objective readability for people with dyslexia.*

H1.2 *Graphical schemes in the text increases objective comprehensibility for people with dyslexia.*

H1.3 *Graphical schemes in the text increases subjective readability for people with dyslexia.*

H1.4 *Graphical schemes in the text increases subjective comprehensibility for people with dyslexia.*

H1.5 *Graphical schemes in the text increases subjective memorability for people with dyslexia.*

H2.1 *Highlighted keywords in the text increases objective readability for people with dyslexia.*

H2.2 *Highlighted keywords in the text increases objective comprehensibility for people with dyslexia.*

H2.3 *Highlighted keywords in the text increases subjective readability for people with dyslexia.*

H2.4 *Highlighted keywords in the text increases subjective comprehensibility for people with dyslexia.*

H2.5 *Highlighted keywords in the text increases subjective memorability for people with dyslexia.*

Based on our results, the main contributions of this chapter are:

– Graphical schemes may help to improve the subjective readability and comprehensibility of people with dyslexia.

– Highlighted keywords in the text increases the comprehension by people with dyslexia, but not its readability.

– For people without dyslexia, no effects on the readability and the comprehension were found for both strategies.

The analyses of this chapter was presented in Rello *et al.* [327] (graphical schemes) and Rello *et al.* [334] (keywords).

13.2 Related Work

We divide the related work in: keywords and graphical schemes education literature for dyslexia, literature about automatic graphical schemes generation, natural language processing (NLP) literature about key-phrase and keyword extraction, and accessibility literature.

13.2.1 Education Literature

The use of graphical schemes has been an extensively recommended pedagogical strategy for dyslexic students [75, 299] as well as for students with reading disabilities [219]. These recommendations are partially related to the importance of visual thinking (visual analysis and pattern recognition) as a coping strategy for people with dyslexia [409]. So far, the inclusion of semantic maps was found to be beneficial for reading comprehension of general disabled readers in [359] and Weaver [405] specifically proposed the inclusion of graphical schemes to improve comprehension for readers with dyslexia.

Also, in education literature, highlighting keywords is a broadly recommended learning strategy [407]. Regarding students with dyslexia, teachers are encouraged to highlight keywords to make texts more accessible [164, 280]. These recommendations are based on qualitative analysis and direct observations with students.

13.2.2 Mind Maps Generation

There is a vast amount of studies that propose automatic or semi-automatic conceptual maps generation. The approaches for the creation of concept maps use statistical methods [77, 78, 162, 393, 401, 425], machine learning [207, 297, 403, 426], dictionaries and ontologies [77, 78, 93, 257, 389], or make use of linguistic tools and techniques [144, 297, 344, 389, 393, 401, 425, 426].

More concretely in natural language processing literature, Wang *et al.* [403] proposed an automatic concept map construction technique to create concept maps from abstracted short texts. Olney *et al.* [257] investigated the extraction of expert skeleton concept map exercises from text books. They extracted concept maps by defining a set of pedagogically relevant key terms, an ontology for linking terms together, and a set of rules for mapping semantic parses to concept map triples. Later, Zubrinic *et al.* [427] presented one method for the creation of concept maps from unstructured textual sources in Croatian language.

13.2.3 Key-phrase and Keyword Extraction

There is a vast amount of NLP literature on key-phrase[1] extraction [139, 191, 416]. The semantic data provided by key-phrase extraction can be used as metadata for refining NLP applications, such as summarization [105, 205], text ranking [235], indexing [232], query expansion [363], or document management and topic search [161].

The closest work to ours is [390] because they highlight key-phrases in the text to facilitate its skimming. They compare the highlighting outputs of two different systems, *Search 97* and *GenEx*, using six corpora belonging to different genre to find out features to extract key-phrases from texts.

13.2.4 Accessibility

Different strategies have been applied for improving readability of people with dyslexia such as the use of different text formats [158], the use of multi-modal information [192] and text to speech technologies [124], among others. The closest work to ours is the incorporation of summaries and graphical schemes in texts. Previous work has shown that the readability of students with dyslexia could be improved by using text summarization [245] and semantic maps [359]. In the case of the use of summaries, the evaluation of comprehension was carried out using questionnaires [245]. Multiple choice questions were applied to measure the incorporation of semantic maps among disable readers

[1] In the works mentioned ahead keyphrase and keywords are used interchangeable.

[359].

In the applications for people with dyslexia, highlighting is used not for keywords or main ideas but to help users for tracking their position when reading such as in *ScreenRuler* [79]. Sometimes highlighting is used simultaneously with text-to-speech technology [79]. In the *SeeWord* tool for *MS Word* [113, 157], highlighting is used on the letters where people with dyslexia normally make mistakes in order to attract the user's attention.

13.2.5 What is Missing?

First, we found no evaluation of the impact of graphical schemes on readability and comprehension of people with dyslexia combining data from eye tracking, questionnaires, and subjective ratings. Second, we did not find any study that measured objectively the impact of highlighting keywords in a text on the readability and comprehensibility for people with dyslexia. Finally, to the best of our knowledge, there are no studies in assistive technology that uses an NLP based engine to highlight keywords or generate graphical schemes for people with dyslexia. Hence, our experiments emulated the output that a potential NLP tool would give for highlighting the main ideas and generate a mind map.

13.3 Methodology

To study the effect of graphical scheme and keywords on text readability and comprehensibility on the screen, we conducted two experiments with Spanish native speakers: *Graphical Schemes* and *Keywords* experiments, with 46 (23 with dyslexia) and 62 participants (31 with dyslexia), respectively. All the participants had to read a set of texts with the presence or the absence of graphical schemes and keywords. Readability and comprehensibility were analyzed via eye tracking and comprehension tests, respectively. Via questionnaires we gathered the participants' subjective ratings of readability and comprehensibility, as well as memorability.

13.3.1 Design

In this section we explain the methodology that is specific to these two experiments. The rest of the methodological details are found in Chapter 3. Refer to Table 13.1 for a summary of the *Graphical Schemes* experiment and Table 13.2 for a summary of the *Keywords* experiment.

Independent Variables

- **Graphical Schemes:** In our experimental design, [±Schemes] served as an *independent variable* with two levels: [+Schemes] and [−Schemes]. The condition [+Schemes] denotes when the text was presented with a graphical scheme in the top, and [−Schemes] denotes the condition when the text was presented without a graphical scheme.

- **Keywords:** [±Keywords] served as an *independent variable* with two levels: [+Keywords] and [−Keywords]. The condition [+Keywords] denotes when main ideas of the text were highlighted in boldface while [−Keywords] denotes the case when the presentation of the text was not modified.

We used a within-subject design, that is, each participant read all the texts, contributing to each of the conditions. To avoid sequence effects, we counter-balanced texts as much as possible.

Dependent Variables

For quantifying readability and comprehension we used the *dependent measures* presented in Tables 13.1 and 13.2. For quantifying objective readability we used *Reading Time* and *Fixation Duration* extracted from the eye tracking data. For quantifying the text comprehension of the texts we used a *Comprehension Score*. To measure the readers subjective perception towards the text we used the *Readability Rating*, the *Comprehensibility Rating*, and the *Subjective Memorability Rating*.

The *Subjective Memorability Rating* is specific of these two experiments. We decided to include it because both strategies and recommended for learning [164, 405] and memorability is a crucial skill in learning processes. The rest of the dependent measures are shared with other experiments. We extensively explained them in Section 3.1.1.

13.3.2 Participants

The details of the participants groups for both experiments are given in Tables 13.1 and 13.2. In both experiments three of the participants with dyslexia were also diagnosed with attention deficit disorder. For more details about the participants please refer to Section 3.1.2. Regarding the reading habits, we believe the participants interpreted reading time as schooling time and that is why the numbers may seem exaggerated.

13.3.3 Materials

Base Texts

For both experiments we picked similar texts from the Spanish corpus Simplext [47]. For each text we matched the readability of the texts by making sure that the parameters commonly used to compute readability had the same or similar values [115].

(a) The texts are written in the same genre (news), and

(b) are about similar topics (culture). In *Graphical Schemes* one text is about the discovery of a supernova and the other text is about the discovery of a new species of fish. In *Keywords* one text was about the opening of a new library with electronic books and the other text is about the work by Picasso;

(c) they have the same number of words (136 words in *Graphical Schemes* and in 158 words in *Keywords*), and

(d) have a similar word length average, 5.06 and 5.12 letters in *Graphical Schemes* and 4.83 and 5.61 letters in *Keywords*;

(e) they are accessible news, readable for the general public so they contained no rare or technical words, which present an extra difficulty for people with dyslexia;

(f) they contain the same number of proper names (one per text in *Keywords*); and the same number of unique named entities (seven in *Graphical Schemes*);

(g) they have the same number of sentences (five per experiment) and similar sentence complexity (three sentences per text contain relative clauses);

(h) each of the texts in *Graphical Schemes* contained one foreign word and one numerical expression. In *Keywords* experiment one text has two numerical expressions and the other has two foreign words [100];

(i) for the *Keywords*, the texts have the same number of highlighted key-phrases.

An example of a text used (see Appendix A.8 for the original in Spanish) is given in Figure 13.1.

Graphical Schemes

For the creation of the graphical schemes[2] we took into account the pedagogical recommendations for dyslexics [299, 75], and the cognitive principles of inductive learning in concept acquisition from scheme theory [11, 12]. Since the texts were going to be read by dyslexics, the graphical schemes were manually created by a dyslexic adult and supervised by a psychologist. The graphical schemes simplify the discourse and highlight the most important information from the title and the content. Each of the graphical schemes shares the following pattern: the first line of the graphical scheme encloses the main words of the title connected by arrows and then, starting from the title, there is a node for each of the sentences of the text. These nodes summarize the most relevant information of the text, as the example translated into English shows in Figure 13.1. We present the original graphical scheme in Spanish in the Appendix A.8.

Keywords

For creating the keywords we highlighted using boldface the words which contained the main semantic meaning (focus) of the sentence. This focus normally corresponds with the direct object and contains the new and most relevant information of the sentence [366]. We

[2]Notice that we distinguish graphical schemes from conceptual graphs [364] or semantic maps [359].

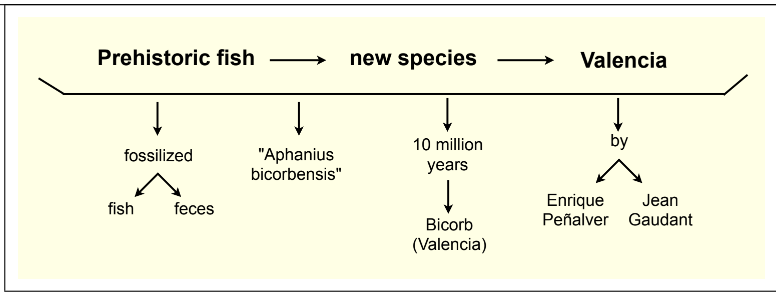


Figure 13.1: Example of a graphical scheme (*Fish*, in English).

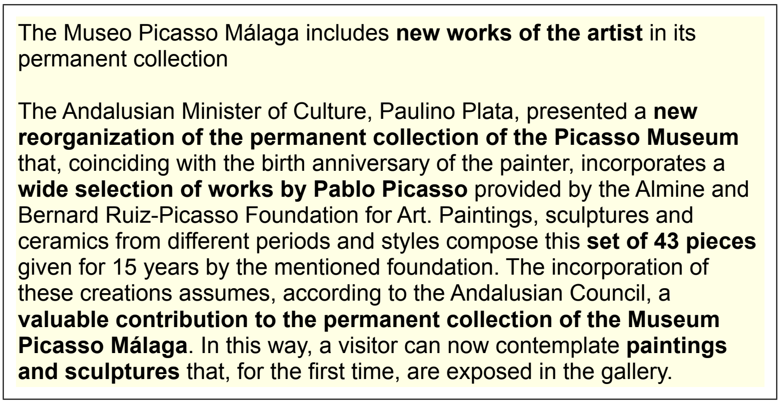


Figure 13.2: Example slide used in the *Keywords* experiment.

only focused on the main sentences; subordinate or relative clauses were dismissed. For the syntactic analysis of the sentences we used Connexor's Machinese Syntax [92], a statistical syntactic parser that employs a functional dependency grammar [375]. We took direct objects parsed by Connexor without correcting the output. An example of a text used (the original in Spanish is in Appendix A.9) is given in Figure 13.2.

Text Presentation

We used the parameters detailed in Section 3.1.3 with black font on creme background and 20 points font size.

Comprehension Questionnaires

We used three inferential items for each of the texts (Tables 13.1 and 13.2). In *Graphical Schemes* each of the items had three an-

El texto habla: *'The text is:'*

☐ Sobre la obra del pintor y escultor Picasso.
'About the work of the painter and sculptor Picasso.'
☐ Sobre la Fundación Almine y Bernard Ruiz-Picasso para el Arte.
'About the Almine and Bernard Ruiz-Picasso Foundation for Arts.'
☐ Sobre incorporación de nuevas obras en el museo Picasso de Málaga.
'About the incorporation of new works in the Picasso Museum of Malaga.'

Figure 13.3: Comprehension questionnaire for *Keywords*.

The presence of keywords in the text helped me to memorize the text:

☐ Strongly disagree.
☐ Disagree.
☐ Neither agree nor disagree.
☐ Agree.
☐ Strongly agree.

Figure 13.4: Memorability questionnaire item.

swers, a correct one, another partially incorrect (normally containing details), and one incorrect. In *Keywords* we also used multiple-choice questions with three possible choices, one correct, and two wrong. See Section 3.1.3 for details about the creation of the comprehension questionnaires. An example question is given in Figure 13.3.

Subjective Questionnaires

To quantify *Readability Rating*, *Comprehensibility Rating*, *Memorability Rating* and *Preferences Rating* we used questionnaires. The participants rated how much did the graphical schemes or keywords helped their reading, their ease to remember the text, and to which extent would they like to find keywords in texts. We used a five-point Likert scale, for more details see Section 3.1.3. Figure 13.4 gives one example scale.

13.4 Results

None of the objective readability data sets (*Fixation Duration* and *Reading Time*) of *Keywords* were normally distributed (Shapiro-Wilk test) and neither of them had a homogeneous variance (Bartlett test). One of the data sets in *Graphical Schemes* had a normal distribution (*Fixation Duration* and *Reading Time* for group C). However only the *Comprehension Scores* presented a homogeneous variance (Bartlett test). We used the Wilcoxon non-parametric test for the analyses of the Likert scales.

13.4.1 Graphical Schemes

Reading Time. In Table 13.3 we present the means, standard deviations and medians of *Reading Time* for both groups.

- **Between Groups:** We found a significant difference between the groups regarding *Reading Time* ($V = 867,0$ $p < 0.001$). Participants with dyslexia had longer reading times than participants without dyslexia.

- **Within Groups:** We did not find a significant effect of [±Schemes] on *Reading Time* for the participants with dyslexia ($V = 176,\ p = 0.113$) nor for the participants without dyslexia ($V = 136,0$ $p = 0.775$).

Fixation Duration. In Table 13.3 we present the means, standard deviations and medians of *Fixation Duration* for both groups.

- **Between Groups:** We found a significant difference between the groups regarding *Fixation Duration* ($V = 680,\ p = 0.004$). Participants with dyslexia had longer fixations than participants from the control group.

- **Within Groups:** Similarly, there were no significant effects of [±Schemes] on *Fixation Duration* for the participants with dyslexia ($V = 158,\ p = 0.144$) or without dyslexia ($V = 127.5,\ p = 0.987$).

  Notice that these positive results are given for the comparison of the texts alone. If we compare the total reading duration of the text alone with the text plus the graphical scheme, it takes in average 18.6% more time to read the whole slide than the text alone.

Comprehension Score. In Table 13.3 we present the means, standard deviations and medians of *Comprehension Score* for both groups.

- **Between Groups:** We did not find any difference between the groups regarding *Comprehension Score* ($V = 18,\ p = 1$).

- **Within Groups:** For the participants with dyslexia, we did not found any significant effect of [±Schemes] on the *Comprehension Score* ($V = 9,\ p = 0.186$). Similarly, for the control group we did not find an effect of [±Schemes] on the *Comprehension Score* ($V = 15,\ p = 0.374$).

Readability Rating. We found significant effects between the groups regarding how much graphical schemes helped them reading the text ($V = 406,\ p = 0.001$). People with dyslexia found that graphical schemes help more to read the text ($\tilde{x} = 4$, $\bar{x} = 3.35, s = 1.15$) than participants without dyslexia ($\tilde{x} = 2$, $\bar{x} = 2.13, s = 1.40$).

Comprehensibility Rating. We found significant effects between the groups regarding how much graphical schemes helped them to understand the text ($V = 375,\ p = 0.011$). People with dyslexia found that graphical schemes help more to understand the text ($\tilde{x} = 4$, $\bar{x} = 3.39, s = 0.94$) than participants without dyslexia ($\tilde{x} = 3$, $\bar{x} = 2.65, s = 0.83$).

Memorability Rating. We found no significant differences between the groups regarding if graphical schemes help to memorize the text ($V = 276,\ p = 0.802$). Both agree that graphical schemes help them to remember the text moderately for the participants with dyslexia ($\tilde{x} = 4$, $\bar{x} = 3.35, s = 1.27$) and for the control group ($\tilde{x} = 4$, $\bar{x} = 3.26, s = 1.21$).

Preference Rating. We significant effects between the groups regarding their preferences in finding graphical schemes in the texts ($V = 398,\ p = 0.002$). People with dyslexia significantly preferred to find text with graphical schemes ($\tilde{x} = 4$, $\bar{x} = 3.52, s = 0.95$) than participants without dyslexia ($\tilde{x} = 3$, $\bar{x} = 2.60, s = 0.84$).

13.4.2 Keywords

Reading Time. In Table 13.4 we present the means, standard deviations and medians of *Reading Time* for both groups.

- **Between Groups:** We found a significant difference between the groups regarding *Reading Time* ($V = 2578.5,\ p < 0.001$). Participants with dyslexia had longer reading times than participants from the control group.

- **Within Groups:** We did not find a significant effect of [±Keywords] on *Reading Time* for the participants with dyslexia ($V = 210,\ p = 0.688$) and for the participants without dyslexia ($V = 702.5,\ p = 0.351$).

Fixation Duration. In Table 13.4 we present the means, standard deviations and medians of *Fixation Duration* for both groups.

- **Between Groups:** We found a significant difference between the groups regarding *Fixation Duration* ($V = 2953,\ p < 0.001$). Participants with dyslexia had longer fixations than participants from the control group.

- **Within Groups:** Similarly, there were found no significant effects of [±Keywords] on *Fixation Duration* for the participants with dyslexia ($V = 259.5,\ p = 0.688$) or without dyslexia ($V = 862,\ p = 0.552$).

Comprehension Score. In Table 13.4 we present the means, standard deviations and medians of *Comprehension Score* for both groups.

- **Between Groups:** We found a significant difference between the groups regarding *Comprehension Score* ($V = 1544,\ p = 0.040$). Participants with dyslexia had lower comprehension scores than participants without dyslexia.

- **Within Groups:** For the participants with dyslexia, we found a significant effect of [±Keywords] on the *Comprehension Score* ($V = 178.5,\ p = 0.022$). Text with highlighted keywords led to significantly higher comprehension scores in this target group. For the control group we did not find an effect of [±Keywords] on the *Comprehension Score* ($V = 740,\ p = 0.155$).

Readability Rating. We found no significant differences between the groups regarding how much highlighting keywords helped them reading the text ($V = 504.5,\ p = 0.316$). Both groups found that keywords can slightly help their reading for the participants with dyslexia ($\tilde{x} = 3,\ \bar{x} = 3.02, s = 1.16$), and for the control group ($\tilde{x} = 3,\ \bar{x} = 2.80, s = 0.97$).

Comprehensibility Rating. We found no significant differences between the groups regarding if highlighting keywords help to understand the text ($V = 484,\ p = 0.493$). Both agree that keywords did not either help of impede them to understand the text for the participants with dyslexia ($\tilde{x} = 2,\ \bar{x} = 2.36, s = 1.00$), and for the control group ($\tilde{x} = 2,\ \bar{x} = 2.55, s = 1.08$).

Memorability Rating. We found no significant differences between the groups regarding if highlighting keywords help to memorize the text ($V = 484,\ p = 0.493$). Both agree that keywords help them to remember the text moderately for the participants with dyslexia ($\tilde{x} = 4,\ \bar{x} = 3.64, s = 1.00$), and for the control group ($\tilde{x} = 4,\ \bar{x} = 3.45, s = 1.09$).

Preference Rating. Also, no differences between groups were found regarding their preferences in finding highlighted keywords in the texts ($V = 46,\ p = 0.73$). Participants with dyslexia would like to find texts including highlighted keywords ($\tilde{x} = 4,\ \bar{x} = 3.64, s = 1.14$), as well as in the control group ($\tilde{x} = 4,\ \bar{x} = 3.60, s = 1.06$).

13.5 Discussion

We found differences between the people with and without dyslexia consistently with eye tracking studies [3, 122, 126, 208, 226].

Since *Graphical Schemes* had no significant effect on objective readability and comprehensibility, we cannot confirm **H1.1** *The presence of graphical schemes in the text increases objective readability for people with dyslexia*, nor **H1.2** *The presence of graphical schemes in the text increases objective comprehensibility for people with dyslexia.*

These results are consistent with some of the opinions that the participants expressed after the session. A few dyslexic participants explained that the graphical scheme actually distracted them from the text content. Another dyslexic participant exposed that the graphical schemes helped her to remember and study texts but not to understand them. The diverse opinions of the participants towards the graphical schemes suggest that normally graphical schemes are highly customized by the person that creates them and therefore a non-customized schema could complicate comprehensibility.

Graphical Schemes had a significant effect on subjective readability and comprehensibility accepting hypothesis **H1.3** and H1.4 . People with dyslexia found that graphical schemes helped them to read and understand the text more than people without dyslexia. Their preferences towards the presence of graphical scheme in the text were also consistent with these subjective perceptions.

This can be explained by the fact that people with dyslexia tend to be strong visual thinkers. Tasks such as visual analysis and pattern recognition are crucial for people with dyslexia to deal with their language difficulties [409].

On the other hand, we reject hypothesis **H1.5:** *The presence of graphical schemes in the text increases subjective memorability for people with dyslexia*, because the presence of graphical schemes did not have an effect on the subjective memorability of the participants.

Shorter reading times and fixation durations are associated with better readability [186]. Since *Keywords* had no significant effect on readability, we cannot confirm **H2.1**: *The presence of highlighted keywords in the text increases readability for people with dyslexia.*

One possible reason for this is that text presentation might only have an impact on readability when the whole text is modified, not only portions of it. Most probably, if one text was presented all in boldface or italics and the other one in roman, significant differences could have been found as in [319], where the effect of different font styles was evaluated. Another explanation could be that the text might look different to what the participants were used to see and participants might need some time to get used to highlighted keywords in the text before testing readability effects.

From the content point of view, the fact that readability did not change is expected, since the content of the text is not modified in any of the conditions.

Because participants with dyslexia had a significantly increase in text comprehension with texts having highlighted keywords, our findings support **H2.2**: *The presence of highlighted keywords in the text increases comprehensibility for people with dyslexia.*

This improvement might be due to the possibility that keywords might help to remember the text better. This is consistent with the pedagogic literature that recommends this strategy for learning and retaining text content [407].

We reject hypotheses **H2.3, H2.4** and **H2.5**. The fact that using keywords for learning is a shared strategy for both groups [407], may explain that no significant differences between groups were found regarding their preference and perception of keywords on readability, comprehensibility, and memorability. Also, highlighted keywords in bold are found in general school books, not only in materials for people with dyslexia, [165], so both groups were familiar with the conditions.

Limitations of the Study

These studies have at least two limitations. First, we did not evaluate automatic strategies but ideal outputs, generated manually. The graphical schemes were manually created. Also the *Keywords* experiment was performed with a manually annotated data set. These annotations were based on the output of the Connexor parser. We have not found any evaluation of Connexor's accuracy when parsing syntactic constituents. Nevertheless, it has been observed that the accuracy for direct objects in Spanish achieves results that varies from 85.7% to 93.1%, depending on the test set [263]. Second, the participants read only two texts because we did not wanted to fatigue participants with dyslexia.

Our results shall be taken with care since readability, especially in people with dyslexia, depends on many factors that are very challenging to control in an experimental setup. These factors include the vocabulary of the participants, their working memory or the different strategies they use to overcome dyslexia.

Graphical Schemes Experiment		
Design	Within-subjects	
Independent Variables	[±Schemes]	[+Schemes] [−Schemes]
Dependent Variables (Sec. 3.1.1)	*Reading Time*	(objective readability)
	Fixation Duration	
	Comprehension Score	(objective comprehensibility)
	Readability Rating	(subjective readability)
	Comprehensibility Rating	(subjective comprehensibility)
	Memorability Rating	(subjective memorability)
	Preference Rating	(subjective preferences)
Participants (Sec. 3.1.2)	Group D (23 participants)	12 female, 10 male **Age:** range from 13 to 37 ($\bar{x} = 20.74$, $s = 8.20$) **Education:** high school (11), university (10), no higher education (2) **Reading:** more 8 than hours (3), 4-8 hours (9), less than 4 hours/day (11)
	Group C (23 participants)	13 female, 9 male **Age:** range from 13 to 35 ($\bar{x} = 20.91$, $s = 6.96$) **Education:** high school (9), university (13), no higher education (1) **Reading:** more than 8 hours (1), 4-8 hours (9), less than 4 hours/day (13)
Materials (Sec. 3.1.3)	Base Texts	2 texts
	Graphical Schemes	2 graphical schemes (1 per text)
	Text Presentation	
	Comprehension Quest.	6 inferential items (3 per text)
	Sub. Readability Quest.	2 Likert scales (1/condition level)
	Sub. Comprehension Quest.	2 Likert scales (1/condition level)
	Sub. Memorability Quest.	2 Likert scales (1/condition level)
	Preferences Quest.	1 item (1 item/condition)
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750	
Procedure (Sec. 3.1.5)	**Steps:** Instructions, demographic questionnaire, reading task (×2) comprehension questionnaires (×2), subjective readability quest. (×2) subjective comprehension questionnaire (×2), and subjective memorability questionnaire (×2).	

Table 13.1: Methodology for the *Graphical Schemes* experiment.

Keywords Experiment		
Design	Within-subjects	
Independent Variables	[±Keywords]	[+Keywords] [−Keywords]
Dependent Variables (Sec. 3.1.1)	*Reading Time*	(objective readability)
	Fixation Duration	
	Comprehension Score	(objective comprehensibility)
	Readability Rating	(subjective readability)
	Comprehensibility Rating	(subjective comprehensibility)
	Memorability Rating	(subjective memorability)
	Preference Rating	(subjective preferences)
Participants (Sec. 3.1.2)	Group D (31 participants)	12 female, 10 male **Age:** range from 13 to 37 ($\bar{x}$ = 21.09, s = 8.18) **Education:** high school (14), university (15), no higher education (2) **Reading:** more 8 than hours (3), 4-8 hours (16), less than 4 hours/day (12)
	Group C (31 participants)	13 female, 9 male **Age:** range from 13 to 40 ($\bar{x}$ = 23.03, s = 7.10) **Education:** high school (11), university (19), no higher education (1) **Reading:** more 8 than hours (1), 4-8 hours (19), less than 4 hours/day (11)
Materials (Sec. 3.1.3)	Base Texts	2 texts
	Graphical Schemes	10 keywords (5 per text)
	Text Presentation	
	Comprehension Quest.	6 inferential items (3 per text)
	Sub. Readability Quest.	2 Likert scales (1/condition level)
	Sub. Comprehension Quest.	2 Likert scales (1/condition level)
	Sub. Memorability Quest.	2 Likert scales (1/condition level)
	Preferences Quest.	1 item (1 item/condition)
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750	
Procedure (Sec. 3.1.5)	**Steps:** Instructions, demographic questionnaire, reading task (×2) comprehension questionnaires (×2), subjective readability quest. (×2) subjective comprehension questionnaire (×2), and subjective memorability questionnaire (×2).	

Table 13.2: Methodological summary for the *Keywords* experiment.

Graphical Schemes	**Group D**		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Reading Time			
[−Schemes]	64.14	76.49 ± 41.06	43.97	45.12 ± 13.35
[+Schemes]	58.96	61.85 ± 19.28	41.88	43.77 ± 14.79
	Fixation Duration			
[−Schemes]	0.24	0.25 ± 0.06	0.21	0.21 ± 0.04
[+Schemes]	0.23	0.23 ± 0.05	0.21	0.21 ± 0.03
	Comprehension Score			
[−Schemes]	100	97.73 ± 10.66	100	95.45 ± 21.32
[+Schemes]	100	86.36 ± 35.13	100	88.64 ± 21.45

Table 13.3: Results for the *Graphical Schemes* experiment.

Keywords	**Group D**		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Reading Time			
[−Keywords]	46.70	53.71 ± 18.42	31.82	33.81 ± 12.82
[+Keywords]	0.43	59.98 ± 25.32	35.44	36.31 ± 15.17
	Fixation Duration			
[−Keywords]	0.23	0.23 ± 0.06	0.20	0.19 ± 0.04
[+Keywords]	0.22	0.22 ± 0.06	0.19	0.18 ± 0.04
	Comprehension Score			
[−Keywords]	100	77.27 ± 42.89	100	94.87 ± 22.35
[+Keywords]	100	100 ± 0	100	100 ± 0

Table 13.4: Results for the *Keywords* experiment.

Part V

INTEGRATION

We have already addressed which text presentations (Part III) and content modifications (Part IV) benefit the reading performance of people with dyslexia. These findings can have great impact on interactive systems that rely on text as the main information medium, such as browsers, PDF viewers, or eBook readers. By applying certain text alterations, these systems could make texts easier to read and understand for people with dyslexia. However, all the existing applications at the moment only modify the design of the text, but not its content.

In this part we first test (Chapter 14) two different interaction strategies to find out how lexical simplification can be helpful for people with dyslexia. Out of the two strategies tested –substitution *vs.* showing synonyms on-demand–, we found that presenting simpler synonyms when the user desires has a better impact than substituting such synonyms. Given the positive results, in Chapter 15 we implemented an algorithm for simpler synonyms generation using the Google Books Ngram Corpus for word sense disambiguation. We evaluated positively the algorithm with two groups, people with dyslexia and strong readers. The output of our lexical simplification algorithm is a resource containing lists of synonyms ranked by their complexity in Spanish, to be further integrated in other systems.

Finally, in Chapter 16 we present a combination of the results of the thesis. These results are summarizes in a set of recommendations for modifying the text presentation and content to make it more accessible for people with dyslexia. Finally, we present four reading applications where the model *DysWebxia* has been successfully integrated.

Chapter 14

Showing Synonyms in a Helpful Way

14.1 Introduction

Previous findings presented in Chapter 10 have shown that the reading performance of people with dyslexia increases when the text has more frequent and shorter words. 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 [79, 157] modify only the text presentation but not its content.

With this idea in mind, in this chapter we used an automatic lexical simplification system, LexSiS [45], to study the impact on readability and comprehensibility of two different 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 (laptop, tablet, smartphone). To the best of our knowledge, this is the first time that an automatic lexical simplification system is evaluated for end-users with dyslexia, combining eye tracking, questionnaires, and the use of different devices. In addition, this is the largest user study of its kind. This chapter presents the following main contributions:

– A new strategy, ShowSyns, which adapts LexSiS, and allows users to interactively request simpler synonyms for complex words.

– Participants with dyslexia found that texts presented with the new strategy were significantly more readable and comprehensible while participants without dyslexia found it significantly more comprehensible.

The findings of this chapter were presented in Rello *et al.* [321].

14.2 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 [228, 109], all the existing applications at the moment only alter the design of the text [157, 79, 188], 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 (NLP) 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.

14.2.1 Lexical Simplification

Automatic text simplification is an NLP task that transforms a text into an equivalent that is easier to read than the original, preserving the original meaning. Text simplification started as a task for making other NLP tasks easier. However, the task of simplifying a text also has a high potential to help people with reading comprehension problems [6, 74].

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, 74, 112] and word length [28] as a measure of lexical complexity. To find appropriate word substitutions they use different resources such as WordNet [74], dictionaries [112], word sense disambiguation [107], thesaurus and lexical ontologies [6], and synonym dictionaries [28].

14.2.2 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 words' processing time [305], and because people with dyslexia specifically encounter problems with less frequent words and long words [174, 322, 339, 358]. 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.* [174] show that low frequency and long words present longer gaze durations and more re-inspections in both groups. Also in Chapter 10 we found that frequent words improve readability and short words improve comprehensibility for people with dyslexia. Also, Rüsseler *et al.* [339] show that it takes more time to recognize infrequent words and this recognition performance is lower in readers with dyslexia. Simmons and Singleton [358] measured comprehension of people with dyslexia who performed significantly poorer on inferential questions.

14.2.3 Accessibility

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 [4], ten participants tested Web navigation using semi-structured interviews. In [199], 27 participants did assignments after reading texts with different presentations. In [96], interviews, questionnaires, log sheets and focus groups are used to explore user behavior and usability issues related to the use of web-based resources by people with disabilities (nine participants with dyslexia); while in [372] six participants performed tasks in a website to explore its design. Hence, our number of participants is much larger.

14.2.4 What is Missing?

An evaluation of different interaction strategies derived from an automatic lexical simplification system for end-users with dyslexia, combining eye tracking, questionnaires, and the use of different devices.

14.3 LexSiS: Simplification Strategies

Here we evaluate two lexical simplification strategies based on the LexSiS algorithm [45]. LexSiS is the first system for the lexical simplification of Spanish text and is being developed in the context of the

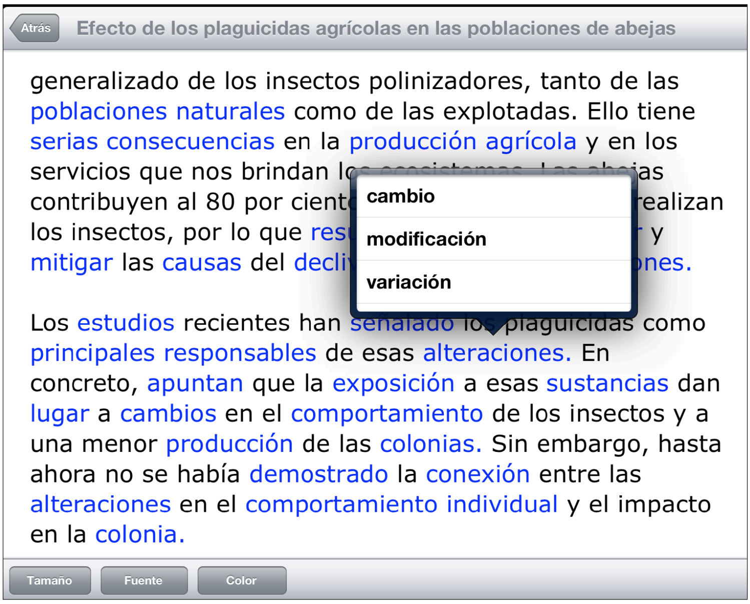


Figure 14.1: Condition ShowSyns mock-up for iPad.

Simplext project [341]. 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.[1]

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 on demand 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 14.1).

[1]Out of the synonyms that LexSiS generates, 65% are simpler than the target word [45].

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.[2] The following is an example of an entry in OpenThesaurus:

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

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 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 words corpus of Spanish Web news. LexSiS uses this model to construct vectors that 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

[2]There is no English version, see `http://openthes-es.berlios.de`.

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 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 [45]). 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.

14.4 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).

14.4.1 Design

In Table 14.1 we show a summary of the methodology of the *Lexical Simplification* experiment.

Independent Variables

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.

Due to the large length of the texts, 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.

Dependent Variables

For quantifying readability and comprehension we used the *dependent measures* presented in Table 14.1. For quantifying objective readability we used *Reading Time* and *Fixation Duration* extracted from the eye tracking data. For quantifying the text comprehension of the texts we used *Comprehension Score*. 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*, we counted the choices as 100%, 50%, and 0%, respectively. To measure the readers' subjective perception towards the text we used the *Readability Rating*, the *Comprehensibility Rating*, and a *Memorability Rating*.

14.4.2 Participants

The details of the participants groups for both experiments are given in Table 14.1. For more details about the participants please refer to Section 3.1.2.

14.4.3 Materials

To study the effects of the simplification strategies, we studied them as part of a text. The rationale behind this is that readability and comprehensibility pertain to longer segments of texts [173]. To isolate the effects of the different strategies, the texts need to be comparable in complexity.

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*. We chose these type of texts because they contain a considerable number of infrequent words, candidates to be simplified. We matched the readability of the texts by making sure that the parameters commonly used to compute readability yielded the same or similar values:

(a) Within each experiment, the texts use the same genre, scientific articles.

(b) 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*.

(c) They have the same number of words: 302 words.

(d) 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.

(e) They contain the same number of sentences: eleven.

(f) They do not contain acronyms or numerical expressions since numerical expressions are processed differently by people with and without dyslexia [329]. Both texts have the same foreign word (*Nature*).

(g) For the SubsBest strategy we made the same numbers of substitutions in both texts: 34.

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 14.2. Please see the Appendix A.10 for the complete lists of lexical simplifications performed for the *Bee* text.

Text Presentation

We used the parameters detailed in Section 3.1.3 with black font on creme background and an almost black font (10% grey scale) on white background. Each line did not exceeded 62 characters and the font size was 20 points.

Comprehension Questionnaires

Each of the questionnaires was composed of three multiple-choice inferential questions (see example in Figure 14.2). We made sure that the questions did not include a synonym that may benefit a particular strategy.

Subjective Questionnaires

To quantify the *Readability*, *Comprehensibility* and *Memorability* ratings we used questionnaires. The participants rated how much did the strategies helped their reading and their ease to remember the text, using a five-point Likert scale. For more details see Section 3.1.3.

El texto trata sobre: *'The text is about:'*

– Un artículo científico sobre el origen de la luz infrarroja.
' 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 su luz infrarroja.
'On the first galaxies in the universe and their infrared light'.

Figure 14.2: Inferential question example.

14.4.4 Equipment

The eye tracker used was a Tobii T50, which has a 17-inch TFT monitor with a resolution of 1024×768 pixels. For more details about the eye tracker measurements and conditions please see Section 3.1.4.

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, Chrome and Safari, for the first two devices and Firefox 16.0.2 for the laptop.

14.4.5 Procedure

This experiment has an specific step in the procedure. After completing the demographic questionnaire, to ensure 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, 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.

14.5 Results

In this section we present the analysis 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 C; for [SubsBest] we had 16 samples for group D and 17 for group C; for [ShowSyns] we had 14 samples for group D and 14 for group C; and for [Gold] we had 15 samples for group D and 14 for group C.

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

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

– One-way analysis of variance (ANOVA) to show effects of the conditions on reading time, fixation duration, and comprehension score within groups.

– Student's independent two tailed t-test to show effects on reading time, fixation duration, and comprehension score among groups D and N.

– 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).

Reading Time. In Table 14.3 and Figure 14.3 we show the averages of the reading times.

- **Between Groups:** Considering all the conditions, we found a significant difference between the groups regarding reading time

($t(67.66) = 4.42,\ p < 0.001$). Participants with dyslexia had significantly longer reading times ($\mu = 132.08$, $s = 51.17$ seconds) than the participants without dyslexia ($\bar{x} = 95.25$, $s = 26.02$ seconds).

- **Within Groups:** We did not find a significant effect of any of the conditions on reading time in group D ($F(2, 44) = 0.18,\ p = 0.841$) or in group C ($F(2, 43) = 2.25,\ p = 0.117$). Also, there was a strong positive correlation between groups ($r = 0.625$). Both groups read faster under the same condition, [Gold].

Fixation Duration. In Table 14.3 and Figure 14.3 we present the average of fixation durations.

- **Between Groups:** Pooling the data together for all the conditions, there was a significant difference between the groups fixation duration ($t(77.16) = 4.08,\ p < 0.001$). Participants with dyslexia had significantly longer fixation times ($\bar{x} = 0.24$, $s = 0.05$ seconds) than the participants without dyslexia ($\bar{x} = 0.20$, $s = 0.03$ seconds).

- **Within Groups:** We did not find a significant effect of any of the conditions on fixation time in group D ($F(2, 44) = 0.06,\ p = 0.94$) or in group C ($F(2, 43) = 0.10,\ p = 0.904$). Again, there was strong positive correlation between groups ($r = 0.994$). See Table 14.3 and Figure 14.3 for the average of fixation durations.

Comprehension Score. In Table 14.3 and Figure 14.4 we present the average the comprehension score.

- **Between Groups:** Considering all the conditions, participants with dyslexia answered less questions correctly ($\bar{x} = 54.5\%$, $s = 45.0\%$) than participants without dyslexia ($\bar{x} = 59.9\%$, $s = 45.9\%$). However, the difference between the groups was not statistically significant ($t(389.36) = -1.18,\ p = 0.239$).

- **Within Groups:** We did not find a significant effect of text simplification on the comprehension score in group D ($F(3, 186) = 0.74,\ p = 0.529$) or in group C ($F(1, 198) = 1.16,\ p = 0.325$). In

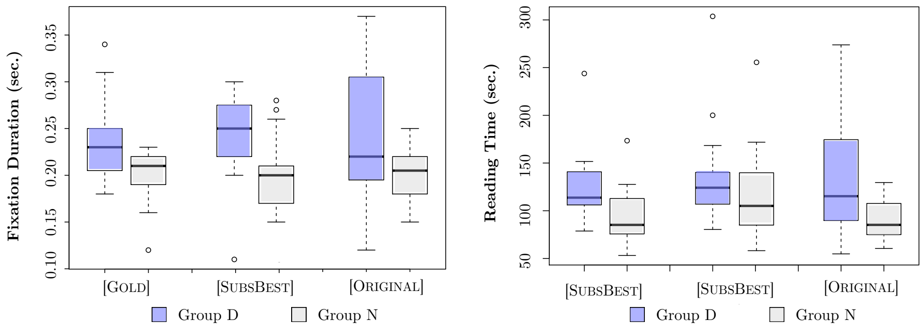


Figure 14.3: *Reading Time* and *Fixation Duration* means in seconds.

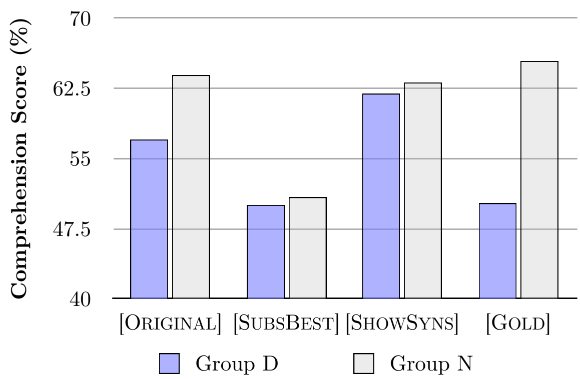


Figure 14.4: *Comprehension Score* means.

this case there was a mild positive correlation between groups ($r = 0.429$). See Table 14.3 and Figure 14.4 for the averages of the comprehension scores.

Readability Rating. In Figure 14.5 we show the histograms of the easiness ratings and in Table 14.4 we show their averages.

- **Between Groups:** There was a small correlation between both groups on the readability ($r = 0.241$).

- **Within Groups:** For the participants with dyslexia, we found a significant effect of the simplification strategy on readability

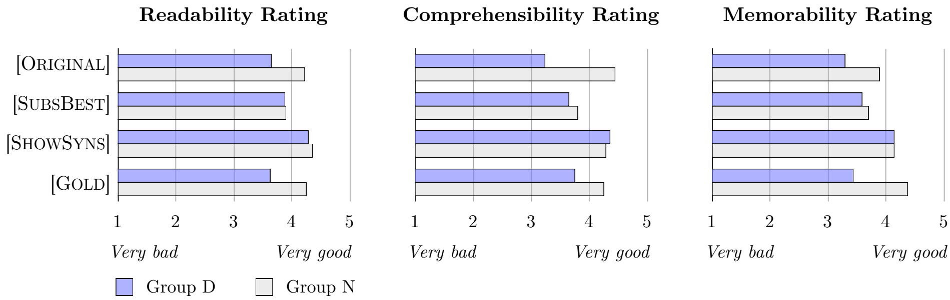


Figure 14.5: Readability, comprehensibility, and memorability ratings means.

ratings ($H(3) = 8.28,\ p = 0.041$). Pairwise comparisons showed that the participants found [ShowSyns] significantly easier to read than [Gold] ($p = 0.034$) and [Orig] ($p = 0.015$).

Comprehensibility Rating. See Figure 14.5 and Table 14.4.

- **Between Groups:** There was no correlation between both groups on the ratings about the comprehensibility of the text ($r = -0.085$).

- **Within Groups:** For group D we found a significant effect of the simplification strategy on comprehensibility ratings ($H(3) = 12.20,\ p = 0.007$). 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 comprehensibility ratings ($H(3) = 9.60,\ p = 0.022$). Pairwise comparisons showed that the participants found [SubsBest] significantly more difficult to understand than [Orig] ($p = 0.003$), [ShowSyns] ($p = 0.047$) and [Gold] ($p = 0.049$).

Memorability Rating. See Figure 14.5 and Table 14.4.

- **Between Groups:** There was a small correlation between both groups on the ease of remembering the text ($r = 0.160$).

- **Within Groups:** For group C we found a significant effect of the simplification strategy on memorability ratings ($H(3) = 9.02$, $p = 0.029$). Pairwise comparisons showed that in the [SubsBest] condition, the participants found texts significantly more difficult to remember than in the [Gold] condition ($p = 0.003$).

14.5.1 Correlations

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 C ($r = 0.928$). The options with a higher comprehension score, [ShowSyns] and [Orig], were also perceived as more comprehensible by both groups. For readability in group D we found a medium 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 C, we found strong negative correlations between the easiness rating for readability and reading time ($r = -0.999$) and fixation duration ($r = -0.554$). 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.

14.6 Discussion

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

Differences between 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 [122]. However, our comprehension results are not consistent with [358] because our participants with dyslexia did not have a significantly poorer understanding of the texts using inferential items.

Objective 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 [174, 322] 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 case 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 [322].

Another possible explanation is that only a relatively small percentage of the words in the text were 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 focused on target words [174]. We analyzed the eye fixation duration and the reading time of the whole text and not target words only as in Hyona *et al.* [174] because we aim to measure text readability and the readability is related to longer text segments [173].

Objective 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 are obtained with [ShowSyns] for group D and [Orig] for group C, that is, options that do not include any lexical substitution 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 [173], 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').*

Subjective 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 that took them longer to read and where the comprehension was poorer.

Lexical Simplification Experiment		
Design	Between-subjects	
Independent Variable	*Lexical Simplification Strategy*	[Orig] [SubsBest] [ShowSyns] [Gold]
Dependent Variables (Sec. 3.1.1)	*Reading Time*	(objective readability)
	Fixation Duration	
	Comprehension Score	(objective comprehensibility)
	Readability Rating	(subjective readability)
	Comprehensibility Rating	(subjective comprehensibility)
	Memorability Rating	(subjective memorability)
Participants (Sec. 3.1.2)	Group D (47 participants)	28 female, 19 male **Age:** range from 13 to 50 ($\bar{x} = 24.36$, $s = 10.19$) **Education:** high school (18), university (26), no higher education (3)
	Group C (49 participants)	29 female, 20 male **Age:** range from 13 to 40 ($\bar{x} = 28.24$, $s = 7.24$) **Education:** high school (16), university (31), no higher education (2)
Materials (Sec. 3.1.3)	Base Texts	2 texts
	Word Substitutions	34 per text (in [SubsBest]), and 40/44 per text (in [Gold])
	Synonyms on-demand	100/110 synonyms for 50/55 words per text (in [ShowSyns])
	Text Presentation	
	Comprehension Quest.	6 inferential items (3 per text)
	Sub. Readability Quest.	2 Likert scales (1/condition level)
	Sub. Comprehension Quest.	2 Likert scales (1/condition level)
	Sub. Memorability Quest.	2 Likert scales (1/condition level)
Equipment (Sec. 3.1.4)	Eye tracker Tobii 1750, Samsung Galaxy Ace S5830, iPad 2, and MacBook Air	
Procedure (Sec. 3.1.5)	**Steps:** Instructions, demographic questionnaire, text choosing, reading task, comprehension questionnaires, subjective readability quest., subjective comprehension quest., and subjective memorability quest.	

Table 14.1: Methodology for the *Lexical Simplification* experiment.

Orig	SubsBest	ShowSyns	Gold
alteración	*cambio*	*cambio, modificación, variación*	*cambio*
'alteration'	*'change'*	*'change, modification, variation'*	*'change'*

Table 14.2: Examples of lexical simplifications.

Lexical Simplification	**Group D**		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Reading Time			
[Orig]	108.74	134.79 ± 63.03	85.20	90.24 ± 20.33
[SubsBest]	124.13	135.77 ± 53.65	105.97	105.77 ± 32.12
[Gold]	113.73	125.86 ± 37.16	85.17	89.08 ± 21.44
	Fixation Duration			
[Orig]	0.22	0.24 ± 0.07	0.21	0.20 ± 0.03
[SubsBest]	0.25	0.24 ± 0.04	0.20	0.20 ± 0.05
[Gold]	0.23	0.24 ± 0.04	0.21	0.20 ± 0.03
	Comprehension Score			
[Orig]	50	57.00 ± 47.39	100	63.89 ± 45.99
[SubsBest]	50	50.00 ± 45.83	50	50.83 ± 47.38
[ShowSyns]	100	61.90 ± 43.91	50	63.10 ± 42.85
[Gold]	50	50.19 ± 42.76	100	65.39 ± 45.54

Table 14.3: Objective measures for the *Lexical Simplification* experiment.

Lexical Simplification	**Group D**		**Group C**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Readability Rating			
[ORIG]	4	3.65 ± 0.61	4	3.24 ± 0.83
[SUBSBEST]	4	3.88 ± 0.49	4	3.65 ± 0.70
[SHOWSYNS]	4	4.29 ± 0.73	4	4.36 ± 0.74
[GOLD]	4	3.63 ± 0.89	4	3.75 ± 1.06
	Comprehensibility Rating			
[ORIG]	3	3.29 ± 0.92	4	4.22 ± 0.65
[SUBSBEST]	4	3.59 ± 0.51	4	3.90 ± 0.72
[SHOWSYNS]	4	4.14 ± 0.77	4	4.36 ± 0.74
[GOLD]	4	3.44 ± 1.03	4	4.25 ± 0.68
	Memorability Rating			
[ORIG]	3	4.44 ± 0.51	4	3.89 ± 0.90
[SUBSBEST]	4	3.80 ± 0.62	4	3.70 ± 0.47
[SHOWSYNS]	4	4.29 ± 0.73	4	4.14 ± 0.77
[GOLD]	4	4.25 ± 0.68	4.5	4.38 ± 0.72

Table 14.4: Subjective measures for the *Lexical Simplification* experiment.

Chapter 15

CASSA: Algorithm and Resource

15.1 Introduction

Regarding the content of the text, previous studies have shown that simpler words lead to a better reading performance. For instance, Hyona *et al.* [174] used eye tracking to show that low frequency and long words present longer gaze durations and more re-inspections. Also in Chapter 10 we found that frequent words improve readability and short words improve comprehensibility for people with dyslexia. Also, Rüsseler *et al.* [339] 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, in Chapter 14 we tested an automatic lexical simplification algorithm for Spanish [45]. There, 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.

Once we know how to present synonyms to people with dyslexia in a helpful way (Chapter 14), we created a synonym generation algorithm for that purpose. Although the new algorithm used for this unique feature is language independent, our first prototype is for Spanish. The goal of this chapter is the evaluation of the quality of the synonyms on demand that included 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 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, which selects the most frequent synonyms of the most common sense. Our results show that the quality of the synonyms generated by the new algorithm outperforms the frequency based baseline. The main contributions are:

– A new algorithm called CASSA for automatic generation of synonyms for people with dyslexia, which output outperforms the frequency based baseline.

– A resource of synonyms called CASSA that contain lists of synonyms ranked by their complexity in Spanish.

The analysis of this chapter was presented in Rello and Baeza-Yates [320].

15.2 Related Work

Related contributions to our work can be found in natural language processing (NLP) literature about lexical simplification, and resources for lexical simplification in Spanish.

15.2.1 Generation of Simpler Synonyms

Lexical simplification is a kind of text simplification (see Section 14.2.1) which aims at the word level. It could 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, many approaches use WordNet [59, 74, 202]. De Belder *et al.* [107] apply explicit word sense disambiguation with a latent words language model. Devlin and Unthank [112] use dictionaries. Aluisio and Gasperin [6] use a thesaurus and lexical ontologies. Bott *et al.* [45] 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.* [421] used 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.* [40] also relied 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 simplest [59, 74, 112, 202], with the exception of [45] that used word frequency and length. In many studies of lexical simplification [40, 45, 112, 421] an algorithm based on the most frequent synonym is a very hard to beat baseline for simpler synonyms generation. For instance, in a shared task for English lexical simplification [365], only one system out of nine outperformed the frequency baseline.

The closest algorithm to ours is LexSiS by Bott *et al.* [45] that presents a lexical simplification algorithm for Spanish and also uses the Spanish OpenThesaurus (we used LexSiS to compare different simplification methods in Chapter 14). 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 this case, we specifically targeted people with dyslexia. LexSiS 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 [45]. For that reason we devised an improved algorithm for this task, CASSA.

15.2.2 Resources for Spanish

One of the current main limitations is the lack of resources for lexical simplification in Spanish.

For instance, there is no Simple Wikipedia in Spanish, while there is Simple English Wikipedia [95] which had lead to new approaches for lexical simplification in English [40, 421].

To the best of our knowledge, the existing resources previously used for lexical simplification in Spanish are the following.

The Simplext Corpus [47] is used by the first lexical simplification system for Spanish [45]. This is a set of 200 news articles of which 40 have been manually simplified. The parallel part of this corpus contains 6,595 words of original and 3,912 words of simplified text. All texts have been annotated using Freeling, including part-of-speech tagging, named entity recognition and parsing [262].

Another language resource used for Spanish lexical simplification [47] is the Spanish OpenThesaurus (SpOT).[1] The SpOT is freely available under the GNU Lesser General Public License, to be used with OpenOffice.org. This thesaurus provides 21,378 target words (lemmas) with a total of 44,348 different word senses for them.

Some approaches to lexical simplification make use of WordNet [236] in order to measure the semantic similarity between lexical items and to find an appropriate substitute. Spanish is one of the languages represented in EuroWordNet [402] and this resource was also used for lexical simplification [343]. The Spanish part of EuroWordNet contains only 50,526 word meanings and 23,370 synsets, in comparison to 187,602 meanings and 94,515 synsets in the English WordNet 1.5.

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

While SpOT is freely available, EuroWordNet is not freely available.

15.2.3 What is Missing?

To the best of our knowledge, there is no existing resource like CASSA containing lists of synonyms ranked by their complexity. Also, there is no evaluation of the quality of the synonyms generated by an automatic lexical simplification algorithm by people with dyslexia.

15.3 The CASSA Algorithm

CASSA (Context Aware Synonym Simplification Algorithm) is a method that generates simpler synonyms of a word. Words can be polysemic,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.

15.3.1 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,[2] and (b) a large n-gram corpus with frequencies, where we used Google Books Ngram Corpus [234]. Next we detail these two resources:

– **Spanish OpenThesaurus** (version 2): See description in Section 15.2.2. The following is the thesaurus entry for *mono*, which is ambiguous, as it could mean *'ape'*, *'overall'*, or the adjective *'cute'*.

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

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

- **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[3] 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 [214].

15.3.2 Algorithm Description

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

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 unique 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. The other tokens are the context of the target word. A context is considered valid if all words, including the target word, consist only of lowercase alphabetic characters, to filter for proper names, and is not a stop word, using a standard list of stop words in Spanish (*e.g. and, of, at, etc.*).

In order to determine the possible lemmas of a target word, a lemmatizer is used. However, in Spanish, in many cases a word can refer to more than one possible lemma, as for instance the word *'sitio'* in most cases refer to the very frequent noun (*'site'*), but could also be an inflection of the less common verb *'sitiar'* (*'besiege'*). To avoid the cases where a verb is wrongly assigned as the lemma for a word that is actually noun, we use the following rules. If the target word is the same as the lemma, the lemma is kept. However, if the lemma differs from the target word and the preceding word is a determiner (by using

[3] `http://books.google.com/ngrams`

a standard list of determiners in Spanish), then the target word should be a noun. If now the last letter of the assigned lemma is an *'r'*, we assume that the lemma is in fact a verb and discard the lemma for this word (in Spanish all verb infinitives end in *'r'*). If still several lemmas for one word remain, later the context is used for disambiguating the lemma.

The lemmatized token is included in the Synonyms List as a target word only if it appears in our List of Senses. 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, *noche* and *fortuna* in the examples below:

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

`de probar` *fortuna* `en el` *('to try fortune in the')*

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 to be more challenging for people with dyslexia, leading to worse reading performance [174, 322, 339]. 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.

That is, we use a parameter k such that if a word is k or more times less frequent than one or more of its synonyms, is considered a complex word. We used k as the default threshold because worked well in practice (27% of the words have simpler synonyms in this way).

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 n-gram corresponding to that (word,context) pair. That is, to disambiguate the sense, CASSA uses the context where the target word appears. If the context is not found, CASSA uses the most frequent sense.

124	fortuna	de probar ~ en el	[sino,estrella,destino]	fortuna
	fortune	*to try ~ in the*	*[destiny,fate,luck]*	*fortune*
48	fortuna	la mala ~ de cruzarse	[sino,estrella,destino]	fortuna
	fortune	*the mis~ of crossing*	*[destiny,fate,luck]*	*fortune*
49	fortuna	desigualdad de ~ en el	[recursos,medios,capital]	fortuna
	fortune	*inequality of ~ in the*	*[resources,means,capital]*	*fortune*
64	fortunas	acumularon las ~ de las	[recursos,medios,capital]	fortuna
	fortune	*accumulated the ~ of the*	*[resources,means,capital]*	*fortune*
56	fortunas	de las ~ amasadas por	[recursos,medios,capital]	fortuna
	fortune	*of the ~ amassed by*	*[resources,means,capital]*	*fortune*
52	fortuna	dueño de ~ y de	[capital,dinero,patrimonio]	fortuna
	fortune	*owner of ~ and of*	*[capital,money,heritage]*	*fortune*
63	fortuna	fue una ~ para el	[gracia,favor,suerte]	fortuna
	fortune	*it was a ~ for him*	*[grace,favor,luck]*	*fortune*
60	fortuna	golpe de ~ que le	[gracia,suerte,dicha]	fortuna
	fortune	*a stroke of ~*	*[grace,luck,bliss]*	*fortune*
46	fortuna	la inmensa ~ de haber	[regalo,paz,suerte]	fortuna
	fortune	*extremely ~ to have*	*[gift,peace,luck]*	*fortune*

Figure 15.1: Example lines extracted from the CASSA resource.

952414	fortuna	[recursos,medios,capital]	fortuna
	fortune	*[resources,means,capital]*	*fortune*
100797	fortunas	[recursos,medios,capital]	fortuna
	fortunes	*[resources,means,capital]*	*fortune*

Figure 15.2: Example lines extracted from Simple CASSA.

15.4 CASSA Resource for Spanish

CASSA resource is a list of 41,106 complex words with their corresponding synonyms (ranging from one to three) depending on the sense of the word. The resource is freely available online[4]

15.4.1 Description

The CASSA resource is composed of two files:

[4] `www.luzrello.com/resources.html` and `http://grupoweb.upf.edu/WRG/CASSA.txt`.

CASSA: This file includes 41,106 complex words in Spanish and 4,229,868 lists of synonyms in context (senses) generated using CASSA with $k = 10$. Depending on the sense of the complex words a different list of synonyms is provided. Each line is composed by four elements: **frequency, complex word, context, simpler synonyms** and **lemma**. Following there is an example of a line and its translation:

```
48,fortuna,golpe de que le, [gracia,suerte,dicha], fortuna
```

Frequency: It is the first element of the lines in the resource –*e.g.* 48–, and it corresponds to the **frequency** of the **complex** in the n-grams with the inflected word –*e. g. fortuna ('fortune')*, second element– in its **context**, *e.g. golpe de que le ('stroke of that his')*, third element.

Complex word: The inflected complex word in the same form it is found in the n-grams.

Context: The context of the inflected complex word, the first, second, fourth and fifth elements of the ngram.

Simpler synonyms: The three most frequent synonyms of the disambiguated senses of the complex word, *e.g. gracia, suerte, dicha, ('grace, luck, happiness'*),

Lemma: The lemma of the complex word, *e.g. fortuna, ('fortune'*).

In Figure 15.1 we find different simpler synonyms for the word *fortuna ('fortune').* For instance in *golpe de fortuna que le ('a stroke of fortune')*, *fortuna* means *'luck' ('grace, luck, happiness')* while in *dueño de fortuna y de ('owner of fortunes and')*, *fortuna* means *'money' ('capital, money, patrimony').*

Simple CASSA: This file includes 40,825 words of simpler synonyms for complex words in Spanish and 135,577 simpler synonyms lists (senses).

However, it is worth mentioning that both resources, Simple CASSA and CASSA, have the same number of different target words

Resource	**CASSA**	**Simple CASSA**
Complex Words	41,106	40,825
Contexts	1,817,069	–
Complex Lemmas	9,928	9,732
Simpler Synonyms	9,345	7,562

Table 15.1: Number of complex inflected words, complex lemmas, and simpler synonyms of CASSA resources.

(43,996). CASSA has more subgroups of synonyms than Simple CASSA depending on the context. In this case the different senses of the complex words are not taken into consideration. Each line is composed by six elements: **frequency, complex word, simpler synonyms** and **lemma**.

```
952414, fortuna, [recursos,medios,capital], fortuna
```

Frequency: The absolute frequency of the complex word in the Web.

Complex word: The inflected complex word in the same form it is found in the n-grams.

Simpler synonyms: The three most frequent synonyms taken within all the senses where the complex word appear in SpOT.

Lemma: The lemma of the complex word.

The number of complex inflected words, complex lemmas and simpler synonyms of the CASSA disambiguated resource and Simple CASSA can be found in Table 15.1.

15.4.2 Coverage

To check the coverage of the synonyms resource, we created a corpus made of 196 classic literature books from the 15th century to the 20th century. We included the books that are compulsory readings for secondary and high school in Spain. All the book titles used for this corpus are given in Appendix A.11.

Case	$k = 10$	$k = 5$	$k = 2$	No k
Complex words	27.16	38.80	54.24	100.00
Simple CASSA (abs.)	24.07	35.32	50.14	84.43
Simple CASSA (rel.)	88.62	91.03	92.04	84.43
Complex contexts	27.95	40.03	55.84	100.00
CASSA (abs.)a	2.67	4.14	6.44	12.14
CASSA (rel.)	9.55	10.34	11.53	12.14

Table 15.2: Coverage of CASSA and Simple CASSA.

This corpus is composed by 16,495,885 tokens and 5,886,366 lexical words (without stop words, proper names and punctuation marks).

The coverage of SpOT in our corpus is 88.34%. This is the maximum that any simplification algorithm that uses SpOT as resource can obtain. In Table 15.2 we present the coverage of Simple CASSA and disambiguated CASSA depending on the threshold k used to decide what is a complex word and hence a complex content, including the absolute percentage as well as the relative percentage with respect to the complex words or contexts.

For smaller k, the coverage of the baseline increases significantly being the maximum possible 84.43% when all words are considered complex (more than three times the default coverage). On the other hand, CASSA does not increase much the coverage as that is limited by the context coverage reaching a maximum of 12.14%, only 27% more than the default case ($k = 10$). This maximum, compared with the baseline is a bit more than 14% of the cases, implying that CASSA is similar to the frequency baseline around 85% of the time.

15.5 Methodology

To evaluate the synonym quality generated by CASSA we conducted an experiment with 32 participants with dyslexia. Using online questionnaires, each participant had to read and rate a set of synonyms generated by CASSA and the baseline.

15.5.1 Design

Independent Variables

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

- [FREQUENCY]: As a baseline we used the Simple CASSA resource explained in 15.4. We showed the top-3 most frequent synonyms in the selected sense. As we mentioned in Section 15.2.1, this baseline is hard to outperform and in our case has been improved by lemmatization.

- [CASSA]: This method is explained in Section 15.3.

- [HIGH] and [LOW] complex word frequency: We divided into two groups the complex words for which synonyms were created. [LOW], that includes very low frequency complex words, and (b) [HIGH], that contains high frequency complex words. The frequency of the [LOW] group ranges from 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.

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.

Dependent Variables

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 level 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 grades are given by using a ten-point rating.

15.5.2 Participants

The details of the participants who volunteered to take the online test are given in Table 15.3. In this experiment as control group we recruited a group of strong readers as a control group to verify that the learning disability does not affect the results of [Cassa]. We consider them strong readers because they all finished post-compulsory schooling[5] and were frequent readers.

15.5.3 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 [223], and second, the comprehension of the text pertain to longer segments, not only words [173]. Following, we describe how we designed the materials that were used in this study.

Evaluation Dataset

We have two evaluation data sets derived from [Cassa] and [Frequency], respectively. Each data set is composed of:

– **Target Words**: We selected 40 target words, 20 [Low] and 20 [High] which are intended to be complex words for each evaluation data set, so 80 target words in total. After defining the frequency ranges for [Low] and 20 [High], we randomly extracted the candidates for target words and selected only polysemic words that have different senses, like *fortuna* (which can

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

mean *'luck'* or *'treasure'*). Monosemous complex words[6] such as infrequent nouns were discarded because for those cases [CASSA] and [FREQUENCY] present the same synonyms.

– **Synonyms**: For each target word there is a set of simple 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.

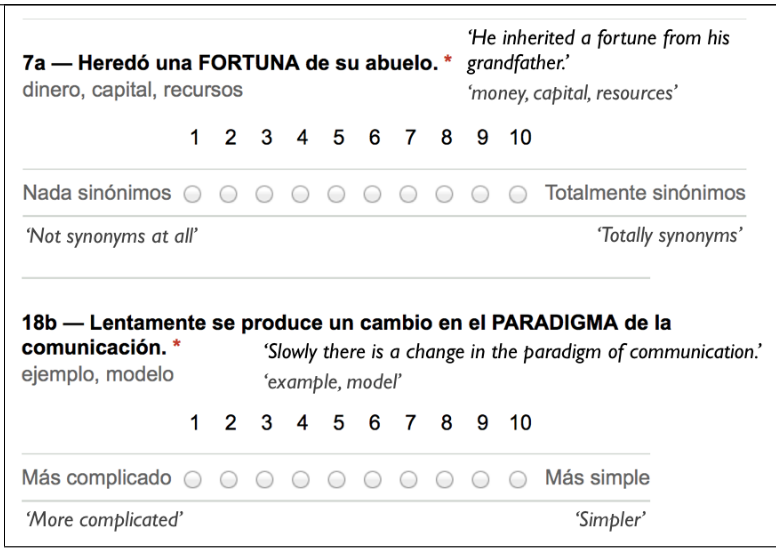
Test

The evaluation data set was integrated in an 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 10-point Likert scales items to rate the *Synonymy* and the *Simplicity* of the set of synonyms in comparison with the target word (see Figure 15.3). There were 160 items and the conditions were counter balanced.

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 judgments were similar between participants. These items were done manually, containing perfect synonyms or antonyms of the target word, and uniformly interspersed in the test.

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 data sets but they were not rated by the target group but by two or three annotators. Yatskar *et al.* [421] used six annotators (three native, three non-native speakers of English) that rated 200 simplification examples in English while Biran *et al.* [40] used 130 examples that were judged by three

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



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 15.3: Two items extracted from the online test.

annotators (native English speakers). In Bott *et al.* [45], three annotators (native speakers of Spanish) rated 69 sentences for each Spanish lexical simplification algorithm evaluated.

15.5.4 Procedure

Depending on the participant the test lasted from 30 to 50 minutes. Eleven participants performed the test at the Madrid for Dyslexia Association[7] supervised by the author of this thesis. The rest of the participants undertook the test at their homes online. In that case the author of this thesis was connected 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. Second, they began with a questionnaire that was designed to collect demographic information. Third, they started the test and rated the first 90 items, followed by a small break and then 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

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

the synonyms affected their reading.

15.6 Results

Now we present the analysis of the data from the tests. First, we checked the validation/calibration items. All participants answered these items 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).

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.

15.6.1 Synonymy

Between Groups

– Regarding the candidates generated by 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 ($\tilde{x} = 5$, $\bar{x} = 6.21, s = 3.16$), ($V = 1,740,194$, $p = 0.013$).

– 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 ($\tilde{x} = 8$, $\bar{x} = 7.02, s = 3.13$), ($V = 430\,363$, $p = 0.029$). 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 ($\tilde{x} = 5$, $\bar{x} = 5.40, s = 2.97$), ($V = 432,193$, $p = 0.047$).

[CASSA] **vs.** [FREQUENCY]

– **Group D:** 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 ($\tilde{x} = 8$, $\bar{x} = 7.02, s = 3.13$) than candidates generated by [FREQUENCY] ($\tilde{x} = 5$, $\bar{x} = 5.40, s = 2.97$).

– **Strong Readers:** 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 of Complex Words**

– **Group D:** 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 ($\tilde{x} = 6$, $\bar{x} = 5.72, s = 3.25$). In Table 15.4 we show the results for the subgroups.

– **Strong Readers:** 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 15.4 we show the results for the subgroups.

15.6.2 Simplicity

Between Groups

– 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]

- **Group D:** 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] ($\tilde{x} = 6$, $\bar{x} = 5.62, s = 2.91$).

- **Strong Readers:** 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 of Complex Words**

- **Group D:** 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 ($\tilde{x} = 6$, $\bar{x} = 5.99, s = 3.19$). In Table 15.4 we show the results for all the subgroups.

- **Strong Readers:** 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 15.4 we show the results for all the subgroups.

15.7 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 [40, 45, 112, 421]. 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 [40, 45, 321], 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 and any comparison shall be taken with care.

Notice that we specifically tested the examples where [Frequency] and [Cassa] gave different synonyms candidates for the complex word, as 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 [45], 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.

Limitations of the Study

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 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 (Chapter 4), or the frequency of the word has a larger effect on reading difficulty for people with dyslexia than for people without dyslexia (Chapter 10).

CASSA Experiment		
Design	Within-subjects	
Independent Variables	*Lexical Simplification Strategy*	[FREQUENCY] [CASSA]
	Word Frequency	[LOW] [HIGH]
Dependent Variables (Sec. 15.5.1)	*Synonymy Rating*	(Synonymy)
	Simplicity Rating	(Simplicity)
Participants (Sec. 15.5.1)	Group D (32 participants)	18 female, 14 male **Age:** range from 6 to 52 ($\bar{x} = 23.15$, $s = 12.07$) **Bilingual:** Catalan (12), English (1) French (1) **Education:** primary school (9), secondary school (7), further studies (16) **Reading:** more 5 than hours (1), 3-5 hours (11), less than 2 hours/day (20)
	Strong Readers (38 participants)	18 female, 14 male **Age:** range from 17 to 69 ($\bar{x} = 38.16$, $s = 13.06$) **Bilingual:** Catalan (11), Galician (5) English (5), French (3), German (2) **Education:** further studies (38) **Reading:** more 5 than hours (14), 3-5 hours (10), less than 2 hours/day (14)
Materials (Sec. 15.5.1)	Evaluation Dataset	40 complex target words (20 [LOW] and 20 [HIGH]) 80 lists of synonyms (40 by [FREQUENCY] and 40 by [CASSA]) 40 sentences
	Online Test	180 items: -80 synonymy items -80 simplicity items -20 validation/calibration items
Procedure (Sec. 15.5.1)	**Steps:** Instructions, questions & answers, demographic questionnaire, first part of the online test, break, second part of the online test, semi-structured interview.	

Table 15.3: Methodology for the CASSA experiment.

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

Table 15.4: Median, mean, and standard deviation for *Synonymy* and *Simplicity* ratings for [CASSA] and [FREQUENCY] for [LOW] and [HIGH] frequency complex words.

Chapter 16

Recommendations and Applications

16.1 Introduction

In this chapter we take all the results of the previous chapters and integrate them in a model called *DysWebxia* which has been integrated in several tools. The model aims to make text more accessible to people with dyslexia and it combines the following innovative contributions:

– The recommendations of the model are grounded in quantitative and qualitative results from our research with people with dyslexia that measures the impact on readability and comprehension of different text alteration strategies using eye tracking.

– This is the first model that integrates both, text alterations regarding the presentation as well as the content of the text.

– Although *DysWebxia* was first implemented for reading text on the Web, the model has been adapted for other platforms where there was no similar reading software for people with dyslexia. So far, four tools include the model: *DysWebxia Reader*, *IDEAL eBook reader*, *Text4All*, and *AccessibleNews DysWebxia*.

The demos of the four applications were presented in ASSETS' 2012 [188], ASSETS' 2013 [325], W4A' 2012 [331], and W4A' 2013 [326].

16.2 Text Presentation

In Table 16.1, we present a set of recommendations for formatting screen text in a more accessible way for people with dyslexia. We have considered both kind of data, the quantitative data (objective readability and comprehensibility) given by the eye-tracker and the user ratings and preferences (subjective readability, comprehensibility and preferences). The order in Table 16.1 is based on the impact of each parameter in objective readability and comprehensibility. The first parameters are the ones that lead to greater differences. Please notice that the results for each of the parameters are independent from each other and no interactions between them where tested in our experiments except from font size and line spacing. Since the user preferences might change with time [21], we gave priority to the objective readability data in the recommendations.

These are only recommendations, the combination of the parameters and the final customization of the text are left to the user. Dyslexia varies from person to person and there are studies that have shown the benefits of self text customization [113, 158]. Furthermore, WebAIM [406], the British Dyslexia Association [52], as well as the Web Accessibility Initiative (WAI) [418] recommend text customization. In Figure 16.1, we show an example which uses our recommendations for font size, column width, and character, line and paragraph spacing (the text and background colors and paragraph was customized according the preferences of the user).

16.2.1 Font Type

In Chapter 9 we compared 12 fonts using eye tracking. Based on the results we recommend fonts and font styles presented in Table 16.1.

For people with dyslexia the fonts that significantly lead to better objective readability were: *Courier*, *Helvetica*, *Arial* (shorter fixation durations), *CMU*, and *Helvetica* (shorter reading times). For people without dyslexia the fonts that lead to better objective readability were: *Arial*, *CMU* (shorter reading times), *Courier*, and *Verdana* (shorter fixation durations). The fonts that people with dyslexia and without dyslexia significantly preferred were the same: *Verdana*, *Helvetica*, and *Arial*. Non-italics fonts (*roman* fonts) lead to better reading

Condition	Measures	Values with positive effects with Dyslexia	Values with positive effects without Dyslexia
Font Type	Obj. Readability	*Arial, Courier CMU, Helvetica*	*Arial, Courier CMU, Helvetica*
	Preferences	*Verdana, Helvetica Arial*	*Verdana, Helvetica Arial*
	Recommendation: Arial, Courier, CMU, Helvetica & Verdana		
	Obj. Readability	*roman* *sans serif* *monospaced*	*roman* *sans serif* *monospaced*
	Preferences	*roman* *sans serif* no effects	*roman* no effects *proportional*
	Recommendation: roman, sans serif, and monospaced		
Font Size	Obj. Readability	18, 22, and 26 points	18, 22, and 26 points
	Obj. Comprehensibility	18, 22, and 26 points	14, 18, 22, and 26 points
	Subj. Readability	18 and 22 points	18 and 22 points
	Subj. Comprehensibility	18, 22, and 26 points	14, 18, 22, and 26 points
	Recommendation: ranging from 18 to 22 points		
Character Spacing	Obj. Readability	+7% and +14%	
		no effects	no effects
	Preferences	no effects	0%
	No recommendations or ranging from +7% to +14%		
Line Spacing	Obj. Readability	no effects	no effects
	Obj. Comprehensibility	0.8, 1, and 1.2 lines	no effects
	Subj. Readability	no effects	no effects
	Subj. Comprehensibility	no effects	1 line
	Recommendation: ranging from 1 to 1.5 lines		
Paragraph Spacing	Obj. Readability	no effects	no effects
	Preferences	no effects	no effects
	No recommendations		
Grey Scale (text)	Obj. Readability	no effects	no effects
	Preferences	0%	0%
Grey Scale (background)	Obj. Readability	no effects	no effects
	Preferences	0%	0%
	Recommendation: black font (0%) on white background, or white font on black background (0%)		
Colors (text/background)	Obj. Readability	no effects	no effects
	Preferences	no effects	no effects
	No recommendations		
Column Width	Obj. Readability	no effects	no effects
	Preferences	no effects	44 char./line
	No recommendations		

Table 16.1: *DysWebxia* text presentation recommendations.

performance to people with dyslexia (shorter fixations) and without dyslexia (shorter reading times). Consistently both groups significantly preferred font in *roman* than fonts in *Italic*. *Sans serif* fonts lead to better reading performance to people with dyslexia (shorter fixations) and without dyslexia (shorter fixations). Only participants

Once upon a time and a very good time it was there was a moocow coming down along the road and this moocow that was down along the road met a nicens little boy named baby tuckoo...

His father told him that story: his father looked at him through a glass: he had a hairy face.

He was baby tuckoo. The moocow came down the road where Betty Byrne lived: she sold lemon platt.

Figure 16.1: Text example using our dyslexic-friendly recommendations.

with dyslexia significantly preferred *sans serif* fonts. *Monospaced* fonts lead to better objective readability of people with dyslexia (shorter reading times and fixations) and without dyslexia (shorter fixations).

Suggestion for best practice: Sans serif, roman and monospaced fonts are good fonts for people with dyslexia specifically *Arial*, *Courier*, *Computer Modern Unicode (CMU)*, *Helvetica*, and *Verdana*.

16.2.2 Font Size

In Chapter 7 we presented a study comparing font sizes ranging from 10 to 26 points using raw text on the screen. Later, in Chapter 8 we tested with 132 participants (28 with dyslexia) the same range of sizes in the context of the Web (Wikipedia). The values which lead to a significant better reading performance are presented in Table 16.1.

For both groups, texts written with 18, 22 and 26 points were the fastest to read (Chapter 7). Also in Chapter 8 font sizes of 18, 22, and 26 points lead to a better objective readability (shorter fixations) than fonts of 10, 12, and 14 points, for both groups. Beyond 18 points there were no further improvements on reading speed. At the same time, fonts of 10 and 12 points lead to a lower objective comprehensibility of people with dyslexia as well as fonts of 10 point lead to significant lower comprehension for people without dyslexia. For people with and without dyslexia the highest rates of subjective readability were with 18 points followed by 22 points. For both groups subjective readability ratings were significantly lower for fonts of 10, 12, 14, and 26 points. Subjective comprehension is higher for the larger font sizes (18, 22, and 26 points) for people with dyslexia and for people without dyslexia (14, 18, 22, and 26 points).

Suggestion for best practice: Font sizes ranging from 18 to 22 points strikes the balance between having the best readability, comprehension, and subjective perception scores.

16.2.3 Line Spacing

In Chapter 7 we tested different line spacing (0.8, 1, 1.2, and 1.4 lines) and found no effects. However, line spacing was found to be strongly negatively correlated with reading performance: the narrower the space between the lines, the slower the participants read. Later, in Chapter 8 we measured the effect of line spacing (0.8, 1, 1.4, and 1.8 lines) and font size in combination, but only font size lead to significant differences on objective readability. On objective comprehensibility 1.8 spacing lead to worse compression scores for people with dyslexia, while no effects were found for people without dyslexia. Similarly, no effects were found on subjective readability for both groups and on subjective comprehensibility for people with dyslexia. People without dyslexia found texts with 1.0 line spacing significantly more comprehensible (see Table 16.1).

Suggestion for best practice: On the basis of our results regarding line spacing, we recommend to use the standard (1.0). However, moderately larger line spacing, such as the widely used 1.5 spacing, might be equally well to ensure readability and comprehension. Although no effects were found for narrower spacing, we do not recommend narrow line spacing that 1.0 because of the strong negatively correlation found for line spacing: the narrower the space between the lines, the slower the reading.

16.2.4 Character Spacing

In Chapter 7 we tested different character spacing (−7%, 0%, +7%, and +14%) and found that both groups presented a better readability (shorter fixations) with +7% and +14% values than with −7%. However the effect were found taking into consideration both groups, we could not find any effect within groups. Also people without dyslexia significantly preferred character spacing of 0% as shown in Table 16.1.

Suggestion for best practice: We recommend to use larger character spacing (+7% and +14%).

16.2.5 Paragraph Spacing

No effects were found on the objective readability and on the preferences for any of the paragraph spacing tested (0.5, 1, 2, and 3 lines) in Chapter 7. Participants with and without dyslexia preferred paragraphs spacing of 1 two 2 lines although these preferences did not yield statistical significance.

Suggestion for best practice: Since no effects were found we do not make any specific recommendation.

16.2.6 Text and Background Color

In Chapter 7 no effects were found on the objective readability and the preferences for any of the groups about any of the text and background colors tested. Some of the color pairs with good readability – no significant effects– were cream/black, yellow/blue, and light mucky green/dark brown.

Suggestion for best practice: Since no effects were found for people with dyslexia we do not make any specific recommendation.

16.2.7 Grey Scales for Text and Background

No effects were found on the objective readability for any of the grey scales tested on the text and background in Chapter 7.

For the different gray scales of the font –the background was always white–, participants found pure black font (0%) significantly more readable than text presented with different grey scales.

For the different gray scales of the background –the font was always white–, participants found pure black background (0%) significantly more readable than the other backgrounds presented with different grey scales.

Suggestion for best practice: On the basis of the subjective results, we recommend using either a black font on white background, or a white font on black background.

16.2.8 Column Width

No effects were found on the objective readability for any of the groups for the column widths tested (22, 44, 66 and 88 characters per line) in Chapter 7. Participants without dyslexia found the option of 66 characters per line significantly more readable. Participants with dyslexia preferred columns from 44 to 66 characters per line although these preferences did not yield significance.

Suggestion for best practice: Since no effects were found for people with dyslexia we do not make any specific recommendation.

16.3 Text Content

In Table 16.2 we present a set of recommendations for modifying screen text in a more accessible way for people with dyslexia. We have considered both kind of data, quantitative and qualitative data.

16.3.1 Word Frequency and Length

In Chapter 10 we studied how word frequency and length influences the objective readability and comprehension. For people with dyslexia more frequent words improved objective readability (shorter fixations and reading times) and shorter words improved their objective comprehensibly. No effects were found for people without dyslexia.

Suggestion for best practice: On the basis of the subjective results, we recommend to include in the text frequent and short words.

16.3.2 Numerical Expressions

In Chapter 11 we compared different numerical expressions representations: digits *vs.* letters, rounded *vs.* not rounded, percentages *vs.* fractions. Texts with numerical representations written in digits were

Condition	Measures	Values with positive effects for people with Dyslexia	without Dyslexia
Word Frequency	Obj. Readability	frequent words	no effects
	Obj. Comprehensibility	no effects	no effects
	Recommendation: frequent words		
Word Length	Obj. Readability	no effects	no effects
	Obj. Comprehensibility	short words	no effects
	Recommendation: short words		
Numerical Expressions	Obj. Readability	digits no effects percentages	digits no effects no effects
	Obj. Comprehensibility	no effects no effects no effects	no effects no effects no effects
	Subj. Readability	digits no effects percentages	digits no effects percentages
	Subj. Comprehensibility	digits no effects percentages	digits no effects percentages
	Recommendation: digits and percentages		
Verbal Paraphrases	Obj. Readability	no effects	no effects
	Obj. Comprehensibility	no effects	no effects
	No recommendations		
Keywords	Obj. Readability	no effects	no effects
	Obj. Comprehensibility	highlighted keywords	no effects
	Recommendation: highlighted keywords		
Graphical Schemes	Obj. Readability	no effects	no effects
	Obj. Comprehensibility	no effects	no effects
	No recommendations		
Lexical Simplification	Obj. Readability	no effects	no effects
	Obj. Comprehensibility	no effects	no effects
	Subj. Readability	synonyms on-demand	no effects no effects
	Subj. Comprehensibility	synonyms on-demand	synonyms on-demand
	Recommendation: synonyms on-demand		

Table 16.2: *DysWebxia* text content recommendations.

more readable for people with dyslexia (shorter fixations and reading times) and for people without dyslexia (shorter reading times). Texts with numerical representations written in percentages instead of fractions benefit the readability of people with dyslexia (shorter fixations and reading times). Subjectively, people with and without dyslexia found significantly more readable and comprehensible numbers written in digits and in percentages. No effects were found for the subjective and the objective measures regarding numerical rounding (Table 16.2).

Suggestion for best practice: On the basis of the results, we recommend presenting numerical expressions in digits and percentages instead of using letters and fractions, respectively.

16.3.3 Verbal Paraphrases

In Chapter 12 we studied how syntactic simplification via verbal paraphrases affected the objective readability and comprehension of people with and without dyslexia. No effects were found for any of the groups.

Suggestion for best practice: Since no effects were found we do not make any specific recommendation.

16.3.4 Keywords and Graphical Schemes

In Chapter 13 we studied how discourse simplification via highlighting keywords and including graphical schemes affected the objective and subjective readability and comprehension of people with and without dyslexia. The only effect found was on the objective comprehensibility of people with dyslexia which improved when the text contained highlighted keywords.

Suggestion for best practice: To highlight the main ideas in the text using boldface.

16.3.5 Lexical Simplification

In Chapter 14 we presented a study that tested different lexical simplification interaction strategies. The results showed that that performing automatic lexical simplifications (substituting complex words by a simpler synonyms) did not improve the objective readability nor the comprehensibility for any of the groups. However, when synonyms were presented on demand participants with dyslexia found texts easier to read and to understand. Similarly, participants without dyslexia found the texts easier to understand using synonyms on demand.

Suggestion for best practice: We recommend to show synonyms on demand since this strategy was found to be beneficial regarding the perception of the complexity of the text.

16.4 Applications

16.4.1 Motivation

There are three reasons motivating the decision to develop dyslexic-friendly readers or to integrate a dyslexic-friendly option in existing readers. First, the increasing growth of ebook usage in the last years, for instance, the Association of American Publishers reported that ebook sales increased by 115.8 percent in January 2011 [15]. Second, the fact that people with dyslexia represent a relatively large group of people, as it is estimated that from 10 to 17.5% of the U.S.A. population has some level of dyslexia [176] (see Chapter 1). Third, the use of accessibility practices for users with dyslexia is beneficial for all, since dyslexic-accessible practices can alleviate difficulties faced by all users including other users with disabilities (see Chapter 1).

According to Sections 16.2 and 16.3 there is a set of parameters which lead to significant benefits on reading for people with dyslexia. These parameters integrated in the *DysWebxia* model shall be included in the reading tools for people with dyslexia if we want to have more accessible text.

In this section we present the tools which have successfully integrated the *DysWebxia* model:

- The iOS reader *DysWebxia Reader* [1] for ebooks (Section 16.4.3).
- The Android *IDEAL eBook reader* [2] for ebooks (Section 16.4.4).
- The web service *Text4All* [3] for existing websites (Section 16.4.5).

[1] Freely available soon in the App Store under the name *DysWebxia Reader*.
[2] `https://play.google.com/store/apps/details?id=org.easyaccess.epubreader`
[3] `http://www.text4all.net/DysWebxia.html`

- The prototype web-based service *AccessibleNews DysWebxia* for displaying news (Section 16.4.6).

Since profiles of users with dyslexia are different, in all these implementations the settings are customizable so the user can override the settings according to their personal reading preferences. In addition, for the first two tools we performed a usability evaluation.

16.4.2 Related Tools

Among the mobile applications for users with dyslexia there are: spell checkers such as *American Wordspeller & Phonetic Dictionary* that converts phonetic spelling to proper spelling; applications that exploit speech recognition such as *Dragon Search* and *Dragon Dictate* to search and dictate email or messages; and software that uses text-to-speech for reading texts to people with dyslexia such as *Web Reader* or *CapturaTalk*. There are other applications for people with dyslexia that we did not consider because they are not reading software *per se* such as *ScreenRuler*,[4] an application that provides an overlay rule to support reading tasks, or *Colour Explorer*,[5] a software to change the colors in the *Windows* operating system. Related contributions to our work can be found in accessibility literature regarding reading tools for people with dyslexia. These are *Firefixia* [347], *SeeWord* [113, 157] and *MultiReader* [286]. A detailed description of these reading software can be found in Section 2.2.2.

16.4.3 DysWebxia Reader

In this section we present *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. We did a study performing a usability evaluation of *DysWebxia* with 12 participants. Our results show that *DysWebxia* is very usable, and we collected several ideas for future improvements. To the best of our

[4] http://www.clarosoftware.com/index.php?cPath=348

[5] http://colour-explorer.software.informer.com/9.0/

knowledge, there are no similar applications that offer an adapted layout to users with dyslexia when reading ebooks in iOS mobile devices. *DysWebxia* was demoed at ASSETS' 2013 [325] and we plan to offer it for free through the App Store in the near future.

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, keywords and numerical expressions.

For the user interface design, we first performed a competitive analysis of existing reading tools 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 mock-ups (see Table 16.3 for a comparison of different readers).

For the implementation we used the Apple iOS SDK, building an iOS 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 16.2.

Our previous research has 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 or showing synonyms on demand. Following, we present the features of the app:

(a) **Font Size**: Sizes ranging from 18 points to 26 points lead to faster reading (Chapters 7 and 8).

(b) **Font Type**: *Sans serif*, *roman*, and *monospaced* lead to a better reading performance for people with dyslexia, specifically, *Helvetica*, *Courier*, *Arial*, *Verdana*, and *Computer Modern Unicode* (Chapter 9).

(c) **Colors**: Good color pairs for readability –no significant effects– (background/font): cream/black, yellow/blue, light mucky green/dark brown, grey (25%) in the background with white font, and grey font (25%) with white background (Chapter 7).

(d) **Character Spacing**: Larger letter spacing which was found to lead to faster reading (Chapter 7).

(e) **Synonyms on Demand:** It includes a unique feature that shows synonyms on demand for complex words (see Figure 16.3 for an example). For this feature, we integrated the CASSA resource in the tool (Chapter 15).

Usability Evaluation

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

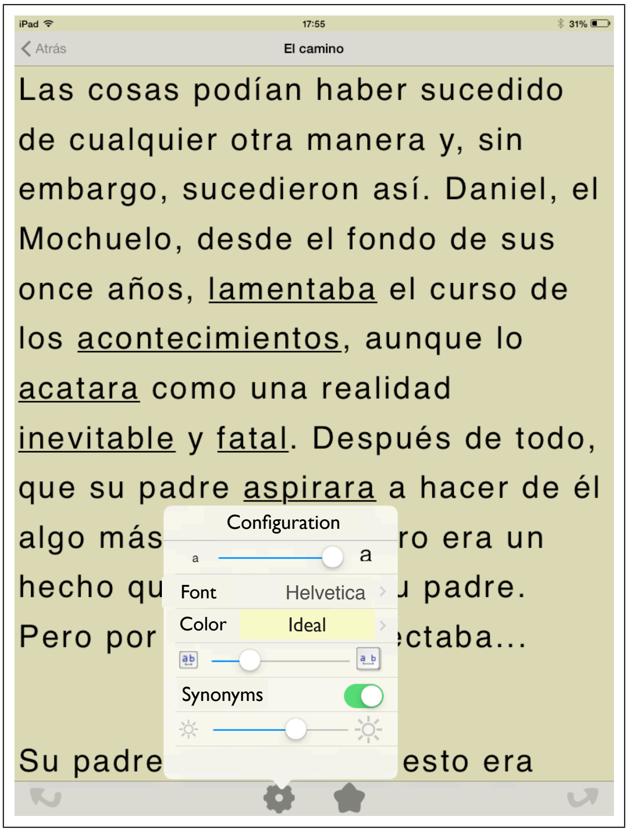


Figure 16.2: *DysWebxia Reader* customization interface.

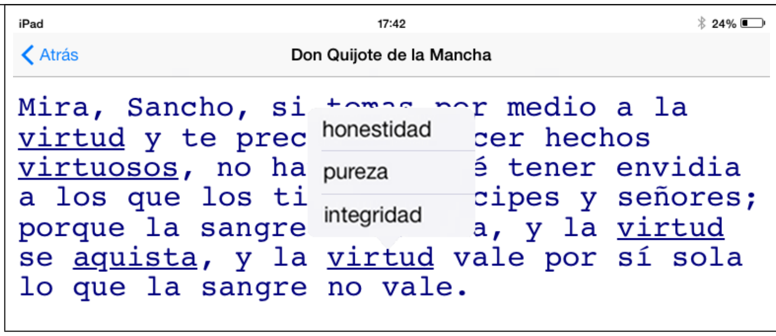


Figure 16.3: *DysWebxia Reader* example showing the synonyms *honesty, purity*, and *integrity* for the complex word *virtue* after the user clicks in it. Synonyms are available for all underlined words.

Design: In a within-subject design, all the participants had to perform some tasks with the tool using the think aloud protocol [211]. They had to choose and customize a text to later read it. They also undertook a questionnaire and a semi-structured interview.

Participants: We recruited 12 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 [253] 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.

Materials:

Texts: For the evaluation we used three texts processed by CASSA. The texts were fragments of compulsory readings[6] from

[6] 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

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 16.3).

Questionnaire: The items of the questionnaire were inspired by the WACG 2.0 [63] and usability principles [254]. 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. (See all the items used in the Appendix A.12)

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.

Device: iPad Mini.

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 author of this thesis. First, we began with a questionnaire that was designed to collect demographic information. Second, to ensure the engagement of the participant while reading, s/he chose the text to read from the bookshelf of the application (see Figure 16.4). 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 16.2). 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.

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

---

of the short story, by Gustavo Adolfo Béquer; and *El camino*, beginning of Chapter 1, by Miguel Delibes.

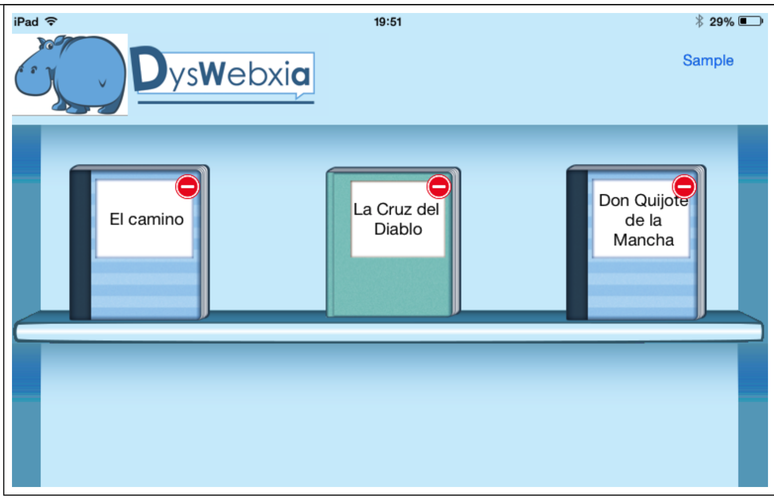


Figure 16.4: Bookshelf of *DysWebxia Reader*.

During the *customization task*, all the participants but one, decided to turn the synonyms option on, so complex words in the text were underlined. During the *reading* task only eight participants (67%) actually made use of the synonyms on demand option.

Only two participants changed the customization settings more than once. In Figure 16.5 we show two of the preferred settings combinations. 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 software. We believe that the positive attitude towards the tool have impacted the answers of the survey we present next.

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

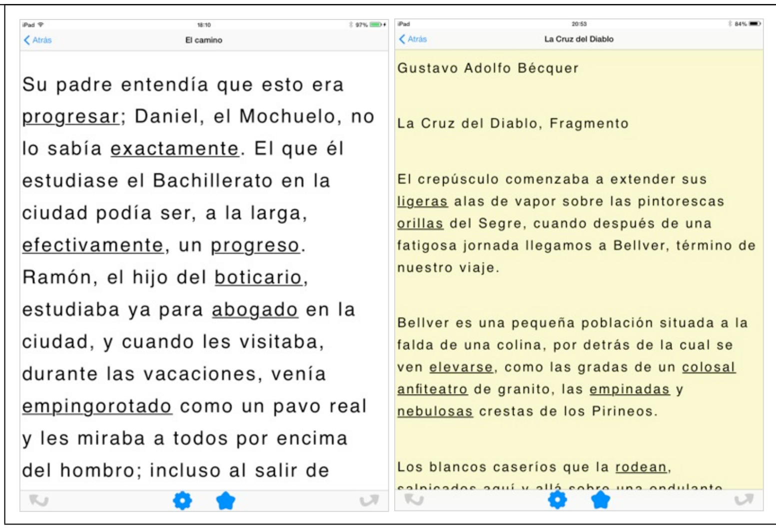


Figure 16.5: Preferred settings of two participants.

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 (see Figure 16.3), two participants found underlining words confusing and would have preferred to see the complex words in boldface or in a different color. Most of the participants found the option very helpful for reading ($\bar{x} = 6.42 \pm 0.79$). The main objection of the participants 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. Finally, one participant would have liked to remember the synonyms that he found more readable for future readings.

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 ($\bar{x} = 6.50 \pm 0.67$).

In summary, our participants found that the app was very usable and gave very good feedback to improve our next prototype. Based on the results, we believe that *DysWebxia* may have a large future impact for people with dyslexia.

16.4.4 IDEAL eBook Reader

In this section we present the *IDEAL eBook reader* for Android which displays ebooks in a more accessible manner for users with dyslexia. The ebook reader combines features that other related tools already have, such as text-to-speech technology, and new features, such as displaying the text with an adapted text layout based on the results of the user studies presented in this thesis (see a feature comparison in Table 16.3). Since there is no universal profile of a user with dyslexia, the layout settings are customizable and users can override the special default layout setting according to their reading preferences. We did a usability evaluation of *IDEAL eBook reader* with 14 participants to collect ideas that were subsequently included in the tool. To the best of our knowledge, there are no similar applications that offer an adapted layout to users with dyslexia when reading ebooks for Android mobile devices. The *IDEAL eBook reader* was demoed at ASSETS' 2012 [188] and it is freely available on Google Play.[7]

Description

The *IDEAL eBook reader* is an ebook reader for Android devices developed by Accessible Systems of India.[8] This application displays ebooks that have been formatted according to ePub, a free and open ebook standard by the International Digital Publishing Forum (IDPF). Epub is a globally adopted set of rules that define how an ebook should be constructed. When ebooks follow this standard, they can be displayed with the same convenience and accessibility on a wide variety of platforms and devices. This way we bring accessibility to mainstream reading environments so users do not have to stay with special DAISY books and devices. We explain below the general features of the *IDEAL eBook reader*:

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

[8] http://www.accessiblesystems.co.in

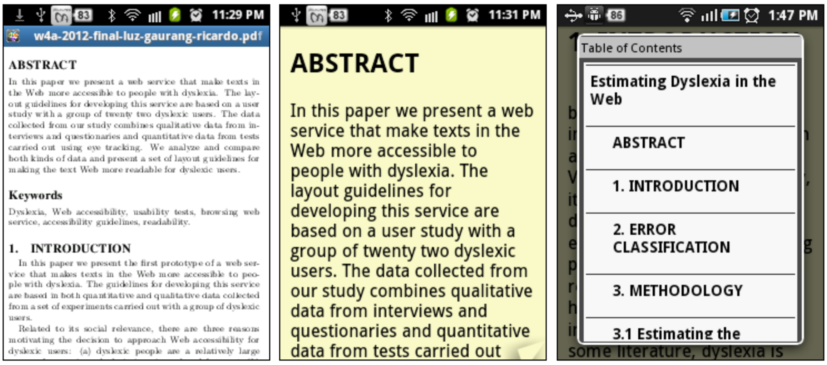


Figure 16.6: Some screenshots of the *IDEAL eBook Reader*: original PDF (left), *DysWebxia* option (middle) and table of contents (right).

Application Features. We explain below the general features of the *IDEAL eBook Reader*. As the features show, the application can be customized depending on the user needs.

(a) It displays the text in a more accessible way with a large font-size, making optimal use of the screen. In Figure 16.6 we can observe a PDF document (left) and the same text displayed by the reader (middle).

(b) It displays a table of contents of the ebook that allows the user to navigate to specific places within the ebook (see Figure 16.6, right).

(c) It also allows a user to customize how the ebook will be displayed. The users can choose the font styles, colors (background and font), brightness contrast, font size, and the character, line and paragraph spacing. Any individual can customize the parameters for greatest comfort while reading. For users with dyslexia there is an option called *DysWebxia* default which sets all the parameters to our dyslexic friendly recommendations (see Figure 16.7, left).

(d) It supports text-to-speech technology that enables users to listen to the ebook content as an audio book. This feature helps

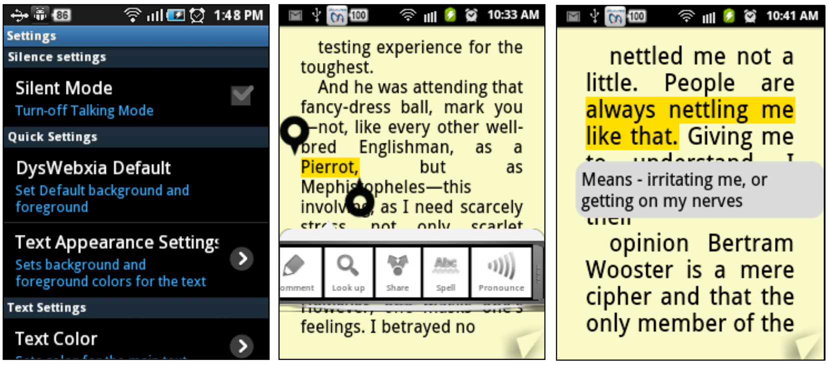


Figure 16.7: Examples of screenshots of the *IDEAL eBook Reader*: Settings of the *IDEAL eBook Reader* (left), highlight options (middle) and making notes (right).

people with vision impairments as well as dyslexia among other disabilities. The tool is compatible with a wide range of text-to-speech engines, and hence multiple languages are also supported, such as English, Spanish, Portuguese, German, French, Italian, Chinese, and Japanese, among others.

(e) The text being read out loud is highlighted to follow the reading easer.

(f) Control of the speech is gesture based, so it is convenient to use and very accessible. It is possible to read even word-by-word or letter-by-letter if the user wishes (see Figure 16.7, middle). That is, the user can select the piece of text to be read. By this means, a person with dyslexia could learn how to read new words.

(g) It allows the user to write a comment over a phrase (see Figure 16.7, right).

Using IDEAL eBook Reader. When the *IDEAL eBook Reader* starts, you can open an ebook from your phone memory, or you can download one from online sites such as Project Gutenberg, Feedbooks, etc. Once a book opens, the user can set the text font size, color, spacing, according to his/her preferences. A dyslexic friendly option can be also selected. For speech, instructions are as follows: to start and stop speech, double tap the screen; to move to the next paragraph, swipe the finger across the screen from left-to-right; and to move to the previous paragraph, swipe the finger across the screen from right-to-left. Volume keys on Android devices can be used for this purpose.

Usability Evaluation

Design: In a within-subject design, all the participants had to perform some tasks with the tool using the think aloud protocol [211]. They had to choose and customize a text to later read it. Then in an open interview they explained their thought about the tool to the interviewer (the author of this thesis).

Participants: We recruited 14 participants with diagnosed dyslexia, all Spanish native speakers. This number of participants is enough to discover most of the usability problems since only five users are enough to find 80% of the usability problems [253].

Materials:

> **Texts:** For the evaluation we used two books in Spanish and English: *Poems* by Edgar Allan Poe, and *The Hound of the Baskerville* by Sir Arthur Conan Doyle.
>
> **Open Interview:** The interview contains questions about their daily difficulties as a person with dyslexia, and their use of technology.
>
> **Device:** Samsung Galaxy Ace S5830.

Procedure: The sessions lasted around 15 minutes and were conducted in a quiet room prepared for the study at Universitat Pompeu Fabra. This usability evaluation followed the same produce conducted for *DysWebxia Reader* explained in Section 16.4.3, with the addition of the last open interview.

Results: Following we describe the observations collected while the participants were performing the task, and the relevant data extracted from the open interview. For instance, one participant would have like to chose the font type and find more colors. They suggested to include

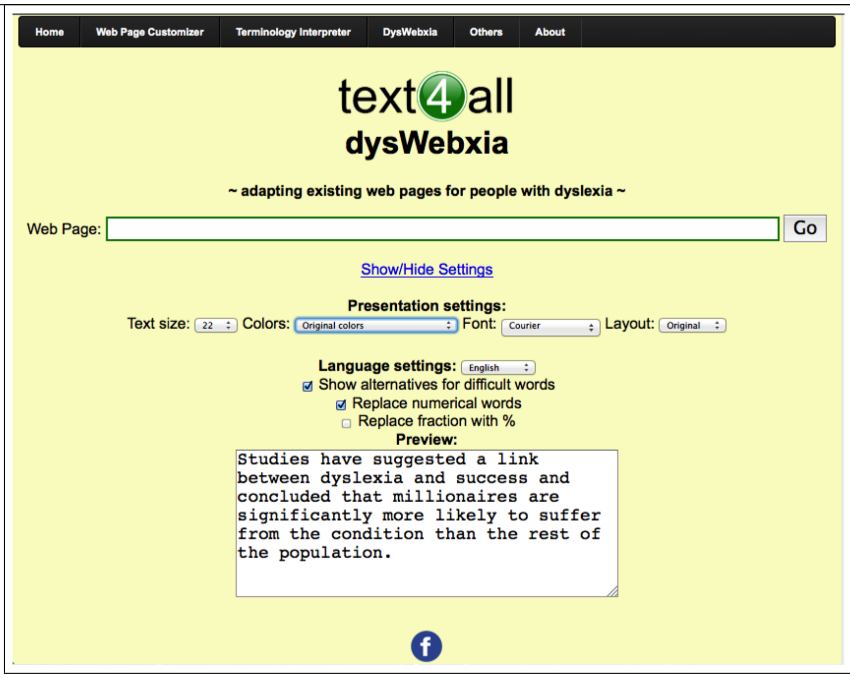


Figure 16.8: Text4all *DysWebxia* web page.

a book mark and to include an option that tells the last opened book. Also participants suggested including a preview of the book in the icons of the library. In general the participants found the *IDEAL eBook Reader* very usable. The suggestions of the participants were subsequently integrated in the application.

16.4.5 Text4all

The web server *text4all*, created by Vasile Topac at Politehnica University of Timisoara, allows users to adapt text from existing web pages both at presentation level and content level. Since *text4all* runs on a server everything works directly into the browser o any device, requiring no setup nor special rights on the client machine [382]. On the presentation level it allows the user to change the presentation of the text providing several default settings, such as settings for low vision or dyslexia, the *DysWebxia* settings. On the content level, *text4all* focuses on specialized languages (currently medical language). It adapts the content by explaining terminology, terminology annotation, and language analysis [384]. The terminology service was evaluated with 41 participants obtaining good ratings [383]. The *text4all* web service with *DysWebxia* settings was demoed at W4A' 2013 [326] and it is available on the Web.[9]

As presented in Figure 16.8, *text4all* is composed of an input area for the target web page address, a settings area for customization, and a live preview area. The *text4all DysWebxia* service adapts the look of the text by overriding the cascade style sheets of the original website. At the language level adaptation, for synonyms some dictionaries are being used, while for replacing numerical expressions with numbers and fractions with percentages basic string parsing and replacing is performed [383].

The *text4all* web service takes into consideration all recommendations from *DysWebxia* model, that is, its combines all the features that, to the best of our knowledge, lead to a significant improvement of the reading performance of people with dyslexia. These are:

(a) adapted color scheme (text and background) (Chapter 7),

(b) text size (Chapter 8),

(c) font type (Chapter 9),

(d) showing synonyms for difficult words (Chapter 14) as in Figure 16.9,

(e) replacing numerical expressions written in letters with digits,

(f) and replacing fractions with percentages, as suggested in Chapter 11. The language level adaptations are currently supported in English, Romanian and Spanish.

Figure 16.9 presents a comparison between the original website and the website adapted using *text4all DysWebxia*. One can notice the additions of tooltips with synonyms for complex words, for example in Figure 16.9 (up) the synonym *native* is proposed for the word *indigenous*. Also the original layout of the web page is preserved, as well as other functionalities like video streaming.

[9] `http://www.text4all.net/DysWebxia.html`

Figure 16.9: Original web page (top) and *text4all DysWebxia* adapted web page (bottom).

16.4.6 AccessibleNews DysWebxia

AccessibleNews DysWebxia is a web-based service for displaying news in a more accessible way for people with dyslexia. It is an extension of the AccessibleNews DAISY software built by Accessible Systems, India [187]. It was demoed at W4A' 2012 [331].

AccessibleNews DysWebxia detects useful text from an article web page, and renders it in a simplified manner. *AccessibleNews DysWebxia* resides on a server, and is accessed using a web browser. The

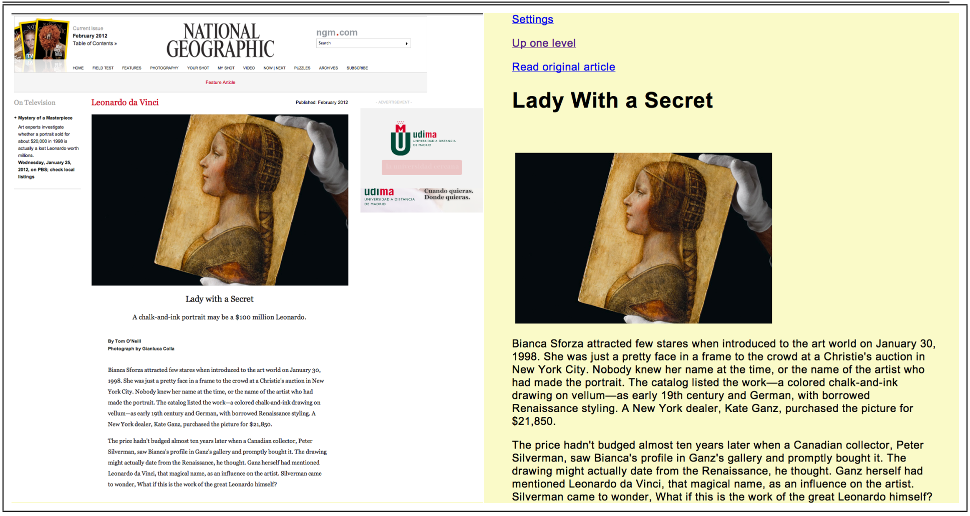


Figure 16.10: Example of an article browsed with *AccessibleNews DysWebxia* (original on the left).

server visits web pages, processes them, and uses machine learning to identify the article of interest from each web page. This article is then displayed in a browser in a plain simple format, devoid of fancy styling. Since the display is browser-based, specific parameters of the text, such as the colors of the font size, can be individually changed using Javascript. As a result, it is possible to create a combination best suited for persons with dyslexia. Further, any individual can customize the parameters for greatest comfort while reading (see Figure 16.11).

Our recommendations for text presentation were incorporated in the service as the default values for various parameters. Figure 16.10 shows an example, with an original article on the left and the same article using *AccessibleNews DysWebxia* on the right. The browser-based user interface of *AccessibleNews DysWebxia* can be accessed not only from a PC or laptop, but also from most smartphones and tablets.

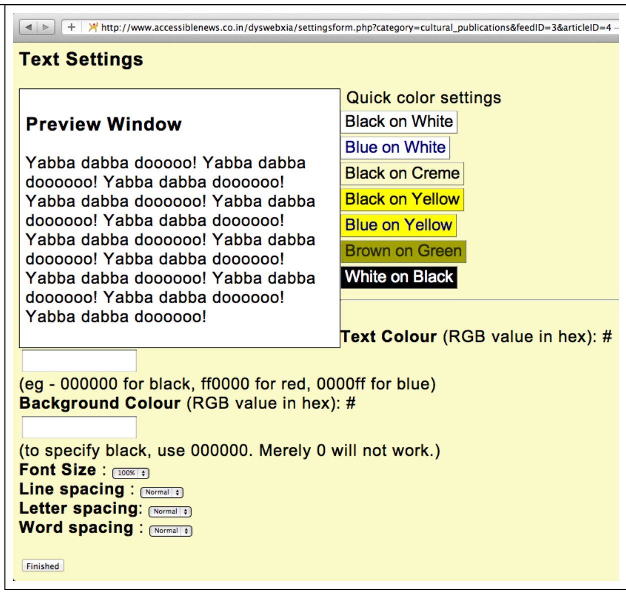


Figure 16.11: Settings of *AccessibleNews DysWebxia*.

16.5 Discussion

16.5.1 Recommendations

There are two surveys about web accessibility and dyslexia, one by McCarthy and Swierenga [228] from 2010 and one by Santana *et al.* [109] from 2012. These surveys also cover the recommendations about text for people with dyslexia. According to these surveys, the main problems that people with dyslexia find with text on the screen are small text, complicated language, and poor colors. Regarding the presentation of the text there are a number of recommendations, some of them based on user studies (see Section 2.2.4). Most of the text presentation recommendations are focused in a few parameters, being the recommendations of the British Dyslexia Association [52] the ones that cover all the text presentation parameters studied in this thesis. The recommendations addressing the content of the text are vague, such as avoiding using complicated language, however, we could not find a concrete definition of "complicated language". We collected the rest of the recommendations regarding text content (keywords and graphical schemes) from educational literature (also in Section 2.2.4).

In general terms, our recommendations for text presentation are consistent with previous suggestions. For font types, *italics* and *serif* fonts are not recommended [52, 217] and most studies recommend *Arial* [52, 130, 217], *Courier* [1], *Comic Sans* [52, 130] or, alternatives to these such as *Verdana* [52]. For font size our results support previous recommendation which suggest using big fonts [52, 113, 256], based on our results we recommend using even bigger fonts (18 points), which were not previously tested. Our recommendation for spacing are consistent with a previous study for character spacing tested on paper [424] and with the recommendations of the British Dyslexia Association for line spacing [52]. Our color recommendations do not fully support previous recommendations which suggest other colors [52, 158, 296], or lower contrasts [51, 388]

Our recommendations for text content provide empirical evidence to the previous recommendations in educational literature regarding graphical schemes [75, 219, 299, 405] and keywords [164, 280]. Our findings are also consistent with cognitive neuroscience literature addressing numerical expressions [80, 203] and more complex words (long and less frequent) [100, 174, 339, 358, 414].

16.5.2 Applications

In Table 16.3 we compare the features of the two most popular reading applications –*Amazon's Kindle reading software*[10] and *Apple's iBooks*[11]– with eight specific reading software for people with dyslexia. Although we could not find research papers about *ClaroRead*,[12] we include it because of its broad commercialization among people with dyslexia. The features 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). Synonyms on demand lead to an increase of subjective readability (Chapter 14), and text-to-speech technology (TTS) have shown gains in word recognition and phonological decoding [260]. We added some features that are included in the British Association of Dyslexia recommendations [52].

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

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

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

Text Presentation									
Software	**Font**	**Size**	Bright.	Color	**Spacing**				Width
					Char.	Word	**Line**	Par.	
Amazon's Kindle	yes	yes	yes	yes	no	no	yes	no	yes
Apple's iBooks	yes	yes	yes	yes	no	no	no	no	no
AccessibleNews	no	yes	no	yes	yes	yes	yes	no	no
ClaroRead	yes	yes	no	yes	yes	no	yes	yes	no
DysWebxia	yes	yes	yes	yes	yes	no	no	no	no
Firefixia	yes	yes	no	yes	yes	no	yes	no	yes
IDEAL	yes	yes	yes	yes	yes	yes	yes	yes	no
MultiReader	no	yes	no	yes	no	no	no	no	no
SeeWord	yes	yes	no	yes	yes	no	yes	no	no
Text4All	yes	yes	no	yes	yes	no	yes	no	no

Text Content					
Software	**Show Synonyms**	**Numerical Expressions**		**Keywords**	**TTS**
		Digit	**Percentage**		
Amazon's Kindle	no	no	no	no	no
Apple's iBooks	no	no	no	no	no
AccessibleNews	no	no	no	no	no
ClaroRead	no	no	no	no	yes
DysWebxia	yes	no	no	no	no
Firefixia	no	no	no	no	no
IDEAL	no	no	no	no	yes
MultiReader	no	no	no	no	yes
SeeWord	no	no	no	no	no
Text4All	yes	yes	yes	no	no

Table 16.3: Feature comparison for reading tools where boldface is used when they improve significantly the reading performance.

The feature for adjusting the brightness contrast is integrated in the reading software of the mobile devices (*Kindle*, *iBooks*, *IDEAL eBook reader* and *DysWebxia Reader*). For the rest, such feature is part of the operating system, not of the reading software itself. *ClaroRead* also includes a feature called 'homophones' which detects words that are pronounced the same, coloring them in the document to make these words less confusable. Similarly, *SeeWord* colors differently similar looking letters such as <d> and <b> [157]. The rationale behind these features is that people with dyslexia specifically encounter problems with words that are phonetically or orthographically similar [125]. *Firefixia* also includes the option of eliminating italics and other features for web text such as choosing the color of the links [347]. Even if *SeeWord* does not allow customizing the column width, it uses another strategy to reduce the visual clutter in the interface: a masking window which leaves only part of the text visible, reducing the problem of visual stress and memory. *Text4All* also includes other language specific domain services (Section 16.4.5). The *IDEAL eBook reader* also includes text-to-speech technology (Section 16.4.4).

The most popular tools, *Kindle* and *iBooks*, allow to customize most of these crucial text presentation parameters, with the exception of character spacing and the text content changes such as the presentation of synonyms on-demand. These mainstream tools with some additional features (which at the moment are only on specialized software for dyslexia) could provide greater accessibility, leading to a more inclusive design where additional software would not be needed for people with dyslexia. However, nowadays, specialized reading software is still needed to cover all features that lead to a more accessible reading for people with dyslexia.

Chapter 17

Conclusions and Future Work

In this chapter we summarize the main conclusions of each part of the thesis and draw future lines of research. One of the strengths of our conclusions, in comparison with other studies, is that in most of the results we are able compare both qualitative and quantitative collected form the experiments.

17.1 Understanding

In Part II *Understanding*, we analyzed dyslexic errors from three different perspectives: reading, writing and their presence in the Web. The results of this part show that errors are a good source of knowledge to understand dyslexia.

First, in Chapter 4, we learnt that people with dyslexia process errors in a different way that people without dyslexia. Indeed, the presence of errors in the text has less impact in the reading performance of people with dyslexia than for people without dyslexia. In some cases people with dyslexia even has better comprehension performance than people without dyslexia. Second, the letter transposition effect produces a similar reading and comprehension performance for both, people with and without dyslexia. That is, both groups read texts similarly with jumbled letters. Third, people with dyslexia are less aware of written errors than people without dyslexia. Finally, we can conclude that in general terms, lexical quality is a good indicator for text readability and comprehensibility –except for people with dyslexia–.

Second, in Chapter 5 we learnt that many errors written by people with dyslexia are different from other written errors. We created the first resource of dyslexic errors in Spanish, *DysList*, and analyzed the written dyslexic errors from a linguistic, phonetic and visual perspective. We found that the different types of dyslexic errors in English and Spanish present similar frequency distributions and the differences are due to the different orthographies of the languages. In Spanish, substitutions are the most frequent errors –followed by letter insertions and deletions– while transpositions are the less frequent type. Second, there are phonetic patterns in errors written by people with dyslexia. The most frequent errors involve letters where the one-to-one correspondence between graphemes and phones is not maintained. Most of the vowel substitutions occur between vowels that share phonetic characteristics, and errors in consonants occur within the same class of consonants. Finally, dyslexic errors are visually motivated. Taking into account handwriting visual features of the letter we found that errors occur more frequently along with letters containing mirror and rotation features as well as with fuzzy letters. We have published in the Web this resource,[1] a list of unique errors annotated with linguistic, phonetic and visual features.

Third, in Chapter 6 we found that errors with dyslexia could be also useful to estimate the presence of dyslexia in the Web. We classified the different errors found in the Web and presented a methodology to estimate the impact of dyslexic errors in the Web. We estimate that the presence of dyslexia in the Web is that around 0.67 and 0.43% of the errors in the Web are dyslexic errors for English and Spanish, respectively. This suggests that the widespread use of spell checkers ameliorates dyslexia in the Web so the prevalence of content with dyslexic errors is a function of both people and technology. Second, even though Spanish orthography is shallower than English, the difference between these two languages in terms of dyslexic error rates is not as substantial as expected. Third, the rate of dyslexic errors is

[1] `www.luzrello.com/resources.html` and `http://grupoweb.upf.edu/WRG/DysList.csv`.

independent from the rate of spelling errors in web pages. A study of the estimation in subsequent years shows that the error percentages are growing. Hence, the increase of size of the Web is not correlated with the improvement of its lexical quality, which can be explained by the fact that each year we have a large number of new users. Finally, the lexical quality of the social media is not as bad as it would be expected given the large amount of user generated content.

Impact

These findings can have an impact in how students with dyslexia are evaluated in the educational system. Written orthographic errors shall penalize less for people with dyslexia since there is scientific evidence that people with dyslexia can see the errors –eye tracking measurements–, but cannot process them consciously –lower error awareness rate– and do not affect their comprehension –no effect in the comprehension score–. Second, dyslexic written errors are very valuable since there are linguistically motivated. Instead of throwing them away, they could be used to learn from them in order to overcome dyslexia. That is, learning from the errors, instead of learning from correct words, such as in the *Dyseggxia* game [330] (*Piruletras* in Spanish).

17.2 Presentation

In Part III *Text Presentation*, we studied how the text presentation can lead to a better reading performance of people with dyslexia. We learnt that some text presentation parameters do have an impact.

First, in Chapter 7, we explored the effect of eight text presentation parameters on readability and preferences for people with and without dyslexia. The main conclusions are that larger font sizes significantly improve readability, especially for people with dyslexia; and that larger character spacing significantly improve readability for people with and without dyslexia. Regarding preferences, both groups found texts with no grey scales and with larger font sizes as significantly more readable.

Second, in Chapter 8, we analyzed the effect of font size and line spacing on objective and subjective readability and comprehensibility of web text (Wikipedia). The main result can be summarized as *size matters, spacing doesn't*. Up to a font size of 18 points, subjective and objective readability and comprehension improved. Beyond 18 points, there were no further increases for the objective measures, and even decreases in the subjective measures. Line spacing, in contrast, had almost no effect. We only found hints that larger line spacing may lead to worse text comprehension. For people without dyslexia very large line spacing (1.8) negatively affected subjective comprehension. Therefore, we conclude that a font size of 18 points ensures optimal both subjective and objective readability and comprehensibility. For line spacing, we suggest to keep the default spacing 1.0, since this is what readers are most used to, since increasing it too much might harm comprehension.

Finally, in Chapter 9, we tested the effect of twelve font types on objective readability and preferences of people with and without dyslexia. The main conclusion is that font types have an impact on readability for people with and without dyslexia. Good fonts for people with dyslexia are *Helvetica*, *Courier*, *Arial*, *Verdana* and *CMU*, taking into consideration both, reading performance and subjective preferences. Also, *sans serif*, *monospaced*, and *roman* font types increased significantly the reading performance, while *italic* fonts decreased reading performance. In particular, *Arial It.* should be avoided since it significantly decreases readability. What is good for people with dyslexia regarding font types is also good for people without dyslexia.

Impact

The findings of Part III can have impact on the recommendations for screen text presentation, and on the text options chosen by developers, designers, or content producers when they target people with dyslexia. Also, these findings can have an impact on interactive systems that rely on text as the main information medium. By applying our suggested text presentation recommendations, these systems could make texts easier to read and understand for people with dyslexia.

17.3 Content

In Part IV *Text Content*, we analyzed how certain text modifications can lead to a better reading performance of people with dyslexia. We learnt that some content modifications matter, especially the ones addressing the lexical and the discourse language level.

First, in Chapter 10 we tested the effect of word length and word frequency on readability and comprehension. Our results show that more frequent words improve readability while shorter words may improve comprehensibility, especially in people with dyslexia. This suggests that people without dyslexia come closer to the ideal reading scenario with faster reading time as well as better text comprehension.

Second, in Chapter 11 we found out that the presence of different representations of numerical expressions in a text impacts the readability for people with and without dyslexia. The main contribution is that numbers represented as digits instead of words, as well as percentages instead of fractions, improve readability of people with dyslexia.

Third, in Chapter 12 we studied the impact of verbal paraphrases in lexical simplification for people with and without dyslexia. The effect of verbal paraphrases is concluded to be insignificant. Our results are negative in the sense that verbal paraphrases neither improved readability nor comprehensibility in our experiment.

Finally, in Chapter 13 we carried out two experiments regarding the use of graphical schemes and highlighting keywords. The main conclusions is that graphical schemes improve the subjective readability and comprehensibility for people with dyslexia, while highlighted keywords increases the objective comprehension of people with dyslexia, but not readability. Also the presence of graphical schemes might motivate people with dyslexia to read more, since they find texts with graphs easier to read and understand. For people without dyslexia no effects were found.

Impact

The findings of Part IV can have impact on current systems for people with dyslexia that modify the text presentation but not its content such as the *Claro ScreenRuler Suite* [79] or the *IDEAL eBook reader* [188]. For instance, digital texts could become easier to understand by people with dyslexia by performing automatic keyword extraction and highlighting the keywords found.

The results from Chapters 10 and 11 provide an empirical basis for the development and refinement of recommendations for text simplification. These recommendations exist in very general form, such as the *European guidelines for the production of easy-to-read information for people with learning disability* [140]. This finding can also have an impact in the assessment of readability since computational models for predicting readability of texts such as FOG [159], Flesch, Flesch-Kincaid [135], and SMOG [231], do not specifically take into account word frequency or numerical expressions –they do account for word length–. These findings also motivate natural language processing work on lexical text simplification and numerical expressions [27, 45, 293].

The results from Chapter 13 motivate natural language processing work on discourse simplification. They also support previous educational recommendations by adding the empirical analysis of the impact of graphical schemes and highlighting keywords using objective measures.

17.4 Integration

In Part V *Integrating Presentation and Content*, we have explored to what extent different automatic lexical simplifications can improve the reading performance of people with dyslexia and the effect of how these lexical simplifications are presented.

First, in Chapter 14 we tested the effect of two lexical simplification strategies on readability, comprehension and subjective ratings. We did not find significant effects of the lexical simplification strategy on objective readability and comprehension. However, we did find significant effects on the participants' subjective 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 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 manually simplified texts. Therefore, a system like ShowSyns that displays 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 stay on a lower reading proficiency level. Therefore, anything that might help them to subjectively perceive reading as being easier, can potentially help them to avoid this vicious circle, even if no significant improvement on readability can be demonstrated.

Second, in Chapter 15 we evaluated the quality of synonyms generated on demand by two algorithms: 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 LexSiS algorithm presented in Chapter 14, the negative results found for readability might be different for this new algorithm. Regarding the resource generated by CASSA (a dictionary of synonyms with its contexts), we have published it in the Web, giving the opportunity to other researchers and developers to enrich their tools.[2]

Finally, in Chapter 16 we took all the results of the previous chapters and combine them in a model called *DysWebxia*. This model proposes a set of recommendations that have been already integrated in several applications. This is the first model that integrates, both, text alterations regarding the presentation and the content of the text. So far, four tools include the model: *DysWebxia Reader*, *IDEAL eBook reader*, *Text4All* and *AccessibleNews DysWebxia*. The results of the usability analyses of two of the tools are very positive.

Impact

The application of the recommendations and the use of these tools could improve the ability to read and access a wider range of information content, empowering people with dyslexia by slightly leveling the reading playground. Based on these results, we believe that *DysWebxia* may have a large future impact for people with dyslexia. As a matter of fact, the *IDEAL eBook reader* has been downloaded more than 35,000 times. The text recommendations for people with dyslexia have the potential to impact different fields, particularly in

[2]`www.luzrello.com/resources.html` and `http://grupoweb.upf.edu/WRG/CASSA.txt`

education. Currently, the text presentation recommendations for people with dyslexia are being adapted to present online exams to children with dyslexia in the schools of the Generalitat de Catalunya *('Government of Catalonia')*.

17.5 Future Work

Part II. To further understand dyslexia, more experiments shall be done addressing how people with dyslexia read errors, using different types of errors (phonetically or visually motivated) and taking into consideration other variables such as word frequency. Accounting for the letter transposition effect, we shall perform further experiments, testing transpositions in different part of the words –not only interior transpositions–. Regarding the information extracted from the written errors of Chapter 5, we plan to use the most relevant phonetic and visual features to tailor the CASSA algorithm. This way we will create a word complexity measure adapted for people with dyslexia. In Chapter 6 we learnt that particular words can be used to detect dyslexic texts. Hence, this could also be applied to detect users with dyslexia. This can be used in the future to improve Web accessibility as well as future spell checkers targeted to users with dyslexia.

Part III. Future work regarding the presentation of the text needs to focus on studying even bigger font sizes and more character spacing values. While our results did not show improvements for font sizes beyond 18 points, we did not find conclusive evidence about the point where increasing the font size leads to reduced readability and comprehensibility. Regarding font type, future challenges involve studying the effect of the font types on the comprehension. Finally, we should also explore the effect of the presentation of the text in different contexts and devices as well as the study of different parameters in combination, as we have not studied all the effects due to the interaction of different presentation parameters.

Part IV. The research field regarding the content of the text is less explored than the one addressing the presentation of the text. More parameters and theirs interaction with text presentation conditions shall be explored. For instance the combination of colors together with highlighting keywords or lexical simplification. It also shall include testing memorability using objective measures in addition to the subjective responses of the participants. More concretely, since there is a correlation between word length and frequency –frequent words tend to be shorter–, future work needs to investigate potential interdependencies between these two factors. Also, a future challenge is the evaluation of other representations for specific numerical expressions, for example to represent time. Another challenge is the study of more types of graphical schemes and keywords and adding to the evaluation a delayed post-test to address the effect of supplemental graphical schemes and keywords on robustness of learning.

From the methodological point of view (Part III and IV), a future challenge in to measure the effect of the strategies in time. For instance, by carrying out a preference questionnaire after some time or by adding to the evaluation a delayed post-test to address the effect of supplemental graphical schemes and keywords on robustness of learning.

Part V. Here future work includes the refinement of CASSA by tailoring the detection of lexical complexity, as already mentioned in Part III. We will consider the orthographic and phonetic similarity of words, because these language features are present in the errors (Chapter 5) and makes words more difficult to recognize for people with dyslexia [100] as well as without dyslexia [238]. This implies defining a new measure of word complexity that takes into account these features. More concretely, regarding the *DysWebxia Reader*, future work include integrating the participants' suggestions. We will add to each suggested synonym a link to search the complex word in Wikipedia and an option to read its definition using text-to-speech. We also plan to add a module with hyperonyms for targeting complex specific words that have no synonyms, such as names of animals or plants. We will also add further options to handle other file formats, in particular HTML, and to customize the highlighting of complex words with boldface, colors, or different kinds of underlining. Further work will also focus in the integration of the model in more systems. For instance, at the moment we are working on the integration of the model in the project Cloud4All in collaboration with the company Technosite, and on the adaptation of the dyslexic-friendly recommendations on school materials of the Generalitat de Catalunya *('Government of Catalonia')*.

Last Words

Accessibility is a continuum. Everyone experiences barriers. Barriers do not only depend on the biological heritage but also on the education, on the situation, and on the time in life. If a European without dyslexia is suddenly dropped in Japan, it becomes letter blind. Around 80% of the elderly population has a disability. We all are very likely to have a disability in the future. The results presented in this thesis have the potential to be extended, not only to other groups –dyslexia difficulties are shared by other people with special needs– but also to cover general usability problems since dyslexia symptoms are common to varying degrees among most people. Further research should focus, in the light of universal accessibility, in how to bridge the gap between specialized and non-specialized software, to empower their inclusion by not making differences between people. Future work should involve other targets groups, to explore which parameters are beneficial not only for people with dyslexia, but also for all.

Bibliography

[1] Ability Net. NHS choices: Help with accessibility, changing fonts on your computer, 2013. `http://www.nhs.uk/accessibilityhelp/index.html`.

[2] J. Acha and M. Perea. The effect of neighborhood frequency in reading: Evidence with transposed-letter neighbors. *Cognition*, 108(1):290–300, 2008.

[3] D. Adler-Grinberg and L. Stark. Eye movements, scan paths, and dyslexia. *American Journal of Optometry and Physiological Optics*, 55(8):557–570, 1978.

[4] 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*, 4554:593–602, 2007.

[5] Alexa Internet. Top 500, February 2013. `http://www.alexa.com/topsites`.

[6] 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.

[7] S. M. Aluísio, L. Specia, T. A. S. Pardo, E. G. Maziero, and R. P. M. Fortes. Towards Brazilian Portuguese automatic text simplification systems. In *Proceeding of the eighth ACM symposium on Document engineering (DocEng '08)*, pages 240–248, New York, NY, USA, 2008. ACM Press.

[8] S. Aluísio, L. Specia, T. A. S. Pardo, E. Maziero, H. M. Caseli, and R. P. M. Fortes. A corpus analysis of simple account texts and the

proposal of simplification strategies: first steps towards text simplification systems. In *Proceedings of the 26th annual ACM international conference on Design of communication (SIGDOC '08)*, pages 15–22, New York, NY, USA, 2008. ACM Press.

[9] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. American Psychiatric Publishing, Arlington, VA, 2000.

[10] A. Andersen and C. Rowland. Improving the outcomes of students with cognitive and learning disabilities: phase I development for a web accessibility tool. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility (ASSETS '07)*, pages 221–222, New York, NY, USA, 2007. ACM Press.

[11] J. R. Anderson, P. J. Kline, and C. M. Beasley. A general learning theory and its application to schema abstraction. *Psychology of learning and motivation*, 13:277–318, 1979.

[12] J. R. Anderson and J. Robert. *Learning and memory*. John Wiley New York, 2000.

[13] I. Androutsopoulos and P. Malakasiotis. A survey of paraphrasing and textual entailment methods. *Journal of Artificial Intelligence Research*, 38:135–187, 2010.

[14] L. E. Aragón and A. Silva. Análisis cualitativo de un instrumento para detectar errores de tipo disléxico (IDETID-LEA) (Qualitative analysis of an instrument to detect dyslexic errors, IDETID-LEA). *Psicothema*, 12(Supl. 2):35–38, 2000.

[15] Association of American Publishers. AAP reports January 2011 book sales, March 2011. `http://es.scribd.com/doc/50948417/AAP-Reports-January-2011-Book-Sales`.

[16] R. Baeza-Yates and L. Rello. Estimating dyslexia in the Web. In *Proceedings of the 8th International Cross-Disciplinary Conference on Web Accessibility (W4A 2011)*, Hyderabad, India, 2011. ACM Press.

[17] R. Baeza-Yates and L. Rello. On measuring the lexical quality of the web. In *The 2nd Joint WICOW/AIRWeb Workshop on Web Quality*, pages 1–6, Lyon, France, 2012.

[18] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition)*. Addison Wesley, Harlow, UK, 2011.

[19] J. Banerjee, D. Majumdar, M. S. Pal, and D. Majumdar. Readability, subjective preference and mental workload studies on young Indian adults for selection of optimum font type and size during on screen reading. *Al Ameen Journal of Medical Sciences*, 4:131–143, 2011.

[20] G. Barnard. Statistical calculation of word entropies for four western languages. *Transactions on Information Theory*, 1(1):49–53, 1955.

[21] C. M. Barnum. *Usability Testing Essentials: Ready, Set... Test!* Morgan Kaufmann, 2010.

[22] M. A. Barrios. El dominio de la funciones léxicas en el marco de la Teoría Sentido-Texto (the domain of lexical functions in the MTT framework). *Estudios de Lingüística del Español (Studies of Spanish Linguistics, ELiEs)*, 30, 2010.

[23] M. A. Barrios, G. Aguado de Cea, and J. A. Ramos. Enriching a lexicographic tool with domain definition problems and solutions. In G. Sierra, M. Pozzi, and J. M. Torres, editors, *1st International Workshop on Definition Extraction, RANLP 09*, Shoumen, Bulgaria, 2009. INCOMA Ltd.

[24] M. A. Barrios Rodríguez and L. Rello. False paraphrase pairs in Spanish for verbs and verb + noun collocations. *Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural (Journal of the Spanish Society for Natural Language Processing)*, 46:107–112, 2011.

[25] R. Barzilay, N. Elhadad, and K. R. McKeown. Inferring strategies for sentence ordering in multi document news summarization. *Journal of Artificial Intelligence Research*, 17:35–55, 2002.

[26] S. Bautista, B. Drndarevic, R. Hervás, H. Saggion, and P. Gervás. Análisis de la Simplificación de Expresiones Numéricas en Español mediante un Estudio Empírico (Analysis of Numerical Expressions Simplification in Spanish through an Empirical Study). *LinguaMATICA*, 4(2):27–41, 2012.

[27] S. Bautista, R. Hervás, P. Gervás, R. Power, and S. Williams. A system for the simplification of numerical expressions at different levels of understandability. In *NAACL Workshop Natural Language Processing for Improving Textual Accessibility (NLP4ITA)*, 2013.

[28] 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.

[29] M. F. Beauvois and J. Derouesne. Phonological alexia: three dissociations. *Journal of Neurology, Neurosurgery & Psychiatry*, 42:1115–1124, 1979.

[30] M. Bernard, C. H. Liao, and M. Mills. The effects of font type and size on the legibility and reading time of online text by older adults. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '01, pages 175–176, New York, NY, USA, 2001. ACM.

[31] M. Bernard, B. Lida, S. Riley, T. Hackler, and K. Janzen. A comparison of popular online fonts: Which size and type is best. *Usability News*, 4(1):2002, 2002.

[32] M. L. Bernard, B. S. Chaparro, M. M. Mills, and C. G. Halcomb. Comparing the effects of text size and format on the readibility of computer-displayed Times New Roman and Arial text. *International Journal of Human-Computer Studies*, 59(6):823–835, 2003.

[33] T. Berners-Lee, J. Hendler, O. Lassila, et al. The semantic web. *Scientific American*, 284(5):34–43, 2001.

[34] V. W. Berninger and B. J. Wolf. *Teaching students with dyslexia and dysgraphia: Lessons from teaching and science*. Paul H Brookes Pub Co., 2009.

[35] V. W. Berninger, K. H. Nielsen, R. D. Abbott, E. Wijsman, and W. Raskind. Writing problems in developmental dyslexia: Under-recognized and under-treated. *Journal of School Psychology*, 46(1):1–21, 2008.

[36] D. Beymer, D. Russell, and P. Orton. An eye tracking study of how font size and type influence online reading. In *Proceedings of the 22nd*

British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 2, BCS-HCI '08, pages 15–18, Swinton, UK, UK, 2008. British Computer Society.

[37] D. Beymer and D. M. Russell. Webgazeanalyzer: a system for capturing and analyzing web reading behavior using eye gaze. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1913–1916, New York, NY, USA, 2005. ACM.

[38] D. Beymer, D. M. Russell, and P. Z. Orton. An eye tracking study of how font size, font type, and pictures influence online reading. *Proceedings INTERACT 2007*, pages 456–460, 2007.

[39] S. K. Bhatia, A. Samal, N. Rajan, and M. T. Kiviniemi. Effect of font size, italics, and colour count on web usability. *International Journal of Computational Vision and Robotics*, 2(2):156–179, 2011.

[40] 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.

[41] L. Bix. The elements of text and message design and their impact on message legibility: A literature review. *Journal of Design Communication*, 4, 2002.

[42] J. L. Black, D. W. K. Collins, J. N. De Roach, and S. Zubrick. A detailed study of sequential saccadic eye movements for normal-and poor-reading children. *Perceptual and motor skills*, 59(2):423–434, 1984.

[43] P. Bohman. Cognitive disabilities, part 1: We still know too little, and we do even less, 2004. `http://www.webaim.org/techniques/articles/cognitive_too_little/`.

[44] C. Boldyreff, E. Burd, J. Donkin, and S. Marshall. The case for the use of plain English to increase Web accessibility. *wse*, page 42, 2001.

[45] 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.

[46] S. Bott and H. Saggion. An unsupervised alignment algorithm for text simplification corpus construction. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pages 20–26, Portland, Oregon, June 2011. Association for Computational Linguistics.

[47] S. Bott and H. Saggion. Text simplification tools for Spanish. In *Proc. LREC'12*, Istanbul, Turkey, May 2012. ELRA.

[48] H. Bouma and C. P. Legein. Foveal and parafoveal recognition of letters and words by dyslexics and by average readers. *Neuropsychologia*, 15(1):69–80, 1977.

[49] D. Bourassa and R. Treiman. Spelling in children with dyslexia: Analyses from the Treiman-Bourassa early spelling test. *Scientific Studies of Reading*, 7(4):309–333, 2003.

[50] D. Boyarski, C. Neuwirth, J. Forlizzi, and S. H. Regli. A study of fonts designed for screen display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, pages 87–94, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co.

[51] J. Bradford. Designing web pages for dyslexic readers, 2011. `http://www.dyslexia-parent.com/mag35.html`.

[52] British Dyslexia Association. Dyslexia style guide, January 2012. `http://www.bdadyslexia.org.uk/`.

[53] B. Brown, G. Haegerstrom-Portnoy, C. Yingling, J. Herron, D. Galin, and M. Marcus. Tracking eye movements are normal in dyslexic children. *American journal of optometry and physiological optics*, 60(5):376, 1983.

[54] M. Bruck. The word recognition and spelling of dyslexic children. *Reading Research Quarterly*, pages 51–69, 1988.

[55] M. Bruck. Word-recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology*, 26(3):439, 1990.

[56] M. Bruck. Component spelling skills of college students with childhood diagnoses of dyslexia. *Learning Disability Quarterly*, pages 171–184, 1993.

[57] N. Brunswick. Unimpaired reading development and dyslexia across different languages. In S. McDougall and P. de Mornay Davies, editors, *Reading and dyslexia in different orthographies*, pages 131–154. Psychology Press, Hove, 2010.

[58] M. Brysbaert. Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: General*, 124(4):434–452, Dec 1995.

[59] 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.

[60] G. Buscher, R. Biedert, D. Heinesch, and A. Dengel. Eye tracking analysis of preferred reading regions on the screen. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, pages 3307–3312. ACM, 2010.

[61] B. Butterworth. Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, 14(12):534–541, 2010.

[62] J. Bybee. *Phonology and language use*, volume 94. Cambridge Univ Pr, 2003.

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

[64] Y. Canning and J. Tait. Syntactic simplification of newspaper text for aphasic readers. *Customised Information Delivery*, page 6, 1999.

[65] A. Caramazza. How many levels of processing are there in lexical access? *Cognitive neuropsychology*, 14(1):177–208, 1997.

[66] A. Caramazza, A. Laudanna, and C. Romani. Lexical access and inflectional morphology. *Cognition*, 28(3):297–332, 1988.

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

[68] M. Carreiras, A. Mechelli, and C. J. Price. Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human brain mapping*, 27(12):963–972, 2006.

[69] M. Carreiras and M. Perea. Naming pseudowords in Spanish: Effects of syllable frequency. *Brain and Language*, 90(1):393–400, 2004.

[70] M. Carreiras, M. Perea, and J. Grainger. Effects of the orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of experimental psychology: learning, memory, and cognition*, 23(4):857, 1997.

[71] M. S. Carrillo, J. Alegría, P. Miranda, and N. Sánchez. 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.

[72] J. Carroll, G. Minnen, Y. Canning, S. Devlin, and J. Tait. Practical Simplification of English Newspaper Text to Assist Aphasic Readers. In *Proc. of AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pages 7–10, 1998.

[73] J. Carroll, G. Minnen, Y. Canning, S. Devlin, and J. Tait. Practical simplification of English newspaper text to assist aphasic readers. In *AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, Wisconsin, 1998.

[74] 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.

[75] B. Chalkley, J. Waterfield, and G. D. Network. *Providing learning support for students with hidden disabilities and dyslexia undertaking fieldwork and related activities.* University of Gloucestershire, Geography Discipline Network, 2001.

[76] C. Chapman. The most popular fonts used by designers, 2011. `http://www.webdesignerdepot.com/2011/08/the-most-popular-fonts-used-by-designers/`.

[77] N.-S. Chen, Kinshukb, C.-W. Wei, and H.-J. Chen. Mining e-learning domain concept map from academic articles. *Computers & Education*, 50(3):1009–1021, 2008.

[78] R. B. Clariana and R. Koul. A computer-based approach for translating text into concept map-like representations. In *Proceedings of the first international conference on concept mapping*, pages 14–17, 2004.

[79] ClaroSoftware. Screenruler, 2012. `http://www.clarosoftware.com/index.php?cPath=348.`

[80] L. Cohen, S. Dehaene, and P. Verstichel. Number words and number non-words: A case of deep dyslexia extending to Arabic numerals. *Brain*, 117:267–279, 1994.

[81] Col·legi de Logopedes de Catalunya. *PRODISCAT: Protocol de detecció i actuació en la dislèxia. Àmbit educatiu (Protocol for detection and management of dyslexia. Educational scope).* Departament d'Educació, Generalitat de Catalunya (Department of Education, Catalan Government), Barcelona, 2011.

[82] C. Coleman, N. Gregg, L. McLain, and L. W. Bellair. A comparison of spelling performance across young adults with and without dyslexia. *Assessment for Effective Intervention*, 34(2):94–105, 2009.

[83] M. Coleman and T. L. Liau. A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60(2):283, 1975.

[84] D. Collier. Colour text legibility, 2012. `http://www.hgrebdes.com/colour/spectrum/colourvisibility.html`.

[85] M. Coltheart. Disorders of reading and their implications for models of normal reading. *Visible Language*, XV:245–286, 1981.

[86] M. Coltheart. Varieties of developmental dyslexia: A comment on Bryant and Impey. *Cognition*, 27:97–101, 1987.

[87] M. Coltheart. Phonological dyslexia: Past and future issues. *Cognitive Neuropsychology*, 13(6):749–762, 1996.

[88] M. Coltheart, E. Davelaar, T. Jonasson, and D. Besner. Access to the internal lexicon. *Attention and Performance VI*, pages 535–555, 1977.

[89] M. Coltheart, J. Masterson, S. Byng, M. Prior, and J. Ridoch. Surface dyslexia. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 35A:469–495, 1983.

[90] M. Coltheart, K. Patterson, and J. C. Marshall. *Deep dyslexia*. Routledge and Kegan Paul, London, 1980.

[91] V. Connelly, S. Campbell, M. MacLean, and J. Barnes. Contribution of lower order skills to the written composition of college students with and without dyslexia. *Developmental Neuropsychology*, 29(1):175–196, 2006.

[92] Connexor Oy. *Machinese language model*. Connexor Oy, Helsinki, Finland, 2006.

[93] J. W. Cooper. Visualization of relational text information for biomedical knowledge discovery. In *Information Visualization Interfaces for Retrieval and Analysis workshop. ACM SIGIR*, 2003.

[94] C. Cornoldi and J. V. Oakhill. *Reading comprehension difficulties: Processes and intervention*. Routledge, 2013.

[95] W. Coster and D. Kauchak. Simple english wikipedia: A new text simplification task. In *ACL (Short Papers)*, pages 665–669, 2011.

[96] J. Craven and H. Booth. Putting awareness into practice: practical steps for conducting usability tests. *Library review*, 55(3):179–194, 2006.

[97] S. A. Crossley, J. Greenfield, and D. S. McNamara. Assessing text readability using cognitively based indices. *Tesol Quarterly*, 42(3):475–493, 2008.

[98] S. A. Crossley, M. M. Louwerse, P. M. McCarthy, and D. S. McNamara. A linguistic analysis of simplified and authentic texts. *The Modern Language Journal*, 91(1):15–30, 2007.

[99] D. Crystal. *English as a Global Language*. Cambridge University Press, Cambridge, 2003.

[100] F. Cuetos and F. Valle. Modelos de lectura y dislexias (Reading models and dyslexias). *Infancia y Aprendizaje (Infancy and Learning)*, 44:3–19, 1988.

[101] L. E. Cutting, A. Materek, C. A. S. Cole, T. M. Levine, and E. M. Mahone. Effects of fluency, oral language, and executive function on reading comprehension performance. *Annals of Dyslexia*, 59(1):34–54, 2009.

[102] F. J. Damerau. A technique for computer detection and correction of spelling errors. *Communications of the A.C.M.*, 7:171–176, 1964.

[103] M. F. Damian. Asymmetries in the processing of Arabic digits and number words. *Memory & Cognition*, 32:164–171, 2004.

[104] I. Darroch, J. Goodman, S. Brewster, and P. Gray. The effect of age and font size on reading text on handheld computers. In *Human-Computer Interaction-INTERACT 2005*, pages 253–266. Springer, 2005.

[105] E. D'Avanzo and B. Magnini. A keyphrase-based approach to summarization: the lake system at duc-2005. In *Proceedings of DUC*, 2005.

[106] A. Davison and G. M. Green. *Linguistic complexity and text comprehension: Readability issues reconsidered.* Lawrence Erlbaum, 1988.

[107] J. De Belder, K. Deschacht, and M.-F. Moens. Lexical simplification. In *Proceedings of LTEC 2010: 1st International Conference on Interdisciplinary Research on Technology, Education and Communication*, 2010.

[108] R. De Leeuw. Special font for dyslexia? *Master's thesis, University of Twente*, 2010.

[109] V. F. de Santana, R. de Oliveira, L. D. A. Almeida, and M. C. C. Baranauskas. Web accessibility and people with dyslexia: a survey on techniques and guidelines. In *Proc. W4A '12*, page 35. ACM, 2012.

[110] S. Dehaene. Varieties of numerical abilities. *Cognition*, 44:1–42, 1992.

[111] G. Deloche, E. Andreewsky, and M. Desi. Surface dyslexia: A case report and some theoretical implications to reading models. *Brain and Language*, 15(1):12–31, 1982.

[112] S. Devlin and G. Unthank. Helping aphasic people process online information. In *Proc. ASSETS '06*, pages 225–226. ACM, 2006.

[113] A. Dickinson, P. Gregor, and A. F. 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.

[114] 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.

[115] B. Drndarevic and H. Saggion. Towards automatic lexical simplification in Spanish: an empirical study. In *Proc. NAACL HLT '12 Workshop PITR '12*, Montreal, Canada, 2012.

[116] B. Drndarevic, S. Stajner, S. Bott, S. Bautista, and H. Saggion. Automatic text simplification in Spanish: A comparative evaluation of complementing components. In *Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING)*, 2013.

[117] M. Du. Approximate name matching. *Master's thesis, Royal Institute of Technology*, 2005.

[118] J. A. Duñabeitia, M. Perea, and M. Carreiras. Eye movements when reading words with \$ymβol and num83r5: There is a cost. *Visual Cognition*, 17(5):617–631, 2009.

[119] J. A. Duñabeitia, M. Perea, and M. Carreiras. There is no clam with coats in the calm coast: Delimiting the transposed-letter priming effect. *The Quarterly Journal of Experimental Psychology*, 62(10):1930–1947, 2009.

[120] Dyslexic.com. Readability for dyslexic people, December 2000.

[121] M. C. Dyson and M. Haselgrove. The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies*, 54(4):585–612, 2001.

[122] G. F. Eden, J. F. Stein, H. M. Wood, and F. B. Wood. Differences in eye movements and reading problems in dyslexic and normal children. *Vision Research*, 34(10):1345–1358, 1994.

[123] C. Elbro, I. Nielsen, and D. K. Petersen. Dyslexia in adults: Evidence for deficits in non-word reading and in the phonological representation of lexical items. *Annals of dyslexia*, 44(1):203–226, 1994.

[124] J. Elkind, K. Cohen, and C. Murray. Using computer-based readers to improve reading comprehension of students with dyslexia. *Annals of Dyslexia*, 43(1):238–259, 1993.

[125] A. W. Ellis. *Reading, writing and dyslexia.* Erlbaum, London, 1984.

[126] R. D. Elterman, L. A. Abel, R. B. Daroff, L. F. Dell'Osso, and J. L. Bornstein. Eye movement patterns in dyslexic children. *Journal of Learning Disabilities*, 13(1):16–21, 1980.

[127] 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.

[128] B. Evans. *Dyslexia and vision*, volume 5. Wiley, 2001.

[129] B. J. Evans, A. Cook, I. L. Richards, and N. Drasdo. Effect of pattern glare and colored overlays on a simulated-reading task in dyslexics and normal readers. *Optometry & Vision Science*, 71(10):619–628, 1994.

[130] 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.

[131] M. E. Farmer and R. M. Klein. The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin & Review*, 2(4):460–493, 1995.

[132] FEDIS. Presidente de la FEDIS (Federación Española de Dislexia) asiste a las jornadas de león, Januray 2008. `http://actualidadfedis.blogspot.com.es/`.

[133] W. Fias. Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research*, 65:250–259, 2001.

[134] W. Fias, M. Brysbaert, F. Geypens, and G. d'Ydewalle. The importance of magnitude information in numerical processing: Evidence from the snarc effect. *Mathematical Cognition*, 2:95–110, 1996.

[135] R. Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221, 1948.

[136] J. M. Fletcher, S. E. Shaywitz, and B. A. Shaywitz. Comorbidity of learning and attention disorders: Separate but equal. *Pediatric Clinics of North America*, 46(5):885–897, 1999.

[137] K. I. Forster. Accessing the mental lexicon. In *New approaches to language mechanisms*, pages 231–256. North Holland, Amsterdam, 1976.

[138] E. Francik, S. Levine, and S. Tremain. Telecommunications problems and design strategies for people with cognitive disabilities. Annotated bibliography and research recommendations. *Rehabilitation Engineering Research Center on Universal Telecommunication Access.*, 1999.

[139] E. Frank, G. W. Paynter, I. H. Witten, C. Gutwin, and C. G. Nevill-Manning. Domain-specific keyphrase extraction. In *Proc. Sixteenth International Joint Conference on Artificial Intelligence (IJCAI 1999).* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[140] G. Freyhoff, G. Hess, L. Kerr, E. Menzel, B. Tronbacke, and K. V. D. Veken. European guidelines for the production of easy-to-read information for people with learning disability, 1998.

[141] R. B. Friedman. Recovery from deep alexia to phonological alexia: Points on a continuum. *Brain and Language*, 52(1):114–128, 1996.

[142] U. Frith. A developmental framework for developmental dyslexia. *Annals of dyslexia*, 36(1):67–81, 1986.

[143] R. Frost. Toward a strong phonological theory of visual word recognition: true issues and false trails. *Psychological bulletin*, 123(1):71, 1998.

[144] K. Furdík, J. Paralič, and P. Smrž. Classification and automatic concept map creation in elearning environment. In *Proc. of the conference Znalosti*, pages 78–89, 2008.

[145] T. N. Gadd. Phonix: The algorithm. *Program: electronic library and information systems*, 24(4):363–366, 1990.

[146] R. Galván Gómez. Dislexia va en aumento (Dyslexia is increasing). *El Peruano (The Peruvian, newspaper)*, page 22, january 2010.

[147] J. E. García-Albea, R. M. Sánchez-Casas, and S. Viso. Efectos de la frecuencia de uso en el reconocimiento de palabras (Effects of usage frequency in word recognition). *Investigaciones psicológicas (Psychological Research)*, 1:24–63, 1982.

[148] C. Gasperin, E. Maziero, L. Specia, T. A. S. Pardo, and S. M. Aluisio. Natural language processing for social inclusion: a text simplification architecture for different literacy levels. *the Proceedings of SEMISH–XXXVI Seminário Integrado de Software e Hardware*, pages 387–401, 2009.

[149] G. Geiger and J. Y. Lettvin. Peripheral vision in persons with dyslexia. *New England Journal of Medicine*, 316(20):1238–1243, 1987.

[150] G. Geiger, J. Y. Lettvin, and M. Fahle. Dyslexic children learn a new visual strategy for reading: A controlled experiment. *Vision Research*, 34(9):1223–1233, 1994.

[151] I. A. Gelman and A. L. Barletta. A "quick and dirty" website data quality indicator. In *The 2nd ACM Workshop on Information Credibility on the Web (WICOW '08)*, pages 43–46, Napa Valley, USA, 2008.

[152] J. Gil. *Fonética para profesores de español: de la teoría a la práctica (Phonetics for teachers of Spanish: from theory to practice)*. Arco/Libros, Madrid, 2007.

[153] F. J. Gómez Guinovart. Bases lingüísticas y computacionales del procesamiento de la impropiedad estilística y la legibilidad (Linguistic and computational foundations for processing stylistic inappropriateness and legibility). *Revista española de lingüística aplicada (Spanish Journal of Applied Linguistics)*, 1:153–174, 1999.

[154] P. Goolkasian and J. King. Letter identification and lateral masking in dyslexic and average readers. *The American journal of psychology*, 1990.

[155] N. Goulandris, editor. *Dyslexia in different languages: Cross-linguistic comparisons*. Whurr Publishers, London, 2003.

[156] J. Grainger. Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, 23(1):1–35, 2008.

[157] 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.

[158] 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.

[159] R. Gunning. *The technique of clear writing*. McGraw-Hill, 1952.

[160] M. Gustafsson. A spell checker with a user model for Swedish dyslexics. *Projektarbeten 2003*, page 69, 2003.

[161] C. Gutwin, G. Paynter, I. Witten, C. Nevill-Manning, and E. Frank. Improving browsing in digital libraries with keyphrase indexes. *Decision Support Systems*, 27(1):81–104, 1999.

[162] M. Hagiwara. Self-organizing concept maps. In *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on*, volume 1, pages 447–451. IEEE, 1995.

[163] B. E. Hagtvet. Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and writing*, 16(6):505–539, 2003.

[164] S. Hargreaves. *Study skills for dyslexic students*. Sage, 2007.

[165] M. Healey. Developing an inclusive curriculum for ALL our students, 2010. Presentation to Macquarie University, Australia.

[166] J. M. Henderson and F. Ferreira. Effects of foveal processing difficulty on the perceptual span in reading: implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3):417, 1990.

[167] A. Herrera and P. Macizo. ¿Cómo leemos los números? (How do we read numbers?). *Ciencia Cognitiva (Cognitive Science)*, 6(2):44–47, 2012.

[168] R. Hillier. Sylexiad: A typeface for the adult dyslexic reader. *Journal of Writing in Creative Practice*, 1(3):275–291, 2008.

[169] C. S. H. Ho, D. W. O. Chan, S. M. Tsang, and S. H. Lee. The cognitive profile and multiple-deficit hypothesis in Chinese developmental dyslexia. *Developmental Psychology*, 38(4):543, 2002.

[170] D. Holbrook. *English for the Rejected: Training Literacy in the Lower Streams of the Secondary School*. Cambridge University Press, Cambridge, 1964.

[171] V. Holmes and J. Carruthers. The relation between reading and spelling in skilled adult readers. *Journal of memory and language*, 39(2):264–289, 1998.

[172] B. Hornsby. *Overcoming dyslexia*. Martin Dunitz, London, 1986.

[173] 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.

[174] J. Hyönä and R. K. 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.

[175] A. W. Inhoff and K. Rayner. Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40(6):431–439, 1986.

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

[177] 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.

[178] International Phonetic Association. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press, Cambridge, 1999.

[179] Internet World Stats. Usage and population statistics, April 2011. `http://www.internetworldstats.com`.

[180] K. Inui, A. Fujita, T. Takahashi, R. Iida, and T. Iwakura. Text simplification for reading assistance: A project note. In *Proceedings of the second international workshop on Paraphrasing-Volume 16*, pages 9–16. Association for Computational Linguistics, 2003.

[181] Y. Ito and T. Hatta. Semantic processing of Arabic, Kanji and Kana numbers: Evidence from interference in physical and numerical size judgments. *Memory & Cognition*, 31:360–368, 2003.

[182] R. Jeanes, A. Busby, J. Martin, E. Lewis, N. Stevenson, D. Pointon, and A. Wilkins. Prolonged use of coloured overlays for classroom reading. *British Journal of Psychology*, 88(4):541–548, 1997.

[183] 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.

[184] R. L. Johnson, M. Perea, and K. Rayner. Transposed-letter effects in reading: Evidence from eye movements and parafoveal preview. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1):209, 2007.

[185] D. Jurafsky, A. Bell, M. Gregory, and W. D. Raymond. Evidence from reduction in lexical production. *Frequency and the emergence of linguistic structure*, 45:229, 2001.

[186] M. A. Just and P. A. Carpenter. A theory of reading: From eye fixations to comprehension. *Psychological review*, 87:329–354, 1980.

[187] G. Kanvinde and S. Gupta. Accessiblenews daisy: newspapers in daisy. In *Proc. International Cross Disciplinary Conference on Web Accessibility (W4A 2011)*, New York, NY, USA, March 2011. ACM Press.

[188] 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.

[189] S. Katusic, R. Colligan, W. Barbaresi, D. Schaid, and S. J. Jacobsen. Incidence of reading disability in a population-based birth cohort, 1976-1982, rochester. *Mayo Clinic Proceedings*, 76(11):1081, 2001.

[190] J. M. Keenan, R. S. Betjemann, and R. K. Olson. Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3):281–300, 2008.

[191] S. N. Kim, O. Medelyan, M.-Y. Kan, and T. Baldwin. Semeval-2010 task 5: Automatic keyphrase extraction from scientific articles. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 21–26. Association for Computational Linguistics, 2010.

[192] J. Kiraly and P. M. Ridge. Method of and apparatus for multi-modal information presentation to computer users with dyslexia, reading disabilities or visual impairment, 2001. US Patent 6,324,511.

[193] R. Klein, G. Berry, K. Briand, B. D'Entremont, and M. Farmer. Letter identification declines with increasing retinal eccentricity at the same rate for normal and dyslexic readers. *Perception & Psychophysics*, 47(6):601–606, 1990.

[194] D. E. Knuth. *The Computer Modern Fonts, Volume E of Computers and Typesetting.* Addison-Wesley, Reading, Massachusetts, 1986.

[195] E. Kolatch. Designing for users with cognitive disabilities. *Stand*, 5:2007, 2000.

[196] T. Korhonen. Adaptive spell checker for dyslexic writers. In *Proceedings of the 11th international conference on Computers Helping People with Special Needs (ICCHP '08)*, pages 733–741, Berlin, Heidelberg, 2008. Springer.

[197] A. W. Kotula. Matching readers to instructional materials: The use of classic readability measures for students with language learning disabilities and dyslexia. *Topics in Language Disorders*, 23(3):190, 2003.

[198] I. Kriss and B. J. W. Evans. The relationship between dyslexia and Meares-Irlen Syndrome. *Journal of Research in Reading*, 28(3):350–364, 2005.

[199] 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.

[200] T. Lachmann and C. Van Leeuwen. Different letter-processing strategies in diagnostic subgroups of developmental dyslexia. *Cognitive Neuropsychology*, 25(5):730–744, 2008.

[201] A. J. Lait and B. Randell. An assessment of name matching algorithms. *Technical Report Series-University of Newcastle Upon Tyne Computing Science*, 1996.

[202] P. Lal and S. Ruger. Extract-based summarization with simplification. In *Proceedings of the ACL 2002 Automatic Summarization / DUC 2002 Workshop*, 2002.

[203] K. Landerl, A. Bevan, B. Butterworth, et al. Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition*, 93(2):99–125, 2004.

[204] K. Landerl, B. Fussenegger, K. Moll, and E. Willburger. Dyslexia and dyscalculia: two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology*, 103(3):309, 2009.

[205] D. Lawrie, W. B. Croft, and A. Rosenberg. Finding topic words for hierarchical summarization. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 349–357. ACM Press, 2001.

[206] M. Leahy. Spelling, spelling-checkers and dyslexia. In *CESI Conference. St. Patrick's College (January 2002)*, 2002.

[207] C.-H. Lee, G.-G. Lee, and Y. Leu. Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning. *Expert Systems with Applications*, 36(2):1675–1684, 2009.

[208] L. A. Lefton, R. J. Nagle, G. Johnson, and D. F. Fisher. Eye movement dynamics of good and poor readers: Then and now. *Journal of Literacy Research*, 11(4):319–328, 1979.

[209] W. J. Levelt and L. Wheeldon. Do speakers have access to a mental syllabary? *Cognition*, 50(1):239–269, 1994.

[210] V. Levenshtein. Binary codes capable of correcting spurious insertions and deletions of ones. *Problems of Information Transmission*, 1:8–17, 1965.

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

[212] M. P. Lewis, editor. *Ethnologue: Languages of the World.* SIL International, Dallas, Texas, sixteenth edition, 2009.

[213] A. Q. Li, L. Sbattella, and R. Tedesco. Polispell: an adaptive spellchecker and predictor for people with dyslexia. In S. Carberry, S. Weibelzahl, A. Micarelli, and G. Semeraro, editors, *User Modeling, Adaptation, and Personalization*, pages 302–309. Springer, Berlin, Heidelberg, 2013.

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

[215] S. A. Lindgrén and M. Laine. Multilingual dyslexia in university students: Reading and writing patterns in three languages. *Clinical Linguistics & Phonetics*, 25(9):753–766, 2011.

[216] J. Llisterri and J. B. Mariño. Spanish adaptation of SAMPA and automatic phonetic transcription. Technical report, ESPRIT project 6819 SAM-A Speech Technology Assessment in Multilingual Applications, 1993.

[217] S. Lockley. Dyslexia and higher education: accessibility issues. *The Higher Education Academy*, 2002.

[218] P. Longman. *Longman dictionary of contemporary English*. Pearson Longman, 2009.

[219] M. R. López Castro. Intervención educativa en un caso real de problemas de comprensión lectora (Educative intervention in a real case of reading comprehension problems). *Hekademos: revista educativa digital (Digital Educative Magazine)*, 6:27–48, 2010.

[220] G. R. Lyon. Toward a definition of dyslexia. *Annals of Dyslexia*, 45(1):3–27, 1995.

[221] G. R. Lyon, J. M. Fletcher, and M. Barnes. *Child psychopathology*, chapter Learning disabilities, pages 520 – 588. Guilford, New York, 2002.

[222] G. R. Lyon, S. E. Shaywitz, and B. A. Shaywitz. A definition of dyslexia. *Annals of Dyslexia*, 53(1):1–14, 2003.

[223] J. Lyons. *Semantics*, volume 2. Cambridge Univ Pr, 1977.

[224] M. J. Machuca. Problemas de pronunciación (Pronunciation problems). In S. Alcoba, editor, *La expresión oral (The oral expression)*, pages 71–88, Barcelona, 2000. Ariel.

[225] N. Madnani and B. J. Dorr. Generating phrasal and sentential paraphrases: A survey of data-driven methods. *Computational Linguistics*, 36(3):341–387, 2010.

[226] F. J. Martos and J. Vila. Differences in eye movements control among dyslexic, retarded and normal readers in the Spanish population. *Reading and Writing*, 2(2):175–188, 1990.

[227] A. Max. Writing for language-impaired readers. *Computational Linguistics and Intelligent Text Processing*, pages 567–570, 2006.

[228] 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.

[229] J. L. McClelland. Preliminary letter identification in the perception of words and non-words. *Journal of Experimental Psychology: Human Perception and Performance*, 2(1):80–91, 1976.

[230] M. McCloskey, A. Caramazza, and A. Basili. Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brian and Cognition*, 4:171–196, 1985.

[231] G. H. McLaughlin. Smog grading: A new readability formula. *Journal of Reading*, 12(8):639–646, 1969.

[232] O. Medelyan and I. H. Witten. Thesaurus based automatic keyphrase indexing. In *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, pages 296–297. ACM Press, 2006.

[233] H. Meng, S. Smith, K. Hager, M. Held, J. Liu, R. Olson, B. Pennington, J. DeFries, J. Gelernter, T. O'Reilly-Pol, S. Somlo, P. Skudlarski, S. Shaywitz, B. Shaywitz, K. Marchione, Y. Wang, P. Murugan, J. LoTurco, P. Grier, and J. Gruen. DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proceedings of the National Academy of Sciences*, 102:17053–17058, November 2005.

[234] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, The Google Books Team, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, and M. Nowak. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182, 2011.

[235] R. Mihalcea and P. Tarau. Textrank: Bringing order into texts. In *Proceedings of EMNLP*, volume 4. Barcelona, Spain, 2004.

[236] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–244, 1990.

[237] D. C. Mitchell and D. W. Green. The effects of context and content on immediate processing in reading. *The Quarterly Journal of Experimental Psychology*, 30:609–636, 1978.

[238] 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.

[239] R. Mitton. *English spelling and the computer.* Longman, Harlow, 1996.

[240] L. C. Moats. Phonological spelling errors in the writing of dyslexic adolescents. *Reading and Writing*, 8(1):105–119, 1996.

[241] K. Mori and R. Komatsu. Reading words with jumbled characters in Japanese. *Perceptual and motor skills*, 112(3):947–950, 2011.

[242] R. K. Morris and K. Rayner. Eye movements in skilled reading: Implications for developmental dyslexia. *Vision and visual dyslexia*, pages 233–242, 1991.

[243] J. Morton. Word recognition. *Psycholinguistic Series*, 2:107–156, 1979.

[244] L. Mosberg and D. Johns. Reading and listening comprehension in college students with developmental dyslexia. *Learning Disabilities Research & Practice*, 1994.

[245] K. Nandhini and S. R. Balasundaram. Improving readability of dyslexic learners through document summarization. In *IEEE International Conference on Technology for Education (T4E)*, pages 246–249. IEEE, 2011.

[246] K. Nation and M. J. Snowling. Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development*, 69(4):996–1011, 1998.

[247] U. Nations. Standard Rules on the Equalization of Opportunities for Persons with Disabilities, 1994.

[248] A. F. Newell and L. Booth. The use of lexical and spelling aids with dyslexics. *Computers and Literacy*, pages 35–44, 1991.

[249] R. I. Nicolson and A. J. Fawcett. Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex*, 47(1):117–127, 2011.

[250] J. Nielsen. Lower-literacy users, 2012. Jakob Nielsen's Alertbox. `http://www.useit.com/alertbox/20050314.html`.

[251] J. Nielsen. F-shaped pattern for reading web content. *Last visited Feb 22, 2013*, 2006.

[252] J. Nielsen. Top 10 mistakes in web design. *Last visited Sep 07, 2013*, 2006.

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

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

[255] J. Nielsen and K. Pernice. *Eye tracking web usability*. New Riders Pub, 2010.

[256] B. A. O'Brien, J. S. Mansfield, and G. E. Legge. The effect of print size on reading speed in dyslexia. *Journal of Research in Reading*, 28(3):332–349, 2005.

[257] A. M. Olney, W. L. Cade, and C. Williams. Generating concept map exercises from textbooks. In *Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 111–119. Association for Computational Linguistics, 2011.

[258] R. K. Olson, R. Kliegl, and B. J. Davidson. Dyslexic and normal readers' eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 9(5):816, 1983.

[259] R. K. Olson, F. A. Conners, and J. P. Rack. Eye movements in dyslexic and normal readers. *Vision and visual dyslexia*, pages 243–250, 1991.

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

[261] M. G. Paciello. *Web accessibility for people with disabilities*. CMP Books, Lawrence, Kansas, 2000.

[262] L. Padró, M. Collado, S. Reese, M. Lloberes, and I. Castellón. Freeling 2.1: Five years of open-source language processing tools. In *Proc. LREC'10*, Valletta, Malta, May 2010.

[263] M. Padró, M. Ballesteros, H. Martínez, and B. Bohnet. Finding dependency parsing limits over a large Spanish corpus. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, Nagoya, Japan, October 2013.

[264] D. G. Paterson and M. A. Tinker. Studies of typographical factors influencing speed of reading. ii. size of type. *Journal of Applied Psychology*, 13(2):120, 1929.

[265] D. G. Paterson and M. A. Tinker. Studies of typographical factors influencing speed of reading. viii. space between lines or leading. *Journal of Applied Psychology*, 16(4):388, 1932.

[266] K. Patterson, J. C. Marshall, and M. Coltheart. *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading.* Lawrence Erlbaum Associates, London, 1985.

[267] K. Patterson, M. S. Seidenberg, and J. L. McClelland. Connections and disconnections: acquired dyslexia in a computational model of reading processes. In R. G. M. Morris, editor, *Parallel distributed processing: Implications for psychology and neurobiology*, Oxford Psychology Series, pages 131–181. Clarendon Press/Oxford University Press, 1989.

[268] K. Patterson, T. Suzuki, T. Wydell, and S. Sasanuma. Progressive aphasia and surface alexia in Japanese. *Neurocase*, 1995.

[269] K. E. Patterson. Neuropsychological approaches to the study of reading. *British Journal of Psychology*, 72(2):151–174, 1981.

[270] K. E. Patterson. The relation between reading and phonological coding: Further neuropsychological observations. *Neurocase*, 1(3):251–257, 1995.

[271] E. Paulesu, J.-F. Démonet, F. Fazio, E. McCrory, V. Chanoine, N. Brunswick, S. F. Cappa, G. Cossu, M. Habib, C. D. Frith, and U. Frith. Dyslexia: Cultural diversity and biological unity. *Science*, 291(5511):2165–2167, November 2001.

[272] G. T. Pavlidis. The dyslexics erratic eye movements: Case studies. *Dyslexia review*, 1:22–28, 1978.

[273] G. T. Pavlidis. Do eye movements hold the key to dyslexia? *Neuropsychologia*, 19(1):57–64, 1981.

[274] G. T. Pavlidis. The “dyslexia syndrome” and its objective diagnosis by erratic eye movements. *Eye movements in reading: Perceptual and language processes*, pages 441–466, 1983.

[275] G. T. Pavlidis. Diagnostic significance and relationship between dyslexia and erratic eye movements. *Vision and visual dyslexia*, pages 263–270, 1991.

[276] J. Pedler. The detection and correction of real-word spelling errors in dyslexic text. In *Proceedings of the 4th Annual CLUK Colloquium*. CLUK, 2001.

[277] J. Pedler. *Computer Correction of Real-word Spelling Errors in Dyslexic Text*. PhD thesis, Birkbeck College, London University, 2007.

[278] J. Pedler. Computer spellcheckers and dyslexics - a performance survey. *British Journal of Educational Technology*, 32(1):23–37, 2001.

[279] M. Pedley. Designing for dyslexics: Part 3, 2006. `http://accessites.org/site/2006/11/designing-for-dyslexics-part-3-of-3`.

[280] L. Peer and G. Reid. *Dyslexia: Successful inclusion in the secondary school*. Routledge, 2001.

[281] M. Perea, J. Acha, and M. Carreiras. Eye movements when reading text messaging (txt msgng). *The Quarterly Journal of Experimental Psychology*, 62(8):1560–1567, 2009.

[282] M. A. Pérez, J. R. Alameda Bailén, and F. Cuetos Vega. Frecuencia, longitud y vecindad ortográfica de las palabras de 3 a 16 letras del diccionario de la lengua española (Frequency, length and orthographical neighborhood of the words from 3 to 16 letters of the dictionary of the Spanish language - RAE, 1992). *REMA: Revista Electrónica de Metodología Aplicada (Electronic Journal of Applied Methodology)*, 8(2):1–10, 2003.

[283] C. Perfetti and L. Hart. *Precursors of functional literacy*, chapter The lexical quality hypothesis, pages 189–213. Amsterdam/Philadelphia: John Benjamins, 2002.

[284] C. Perron. Colour choices on web pages: Contrast vs readability, 2012. `http://www.writer2001.com/index.htm`.

[285] A. R. Perry, W. N. Dember, J. S. Warm, and J. G. Sacks. Letter identification in normal and dyslexic readers: A verification. *Bulletin of the Psychonomic Society*, 27(5):445–448, 1989.

[286] H. L. 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.

[287] F. J. Pirozzolo and K. Rayner. The neural control of eye movements in acquired and developmental reading disorders. *Studies in Neurolinguistics*, 4:1–27, 1978.

[288] F. J. Pirozzolo and K. Rayner. The neural control of eye movements in acquired and developmental reading disorders. *Studies in Neurolinguistics*, 4:97–123, 1979.

[289] J. Piskorski, M. Sydow, and D. Weiss. Exploring linguistic features for web spam detection: a preliminary study. In *Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb '08)*, pages 25–28, New York, NY, 2008. ACM Press.

[290] D. Pollak. *Dyslexia, the self and higher education: Learning life histories of students identified as dyslexic.* Trentham Books Ltd, 2005.

[291] A. Pollatsek. What can eye movements tell us about dyslexia. *Eye movements in reading: Perceptual and language processes*, pages 511–522, 1983.

[292] C. Power, H. Petrie, D. Swallow, E. Murphy, B. Gallagher, and C. A. Velasco. Navigating, discovering and exploring the web: Strategies used by people with print disabilities on interactive websites. In *Human-Computer Interaction–INTERACT 2013*, pages 667–684. Springer, 2013.

[293] R. Power and S. Williams. Generating numerical approximations. *Computational Linguistics*, 38(1), 2012.

[294] R Development Core Team. A language and environment for statistical computing (computer software)., 2011. `http://www.R-project.org`.

[295] J. P. Rack, M. J. Snowling, and R. K. Olson. The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly*, pages 29–53, 1992.

[296] P. Rainger. A dyslexic perspective on e-content accessibility, 2012. `www.texthelp.com/media/39360/USDyslexiaAndWebAccess.PDF`.

[297] K. Rajaraman and A.-H. Tan. Knowledge discovery from texts: a concept frame graph approach. In *Proceedings of the eleventh international Conference on Information and Knowledge Management*, pages 669–671. ACM, 2002.

[298] F. Ramírez and E. López. Spelling error patterns in Spanish for word processing applications. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 93–98, Genoa, Italy, 2006.

[299] D. M. Ramírez Sánchez. Estrategias de intervención educativa con el alumnado con dislexia (Strategied for educative intervention for dyslexic students). *Innovación y experiencias educativas (Innovation and educative experiences)*, 49, 2011.

[300] G. E. Raney and K. Rayner. Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology*, 49(2):151, 1995.

[301] G. Rawlinson. Reibadailty. *New Scientist*, 162(2188):55–55, 1999.

[302] G. E. Rawlinson. *The Significance of Letter Position in Word Recognition*. PhD thesis, University of Nottingham, 1976.

[303] J. E. Raymond, N. A. Ogden, J. E. Fagan, and B. J. Kaplan. Fixational instability in dyslexic children. *American Journal of Optometry and Physiological Optics*, 65:174–181, 1988.

[304] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372–422, 1998.

[305] K. Rayner and S. A. Duffy. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3):191–201, 1986.

[306] K. Rayner. Eye movements, perceptual span, and reading disability. *Annals of Dyslexia*, 33(1):163–173, 1983.

[307] K. Rayner. The role of eye movements in learning to read and reading disability. *Remedial and Special Education*, 6(6):53–60, 1985.

[308] K. Rayner. Eye movements and the perceptual span in beginning and skilled readers. *Journal of experimental child psychology*, 41(2):211–236, 1986.

[309] K. Rayner, L. A. Murphy, J. M. Henderson, and A. Pollatsek. Selective attentional dyslexia. *Cognitive Neuropsychology*, 6(4):357–378, 1989.

[310] K. Rayner, A. Pollatsek, and A. B. Bilsky. Can a temporal processing deficit account for dyslexia? *Psychonomic Bulletin & Review*, 2(4):501–507, 1995.

[311] K. Rayner and G. E. Raney. Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, 3(2):245–248, 1996.

[312] K. Rayner, S. J. White, R. L. Johnson, and S. P. Liversedge. Raeding wrods with jubmled lettres there is a cost. *Psychological Science*, 17(3):192–193, 2006.

[313] Real Academia Española. *Diccionario de la lengua española (Dictionary of the Spanish Language)*. Espasa-Calpe, Madrid, 22nd edition, 2001.

[314] Real Academia Española. *Diccionario panhispánico de dudas (Pan-Hispanic Dictionary of Doubts)*. Santillana, Madrid, 2005.

[315] L. Rello and R. Baeza-Yates. Lexical quality as a measure for textual web accessibility. In *Online Symposium on Website Accessibility Metrics.*, December 2011.

[316] 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.

[317] 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.

[318] L. Rello and R. Baeza-Yates. Social media is not that bad! The lexical quality of social media (short paper). In *The International AAAI Conference on Weblogs and Social Media (ICWSM).*, pages 559–562, Dublin, Ireland, June 2012.

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

[320] L. Rello and R. Baeza-Yates. Evaluation of Dyswebxia: A reading app designed for people with dyslexia. In *Proc. W4A '14*, Seoul, Korea, 2014.

[321] 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.

[322] 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.

[323] L. Rello, R. Baeza-Yates, and J. Llisterri. Dyslist: An annotated resource of dyslexic errors. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, pages 1289–1296, Reykjavik, Iceland, May 2014.

[324] L. Rello, R. Baeza-Yates, and H. Saggion. The impact of lexical simplification by verbal paraphrases for people with and without dyslexia. In *Proc. CICLing'13*, pages 501–512. Springer Berlin Heidelberg, 2013.

[325] 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.

[326] L. Rello, R. Baeza-Yates, H. Saggion, S. Bott, R. Carlini, C. Bayarri, A. Gòrriz, 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.

[327] L. Rello, R. Baeza-Yates, H. Saggion, and E. Graells. Graphical schemes may improve readability but not understandability for people with dyslexia. In *Proceedings of the NAACL HLT 2012 Workshop Predicting and improving text readability for target reader populations (PITR 2012)*, Montreal, Canada, 2012.

[328] L. Rello, R. Baeza-Yates, H. Saggion, and J. Pedler. A first approach to the creation of a Spanish corpus of dyslexic texts. In *LREC Workshop Natural Language Processing for Improving Textual Accessibility (NLP4ITA)*, pages 22–27, Istanbul, Turkey, May 2012.

[329] 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.

[330] L. Rello, C. Bayarri, and A. Gorriz. What is wrong with this word? Dyseggxia: a game for children with dyslexia (demo). In *Proceedings of the 14th international ACM SIGACCESS conference on computers and accessibility (ASSETS '12)*, pages 219–220, Boulder, USA, October 2012. ACM Press.

[331] 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.

[332] L. Rello and J. Llisterri. There are phonetic patterns in vowel substitution errors in texts written by persons with dyslexia. In *21st Annual World Congress on Learning Disabilities (LDW 2012)*, pages 327–338, Oviedo, Spain, September 2012.

[333] 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.

[334] L. Rello, H. Saggion, and R. Baeza-Yates. Keyword highlighting improves comprehension for people with dyslexia. In *Proceedings of the EACL 2014 Workshop Predicting and improving text readability for target reader populations (PITR 2014)*, Gothenburg, Sweden, 2014.

[335] L. Rello and M.-C. Marcos. An eye tracking study on text customization for user performance and preference. In *Eighth Latin American Web Congress (LA-WEB)*, pages 64–70. IEEE, 2012.

[336] C. Ringlstetter, K. U. Schulz, and S. Mihov. Orthographic errors in web pages: Towards cleaner web corpora. *Computational Linguistics*, 2006.

[337] C. Romani, J. Ward, and A. Olson. Developmental surface dysgraphia: What is the underlying cognitive impairment? *The Quarterly Journal of Experimental Psychology*, 52(1):97–128, 1999.

[338] H. Rubenstein, L. Garfield, and J. A. Millikan. Homographic entries in the internal lexicon1. *Journal of Verbal Learning and Verbal Behavior*, 9(5):487–494, 1970.

[339] J. Rüsseler, S. Probst, S. Johannes, and T. F. 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.

[340] R. C. Russell and M. Odell. The Soundex coding system. *US Patents*, 1261167, 1918.

[341] 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 (Journal of the Spanish Society for Natural Language Processing)*, 47, 2011.

[342] H. Saggion, E. G. 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 (Journal of the Spanish Society for Natural Language Processing)*, 47:341–342, 2011.

[343] H. Saggion, S. Bott, and L. Rello. Comparing resources for Spanish lexical simplification. In *SLSP 2013: Statistical Language and Speech Processing*, volume 7978 of *Lecture Notes in Computer Science*, pages 236–247, 2013.

[344] H. Saito, T. Ohno, F. Ozaki, K. Saito, T. Maeda, and A. Ohuchi. A semi-automatic construction method of concept map based on dialog contents. In *International conference on computers in education*, 2001.

[345] M. P. Salguero and J. R. Alameda. El procesamiento de los números y sus implicaciones educativas (Processing numbers and its educative implications). *XXI Revista de Educación (Education Journal)*, 5:181–189, 2003.

[346] L. Sánchez Sampedro. *Impostores (Impostors)*. Crowdfunding, Madrid, Spain, 2012. `http://www.impostores.es`.

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

[348] M. H. Schneps, J. M. Thomson, G. Sonnert, M. Pomplun, C. Chen, and A. Heffner-Wong. Shorter lines facilitate reading in those who struggle. *PloS ONE*, 8(8):e71161, 2013.

[349] R. E. Schuberth and P. D. Eimas. Effects of context on the classification of words and non-words. *Journal of Experimental Psychology: Human Perception and Performance*, 3(1):27–36, 1977.

[350] G. Schulte-Körne, W. Deimel, K. Müller, C. Gutenbrunner, and H. Remschmidt. Familial aggregation of spelling disability. *Journal of Child Psychology and Psychiatry*, 37(7):817–822, 1996.

[351] R. J. Senter. Automated readability index. Technical report, DTIC Document, 1967.

[352] S. C. Sereno and K. Rayner. Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7(11):489–493, 2003.

[353] F. Serrano and S. Defior. Dyslexia speed problems in a transparent orthography. *Annals of Dyslexia*, 58(1):81–95, 2008.

[354] P. H. K. Seymour, M. Aro, and J. M. Erskine. Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2):143–174, 2003.

[355] A. D. Shaikh. The effects of line length on reading online news. *Usability News*, 7(2):1–4, 2005.

[356] S. E. Shaywitz and B. A. Shaywitz. Dyslexia (specific reading disability). *Biological Psychiatry*, 57(11):1301–1309, 2005.

[357] S. Shaywitz, M. Escobar, B. Shaywitz, J. Fletcher, and R. Makuch. Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, 326(3):145–150, 1992.

[358] F. Simmons and C. Singleton. The reading comprehension abilities of dyslexic students in higher education. *Dyslexia*, 6(3):178–192, 2000.

[359] R. C. Sinatra, J. Stahl-Gemake, and D. N. Berg. Improving reading comprehension of disabled readers through semantic mapping. *The Reading Teacher*, 38(1):22–29, 1984.

[360] L. Sitbon and P. Bellot. A readability measure for an information retrieval process adapted to dyslexics. In *Second international workshop on Adaptive Information Retrieval (AIR 2008)(in conjunction with IIiX 2008)*, pages 52–57, 2008.

[361] W. Slaghuis, W. Lovegrove, and J. Freestun. Letter recognition in peripheral vision and metacontrast making in dyslexic and normal readers. *Clinical vision sciences*, 7(1):53–65, 1992.

[362] I. Smythe. *Dyslexia in the Digital Age: Making IT Work*. Continuum, London, New York, 2010.

[363] M. Song, I. Y. Song, R. B. Allen, and Z. Obradovic. Keyphrase extraction-based query expansion in digital libraries. In *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, pages 202–209. ACM Press, 2006.

[364] J. F. Sowa. *Conceptual structures: information processing in mind and machine*. Addison-Wesley Pub., Reading, MA, 1983.

[365] 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.

[366] D. Sperber and D. Wilson. *Relevance: Communication and cognition*, volume 142. Harvard University Press Cambridge, MA, 1986.

[367] D. Spinelli, M. De Luca, G. Di Filippo, M. Mancini, M. Martelli, and P. Zoccolotti. Length effect in word naming in reading: Role of reading experience and reading deficit in italian readers. *Developmental Neuropsychology*, 27(2):217–235, 2005.

[368] R. Spooner. *A spelling aid for dyslexic writers*. PhD thesis, University of York, 1998.

[369] G. Stanley, G. A. Smith, and E. A. Howell. Eye-movements and sequential tracking in dyslexic and control children. *British Journal of Psychology*, 74(2):181–187, 1983.

[370] G. Stanley. *Eye movements in reading*, chapter Eye movements in dyslexic and normal children, pages 261–271. Pergamon Press, Oxford, England, 1994.

[371] C. Sterling, M. Farmer, B. Riddick, S. Morgan, and C. Matthews. Adult dyslexic writing. *Dyslexia*, 4(1):1–15, 1998.

[372] S. J. 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.

[373] J. Sykes. *Dyslexia, design and reading: Making print work for college students with dyslexia. A qualitative interaction design study*. ProQuest, 2008.

[374] K. Taghva and E. Stofsky. Ocrspell: an interactive spelling correction system for ocr errors in text. *International Journal on Document Analysis and Recognition*, 3(3):125–137, 2001.

[375] P. Tapanainen and T. Järvinen. A non-projective dependency parser. In *Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP-97)*, pages 64–71, 1997.

[376] C. M. Temple. Developmental analogues to acquired phonological dyslexia. In *Dyslexia: a global issue*. Martinus Nijhoff Publishers, La Haya, 1984.

[377] C. M. Temple and J. C. Marshall. A case study of developmental phonological dyslexia. *British Journal of Psychology*, 74(4):517–533, 1983.

[378] M. A. Tinker. The study of eye movements in reading. *Psychological Bulletin*, 43(2):93, 1946.

[379] M. A. Tinker. Recent studies of eye movements in reading. *Psychological Bulletin*, 55(4):215, 1958.

[380] M. A. Tinker. Influence of simultaneous variation in size of type, width of line, and leading for newspaper type. *Journal of Applied Psychology*, 47(6):380, 1963.

[381] Tobii Technology. Product description of Tobii 50 Series, 2005.

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

[383] V. Topac. *Improving text accessibility and understanding of domain-specific information*. PhD thesis, Universitatea Politehnica Timişoara, 2014.

[384] V. Topac and V. Stoicu-Tivadar. Patient empowerment by increasing the understanding of medical language for lay users. *Methods Inf. Med.*, 52(5):454–62, 2013.

[385] J. Toro and M. Cervera. *TALE: Test de Análisis de Lectoescritura (TALE: Literacy Analysis Test)*. Visor, Madrid, 1984.

[386] R. Treiman. Spelling in normal children and dyslexics. In B. A. Blachman, editor, *Foundations of reading acquisition and dyslexia: Implications for early intervention*, pages 191–218. Lawrence Erlbaum, Mahwah, NJ, 1997.

[387] P. E. Tressoldi, G. Stella, and M. Faggella. The development of reading speed in Italians with dyslexia a longitudinal study. *Journal of learning disabilities*, 34(5):414–417, 2001.

[388] A. Tseng. Six surprising bad practices that hurt dyslexic users, 2012. `http://uxmovement.com/content/6-surprising-bad-practices-that-hurt-dyslexic-users/`.

[389] Y.-H. Tseng, C.-Y. Chang, S.-N. C. Rundgren, and C.-J. Rundgren. Mining concept maps from news stories for measuring civic scientific literacy in media. *Computers & Education*, 55(1):165–177, 2010.

[390] P. Turney. Learning to extract keyphrases from text. In *National Research Council, Institute for Information Technology, Technical Report ERB-1057*, 1999.

[391] N. R. Underwood and D. Zola. The span of letter recognition of good and poor readers. *Reading Research Quarterly*, pages 6–19, 1986.

[392] United Nations Committee of the General Assembly. *Convention on the Rights of Persons with Disabilities.* United Nations, New York, 2006.

[393] D. Valerio and A. Leake. Jump-starting concept map construction with knowledge extracted from documents. In *Proceedings of the Second International Conference on Concept Mapping (CMC*, 2006.

[394] F. Valle. Reading errors in Spanish. In *Reading and writing problems in different systems.* Kluwer, 1988.

[395] W. J. B. Van Heuven, T. Dijkstra, and J. Grainger. Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39(3):458–483, 1998.

[396] H. Velan and R. Frost. Cambridge university versus Hebrew university: The impact of letter transposition on reading English and Hebrew. *Psychonomic Bulletin & Review*, 14(5):913–918, 2007.

[397] F. R. Vellutino. *Dyslexia. Theory and Research.* The MIT Press, Cambridge, MA, 1979.

[398] F. R. Vellutino, J. M. Fletcher, M. J. Snowling, and D. M. Scanlon. Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45(1):2–40, 2004.

[399] F. R. Vellutino. Dyslexia. *Scientific American*, 1987.

[400] F. R. Vellutino and D. M. Scanlon. Verbal processing in poor and normal readers. In *Verbal processes in children*, pages 189–264. Springer, 1982.

[401] J. Villalón and R. A. Calvo. Concept extraction from student essays, towards concept map mining. In *Ninth IEEE International Conference on Advanced Learning Technologies, 2009. (ICALT 2009)*, pages 221–225. IEEE, 2009.

[402] P. Vossen. EuroWordNet: A multilingual database with lexical semantic networks. *International Journal of Lexicography*, 17(2):161–173, 2004.

[403] W. M. Wang, C. F. Cheung, W. B. Lee, and S. K. Kwok. Mining knowledge from natural language texts using fuzzy associated concept mapping. *Information Processing & Management*, 44(5):1707–1719, 2008.

[404] E. Warrington and O. L. Zangwill. A study of dyslexia. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20(3):208, 1957.

[405] P. A. Weaver. Comprehension, recall, and dyslexia: A proposal for the application of schema theory. *Annals of Dyslexia*, 28(1):92–113, 1978.

[406] WebAIM. Web accessibility in mind, February 2014. `http://webaim.org/`.

[407] C. E. Weinstein and R. E. Mayer. The teaching of learning strategies. *Handbook of research on teaching*, 3:315–327, 1986.

[408] J. C. Wells. *Computer-coding the IPA: a proposed extension of SAMPA*. Division of Psychology and Language Sciences, University College London, London, 2000.

[409] T. G. West. *In the Mind's Eye: Creative Visual Thinkers, Gifted Dyslexics, and the Rise of Visual Technologies*. Prometheus Books, 2009.

[410] S. J. White, R. L. Johnson, S. P. Liversedge, and K. Rayner. Eye movements when reading transposed text: the importance of word-beginning letters. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5):1261, 2008.

[411] Wikipedia. The free encyclopedia, April 2011. `http://www.wikipedia.org`.

[412] S. Williams and E. Reiter. Generating basic skills reports for low-skilled readers. *Natural Language Engineering*, 1:1–32, 2008.

[413] S. Williams, E. Reiter, and L. Osman. Experiments with discourse-level choices and readability. In *Proc. ENLG '03)*, Budapest, Hungary, 2003.

[414] H. Wimmer. Characteristics of developmental dyslexia in a regular writing system. *Applied psycholinguistics*, 14(1), 1993.

[415] H. Wimmer. The nonword reading deficit in developmental dyslexia: Evidence from children learning to read German. *Journal of Experimental Child Psychology*, 61(1):80–90, 1996.

[416] I. H. Witten, G. W. Paynter, E. Frank, C. Gutwin, and C. G. Nevill-Manning. Kea: Practical automatic keyphrase extraction. In *Proceedings of the fourth ACM conference on Digital Libraries*, pages 254–255. ACM Press, 1999.

[417] World Health Organization. *International statistical classification of diseases, injuries and causes of death (ICD-10).* World Health Organization, Geneva, 10th edition, 1993.

[418] World Wide Web Consortium (W3C). Web accessibility initiative (WAI), July 2011. `http://www.w3.org/WAI/`.

[419] E. J. Yannakoudakis and D. Fawthrop. The rules of spelling errors. *Information Processing & Management*, 19(2):87–99, 1983.

[420] T. Yarkoni, D. Balota, and M. Yap. Moving beyond Coltheart's *n*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15(5):971–979, 2008.

[421] 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.

[422] O. L. Zangwill and C. Blakemore. Dyslexia: Reversal of eye-movements during reading. *Neuropsychologia*, 10(3):371–373, 1972.

[423] V. Zarach. Ten guidelines for improving accessibility for people with dyslexia, 2012. CETIS University of Wales, Bangor. http://wiki.cetis.ac.uk/Ten_Guidelines_for_Improving_Accessibility_for_People_with_Dyslexia.

[424] 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.

[425] A. Zouaq, D. Gasevic, and M. Hatala. Ontologizing concept maps using graph theory. In *Proceedings of the 2011 ACM Symposium on applied computing*, pages 1687–1692. ACM, 2011.

[426] A. Zouaq and R. Nkambou. Building domain ontologies from text for educational purposes. *Learning Technologies, IEEE Transactions on*, 1(1):49–62, 2008.

[427] K. Zubrinic, D. Kalpic, and M. Milicevic. The automatic creation of concept maps from documents written using morphologically rich languages. *Expert Systems with Applications*, 39(16):12709–12718, 2012.

Appendix

A.1 Errors used for *Errors I* and *Errors II*

**treatro (teatro, 'theatre') (×2), *culturaleres (culturales, 'cultural'), *cerciendo (crescendo, 'growing'), *económinca (económica, 'economic'), *tendecia (tendencia, 'tendency'), *sumarion (sumaron, 'sum up'), *sectiores (sectores, 'sectors'), *fuertentemnte (fuertemente, 'strongly'), *criris (crisis, 'crisis'), *indutria (industria, 'industry'), *desperden (desprenden, 'clear'), *audiovisulares (audiovisuales, 'audiovisual'), *expectadores (espectadores, 'spectators') (×2), *venbidas (vendidas, 'sold'), *eurores (euros, 'euros'), *projectos (proyectos, 'projects'), *major (mayor, 'mayor'), *lso (los, ''the), *grades (grandes, 'big'), *esplicó (explicó, 'explained'), *segretario (secretario, 'secretary')*, and **genenral (general, 'general').*

A.2 Samples W_D and W_E

Sample W_D

For the Sample W_D show the error kind, the error, the target word and the source (in parenthesis) from where the error was extracted.
English Sample $W_D en$:

1. Simple errors:
    (a) Substitution: **studends (students)* [371].
    (b) Insertion: **promblem (problem)*, and **deleteing (deleting)* [277].
    (c) Omission: **approch (approach)* [277], **carful (careful)* [278], and **constrution (construction)* [371].
    (d) Transposition: **worng (wrong)*, **artcile (article)* [277], **childern (children)*, and **poeple (people)* [277].
2. Multi-errors: **situartion (situation)* [277]; **exaplin (explain)*, **confusetion (confusion)*, **torromow (tomorrow)*, **knwolegde*

(knowledge), **comaprsion (comparison)*, **intersenting (interesting)*, **worires (worries)*, **understangind (understanding)*, and **inpossbile (impossible)* (dyslexic subject).

Spanish Sample $W_D sp$:

1. Simple errors:
   (a) Substitution: **probrema (problema, 'problem')* (dyslexic subject).
   (b) Insertion: **docotorado (doctorado, 'PhD')*, and **escribies (escribes, 'write')* (dyslexic subject).
   (c) Omission: **gande (grande, 'big')*, **hombes (hombres, 'men')*, and **pegunta (pregunta, 'question')* [14].
   (d) Transposition: **cambaido (cambiado, 'changed')*, **hablamso (hablamos, 'speak')*, **necestio (necesito, 'need')*, and **tmabién (también, 'too')* (dyslexic subject).
2. Multi-errors: **entenmiento (entendimiento, 'entertainment')* and **sechora (señora, 'woman')* [14]; **felicdidad (felicidad, 'happiness')*, **incleibre (increíble, 'incredible')*, **respondodido (respondido, 'answered')*, **cominucaion (comunicación, 'communication')*, **contimigo (conmigo, 'with you')*, **sufieicnte (suficiente, 'sufficient')*, **tambpo (tampoco, 'either')*, and **temriando (terminando, 'changed')* (dyslexic subject).

Sample W_E

For the Sample W_E show the target word, the error kind and the variants of each error type.

English Sample $W_E en$:

1. Target word: *comparison.*
   (a) Dyslexic: **comaprsion.*
   (b) Spelling: **comparisision, *conparison* and **comprison.*
   (c) Typo: **vomparison, *xomparison, *cimparison, *cpmparison, *conparison, *comoarison, *comprison, *compsrison, *compaeison, *compatison, *comparuson, *comparoson, *compariaon, *comparidon, *comparisin, *comparispn, *comparisob,* and **comparisom.*

   (d) OCR: **compaiison*, and **comparisom.*
   (e) Foreign: **comparition*, and **comparizon.*

2. Target word: *confusion.*
   (a) Dyslexic: **confusetion.*
   (b) Spelling: **confussion.*
   (c) Typo: **xonfusion, *vonfusion, *cinfusion, *cpnfusion, *cobfusion, *comfusion, *condusion, *congusion, *confysion, *confision, *confuaion, *confudion, *confusuon, *confusoon, *confusiin, *confusipn, *confusiob*, and **confusiom.*
   (d) OCR: **coniusion*, and **confuslon.*
   (e) Foreign: **confution.*

3. Target word: *explain.*
   (a) Dyslexic: **exaplin.*
   (b) Spelling: **explane.*
   (c) Typo: **wxplain, *rxplain, *ezplain, *ecplain, *exolain, *expkain, *explin, *explsin, *explaun, *explaon, *explaib*, and **explaim.*
   (d) OCR: **explaln.*
   (e) Foreign: **esplain.*

4. Target word: *impossible.*
   (a) Dyslexic: **inpossbile.*
   (b) Spelling: **umpossible, inpossible*, and **anpossible..*
   (c) Typo: **umpossible, *ompossible, *imoossible, *impissible, *imppssible, *impoaaible, *impoddible, *impossuble, *impossoble, *impossivle, *impossinle, *impossibke, *impossiblw*, and **impossiblr.*
   (d) OCR: **imposlble, *impossible, *inpossible*, and **impossibie.*
   (e) Foreign: **imposible*, and **impozible.*

5. Target word: *interesting.*

(a) Dyslexic: **intersenting.*

(b) Spelling: **intresting.*

(c) Typo: **unteresting, *onteresting, *ibteresting, *imteresting, *inreresting, *inyeresting, *intwresting, *intrresting, *inteeesting, *intetesting, *interwsting, *interrsting, *intereating, *interedting, *interesring, *interesying, *interestung, *interestong, *interestibg, *interestimg, *interestinf*, and **interestinh.*

(d) OCR: **interesiing, *inieresting*, and **inleresting.*

(e) Foreign: **intrestin.*

6. Target word: *knowledge.*

(a) Dyslexic: **knwolegde.*

(b) Spelling: **nowledge.*

(c) Typo: **jnowledge, *lnowledge, *kbowledge, *kmowledge, *kniwledge, *knpwledge, *knoqledge, *knoeledge, *knowkedge, *knowlwdge, *knowlrdge, *knowlesge, *knowlefge, *knowledfe, *knowledhe, *knowledgw*, and **knowledgr.*

(d) OCR: **knowiedge*, and **knowledqe.*

(e) Foreign: **knowlegde*, and **noledge.*

7. Target word: *situation.*

(a) Dyslexic: **situartion.*

(b) Spelling: **situacion.*

(c) Typo: **aituation, *dituation, *sutuation, *sotuation, *siruation, *siyuation, *sityation, *sitiation, *sitution, *situstion, *situarion, *situayion, *situatuon, *situatoon, *situatiin, *situatipn, *situatiob*, and **situatiom.*

(d) OCR: **situaiion*, and **siluation.*

(e) Foreign: **situasion.*

8. Target word: *tomorrow.*

(a) Dyslexic: **torromow.*

(b) Spelling: **toomorrow.*

(ac) Typo: **romorrow, *yomorrow, *timorrow, *tpmorrow, *tonorrow, *tomirrow, *tomprrow, *tomoeeow, *tomottow, *tomorriw, *tomorrpw, *tomorroq*, and **tomorroe.*

(d) OCR: **tomorrov, *tamarraw*, and **tonorrow.*

(e) Foreign: **tomorow*, and **tomorou.*

9. Target word: *understanding.*

   (a) Dyslexic: **understangind.*

   (b) Spelling: **understend*, and **understandin.*

   (c) Typo: **ynderstanding, *inderstanding, *ubderstanding, *umderstanding, *unserstanding, *unferstanding, *undwrstanding, *undrrstanding, *undeestanding, *undetstanding, *undertanding, *underatanding, *undersranding, *undersyanding, *understnding, *understsnding, *understabding, *understamding, *understansing, *understanfing, *understandung, *understandong, *understandibg, *understandimg, *understandinf*, and **understandinh.*

   (d) OCR: **undersianding*, and **understandinq.*

   (e) Foreign: **underestanding*, and **underestandin.*

10. Target word: *worries.*

    (a) Dyslexic: **worires.*

    (b) Spelling: **worrys.*

    (c) Typo: **qorries, *eorries, *wirries, *wprries, *woeeies, *wotties, *worrues, *worroes, *worriws, *worrirs, *worriea*, and **worried.*

    (d) OCR: **woiiies.*

    (e) Foreign: **worryes.*

Spanish Sample W_Esp:

1. Target word: *comunicación.*

   (a) Dyslexic: **cominucaion.*

   (b) Spelling: **comunicasion.*

   (c) Typo: **xomunicación, *vomunicación, *cimunicación, *cpmunicación, *conunicación, *comynicación, *comumicación, *comunocación, *comunixación, *comunivación, *comunicsción, *comunicaxión, *comunicavión, *comunicacuón, *comunicacoón, *comunicaciin, *comunicacipn, *comunicaciób*, and **comunicacióm.*

   (d) OCR: **comunicadón, *cornunicación*, and **comunlcación.*

   (e) Foreign: **conmunicación, *cominicación*, and **comunucación.*

2. Target word: *conmigo.*

   (a) Dyslexic: **contimigo.*

   (b) Spelling: **cinmigo, *comego*, and **commigo.*

   (c) Typo: **xonmigo, *vonmigo, *cinmigo, *cpnmigo, *cobmigo, *conmugo, *conmofo, *conmigi*, and **conmigp.*

   (d) OCR: **conmlgo.*

   (e) Foreign: **conmiguo, *connigo*, and **conmiho.*

3. Target word: *entendimiento.*

   (a) Dyslexic: **entenmiento.*

   (b) Spelling: **entendimento*, and **entindimiento.*

   (c) Typo: **wntendimiento, *rntendimiento, *ebtendimiento, *emtendimiento, *enrendimiento, *enyendimiento, *entwndimiento, *entrndimiento, *entebdimiento, *entemdimiento, *entensimiento, *entenfimiento, *entendumiento, *entendomiento, *entendiniento, *entendimuento, *entendimoento, *entendimiwnto, *entendimirnto, *entendimiebto, *entendimiemto, *entendimienro, *entendimienyo, *entendimienti*, and **entendimientp.*

   (d) OCR: **entcndimiento*, and **entendimicnto.*

   (e) Foreign: **intendimiento*, and **entendimient.*

4. Target word: *felicidad.*

   (a) Dyslexic: **felicdidad.*

   (b) Spelling: **felizidad.*

(c) Typo: _*delicidad, *gelicidad, *fwlicidad, *frlicidad, *fekicidad, *felucidad, *felocidad, *felividad, *felicudad, *felicodad, *felicisad, *felicifad, *felicidsd_, and _*felicidaf_.

(d) OCR: _*fellcidad_, and _*felieidad_.

(e) Foreign: _*felisidad, *felicidaz_, and _*felicidas_.

5. Target word: _increíble_.

   (a) Dyslexic: _*incleibre_.

   (b) Spelling: _*increible_.

   (c) Typo: _*uncreíble, *oncreíble, *ibcreíble, *inxreíble, *invreíble, *inceeíble, *incteíble, *incrwíble, *incrríble, *increuble, *increívle, *increínle, *increíbke, *increíblw_, and _*increíblr_.

   (d) OCR: _*increíbie_, and _*lncreíble_.

   (e) Foreign: _*hincreible, *imcreíble_, and _*incraible_.

6. Target word: _respondido_.

   (a) Dyslexic: _*respondodido_.

   (b) Spelling: _*repondio_.

   (c) Typo: _*eespondido, *tespondido, *rwspondido, *rrapondido, *reapondido, *redpondido, *resoondido, *respindido, *resppndido, *respobdido, *respomdido, *responsido, *responfido, *respondudo, *respondodo, *respondiso, *respondifo, *respondidi_, and _*respondidp_.

   (d) OCR: _*rcspondido_, and _*respondiclo_.

   (e) Foreign: _*respodido_.

7. Target word: _señora_.

   (a) Dyslexic: _*sechora_.

   (b) Spelling: _*siñora_.

   (c) Typo: _*aeñora, *deñora, *swñora, *srñora, *sebora, *señira, *señpra, *señoea_, and _*señota_.

   (d) OCR: _*scñora_, and _*señom_.

   (e) Foreign: _*segnora_, and _*sennora_.

8. Target word: *suficiente*.
   (a) Dyslexic: **sufieicnte*.
   (b) Spelling: **sufuciente*, and **sificiente*.
   (c) Typo: **eespondido, *tespondido, *rwspondido, *rrapondido, *reapondido, *redpondido, *resoondido, *respindido, *resppndido, *respobdido, *respomdido, *responsido, *responfido, *respondudo, *respondodo, *respondiso, *respondifo, *respondidi*, and **respondidp*.
   (d) OCR: **sufidente*, and **suficlente*.
   (e) Foreign: **sificiente*.
9. Target word: *tampoco*.
   (a) Dyslexic: **tambpo*.
   (b) Spelling: **tanpoco*.
   (c) Typo: **rampoco, *yampoco, *tsmpoco, *tamooco, *tamppco, *tampoxo, *tampovo, *tampoci*, and **tampocp*.
   (d) OCR: **tampoeo*.
   (e) Foreign: **tanpoko*.
10. Target word: *terminando*.
   (a) Dyslexic: **temriando*.
   (b) Spelling: **terminao*.
   (c) Typo: **rerminando, *yerminando, *twrminando, *trrminando, *teeminando, *tetminando, *terninando, *termunando, *termonando, *termibando, *termimando, *terminsbdo, *terminamdo, *terminanso, *terminanfo, *terminandi*, and **terminandp*.
   (d) OCR: **terrninando, *termlnando, *terminanclo*, and **tcrminando*.
   (e) Foreign: **termenando*, and **tirminando*.

A.3 *Text Presentation* Texts

Los encuentros de un caracol aventurero
(Federico García Lorca)

Hay dulzura infantil
en la mañana quieta.
Los árboles extienden
sus brazos a la tierra.
Un vaho tembloroso
cubre las sementeras,
y las arañas tienden
sus caminos de seda
–rayas al cristal limpio
del aire–.
En la alameda
un manantial recita
su canto entre las hierbas.

Y el caracol, pacífico
burgués de la vereda,
ignorado y humilde,
el paisaje contempla.
La divina quietud
de la Naturaleza
le dio valor y fe,
y olvidando las penas
de su hogar, deseó
ver el fin de la senda.

Echó a andar e internose
en un bosque de yedras
y de ortigas. En medio
había dos ranas viejas
que tomaban el sol,
aburridas y enfermas.

"Esos cantos modernos
–murmuraba una de ellas–
son inútiles". "Todos,
amiga –le contesta
la otra rana, que estaba
herida y casi ciega–.
Cuando joven creía
que si al fin Dios oyera
nuestro canto, tendría
compasión. Y mi ciencia,
pues ya he vivido mucho,
hace que no lo crea.
Yo ya no canto más..."

Las dos ranas se quejan
pidiendo una limosna
a una ranita nueva
que pasa presumida
apartando las hierbas.

Ante el bosque sombrío
el caracol se aterra.
Quiere gritar. No puede.
Las ranas se le acercan.

"¿Es una mariposa?",
dice la casi ciega.
"Tiene dos cuernecitos
-la otra rana contesta-.
Es el caracol. ¿Vienes,
caracol, de otras tierras?"

"Vengo de mi casa y quiero
volverme muy pronto a ella".
"Es un bicho muy cobarde
-exclama la rana ciega-.
¿No cantas nunca?" "No canto",
dice el caracol. "¿Ni rezas?"
"Tampoco: nunca aprendí".
"¿Ni crees en la vida eterna?"
"¿Qué es eso?
"Pues vivir siempre
en el agua más serena,
junto a una tierra florida
que a un rico manjar sustenta".

"Cuando niño a mí me dijo

un día mi pobre abuela
que al morirme yo me iría
sobre las hojas más tiernas
de los árboles más altos”.

“Una hereje era tu abuela.
La verdad te la decimos
nosotras. Creerás en ella”,
dicen las ranas furiosas.

“¿Por qué quise ver la senda?
–gime el caracol–. Sí creo
por siempre en la vida eterna
que predicáis...”
Las ranas,
muy pensativas, se alejan.
y el caracol, asustado,
se va perdiendo en la selva.

Las dos ranas mendigas
como esfinges se quedan.
Una de ellas pregunta:
“¿Crees tú en la vida eterna?”
“Yo no”, dice muy triste
la rana herida y ciega.
“¿Por qué hemos dicho, entonces,
al caracol que crea?”
“Por qué... No sé por qué
–dice la rana ciega–.
Me lleno de emoción
al sentir la firmeza
con que llaman mis hijos
a Dios desde la acequia...”

El pobre caracol
vuelve atrás.

Ya en la senda
un silencio ondulado
mana de la alameda.
Con un grupo de hormigas
encarnadas se encuentra.
Van muy alborotadas,
arrastrando tras ellas
a otra hormiga que tiene
tronchadas las antenas.
El caracol exclama:
“Hormiguitas, paciencia.
¿Por qué así maltratáis
a vuestra compañera?
Contadme lo que ha hecho.
Yo juzgaré en conciencia.
Cuéntalo tú, hormiguita”.

La hormiga, medio muerta,
dice muy tristemente:
“Yo he visto las estrellas”.

“¿Qué son las estrellas?”, dicen
las hormigas inquietas.
Y el caracol pregunta
pensativo: “¿Estrellas?”
“Sí –repite la hormiga-,
he visto las estrellas,
subí al árbol más alto
que tiene la alameda
y vi miles de ojos
dentro de mis tinieblas”.
El caracol pregunta:
“¿Pero qué son las estrellas?”
“Son luces que llevamos
sobre nuestra cabeza”.
“Nosotras no las vemos”,
las hormigas comentan.
Y el caracol: “Mi vista
sólo alcanza a las hierbas.”

Las hormigas exclaman
moviendo sus antenas:
“Te mataremos; eres
perezosa y perversa.

El trabajo es tu ley."

"Yo he visto a las estrellas",
dice la hormiga herida.
Y el caracol sentencia:
"Dejadla que se vaya.
seguid vuestras faenas.
Es fácil que muy pronto
ya rendida se muera".

Por el aire dulzón
ha cruzado una abeja.
La hormiga, agonizando,
huele la tarde inmensa,
y dice: "Es la que viene
a llevarme a una estrella".

Las demás hormiguitas
huyen al verla muerta.

El caracol suspira
y aturdido se aleja
lleno de confusión
por lo eterno. "La senda
no tiene fin –exclama–.
Acaso a las estrellas
se llegue por aquí.
Pero mi gran torpeza
me impedirá llegar.
No hay que pensar en ellas".

Todo estaba brumoso
de sol débil y niebla.
Campanarios lejanos
llaman gente a la iglesia,
y el caracol, pacífico
burgués de la vereda,
aturdido e inquieto,
el paisaje contempla.

¿Soy dix-leso? de la serie Papelucho (Marcela Paz)

Por la tarde fuimos al doctor. Era un señor bastante preguntón, que se hacía el simpático por fuera, pero se notaba que era malo por dentro. Me martilló las costras y otras cuestiones con un martillito lindo. Y mientras hablaba y hablaba con la mamá se martillaba su otra mano gorda.

Yo pensaba ¿qué pasaría si en vez de su mano gorda se martillara el tremendo grano que tenía en la nariz? Pero apenas se lo rascó y siguió dale que dale hablando de "este niño". Y "este niño" por aquí y "este niño" por allá.

Traté de entender lo que decían.

Casi lo entendí. No estoy seguro si la cosa es que soy superdotado o viceversa. Menos mal que además parece que soy dixleso, que es algo muy guay y como distinto. Y tampoco me importa mucho ser así.

Mis padres fueron al colegio a hablar con mi profe y volvieron furiosos.

De todos modos yo tengo mi enfermedad propia y nadie me la quita.

Pero en la noche, me desvelé. Porque claro, en el día a uno le gusta ser enfermo y en la noche no. Así que desperté a mi padre apretetándole la nariz porque es el único modo de despertarlo.

A.4 Color and Brightness Differences

The CYMK/RGB codes for the colors, and contrast used in the *Text Presentation* experiment.

- **Text Contrast:**
  - 0% (pure black font, 000000/0,0,0): Brightness text: 255; Brightness background: 0; Brightness difference: -255; Color difference: 765.
  - 25% (404040/64,64,64): Brightness text: 255; Brightness background: 63; Brightness difference: 192; Color difference: 573.
  - 50% (7E7E7E/126,126,126): Brightness text: 255; Brightness background: 126; Brightness difference: 129; Color difference: 387.
  - 75% (BFBFBF/191,191,191): Brightness text: 255; Brightness background: 191; Brightness difference: 64; Color difference: 191.
- **Background Contrast:**
  - 100% (pure black background, 000000/0,0,0): Brightness text: 0; Brightness background: 255; Brightness difference: -255; Color difference: 765.
  - 25% (404040/64,64,64): Brightness text: 63; Brightness background: 255; Brightness difference: -192; Color difference: 573.
  - 50% (7E7E7E/126,126,126): Brightness text: 126; Brightness background: 255; Brightness difference: -129; Color difference: 387.
  - 75% (BFBFBF/191,191,191): Brightness text: 191; Brightness background: 255; Brightness difference: -64; Color difference: 191.

The CYMK codes for the colors and contrast used are the following:

- **Colors:**
  - black (000000/0,0,0) / white (FFFFFF/255,255,255): Color difference: 765, Brightness difference: 255;

- off-black (0A0A0A/10,10,10) / off-white (FFFFE5/255,255,259): Color difference: 735, Brightness difference: 245;
- black (000000/10,10,10) / yellow (FFFF00/255,255,0): Color difference: 510, Brightness difference: 226;
- blue (00007D/0,0,125) / white (FFFFFF/255,255,255): Color difference: 640, Brightness difference: 241;
- dark brown (1E1E00/30,30,0) / light mucky green (B9B900/185,185,0): Color difference: 310, Brightness difference: 137
- brown (282800/40,40,0) / mucky green (A0A000/160,160,0): Color difference: 240, Brightness difference: 107
- black (000000/0,0,0) / creme (FAFAC8/250,250,200): Color difference: 700, Brightness difference: 244;
- blue (00007D/0,0,125) / yellow (FFFF00/255,255,0): Color difference: 635, Brightness difference: 212

A.5 Synonyms Pairs

The list of the unique pairs of synonyms used in the *Word Frequency* experiment is shown in Table A.1. The frequency ratio appears in parenthesis after the [±Frequent] word.

The list of the unique pairs of synonyms used in the *Word Length* experiment is shown in Table A.2 The frequency ratio appears in parenthesis after the [+Long] word. The length ratio is shown after the frequency ratio.

A.6 *Numerical Representation* Texts

Experiments *Digits vs. Words* and *Rounding vs. Decimals.* The set of target numerical expression used for each experiment are written in brackets, that is, [Digits / Words – Decimals / Rounding].

Composición de una hamburguesa

El pan supone entre el [30% / treinta por ciento – 18,53% / casi el 20%] y el [50% / cincuenta por ciento – 29,57% / casi el 30%] del peso de una hamburguesa. La hamburguesa tiene un valor energético que oscila entre las [250 / doscientas cincuenta – 297 / casi 300] y [300 / trescientas – 398 / casi 400] kilocalorías. Un adulto con actividad moderada necesita en torno a [2.500 / dos mil quinientas – 2.489 / unas 2.500] kilocalorías diarias, por lo

Pakistan		*Somalia*	
[+Frequent]	[−Frequent]	[+Frequent]	[−Frequent]
ataques *('attacks')*	refriegas (474)	personas *('people')*	individuos (26)
sequía *('drought')*	agostamiento (903)	casas *('houses')*	moradas (173)
entrega *('delivery')*	avituallamiento (787)	inundaciones *('flood')*	aluviones (123)
casa *('houses')*	morada (67)	lluvias *('rains')*	diluvios (195)
personas *('people')*	individuos (26.3)	verano *('summer')*	canícula (21,140)
ciudades *('cities')*	urbes (106)	desastre *('disaster')*	hecatombe (31)
ejército *('army')*	hueste (209)	parte *('part')*	porción (58)
rebeldes *('rebels')*	insubordinados (511)	país *('country')*	territorio (6)
ciudad *('capital')*	capitalidad (484)	estado *('estate')*	tesitura (741)
producción *('production')*	obtención (39)	comunidad *('community')*	colectividad (89)
alimentos *('food')*	sustentos (968)	generosidad *('generosity')*	dadivosidad (285)
precios *('prices')*	valías (5,897)	velocidad *('velocity')*	apresuramiento (491)
gente *('people')*	muchedumbre (107)	necesidades *('needs')*	indigencias (19)
productos *('products')*	manufacturas (133)		
mercado *('market')*	baratillo (888)		

Table A.1: List of unique pairs of synonyms for *Word Frequency*.

Car		*Wizard*	
[−Long]	[+Long]	[−Long]	[+Long]
huída *('run away')*	escabullida (81; 2.2)	piso *('flat')*	apartamento (2; 2.75)
parking *('parking place')*	estacionamiento (15; 2.14)	mago *('wizard')*	prestidigitador (87; 3.75)
pavor *('fear')*	sobrecogimiento (13; 2.8)	raro *('strange')*	estrambótico (10; 3.5)
auto *('car')*	autómovil (65; 2.25)	aumento *('increase')*	acrecentamiento (6; 2.14)
raro *('strange')*	extravagante (7; 3)		
cara *('face')*	semblante (276; 2.25)		
pálida *('pale')*	emblanquecido (93; 2.16)		

Table A.2: List of unique synonym pairs for *Word Length*.

que una hamburguesa a la semana no desequilibra ninguna dieta ni siquiera incorporándole un sobre de [11 / once – 11,8 / casi 12] gramos de ketchup, que contiene [70 / setenta – 70,8 / un poco más de 70] kilocalorías.

'**Composition of a burger**

The bread is between [30% / thirty percent – 18.53% / almost 20%] and [50% / fifty percent – 29.57% / almost 30%] of the weight of a hamburger. The burger has an energy value of between [250 / two hundred fifty – 297 / almost 300] and [300 / three hundred – 398 / almost 400] kilocalories. A moderately active adult needs about [2,500 / two thousand five hundred – 2,489 / around 2,500] kilocalories a day, so a burger a week not even unbalanced diet incorporating any one of [11 / eleven – 11.8 / almost 12] grams of ketchup, which contains [70 / seventy – 70.8 / a little more than 70] calories.'

Composición de las patatas fritas

Las patatas fritas, a pesar de tener [3,6 / tres coma seis – 3,67 / casi 4] gramos de grasa y [234 / doscientas treinta y cuatro – 214 / un poco más de 200] kilocalorías por cada [100 / cien – 89 / casi 100] gramos, son un alimento muy energético por lo que ha de consumirse en pequeñas cantidades y esporádicamente. Contienen [11 / once – 11,82 / un poco más de 10] gramos de hidratos de carbono con un índice glucémico de [70 / setenta – 68,67 / casi 70]. Es decir, suponen un [50% / cincuenta por ciento – 58% / casi un 60%] del consumo diario recomendado de hidratos en mujeres y el [30% / treinta por ciento – 29,12% / casi un 30%] en hombres.

'Composition of french fries

French fries, despite having [3.6 / three point six – 3.67 / almost 4] grams of fat and [234 / two hundred thirty-four – 214 / a little more than 200] kilocalories per [100 / hundred – 89 / almost 100] grams, are a very energetic food which has to be consumed in small quantities and sporadically. Containing [11 / eleven – 11.82 / a little more than 10] grams of carbohydrates with a glycemic index of [70 / seventy – 68.67 / almost 70] That is, suppose [50% / fifty percent – 58% / almost 60%] of the recommended daily intake of carbohydrates in women and [30% / thirty percent – 29.12% / almost 30%] in men.'

Experiment: *Percentages vs. Fractions*

Composición de una hamburguesa

El pan supone entre el [25% / $^{1}/_{4}$] y el [50% / $^{1}/_{2}$] del peso de una hamburguesa incluyendo el [75% / $^{3}/_{4}$] de los hidratos de carbono de esta. Estos hidratos suponen el [20% / $^{1}/_{5}$] del consumo diario recomendado para un adulto con actividad moderada.

'Composition of a burger

The bread is between [25% / $^{1}/_{4}$] and [50% / $^{1}/_{2}$] by weight of a hamburger including [75% / $^{3}/_{4}$] of this carbohydrate. These hydrates represent [20% / $^{1}/_{5}$] of the recommended daily intake for an adult with moderate activity.'

Composición de las patatas fritas

Alrededor del [50% / $^{1}/_{2}$] de los componentes de las patatas fritas son hidratos de carbono con un índice glucémico del [75% / $^{3}/_{4}$]. Es decir, suponen un [25% / $^{1}/_{4}$] del consumo diario recomendado de hidratos en mujeres y el [20% / $^{1}/_{5}$] en hombres.

'Composition of french fries

Approximately [50% / $^{1}/_{2}$] of the components of french fries are carbohydrates with a low glycemic index of [75% / $^{3}/_{4}$]. That is, suppose [25% / $^{1}/_{4}$] of the recommended daily intake of carbohydrates in females and [20% $^{1}/_{5}$] in males.'

[−Simple]	[+Simple]
otorga un premio	premia (×2)
(‘to give an award’)	*(‘to award’)* (×2)
tenían la confianza	confiaban
(‘to trust’)	
tuviera ambición de	ambicionara
(‘have the ambition’)	*(‘to desire’)* (×2)
hacer aparición	aparecer
(‘to appear’)	
otorgó el galardón	ha galardoneado
(‘to give an award’)	*(‘to award’)*
ha hecho una contribución	ha contribuido (×2)
(‘to make a contribution’)	*(‘to contribute’)*
ha prestado atención	ha atendido
(‘to pay attention’)	*(‘listen’)*
dio comienzo	comenzó
(‘to start’)	
impusieron censura	censuraron
(‘to impose censorship’)	*(‘to censor’)*
hizo la concesión del	concedió el
(‘to make a concession’)	*(‘to concede’)*
dar una recompensa	recompensar
(‘to give a reward’)	*(‘to reward’)*
puesta en manifiesto	manifestado
(‘to manifest’)	
ofrece un reconocimiento	reconoce
(‘to provides a acknowledgment’)	*(‘to recognize’)*
dió valor	valoró
(‘to give value’)	*(‘to value’)*
ha aportado riqueza	enriquecido
(‘to bring wealth’)	*(‘to enrich’)*
poner más cerca	acercar
(‘to bring closer’)	

Table A.3: Paraphrases pairs.

A.7 *Verbal Paraphrases* Pairs

The corresponding paraphrases pairs used are shown in Table A.3.

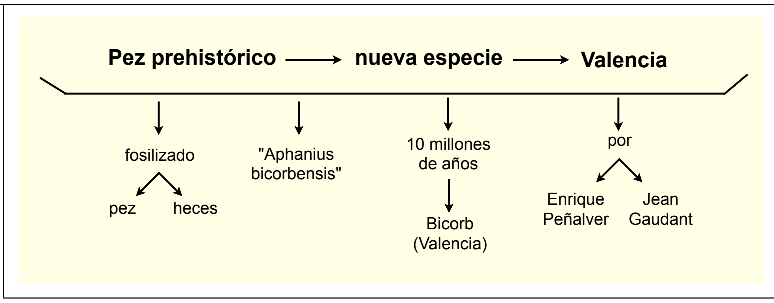


Figure A.1: Example of a graphical scheme (*Fish*), English translation.

A.8 *Graphical Schemes* Texts

Below we present Text 2 (*Fish*) *Graphical Schemes* and its translation to English. The graphical scheme for *Fish* is presented in Figure A.1.

Descubren en Valencia una nueva especie de pez prehistórico

El estudio de un lago salino que existió hace 10 millones de años en Bicorb (Valencia) ha permitido descubrir el fósil de una nueva especie de pez prehistórico y de sus heces. Según informó este martes el Instituto Geológico y Minero de España, este pez depredador ha sido bautizado por los investigadores como “Aphanius bicorbensis”, en honor a la población de Bicorb donde ha sido encontrado. La investigación ha sido realizada por Enrique Peñalver, experto en insectos fósiles del Instituto Geológico y Minero, y por Jean Gaudant, especialista en peces fósiles del Museo Nacional de Historia Natural de París, gracias a la financiación de la Consejería de Cultura de la Generalitat Valenciana. El estudio del contenido de las heces de estos peces, que también quedaron fosilizadas en la roca, ha permitido a los investigadores saber que este depredador se alimentaba de los foraminíferos y de las larvas de mosquito, especialmente abundantes en el lago.

'A new species of a prehistoric fish is discovered in Valencia

The study of a saline lake that existed 10 million years ago in Bicorb (Valencia) has uncovered the fossil of a new species of prehistoric fish and their feces. The Geological and Mining Institute of Spain informed last Tuesday that this predatory fish has been named by the researchers as "Aphanius bicorbensis" in honor of the town of Bicorb where was found. The research was conducted by Enrique Peñalver, an expert on insect fossils of the Geological and Mining Institute, and Jean Gaudant, a specialist in fossil fishes of the National Museum of Natural History in Paris, thanks to funding from the Council of Culture of the Government of Valencia. The study of the content of the feces of these fishes, which were also fossilized in the rock, has allowed researchers to know that this predator was feeding on foraminifera and mosquito larvae, especially abundant in the lake.'

A.9 *Keywords* Texts

Translation of the text example used in the *Keywords* experiment.

El Museo Picasso Málaga expone **nuevas obras del artista** junto a su colección permanente

El consejero andaluz de Cultura, Paulino Plata, presentó **una nueva reorganización de la colección permanente del Museo Picasso** que, coincidiendo con el aniversario del nacimiento del pintor, incorpora **una amplia selección de obras de Pablo Picasso** cedidas por la Fundación Almine y Bernard Ruiz-Picasso para el Arte. Pinturas, esculturas y cerámicas de diferentes periodos y estilos del artista conforman **este conjunto de 43 piezas** cedidas por 15 años por la citada fundación. La incorporación de estas creaciones supone así, según la Junta de Andalucía, **una valiosa aportación a la colección permanente del Museo Picasso Málaga**. De esta forma, el visitante puede contemplar desde ahora **óleos y esculturas** que, por primera vez, se exponen en la pinacoteca.

'The Museo Picasso Málaga includes **new works of the artist** *in its permanent collection*

The Andalusian Minister of Culture, Paulino Plata, presented **a new reorganization of the permanent collection of the Picasso Museum**, *for the birth anniversary of the painter. This incorporates* **a wide selection of works by Pablo Picasso** *provided by Almine and Bernard Ruiz-Picasso Foundation for Art. Paintings, sculptures and ceramics from different periods and styles compose* **this set of 43 pieces**.

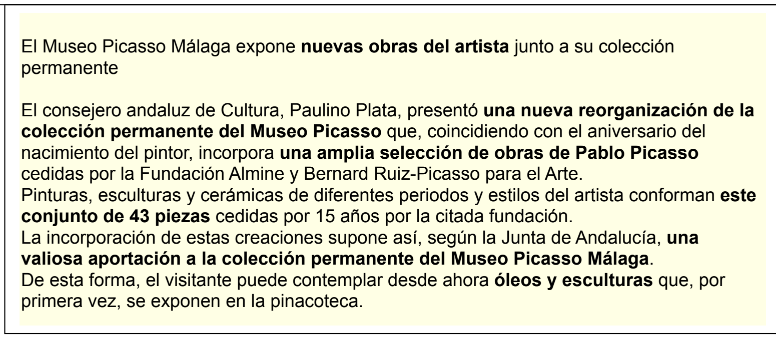


Figure A.2: Example slide used in the *Keywords* experiment.

They were granted for 15 years by that foundation. The incorporation of these creations and means, according to the Andalusian Council, **a valuable contribution to the permanent collection of the Museo Picasso Málaga.** *Thus, visitors can see* **paintings and sculptures** *now, for the first time, shown in the gallery.'*

A.10 Lexical Simplifications

The list of the unique lexical simplification alterations for Spanish in the text *Bee* (*'Effect of agricultural pesticides in bee populations'*) is given in Tables A.4 and A.5.

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,	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

Table A.4: Examples of lexical simplifications in Spanish (1).

A.11 Corpus used for CASSA

Alas Clarín, Leopoldo – *La regenta*
Allende, Isabel – *La casa de los espíritus*

Orig	SubsBest	ShowSyns	Gold
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 A.5: Examples of lexical simplifications in Spanish (2).

Anonymous – *Auto de los Reyes Magos*
Anonymous – *Cantar de Mío Cid*
Anonymous – *Dioses y leyendas griegas*
Anonymous – *Lazarillo de Tormes*
Anonymous – *Lazarillo de Tormes, Part II*
Anonymous – *Libro de Alexandre*
Anonymous – *Libro de Apolonio*
Anonymous – *Mil y una noches*

Austen, Jane – *Orgullo y prejuicio*
Balzac, Honoré de – *El elixir de larga vida*
Baroja, Pío – *Las inquietudes de Shanti Andía*
Baroja, Pío – *Zalacaín el aventurero*
Baudelaire, Charles – *Las flores del mal*
Bécquer, Gustavo Adolfo – *Leyendas*
Benedetti, Mario – *La tregua*
Boccaccio, Giovanni – *Decamerón*
Borges, Jorge Luis – *El aleph*
Borges, Jorge Luis – *Ficciones*
Brontë, Emily – *Cumbres borrascosas*
Bulgákov, Mijaíl – *El maestro y Margarita*
Calderón de la Barca, Pedro – *El gran teatro del mundo*
Calderón de la Barca, Pedro – *La vida es sueño*
Camus, Albert – *El extranjero*
Capote, Truman – *A sangre fría*
Carroll, Lewis – *Alicia en el país de las maravillas*
Cela, Camilo José – *La familia de Pascual Duarte*
Cervantes Saavedra, Miguel de – *Don Quijote, Part I*
Cervantes Saavedra, Miguel de – *Don Quijote, Part II*
Chaucer, Geoffrey – *Cuentos de Canterbury*
Chejov, Antón – *Tío Vania*
Christie, Agatha – *Diez negritos*
Conrad, Joseph – *El corazón de las tinieblas*
Cortázar, Julio – *Rayuela*
Dahl, Roald – *Historias extraordinarias*
Dante Alighieri – *Divina comedia*
Defoe, Daniel – *Aventuras de Robinson Crusoe*
Delibes, Miguel – *El hereje*
Dickens, Charles – *David Copperfield*
Dickens, Charles – *Grandes esperanzas*
Dickens, Charles – *Historia de dos ciudades*
Dickens, Charles – *Oliver Twist*
Dostoievski, Fiódor – *Crimen y castigo*
Dumas, Alexandre – *El conde de Montecristo*
Dumas, Alexandre – *El tulipán negro*
Dumas, Alexandre – *La dama de las camelias*
Dumas, Alexandre – *Los tres mosqueteros*
Eco, Umberto – *El nombre de la Rosa*

Eco, Umberto – *El péndulo de Foucault*
Eliot, George – *El velo descubierto*
Ende, Michael – *El espejo en el espejo*
Ende, Michael – *La historia interminable*
Ende, Michael – *Momo*
Espronceda, José de – *Antología poética*
Espronceda, José de – *El estudiante de Salamanca*
Eurípides – *Medea*
Faulkner, William – *El oso*
Fernández de Moratín, Leandro – *El sí de las niñas*
Flaubert, Gustave – *Madame Bovery*
Follett, Ken – *Los pilares de la tierra*
Frank, Ana – *Diario*
Fuentes, Carlos – *Aura*
Fulcanelli – *El misterio de las catedrales*
Gala, Antonio – *La pasión turca*
García Márquez, Gabriel – *Cien años de soledad*
García Márquez, Gabriel – *El amor en los tiempos del cólera*
Giner de los Ríos, Francisco – *Espíritu y naturaleza*
Goethe, Johann Wolfgang von – *Fausto*
Goethe, Johann Wolfgang von – *Werther*
Golding, William – *El señor de las moscas*
Góngora, Luis de – *Fábula de Polifemo y Galatea*
Góngora, Luis de – *Soledades*
Grimm, brothers – *Blancanieves*
Grimm, brothers – *Hansel y Gretel*
Harris, Thomas – *El silencio de los inocentes*
Hawking, Stephen – *Historia del tiempo*
Hemingway, Ernest – *El viejo y el mar*
Hernández, José – *El gaucho Martín Fierro*
Hernández, Miguel – *Cancionero y romancero de ausencias*
Hernández, Miguel – *El rayo que no cesa*
Homero – *La Ilíada*
Homero – *La Odisea*
Hugo, Victor – *Los miserables*
Hugo, Victor – *Nuestra señora de París*
Huxley, Aldous – *Un mundo feliz*
Irving, Washington – *Cuentos de la Alhambra*
Jiménez, Juan Ramón – *Antología Personal*

Joyce, James – *Retrato del artista adolecente*
Joyce, James – *Ulises*
Kafka, Franz – *La metamorfosis*
Khalil Gibrán – *El profeta*
King, Stephen – *Carrie*
Kipling, Rudyard – *El libro de la selva*
Kipling, Rudyard – *Kim*
Lindo, Elvira – *Cómo molo*
Lindo, Elvira – *Manolito Gafotas*
Lope de Vega, Félix – *Amar sin saber a quién*
Lope de Vega, Félix – *El caballero de Olmedo*
Lope de Vega, Félix – *El castigo sin venganza*
Lope de Vega, Félix – *El perro del hortelano*
Lope de Vega, Félix – *Fuenteovejuna*
Lope de Vega, Félix – *La dama boba*
Lope de Vega, Félix – *Bodas de sangre*
Lope de Vega, Félix – *La casa de Bernarda Alba*
Lope de Vega, Félix – *Libro de poemas*
Lope de Vega, Félix – *Romancero Gitano*
Luca de Tena, Torcuato – *Los renglones torcidos de Dios*
Machado, Antonio – *Antología*
Mann, Thomas – *Muerte en Venecia*
Manrique, Jorge – *Coplas a la muerte de su padre*
Marcela, Paz – *Papelucho*
Marsé, Juan – *La oscura historia de la prima Montse*
Martín Gaite, Carmen – *El cuarto de atrás*
Matute, Ana María – *Pequeño teatro*
Melville, Herman – *Moby Dick*
Mendoza, Eduardo – *El laberinto de las aceitunas*
Molière – *El médico a palos*
Molina, Tirso de – *Don Gil de las calzas verdes*
Molina, Tirso de – *El burlador de Sevilla*
Neruda, Pablo – *Canto general*
Neruda, Pablo – *Cien sonetos de amor*
Onetti, Juan Carlos – *Cuentos*
Ortega y Gasset, José – *La rebelión de las masas*
Orwell, George – *1984*
Orwell, George – *Rebelión en la granja*
Ovidio – *Metamorfosis*

Pardo Bazán, Emilia – *Los pazos de Ulloa*
Paz, Octavio – *El arco y la lira*
Pérez Galdós, Benito – *Doña Perfecta*
Pérez Galdós, Benito – *Fortunata y Jacinta*
Pérez Galdós, Benito – *Misericordia*
Pessoa, Fernando – *Algunos versos*
Platón – *Cratilo*
Platón – *La República*
Poe, Edgar Allan – *Cuentos*
Poe, Edgar Allan – *El cuervo y otros poemas*
Proust, Marcel – *En busca del tiempo perdido I*
Proust, Marcel – *En busca del tiempo perdido II*
Proust, Marcel – *En busca del tiempo perdido III*
Proust, Marcel – *En busca del tiempo perdido IV*
Proust, Marcel – *En busca del tiempo perdido V*
Proust, Marcel – *En busca del tiempo perdido VI*
Proust, Marcel – *En busca del tiempo perdido VII*
Pushkin, Alexander – *Boris Godunov*
Quevedo, Francisco de – *Historia y vida del buscón*
Quevedo, Francisco de – *Poemas*
Rimbaud, Arthur – *Iluminaciones*
Rojas, Fernando de – *La celestina*
Rousseau, Jean-Jacques – *El contrato social*
Rowling, Joanne K. – *Harry Potter y la piedra filosofal*
Rowling, Joanne K. – *Harry Potter y la cámara secreta*
Rowling, Joanne K. – *Harry Potter y el prisionero de Azkaban*
Rowling, Joanne K. – *Harry Potter y el cáliz de fuego*
Rowling, Joanne K. – *Harry Potter y la orden Fénix*
Rowling, Joanne K. – *Harry Potter y el príncipe mestizo*
Ruiz, Juan (Arcipreste de Hita) – *Libro de buen amor*
Rulfo, Juan – *Pedro Páramo*
Sábato, Ernesto – *El túnel*
Safo – *Poemas*
Salinger, Jerome David – *El guardián entre el centeno*
San Agustín – *Confesiones*
San Juan de la Cruz – *El cántico espiritual*
Santa Teresa de Jesús – *Camino de perfección*
Saramago, Jose – *La caverna*
Shakespeare, William – *Hamlet*

Shakespeare, William – *Julio César*
Steinbeck, John – *La perla*
Stendhal – *Rojo y negro*
Stevenson – *El Dr. Jekyll y Mr. Hyde*
Stevenson, Robert Louis – *La isla del tesoro*
Stoker, Bram – *Drácula*
Süskind, Patrick – *El perfume*
Swift, Jonathan – *Los viajes de Gulliver*
Tolkien, John Ronald Reuel – *El señor de los anillos 1: La comunidad del anillo*
Tolkien, John Ronald Reuel – *El señor de los anillos 2: Las dos torres*
Tolkien, John Ronald Reuel – *El señor de los anillos 3: El retorno del rey*
Tolstoi, León – *Ana Karenina*
Tolstoi, León – *Guerra y paz*
Torrente Ballester, Gonzalo – *Gozos y las sombras 1: El señor llega*
Torrente Ballester, Gonzalo – *Gozos y las sombras 2: Donde da la vuelta el aire*
Torrente Ballester, Gonzalo – *Gozos y las sombras 3: La pascua triste*
Twain, Mark – *Las aventuras de Huckleberry Finn*
Unamuno, Miguel de – *Niebla*
Valèry, Paul – *El cementerio marino*
Valle Inclán, Ramón María del – *Luces de Bohemia*
Vallejo, César – *Poesía completa*
Vargas Llosa, Mario – *La fiesta del chivo*
Vega, Garcilaso de la – *Poemas, églogas, canciones*
Virgilio – *Eneida*
Wallace, Lewis – *Ben Hur*
Wells, Herbert George – *La guerra de los mundos*
Whitman, Walt – *Selección poética*
Wilde, ós, car – *El retrato de Dorian Gray*
Woolf, Virginia – *La señora Dalloway*
Zorrilla, José – *Don Juan Tenorio*

A.12 Usability Questionnaire

1. Is the language used in the app descriptive?

nothing ⟷ much

(in the next items "..." represents a seven-point Likert scale)

2. Are you familiar with the vocabulary used in the app?
   nothing ... much

3. Is there anything that confused you?

4. How difficult did you find navigating in the bookshelf?
   nothing ... much

5. Did you understand the symbols and figures used (the star, the knob, the sun, etc.)?
   nothing ... much

6. Is there anything that confused you?

7. How difficult did you find navigating through the book?
   nothing ... much

8. How difficult did you find customizing the text?
   nothing ... much

9. By customizing the text I can read...
   much worse ... much better

10. Why?

11. How helpful did you find accessing to synonyms in the text?
    nothing ... much

12. By accessing the synonyms I can read...
    much worse ... much better

13. Why?

14. Would you use the option “favorites style” for reading?
nothing … much

15. The general usage of the app is:
very complicated … very easy

16. Is there anything you did not like or would change?

17. I would use this app to read:
never … always

18. I would recommend this app to read:
never … always

19. Which is the app of your dreams?