Anita: An Intelligent Text Adaptation Tool

Gustavo Henrique Paetzold and **Lucia Specia**
Department of Computer Science
University of Sheffield, UK
`{g.h.paetzold,l.specia}@sheffield.ac.uk`

Abstract

We introduce Anita: a flexible and intelligent Text Adaptation tool for web content that provides Text Simplification and Text Enhancement modules. Anita's simplification module features a state-of-the-art system that adapts texts according to the needs of individual users, and its enhancement module allows the user to search for a word's definitions, synonyms, translations, and visual cues through related images. These utilities are brought together in an easy-to-use interface of a freely available web browser extension.

1 Introduction

Readers who suffer from reading impairments find it difficult to understand certain types of texts which, to an average reader, would not pose any challenge. Low literacy readers and second language learners, for example, often have very limited vocabulary (Watanabe et al., 2009; Aluisio and Gasperin, 2010), while those with Dyslexia may have problems understanding the meaning of rare and/or long words (Ellis, 1993; Rello et al., 2013b). Other notable examples of such conditions are Aphasia and some forms of Autism, which can also hinder the patient's capability of comprehending sentences made up of a large amount of words and/or complex syntactic constructs (Devlin and Tait, 1998; Barbu et al., 2015).

Previous work has proposed a wide array of approaches that aim to adapt texts for these audiences. Text Simplification strategies are good examples of that. While Lexical Simplification approaches handle vocabulary limitations by replacing complex words with simpler alternatives (Devlin and Tait, 1998; Paetzold and Specia, 2016a), Syntactic Simplification approaches address the problem of long, complex syntactic constructs by re-structuring them (Siddharthan, 2006; Paetzold and Specia, 2013). Text Enhancement approaches can also help: Devlin and Unthank (2006), Watanabe et al. (2009) and Azab et al. (2015) adorn the words of a text with definitions, images and synonyms in order to facilitate their comprehension. Rello et al. (2013a) reveal that while simplification tends to increase a document's readability, enhancement tends to improve its comprehensibility.

One important limitation of the state of the art Text Adaptation systems is that they are not available for download and/or use. Online demos are provided for some, but they only allow the processing of small snippets of text through online interfaces. Another limitation is that the adaptations made by these systems are not personalised i.e. they will be the same for each and every user, regardless of their profiles and backgrounds.

There are, however, commercial reading/writing assistance tools, such as Simplish[1], texthelp[2] and Fast ForWord[3], which provide high quality services for those with reading difficulties. These tools are not free and most of them focus on text-to-speech capabilities, which makes them limited in functionality and inaccessible to the wider public.

In this contribution, we introduce Anita: a freely available Text Adaptation tool that, unlike previous work, tailors the provided assistance with respect to the needs of each user. In the sections that follow,

---

[1]http://www.simplish.org
[2]https://www.texthelp.com
[3]http://www.scilearn.com/products/reading-assistant

(a) User profile interface
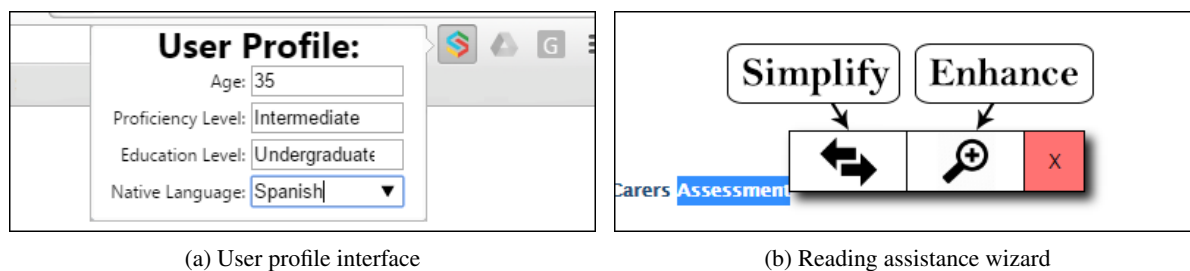

(b) Reading assistance wizard

Figure 1: Interface components of Anita

we describe Anita and its two main modules.

2 System Description

Anita is a Google Chrome extension that aims to assist individuals from various target audiences, such as non-native speakers and the poorly literate, to read and understand the content of web pages. Anita innovates by providing a minimalistic, user-friendly interface, as well as a wide array of state of the art Text Adaptation solutions for English, including an intelligent Lexical Simplification module.

In order to use Anita, the user must download and install the extension. Once installed, the tool can be configured with respect to the user's profile information, which will help in the assistance customisation process. Figure 1a illustrates the tool's profiling interface.

With a profile at hand, the tool is ready to provide personalised reading assistance. To launch it, the user must select a word they do not understand. The reading assistance wizard depicted in Figure 1b will then pop-up. Anita currently offers two types of adaptation: Simplification and Enhancement.

3 Simplification Module

Anita's simplification module attempts to replace the selected word with a simpler alternative. To do so, Anita first finds the sentence containing the selected word and then sends this information to the remote server where Anita's Lexical Simplification engine is running. The engine runs a state-of-the-art Lexical Simplification system powered by the LEXenstein framework (Paetzold and Specia, 2015). The strategy used here has been shown to outperform all other simplifiers from previous work (Paetzold and Specia, 2016a). Upon receiving a simplification request for a word, Anita's simplifier performs the following steps:

1. **Generation:** A context-aware word embeddings model trained over 7 billion words which accounts for grammatical information (Paetzold and Specia, 2016b) is used to produce candidate substitutions for the word.

2. **Selection:** The Unsupervised Boundary Ranking approach (Paetzold and Specia, 2016b) is used to select the candidates that best fit the context of the complex word.

3. **Ranking:** The selected candidates are ranked using a Supervised Boundary Ranking approach (Paetzold and Specia, 2015). The ranker is trained over a dataset composed of simplicity rankings produced by hundreds of non-native English speakers with different backgrounds, and checks the user's profile to determine which candidate best fits the user's simplification needs.

4. **Replacement:** Finally, the simplifier returns a response from the remote server to the chrome extension with the highest ranked candidate. The chrome extension then temporarily modifies the website's content for the user by replacing the selected word with the alternative provided.

Once the word is simplified, Anita highlights it, as illustrated in Figure 2a. If the highlighted word is selected again, the user will be presented with the interface in Figure 2b, which allows to either undo the simplification or enhance it. If the user chooses to undo the simplification, Anita will send a report to the

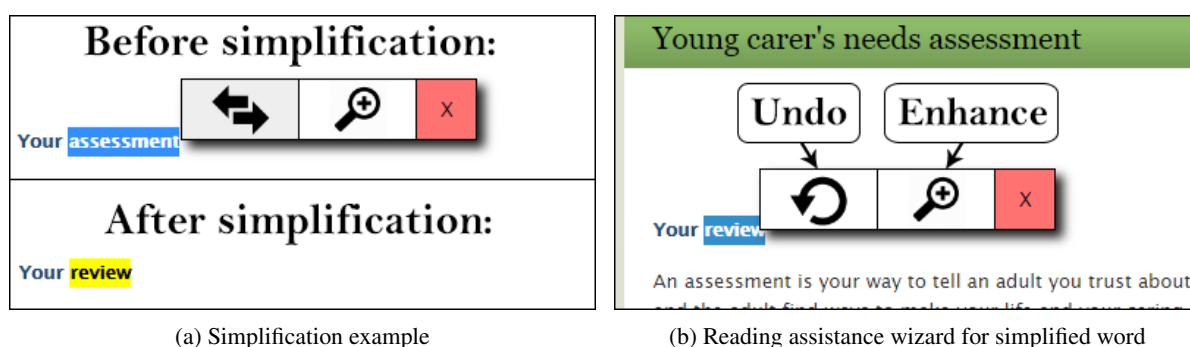(a) Simplification example   (b) Reading assistance wizard for simplified word

Figure 2: Anita's Simplification module

remote server containing the rejected simplification. The content of the report is then fed as a negative training instance to Anita's Lexical Simplification system, which is periodically re-trained to improve prediction for specific users' needs.

4 Enhancement Module

Anita's enhancement module allows the users to learn more about the words they find complex. Unlike in simplification, enhancement does not require the website's content to be modified in any way. As discussed in (Devlin and Unthank, 2006; Rello et al., 2013a), enhancements can help "jog" the memory of the user, and consequently increase comprehensibility. They are also preferred by many as a better alternative to "dumbing down" the language: by explaining a complex concept instead of replacing it by a simpler one, the reader is given the opportunity to learn about a new concept.

When the user requests an enhancement, they will be presented with the interface illustrated in Figures 3a through 3d. The enhancement interface offers the following services, all of which can be customised:

- **Definitions:** Shows dictionary definitions of the word. In the demo version of the tool, the definitions are queried from the Merriam Dictionary and Thesaurus[4], which provides a free API.

- **Synonyms:** Shows synonyms of the word. In the demo version of the tool, synonyms are also queried from the Merriam Dictionary.

- **Translations:** Shows machine translations of the word and of its synonyms. The language in which the translations are presented is defined in the User Profile interface (Section 2). Translations are currently produced by the Yandex API[5], which is also free.

- **Images:** Shows images related to the word. In the demo version of the tool, the images are freely available thumbnails queried from the Getty Images API [6].

Anita's word enhancements are customised through ranking: definitions, synonyms and translations are ranked so that the most helpful among them are featured at the top of the list. To do so, Anita employs the same Supervised Boundary Ranking strategy used in its Simplification module to rank synonyms and their translations by simplicity, based on the user's profile. In order to rank definitions, it first ranks the simplicity of all words in every one of them, then places definitions with the highest average word simplicity at the top of the list.

5 Final Remarks

We introduced Anita, an intelligent Text Adaptation tool composed of a Simplification module, which continuously learns how to adapt its simplifications to the users' needs, and an Enhancement module,

---

[4]http://www.merriam-webster.com
[5]https://www.yandex.com
[6]http://www.gettyimages.com

(a) Examples of word definitions

(b) Examples of word synonyms

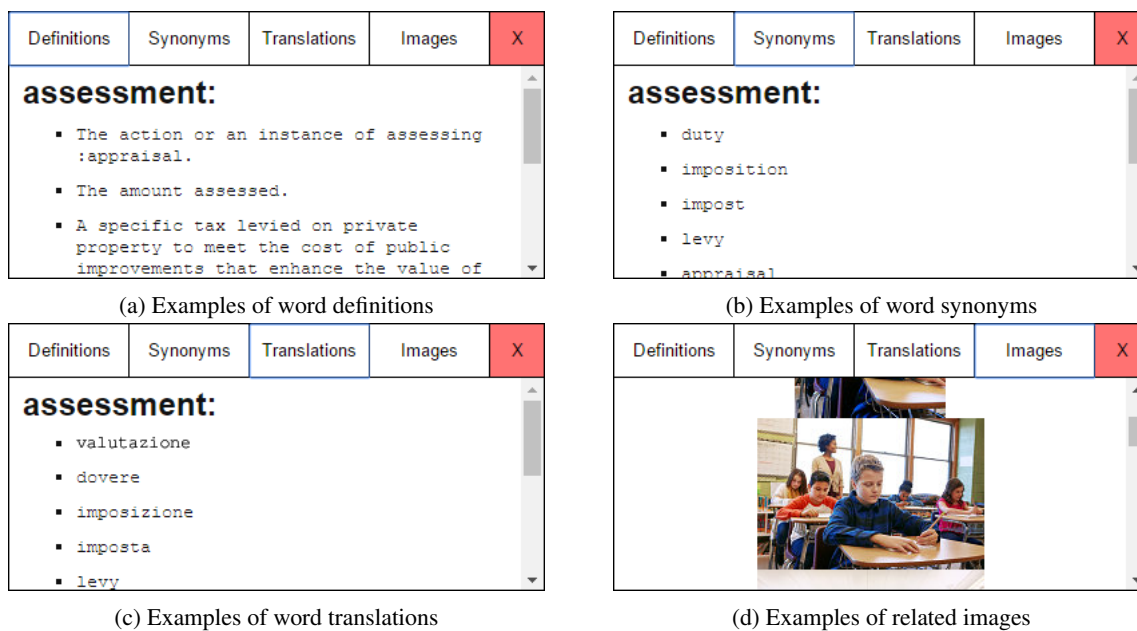(c) Examples of word translations

(d) Examples of related images

Figure 3: Anita's Enhancement module

which allows the user to query for different types of information about words. Anita is an open-source tool from the SIMPATICO project[7] that will be released under a permissive BSD license. In the future, we will extend the tool by adding syntactic and semantic simplification functionalities. We also intend to follow the efforts of (Rello et al., 2013a) and conduct studies in order to investigate how the Anita utilities affect readability and comprehensibility for users.

Acknowledgements

References

Sandra Aluisio and Caroline Gasperin. 2010. Fostering digital inclusion and accessibility: The porsimples project for simplification of portuguese texts. In *Proceedings of the 2010 NAACL*, pages 46–53.

Mahmoud Azab, Chris Hokamp, and Rada Mihalcea. 2015. Using word semantics to assist english as a second language learners. In *Proceedings of NAACL*, pages 116–120.

Eduard Barbu, M Teresa Martín-Valdivia, Eugenio Martínez-Cámara, and L Alfonso Ureña-López. 2015. Language technologies applied to document simplification for helping autistic people. *Expert Systems with Applications*, 42:5076–5086.

Siobhan Devlin and John Tait. 1998. The use of a psycholinguistic database in the simplification of text for aphasic readers. *Linguistic Databases*, pages 161–173.

Siobhan Devlin and Gary Unthank. 2006. Helping aphasic people process online information. In *Proceedings of the 8th SIGACCESS*, pages 225–226.

A.W. Ellis. 1993. *Reading, Writing and Dyslexia: A Cognitive Analysis*. Open University Press.

Gustavo H. Paetzold and Lucia Specia. 2013. Text simplification as tree transduction. In *Proceedings of the 9th STIL*, pages 116–125.

Gustavo Henrique Paetzold and Lucia Specia. 2015. Lexenstein: A framework for lexical simplification. In *Proceedings of The 53rd ACL*, pages 85–90.

---

[7]http://www.simpatico-project.eu

Gustavo Henrique Paetzold and Lucia Specia. 2016a. Benchmarking lexical simplification systems. In *Proceedings of the 10th LREC*.

Gustavo Henrique Paetzold and Lucia Specia. 2016b. Unsupervised lexical simplification for non-native speakers. In *Proceedings of The 30th AAAI*, pages 3761–3767.

Luz Rello, Ricardo Baeza-Yates, Stefan Bott, and Horacio Saggion. 2013a. Simplify or help?: text simplification strategies for people with dyslexia. In *Proceedings of the 10th W4A*, pages 1–10.

Luz Rello, Ricardo Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion. 2013b. Frequent words improve readability and short words improve understandability for people with dyslexia. *Human-Computer Interaction*, pages 203–219.

Advaith Siddharthan. 2006. Syntactic simplification and text cohesion. *Research on Language and Computation*, 4(1):77–109, March.

Willian Massami Watanabe, Arnaldo Candido Junior, Vinícius Rodriguez Uzêda, Renata Pontin de Mattos Fortes, Thiago Alexandre Salgueiro Pardo, and Sandra Maria Aluísio. 2009. Facilita: reading assistance for low-literacy readers. In *Proceedings of the 27th ACM*, pages 29–36.