A Plug-in to Aid Online Reading in Spanish

Luz Rello
Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
luzrello@cs.cmu.edu

Roberto Carlini
NLP Research Group
Universitat Pompeu Fabra
roberto.carlini@upf.edu

Ricardo Baeza-Yates
Yahoo Labs & Web Research Group
Universitat Pompeu Fabra
rbaeza@acm.org

Jeffrey P. Bigham
HCI and LT Institutes
School of Computer Science
Carnegie Mellon University
jbigham@cmu.edu

ABSTRACT

Reading text on the Web is a challenging task for many people, such as those with cognitive impairments, reading difficulties or people who are learning a new language. In this paper we present a web browser plug-in to help with reading Spanish text on the Web. The plug-in is freely available for Chrome and presents definitions and simpler synonyms on demand for the selected web text. The tool was modified following the suggestions of 5 people (2 with diagnosed dyslexia) who tested the tool using the think aloud protocol and undertook a subsequent interview.

Categories and Subject Descriptors

K.4.2 [**Computers and Society**]: Social Issues – Assistive technologies for persons with disabilities.

Keywords

Text simplification, lexical simplification, readability, synonyms, definitions, plug-in, Chrome.

1. INTRODUCTION

In 1994, the United Nations [20] proposed a set of standard rules for document accessibility to equalize the opportunities for people with disabilities. These rules were established because reading can be a very challenging task for some populations such as people with autism spectrum disorder, Down syndrome, aphasia, or dyslexia. At the same time, the essential property of the Web is its universality as it is fundamentally designed to work for all people [8].

While we know that there are 285 million people have vision impairments [35] or 642 million people that suffer from hearing loss [36], it is difficult to estimate how many people with cognitive disabilities use the Web. It is also challenging for a country to track their population with cognitive disabilities and estimations vary greatly. There is also no agreement on what conditions enter into this category. For instance, dyslexia is considered a cognitive disability by the WACG 2.0 [8] but it is defined as a learning disability in the *International Statistical Classification of Diseases (ICD-10)* [34]. Sometimes dyslexia is referred to as only a learning difficulty [6]. Most studies find that about 20% of the population has some kind of disability [33]. The U.S. Census Bureau's estimated that 27% of 53 million people with disabilities have a cognitive disability [32]. The estimation of the presence of these population in the Web is also a challenge. In 2011, we presented an estimation of dyslexia on the Web, which established a lower bound of 0.63% of Internet users having dyslexia [2]. This is presumably an underestimate because 10% of the general population is estimated to have dyslexia [15].

Text content transformations have the potential to help this web population with reading difficulties. In fact previous research have shown how content transformations on the text can significantly improve the reading perception of people with dyslexia [24]. In this paper we present a plug-in that aims to leverage the access content text by showing definitions and simpler synonyms on-demand for complex words in web text written in Spanish. Based of previous studies this tool has the potential to help people with different cognitive impairments of difficulties to access web text.

2. RELATED WORK

We divide related work into studies related to lexical simplification and other available tools similar to our plug-in.

2.1 Approaches to Lexical Simplification

There have been number of efforts to simplify text automatically for different populations. *Lexical simplification* is a kind of text simplification that aims to the word level. It can be performed through the substitution of words by simpler synonyms, by adding a definition, or by showing simpler synonyms. Most of the approaches aim at the substitution of complex words.

To find appropriate synonyms, many approaches use WordNet [7, 10, 17]. De Belder *et al.* [12] apply explicit word

W4A'15 May 18 - 20, 2015, Florence, Italy.

http://dx.doi.org/10.1145/2745555.2746661

Figure 1: Screenshot of the CASSA plug-in (left) and different options for highlighting complex words (right): colors, square, underline, and no highlighting.

sense disambiguation with a latent words language model. Devlin and Unthank [13] use dictionaries. Aluisio and Gasperin [1] use a thesaurus and lexical ontologies. Bott *et al.* [5] make use of the Spanish OpenThesaurus and a simplification corpus. More recently, Biran *et al.* [4] and Yatskar *et al.* [37] used Simple English Wikipedia, in combination with the standard English Wikipedia for their lexical simplification algorithms.

There are also machine translation based approaches [11, 30] as well as the hybrid approaches [19, 29] that are also able to handle lexical simplification, since the translation model maps words from the non-simplified language to words of the simplified language.

Previous approaches on automatic text simplification have focused on people with autism spectrum disorder [14, 22], people with Down syndrome [28], people with dyslexia [24, 23], and people with aphasia [9, 13]. However, automatic lexical simplification is still in an early stage and still is not useful for the target users (see conclusions of Evans *et al.* [14, 22] for people with autism spectrum disorder, and Saggion *et al.* [28] for people with Down syndrome).

For the case of people with dyslexia, even if they read significantly faster more frequent words [25], performing automatic lexical substitution did not improve the readability of the texts [24]. Nonetheless, when these synonyms were presented on demand to the user, texts were perceived as significantly simpler [24]. Therefore, tools that present synonyms for complex words can help people with dyslexia and potentially other populations.

2.2 Similar Tools

There are several Chrome extensions that offer synonyms in English such as *Thesaurus: Synonym 4 Right Click*,[1] or *Instant Thesaurus*.[2] For Spanish we found synonyms dictionaries such as *El Sinónimo*,[3] or *Sinónimos*.[4] Our plug-in differs from the Spanish ones in that the user do not need to type the word avoiding typing errors and saving time. It also differs because it does not show regular synonyms, but only synonyms that were found to be simpler according a study with strong readers and people with dyslexia [23].

This synonym resource is also used by the web server *Text4all* [31],[5] and *IDEAL eBook reader*[6] for Android [16], *DysWebxia reader* [23] for iOS and *Cloud4All* project[7].

3. CASSA PLUG-IN

CASSA plug-in[8] is an extension for Chrome that processes a selected web text and shows synonyms and definitions on-demand when hovering the mouse over complex words in a selected web text. See Figure 1 (left) for an example. First, for the synonyms on-demand, the plug-in parses the Spanish web text and verifies if a word within their context (4 words, 2 on the right and 2 on the left), is a complex word. For this step the plug-in uses the CASSA resource generated by CASSA (Context Aware Synonym Simplification Algorithm) [3]. This algorithm makes use of two free resources to find simpler synonyms: Google Books Ngram Corpus for Spanish and the Spanish OpenThesaurus. The system improves upon the state of the art for lexical simplification in Spanish [5] and the established simplification baseline, *i.e.*, the most frequent synonym, in several aspects: complex word detection, meaning preservation, and simplicity. The CASSA resource is composed by 41,106 complex words in Spanish and 4,229,868 lists of synonyms within a given context. The contexts are used to disambiguate the different senses of the word.

Second, the *CASSA plug-in* detects when a complex word does not have synonyms (*e.g.* does not appear in the thesaurus or the current context does not exist in the resource for that word) to show definitions instead. A list of complex words is checked to determine if the word is complex. This list is a compilation of all the inflected words and conjugated verbs in Spanish and their frequencies in the Web, considering a word complex if it appears less than 2,000 times in the Web according to a major search engine.

Instructions. After installing it by dragging it over `chrome://extensions/`, a blue hippo will appear on the top right of the browser. Then, when navigating, the user can select portions of difficult text, click on the hippo button (see

[1] `https://chrome.google.com/webstore/detail/thesaurus-synonym-4-right/lpkpcliecpgjbkffooidajhakoidhidh`

[2] `https://chrome.google.com/webstore/detail/instant-thesaurus/jambjfjiligfjpeljnkkcihooobfnllm`

[3] `https://chrome.google.com/webstore/detail/el-sinonimo/fpffndbkemhikfhbncibifokjgkjalcf`

[4] `https://chrome.google.com/webstore/detail/sin\%C3\%B3nimos/iekohmkpfbbmgnpmojchnhgepopgemda?hl=es`

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

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

[7] `http://cloud4all.info/`

[8] Available at: `http://www.wrg.upf.edu/WRG/DysWebxia.php?lang=en`

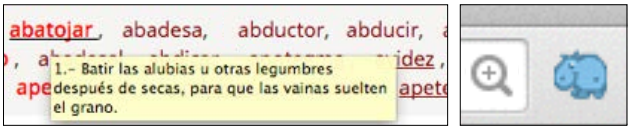

Figure 2: Screenshot CASSA plug-in (right) showing a definition and the plug-in hippo button.

Figure 2, right) and complex words of the text will be highlighted. There are various highlighting options: using colors (green or magenta), boldface, small figures (square, triangle) or the absence of highlighting. See different highlighting options in Figure 1 (right). When hovering the mouse over the complex words, possible synonyms or a definition, if synonyms are not available, appear.

4. USER TESTING

4.1 Methodology

Design. In a within-subject design, all the participants had to perform some tasks with the tool using the think aloud protocol [18].

Participants. We recruited 5 participants, two with diagnosed dyslexia (3 female, 2 male). They were all native Spanish speakers, two bilingual in Catalan. They all had either university degrees or were studying at the university. Their ages ranged from 18 to 29 years old, with a mean of 22.40 years ($s = 4.39$).

Procedure. The participants were provided with the plug-in and a set the instructions to install and test it. They had to choose three texts in the Web to read and test synonyms and definitions. The three texts were: one Wikipedia page (the one they felt like exploring), one newspaper article (from the newspaper they usually read online), and one free choice text. After these tasks the first author carried out an open interview.

Materials. The interview was inspired by the WACG 2.0 [8] and usability principles [21]. It included four open questions about the (a) the language used in the instructions, (b) navigation and control, (c) functionality of the application, and (d) personal opinions.

(a) *Do you find that the language used in the instructions is clear and descriptive?*

(b) *Which problems did you find using the plug-in?*

(c) *How useful you find the functionality of the plug-in?*

(d) *Did you find any other problems? What would you change?*

4.2 Results

We present the changes we made to the tool together with the suggestions of the participants collected during the testing sessions.

Instructions. Since none of the participants had installed a Chrome extension before, the fist step (installation) was the most challenging. They were not familiarized with the interface of Google Chrome (all of them declared that they normally use other browsers; one participant downloaded it specifically for the study). They gave feedback on how to further clarify the language of the instructions. Two of them did not know what the word "extension" meant.

Navigation and Control. The main problem that the participants found during the navigation process was returning to `chrome://extensions/` in order to change the highlighting options, *e.g.* the color in which the complex words were shown. One participant suggested to include an image in the instructions to explain where the options of the extension were placed. Another issue pointed out by the participants was that sometimes the waiting time was long (seconds) if the selected text to show synonyms was a large portion. During the experiments we found small bugs that we could fix, *e.g.* in some websites the plug-in did not work, except for websites starting with `https` where the plug-in does not work due to security restrictions

Functionality. In general the participants were satisfied with the synonyms that the plug-in presented. But three of them commented that there were difficult words with no synonyms. Accordingly to these suggestions we added the show definition option in the plug-in, since in its first version the plug-on only showed synonyms, not definitions.

Personal Opinions. In general they found the plug-in useful and easy to use (select text and press one button to get synonyms). One participant commented that he would like to use it for other languages he is learning when reading on the Web. They suggested changes mentioned before related with the difficulties they experienced during the installation and navigation. We included all of them except changing the place of highlighting options because we preferred to keep the use of the plug-in as simple as possible.

Future Work. In future work we plan to make the plug-in faster, add a module that simplifies numerical expressions [27], include new languages to *CASSA plug-in*, and be able to activate the plug-in automatically by detecting users with reading disorders [26].

Acknowledgments

This paper was developed under a grant from the Department of Education, NIDRR grant number H133A130057. We thank Azuki Gòrriz for creating the logo of the tool.

5. REFERENCES

[1] S. M. Aluísio and C. Gasperin. Fostering digital inclusion and accessibility: the PorSimples project for simplification of Portuguese texts. In *Proc. NAACL HLT '10 Workshop YIWCALA '10*, pages 46–53, Stroudsburg, PA, USA, 2010.

[2] R. Baeza-Yates 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.

[3] R. Baeza-Yates, L. Rello, and J. Dembowski. A context-aware synonym simplification algorithm: Cassa. In *Proc. NAACL '15*, Denver, Colorado, USA, 2015. ACM.

[4] 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.

[5] 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.

[6] British Dyslexia Association. Dyslexia style guide, January 2012. `http://www.bdadyslexia.org.uk/`.

[7] 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.

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

[9] 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.

[10] 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.

[11] W. Coster and D. Kauchak. Learning to simplify sentences using Wikipedia. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pages 1–9. Association for Computational Linguistics, 2011.

[12] 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.

[13] S. Devlin and G. Unthank. Helping aphasic people process online information. In *Proc. ASSETS '06*, pages 225–226. ACM, 2006.

[14] R. Evans, C. Orasan, and I. Dornescu. An evaluation of syntactic simplification rules for people with autism. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR) at EACL*, pages 131–140, 2014.

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

[16] 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.

[17] P. Lal and S. Ruger. Extract-based summarization with simplification. In *Proceedings of the ACL 2002 Automatic Summarization / DUC 2002 Workshop*, 2002.

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

[19] S. Narayan and C. Gardent. Hybrid simplification using deep semantics and machine translation. In *ACL'14*, pages 435–445, 2014.

[20] U. Nations. Standard Rules on the Equalization of Opportunities for Persons with Disabilities, 1994.

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

[22] C. Orasan, R. Evans, and I. Dornescu. *Towards Multilingual Europe 2020: A Romanian Perspective*, chapter Text Simplification for People with Autistic Spectrum Disorders, pages 287–312. Romanian Academy Publishing House, Bucharest, 2013.

[23] L. Rello. *DysWebxia. A Text Accessibility Model for People with Dyslexia*. PhD thesis, Universitat Pompeu Fabra, 2014.

[24] 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.

[25] 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.

[26] L. Rello and M. Ballesteros. Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proc. W4A '15*, Florence, Italy, 2015. ACM.

[27] 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.

[28] H. Saggion, S. Štajner, S. Bott, S. Mille, L. Rello, and B. Drndarevic. Making it simplext: Implementation and evaluation of a text simplification system for spanish. *ACM Transactions on Accessible Computing (TACCESS)*, In Press, 2015.

[29] A. Siddharthan and M. Angrosh. Hybrid text simplification using synchronous dependency grammars with hand-written and automatically harvested rules. In *Proc. EACL '14, Gothenburg, Sweden,*, pages 722–731, 2014.

[30] L. Specia. Translating from Complex to Simplified Sentences. In *PROPOR*, pages 30–39, 2010.

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

[32] U.S. Census Bureau. *Nearly 1 in 5 Americans has some level of disability*. U.S. Census Bureau, Public Information Office, 2001. `http://www.census.gov/Press-Release/www/2001/cb01-46.html`.

[33] WebAIM. Web accessibility in mind, February 2014. `http://webaim.org/`.

[34] World Health Organization. *International statistical classification of diseases, injuries and causes of death (ICD-10)*. World Health Organization, Geneva, 10th edition, 1993.

[35] World Health Organization. Visual impairment and blindness, December 2014. `http://www.who.int/mediacentre/factsheets/fs282/en/`.

[36] World Wide Hearing. Hearing loss facts, December 2014. `http://www.wwhearing.org/`.

[37] 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.