A Review of Research-based Automatic Text Simplification Tools

Isabel Espinosa-Zaragoza[1], **José Abreu-Salas**[2], **Elena Lloret**[3], **Paloma Moreda**[3] **and Manuel Palomar**[3]
[1] Center of Digital Intelligence, University of Alicante
isabel.espinosa@ua.es
[2] University Institute for Computing Research, University of Alicante
ji.abreu@ua.es
[3] Department of Computing and Information Systems, University of Alicante
{elloret,paloma,mpalomar}@dlsi.ua.es

Abstract

In the age of knowledge, the democratisation of information facilitated through the Internet may not be as pervasive if written language poses challenges to particular sectors of the population. The objective of this paper is to present an overview of research-based automatic text simplification tools. Consequently, we describe aspects such as the language, language phenomena, language levels simplified, approaches, specific target populations these tools are created for (e.g. individuals with cognitive impairment, attention deficit, elderly people, children, language learners), and accessibility and availability considerations. The review of existing studies covering automatic text simplification tools is undergone by searching two databases: Web of Science and Scopus. The eligibility criteria involve text simplification tools with a scientific background in order to ascertain how they operate. This methodology yielded 27 text simplification tools that are further analysed. Some of the main conclusions reached with this review are the lack of resources accessible to the public, the need for customisation to foster the individual's independence by allowing the user to select what s/he finds challenging to understand while not limiting the user's capabilities and the need for more simplification tools in languages other than English, to mention a few.

1 Introduction

In the age of knowledge and information, the democratisation of information facilitated through the Internet may not be as pervasive owing to potential challenges posed by written language, particularly among specific segments of the population. A great deal of the daily life processes are written and may produce lexical, syntactic and/or semantic difficulties in general, but particularly for those most vulnerable, such as people with cognitive disabilities, autism spectrum disorders, non-native speakers, children, and others. The guidelines provided by organisations like the Plain Language Association International (PLAIN)[1] and easy-to-read movement (AENOR, 2018) already highlight both the need for and the promotion of text understandability via the simplification of specific language phenomena. Therefore, enhancing text readability and comprehensibility becomes essential to uphold the right to cognitively accessible texts. Currently, these simplification tasks are laborious and time-consuming as they are conducted manually. Thus, Natural Language Processing (NLP) techniques, particularly Automatic Text Simplification (ATS), are demanded by society to address this issue.

The objective of this paper is to present the existing tools for ATS, paying particular attention to those whose target audience is a specific group of people with special needs. Consequently, an analysis of these tools is conducted to determine the specific languages, language phenomena and linguistic levels they simplify; the approaches followed; their intended target audience (i.e. individuals with cognitive impairment, language difficulties, attention deficit, and others); and other relevant aspects.

This study is framed as part of a larger project, the ClearText project[2], that aims at the creation of a text simplifying tool for the simplification of Spanish texts from the public administration to help people with mild to moderate cognitive impairment. In order to accomplish our goal, a preliminary assessment of the existing ATS tools is required to ascertain the advancements made, methodologies employed, and potential areas for refinement in our own simplification tool.

[1] https://plainlanguagenetwork.org/
[2] https://cleartext.gplsi.es/

Proceedings of Recent Advances in Natural Language Processing, pages 321–330
Varna, Sep 4–6, 2023
https://doi.org/10.26615/978-954-452-092-2_036

2 On the Right to Understand

The inherent difficulty in certain written texts has caused society to demand more transparent and accessible texts. This has resulted in several movements, like the plain language movement and the easy-to-read movement.

The plain language movement defends understandable language that ensures the fulfilment of the text's purpose. In fact, Eagleson (1997) even affirms that "[...] it is the writer's responsibility to be clear. It is not the reader's responsibility to understand". As this is not always the case, ATS tools provide citizens with the necessary means to access otherwise unreachable information.

While the plain language movement has the entire society as target audience, the easy-to-read movement is concerned with increasing both the reading and comprehension of texts for those more vulnerable. The individuals that may benefit from easy-to-read materials may be subsumed under two categories: (1) people with disabilities and (2) readers with a limited language proficiency (Nomura et al., 2010). The former category encompasses individuals with conditions such as aphasia, dementia, autism, intellectual disabilities (spanning mild to moderate and profound), neuropsychiatric disabilities (e.g., attention deficit hyperactivity disorder (ADHD)), deafblindness, deafness or hearing impairments (DHH), Asperger syndrome, Tourette syndrome, dyslexia, and other reading difficulties. The latter category comprises non-native speakers, individuals with limited reading abilities, and children.

3 Automatic Text Simplification

Automatic Text Simplification (ATS) can be defined as "a technology to produce adaptable text by reducing their syntactic and lexical complexity so that they become readable for a target user group" (Bott and Saggion, 2012).

3.1 Levels of Simplification According to Language Phenomena

Simplification tools primarily focus on addressing lexical and/or syntactic language phenomena to enhance readability and comprehensibility although, in some cases, stylistic modifications are also employed. According to Chen et al. (2017), ATS is composed of lexical, syntactic and discourse simplification levels.

Lexical simplification entails the identification of complex words i.e. infrequent, technical, abstract and others, and replacing them with simpler, more general, frequent and concrete synonyms. It can also be solved by enriching or enhancing the text by providing a definition, image or video, among others. Implicit in this step is the disambiguation task, which entails selecting the most prevalent meaning among the list of synonyms available. Presently, relying solely on the most frequent sense of a word can engender issues that require further solutions in future ATS research endeavors.

Syntactic simplification involves the reduction of sentence structure complexity i.e. passive constructions, long sentences, appositions, relative clauses. As a result, this process includes sentence structure reordering, splitting, and adjustment, as well as the reduction of grammar complexity and the elision of unnecessary information.

Discourse simplification is concerned with ascertaining that no information is lost in the previous lexical and syntactic simplifications, especially pronouns. Hence, discourse simplification is a step that tackles coreference and coherence aspects, like anaphora resolution, replacing new or repeated entities or making noun phrases more accessible (Todirascu et al., 2022).

Regarding **stylistic simplification** and interface design, in other words, how the textual elements are presented to the user, visual design and layout also affect text readability. Works covering font size and line spacing (Rello et al., 2016), highlighting paragraphs (Kobayashi and Kawashima, 2019), or having whitespace between paragraphs to enhance webpage readability (Yu and Miller, 2010), among others, support this view. Additionally, the guidelines provided by the entities and organisations mentioned in Section 2 also cover stylistic aspects. While we acknowledge that it is not the primary objective of ATS to perform this specific task, we have chosen to include it due to the availability of such stylistic options in certain tools.

3.2 Tool Approaches

As indicated by Al-Thanyyan and Azmi (2021), ATS has followed three different approaches:

(1) A rule-based approach (Siddharthan, 2006) involves a significant amount of handcrafted rules where certain linguistic phenomena are located and replaced. For instance, identifying complex words

and replacing them with simpler, shorter, and more frequent synonyms; using active voice instead of passive voice, among others. This represents the conventional approach within ATS for languages lacking extensive parallel corpora comprising original text and its corresponding simplified version.

(2) A data-driven approach, also regarded as corpus-driven approach or machine learning-based approach, like in Zhu et al. (2010) and Kauchak (2013), is characterised by the use of large parallel data resources through the deployment of machine learning or deep learning techniques, such as neural networks and word embeddings. For instance, LexSiS is a lexical simplification algorithm for Spanish (Bott et al., 2012a).

(3) A hybrid approach, combines the previous two, like in Siddharthan and Mandya (2014) and Bott et al. (2012b).

3.3 Target Users

Several ATS projects have been created with the end user in mind, such as the PSET project (Practical Simplification of English Texts) (Carroll et al., 1998), intended for people with aphasia, which later resulted in the HAPPI project (Devlin and Unthank, 2006); the PorSimples Project (Aluisio et al., 2010), for low literacy individuals; the Simplext project (Saggion et al., 2015b) and the Able2Include project (Saggion et al., 2017) for people with intellectual disabilities; and the FIRST project (Valdivia et al., 2014) for people with autism. Although it must be pointed out that some of them do not offer a corresponding simplification tool.

4 Methodology

This tool review was carried out by following a five-step methodology detailed below. A systematic review of studies was undergone by searching two databases: Web of Science[3] and Scopus[4].

Step 0. Research scope definition and eligibility criteria. We are not concerned with an exhaustive analysis of ATS tools but rather with those tools which are (1) ATS tools with (2) a scientific background, in other words, the tool is supported by a research group. Thus, papers dealing with other simplification aspects, i.e. simplification tool metrics, datasets or corpora, tools for automatic assessment of conceptual text complexity, methods, individual parsers, paraphrasing, lexical resources, tools to enhance readability, etc., are not considered.

Step 1. Search method and bibliographic database query. This step entails the initial search of generic terms dealing with ATS until April of 2023. For this purpose, and as we previously mentioned, Scopus and Web of Science were the selected databases we used. The query utilised was "text simplification" AND "tool" for both databases, which yielded 115 papers: all fields included in case of Web of Science produced 31 results and only article title, abstract and keywords in Scopus provided 84 results.

Step 2. Result fine-grain filtering. This step consists of selecting the papers that are within our scope (i.e. papers presenting a simplification tool) and dismissing those beyond our scope. For instance, the paper dealing with the *Alector* parallel corpus (Gala et al., 2020) or *CoCo*, a tool for the assessment of conceptual complexity (Štajner et al., 2020), were discarded. In addition, preliminary studies where the tool is a prototype not yet developed (i.e. the tool is not named and the simplification levels are not explained) were also not taken into account, as for instance the case of Moen et al. (2018) or Kandula et al. (2010). Repeated papers in both databases and tools presented by several papers were considered only once. After this step, 8 papers were selected and 8 tools were obtained.

Step 3. Result checking and recovery. Finally, this step involves the addition of the papers dealing with ATS in general which were dismissed in the previous step because they do not present a simplification tool. Upon closer revision and examination, they mention one or several ATS tools, mainly in the state of the art section. This step added 19 more papers covering 19 tools. Given that these findings double the results of Step 2, we revisited the underlying cause for the absence of those papers in our query results: it is attributable to the omission of the term "tool" in the titles, abstracts or keywords in those papers. Consequently, our method, far from being erroneous, effectively captures and retrieves ATS tools that would have otherwise been overlooked.

Step 4. Tool analysis. In total, 27 tools were selected after this process. The list of selected tools yielded was analysed to determine the following: (1) the language simplified, the language phenomena tackled, the language level simplified and the

[3] https://www.webofscience.com
[4] https://www.scopus.com/

specific domain (if any); (2) the tool's approach; (3) the specific target audience of the tool; and (4) whether or not these tools are accessible and operative at the moment (i.e. the tool includes an interface and allows the text simplification process) and if they are open-source (i.e. made freely available for the rest of researchers).

5 Simplification Tools Review

As mentioned previously in Step 0, commercial tools were discarded. Although some deductions of what these tools are able to do can be ascertained, there is no way to know which operations (i.e. split, replace, reorder, etc.) the text has undergone in the simplification process. Nonetheless, we acknowledge the usefulness of such tools for the general population, regardless of the shortcomings these often might have: character limitation, payment access restrictions and others. As a way of example, some commercial tools that help users in text simplification without any character limitations are *SIMPLISH*[5] and *Rewordify*[6].

Next, we present the tools selected following the previously explained methodology and analyse the language, language levels and domains they simplify, as well as their respective approaches, intended target users, and accessibly and availability considerations.

5.1 Languages, Language Levels Simplified and Specific Domains

Efforts have been made to create monolingual text simplification tools, especially in English, with 12 out of 27 (44.44%) tools analysed being in English (see Table 1). Nevertheless, Romance languages like Spanish, French, Italian or Portuguese are also present. We can observe a lack of multilingual simplification tools, with only two exceptions: *MUSST*, for English, Spanish and Italian, and *Open Book*, for English, Spanish and Bulgarian.

Concerning the language level simplified by these tools, the vast majority (23, 85.19%) perform **lexical simplifications**, with 11 tools exclusively simplifying at this particular level. This is usually carried out by means of providing more frequent or accessible synonyms, but it may also be solved by enriching the text by offering a definition, a link to Wikipedia or similar sources, and audiovisual aids like pictures or videos. These simplifications are implemented by means of dictionaries of synonyms and databases with the most frequent word sense. For instance, *NavegaFácil* provides definitions, synonyms and antonyms, lemmatisations, images, Google search, Wikipedia, translation and text to voice.

Syntactic simplification is implemented in roughly half of the tools analysed (14, 51.85%). The fact that not all the tools simplify at this level undermines the overall quality of the simplified text. Some other tools only simplify at a syntactic level, like *MUSST*, *Split* and *EuTS*. In fact, *EuTS* tackles a superficial syntactic simplification but maintaining the general structure of the original text. In addition, *FACILITA* uses summarisation and simplification techniques and its syntactic simplification consists of sentence splitting, change of discourse markers, passive to active voice, inversion of clause order, SVO order (subject-verb -object) and (de)topicalisation.

Regarding **discourse simplification**, 5 tools (18.52%) tackle issues related to discourse. For instance, *ERNESTA* addresses anaphora resolution combined with syntactic simplification. *HECTOR* adjusts the coreference chains during the syntactic transformations and, in this way, replaces new or repeated entities, specifies entities, makes noun phrases more accessible. And *ArText* includes discourse-based recommendations, like varying discourse markers.

Lastly, **stylistic changes** are undergone by adapting the typography (e.g. font size, font and background colour, and others) to maximise the understanding of the message and minimise the effort made by the reader. Simplification tools that also modify the font and other stylistic-related aspects are *NavegaFácil*, *FRIENDLYREADER* and *DysWebsia*.

If we consider the entire palette of simplification levels (i.e. lexical, syntactic, discursive and stylistic), only *FRIENDLYREADER* covers all of these levels of simplification (3.70%), whereas *ArText*, *HECTOR* and *Open Book* incorporate 3 out of 4 levels (11.11%). The rest of the ATS tools examined either simplify at one level (14, 51.85%) or two levels (9, 33.33%).

With respect to the specific language domain, even though the majority of tools (22, 81.48%) have a generalist approach, there are tools devoted to the medical field (2, 7.41%), such as *Medical*

[5] https://www.simplish.org/
[6] https://rewordify.com/

Tool	Reference	Language	Level	Approach	User	Access and code
AI-Baseet	(Al-Subaihin and Al-Khalifa, 2011)	AR	LX, SN	H	M	-
ALTER	(Xu et al., 2019)	EN	LX	DD	-	-+
Anita*	(Paetzold and Specia, 2016)	EN	LX	DD	S	- +
ArText	(da Cunha Fanego et al., 2017)	ES	DIS, LX, SN	RB	M	O
CASSA plug-in*	(Rello et al., 2015)	EN	LX	RB	S	I
DysWebxia	(Rello et al., 2013)	ES	LX, ST	-	S	I
EASIER	(Alarcón et al., 2021)	ES	LX	DD	M	O +
ERNESTA	(Barlacchi and Tonelli, 2013)	IT	DIS, SN	H	S	I
EuTS	(Gonzalez-Dios, 2017)	EU	SN	RB	-	-
FACILITA*	(Watanabe et al., 2009)	PT	LX, SN	RB	S	I
FrenLys	(Rolin et al., 2021)	FR	LX	DD	-	I
FRIENDLYREADER	(Rennes et al., 2022)	SV	DIS, LX, SN, ST	H	M	O
HECTOR	(Todirascu et al., 2022)	FR	DIS, LX, SN	H	M	-
Lexi*	(Bingel et al., 2018)	DA	LX	DD	S	I +
LexSiS	(Bott et al., 2012a)	ES	LX	DD	-	-
MTST	(Kauchak and Leroy, 2020)	EN	LX, SN	DD	S	-
MUSST	(Scarton et al., 2017)	EN/ES/IT	SN	RB	M	- +
NavegaFácil	(Bautista et al., 2018)	ES	LX, ST	H	M	- +
Open Book	(Barbu et al., 2015)	BG/EN/ES	DIS, LX, SN	RB	S	I
SALSA	(Azab et al., 2015)	EN	LX	RB	S	-
SIMPLE	(MacMahon et al., 2019)	EN	LX	RB	S	I
Simplext	(Saggion et al., 2015a)	ES	LX, SN	H	S	O
SIMPLIFICA	(Candido Jr et al., 2009)	PT	LX, SN	RB	M	I+
Split*	(Hervás et al., 2014)	EN	SN	RB	-	-+
Synonyms*	(Hervás et al., 2014)	EN	LX	RB	-	- +
Text Adaptation	(Burstein et al., 2007)	EN	LX	RB	S	I
YATS	(Ferrés et al., 2016)	EN	LX, SN	H	-	-

Table 1: Summary of the simplification tools analysed. In accordance with the column information, the first column includes the tools analysed. The ones that include an asterisk are also plug-ins. The language abbreviations in the third column "AR", "BG", "DA", "EN", "ES", "EU", "FR", "IT", "PT", and "SV" correspond to Arabic, Bulgarian, Danish, English, Spanish, Basque, French, Italian, Portuguese and, Swedish respectively, progressing from top to bottom. The abbreviations dealing with the language levels simplified that appear in the fourth column, "DIS", "LX", "SN", and "ST" stands for "discourse", "lexical", "syntactic", and "stylistic", respectively. The user abbreviations employed in the fifth column are "M" and "S", denoting "multiple" and "specific" correspondingly. Regarding the approaches, "DD", "RB", and "H" stands for data-driven, rule-based, and hybrid, respectively. Only one of the tools, *DysWebsia*, remains unknown. Lastly, in the final column assessing tool accessibility and their open-source code, "I" and "O" represent "inoperative" and "operative" in relation to the tool's access link, while a "+" symbol signifies open-source code.

Text Simplification Tool[7] and *SIMPLE*; for educational purposes (2, 7.41%), like *SALSA* and *Text Adaptation*; or for public administration users (1, 3.70%), such as *ArText*.

5.2 Technical Approach for Simplification

In this section, we analyse the approach taken for text simplification. In general, the automatic simplification process comprises two stages (Cripwell et al., 2023): (1) the simplification plan, which refers to the decision about what linguistic aspect to simplify, for instance, identifying complex words or sentences; and (2) the simplification stage, when the plan to produce the simplified content is applied, e.g., splitting long sentences. It is worth noting that a system may perform these tasks holistically without a clear distinction between stages, as in neural generative models (Ondov et al., 2022).

There are three common approaches to solving tasks at each step (Al-Thanyyan and Azmi, 2021). On the one hand, the **rule-based** approach relies on linguistic expertise that is algorithmised enabling the system to perform the task. One example is *SIMPLIFICA* where a set of rules involving PoS tagging, disambiguation algorithms, and dictionaries of complex words are used for lexical simplification. On the other hand, **data-driven** approaches may leverage different corpora to learn how to perform different tasks. Just to illustrate, Sheang and Saggion (2023) and Qiang et al. (2021) trained language models to generate substitution candidates for lexical simplification. Finally, **hybrid** systems may leverage both data-driven and rule-based approaches.

Table 1 shows the following findings regarding the tool approaches: the majority of tools are rule-based (12, 44.44%), whereas 7 are data-driven (25.93%), 7 are hybrid tools (25.93%) and one, *DysWebsia*, is not specified (3.70%). Most data-driven approaches focus on lexical simplification either for complex word identification, such as *Lexi* or *EASIER*, or substitution generation, as in the case of *Anita*. Another aspect worth discussing is the lack of tools leveraging recent advances in large language models (LLM), even for lexical simplification, although there are exceptions such as Rolin et al. (2021) using CamenBERT (Martin et al., 2020). Again, other proposals outside this review, such as Qiang et al. (2021), explored LLMs but without developing a tool.

[7]Onwards referred to as MTST in Table 1 for brevity.

5.3 Target Users

Regarding the target users of the analysed tools, these usually have either (1) a generalist approach with multiple target users or (2) a more specific or specialised approach, by targeting particular target groups like dyslexic people. However, some tools do not explicitly mention whether they were conceived with a target user in mind (see Table 1).

On the one hand, 12 tools (44.44%) have a **specific target audience**. For instance, *SALSA*, aimed at English as a second and foreign language students; *FACILITA*, intended for low literacy readers; *ERNESTA*, created for children with low reading skills; *Open Book*, designed for autistic people; or *DysWebsia*, developed for dyslexic individuals. In addition, under specific target audiences are also subsumed other personalised tools, like *Lexi* and *Medical Text Simplification Tool*, that are customised according to the individual's particular needs.

On the other hand, some other tools have **multiple target audiences** (8, 29.63%): those tools aimed at a wider audience and considered a one-size-fits-all approach by (Bingel et al., 2018), such as people with cognitive disabilities in general, like *NavegaFácil* or *EASIER*; or varied audiences like poor literate individuals, language learners and children (*AI-Baseet*); teachers, publishers, journalists, companies, and others (*SIMPLIFICA*); people with aphasia, dyslexia, intellectual disability, deaf or hard-of-hearing (DHH), second language learners and children (*FRIENDLYREADER*); or specialists, medicine and tourism university students, laypeople and public administration (*ArText*).

Lastly, there are 7 tools (25.93%) that **do not specify** whether they were conceived with a specific target in mind (see Table 1).

5.4 Accessibility and Availability

The vast majority of the simplification tools analysed (23 out of 27, 85.19%) are currently inaccessible either because (1) the link is not working and, therefore, they are inoperative[8] at the moment of the analysis or (2) the link to the tools is not provided and left unspecified[9] in the paper (see Table 1). This means that only four (14.81%) of the tools examined are currently functional and accessible for use[10]: *ArText*, which instead of out-

[8]Indicated with I in Table 1.

[9]Indicated with a hyphen in Table 1.

[10]Indicated with O in Table 1.

putting a simplified text, it identifies the complex language phenomena and recommends solutions; *EASIER*, which identifies complex words in a text and provides a definition; and *FRIENDLYREADER* and *Simplext*, which output simplified text. These results evidence the need to maintain these simplification resources, both technically and in financing terms, so that they fulfil their intended purpose.

Respecting the tool's open-source nature[11], less than half of the tools explicitly acknowledge the availability of their open-source code in their respective papers (see Table 1).

6 Conclusions and Future Work

In this paper we conducted a review of research-based ATS tools to determine which language they simplify, what simplifications are applied, which approaches are followed, who are the target users (e.g. people with disorders and disabilities, students, children, and others) and whether or not these tools are accessible to the public and available for researchers. From this analysis some general conclusions are reached concerning what these tools have to offer, what they are lacking and other future considerations in NLP:

- **Languages simplified and language level simplification.** ATS is an area with a promising future as many languages are still underrepresented in the results derived from this study. If the objective is to create a tool that truly helps people with written comprehension, all levels of simplification must be taken into consideration.

- **Multioption and customisation.** ATS tools should offer multiple options or solutions for the technical and/or complex vocabulary, such as synonyms, definitions, images, links to explanatory webpages, text-to-speech, and translation, to name a few, in order to enrich the text and cater to the different users' needs. A one-size-fits-all simplification approach is not the ideal way of creating simplification tools. These should foster the individual's independence by allowing the user to select what s/he finds challenging to understand and not limiting the user's capabilities.

- **Approaches.** There is a lack of tools based on neural or other data-driven holistic approaches, e.g. performing different types of simplifications at once, after learning from examples of complex/simple text (Ondov et al., 2022). Moreover, we did not detect any tool leveraging advances in LLMs —with some exceptions— but we expect this area to be explored in the future.

- **Target audience.** We understand that the targets' needs are different and, consequently, the text simplifications they require ought to be different as well. Evidently, tools that adopt a generalist approach, albeit targeting a broader range of population, do not refine the simplification depending on the user's needs to the same extent as individualist tools do.

- **Accessibility and availability**. While a substantial amount of research is dedicated to ATS, the full accessibility and functionality of ATS tools is crucial so that the valuable efforts made by the scientific community are effectively disseminated to society.

After this preliminary study, the results indicate different paths that research groups could improve upon, like simplifying more language levels, customising simplifications by having into account the user's needs, maintaining tool accessibility and including other languages that still require simplification tools, among others. Thus, we encourage to continue researching, implementing and providing robust ATS tools to facilitate access information to society at large.

In the context of the ClearText project, the goal is a two-fold simplification approach by addressing both disability-related and individual-specific language obstacles. In this way, we enable users to determine the extent to which they address the language obstacles associated with their specific disabilities, while considering that each individual exhibits unique idiosyncrasies and varying impairment degrees.

Acknowledgments

This research was conducted as part of the ClearText project (TED2021- 130707B-I00), funded by MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR.

[11] Indicated with a + in Table 1.

References

AENOR. 2018. Norma española experimental une 153101 ex. lectura fácil: Pautas y recomendaciones para la elaboración de documentos.

Afnan A Al-Subaihin and Hend S Al-Khalifa. 2011. Albaseet: A proposed simplification authoring tool for the Arabic language. In *2011 International Conference on Communications and Information Technology (ICCIT)*, pages 121–125. IEEE.

Suha S Al-Thanyyan and Aqil M Azmi. 2021. Automated text simplification: a survey. *ACM Computing Surveys (CSUR)*, 54(2):1–36.

Rodrigo Alarcón, Lourdes Moreno, and Paloma Martínez. 2021. Lexical simplification system to improve web accessibility. *IEEE Access*, 9:58755–58767.

Sandra Aluisio, Lucia Specia, Caroline Gasperin, and Carolina Scarton. 2010. Readability assessment for text simplification. In *Proceedings of the NAACL HLT 2010 fifth workshop on innovative use of NLP for building educational applications*, pages 1–9.

Mahmoud Azab, Chris Hokamp, and Rada Mihalcea. 2015. Using word semantics to assist English as a second language learners. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 116–120.

Eduard Barbu, María Teresa Martín-Valdivia, Eugenio Martínez-Cámara, and Luis Alfonso Ureña López. 2015. Language technologies applied to document simplification for helping autistic people. *Expert Systems with Applications*, 42(12):5076–5086.

Gianni Barlacchi and Sara Tonelli. 2013. Ernesta: A sentence simplification tool for children's stories in Italian. In *Computational Linguistics and Intelligent Text Processing - 14th International Conference, CICLing 2013, Samos, Greece, March 24-30, 2013, Proceedings, Part II*, volume 7817 of *Lecture Notes in Computer Science*, pages 476–487. Springer.

Susana Bautista, Raquel Hervás, Pablo Gervás, Axel Bagó, and Javier García-Ortiz. 2018. Taking text simplification to the user: integrating automated modules into a web browser. In *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, pages 88–96.

Joachim Bingel, Gustavo Paetzold, and Anders Søgaard. 2018. Lexi: A tool for adaptive, personalized text simplification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 245–258.

Stefan Bott, Luz Rello, Biljana Drndarevic, and Horacio Saggion. 2012a. Can Spanish be simpler? lexsis: Lexical simplification for Spanish. In *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, 8-15 December 2012, Mumbai, India*, pages 357–374. Indian Institute of Technology Bombay.

Stefan Bott and Horacio Saggion. 2012. Automatic simplification of Spanish text for e-accessibility. In *Computers Helping People with Special Needs: 13th International Conference, ICCHP 2012, Linz, Austria, July 11-13, 2012, Proceedings, Part I 13*, pages 527–534. Springer.

Stefan Bott, Horacio Saggion, and David Figueroa. 2012b. A hybrid system for Spanish text simplification. In *Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies*, pages 75–84.

Jill Burstein, Jane Shore, John Sabatini, Yong-Won Lee, and Matthew Ventura. 2007. The automated text adaptation tool. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, April 22-27, 2007, Rochester, New York, USA*, pages 3–4. The Association for Computational Linguistics.

Arnaldo Candido Jr, Erick Galani Maziero, Lucia Specia, Caroline Gasperin, Thiago Pardo, and Sandra Aluisio. 2009. Supporting the adaptation of texts for poor literacy readers: a text simplification editor for brazilian Portuguese. In *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 34–42.

John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998. Practical simplification of English newspaper text to assist aphasic readers. In *Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pages 7–10. Association for the Advancement of Artificial Intelligence.

Ping Chen, John Rochford, David N Kennedy, Soussan Djamasbi, Peter Fay, and Will Scott. 2017. Automatic text simplification for people with intellectual disabilities. In *Artificial Intelligence Science and Technology: Proceedings of the 2016 International Conference (AIST2016)*, pages 725–731. World Scientific.

Liam Cripwell, Joël Legrand, and Claire Gardent. 2023. Document-level planning for text simplification. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 993–1006.

Iria da Cunha Fanego, M Amor Montané March, and Luis Hysa. 2017. The artext prototype: An automatic system for writing specialized texts. In *Martins A, Peñas A, editors. EACL 2017. 15th Conference of the European Chapter of the Association for Computational Linguistics. Proceedings of the Software Demonstrations; 2017 Apr 3-7; Valencia, Spain. Stroudsburg (PA): ACL; 2017. p. 57-60*. ACL (Association for Computational Linguistics).

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

Robert Donn Eagleson. 1997. *Writing in plain English*. Australian Government Public Service.

Daniel Ferrés, Montserrat Marimon, Horacio Saggion, and Ahmed AbuRa'ed. 2016. Yats: yet another text simplifier. In *Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings 21*, pages 335–342. Springer.

Núria Gala, Anaïs Tack, Ludivine Javourey-Drevet, Thomas François, and Johannes C Ziegler. 2020. Alector: A parallel corpus of simplified French texts with alignments of misreadings by poor and dyslexic readers. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1353–1361.

Itziar Gonzalez-Dios. 2017. Análisis de la complejidad y simplificación automática de textos. el análisis de las estructuras complejas en euskera. *Procesamiento del Lenguaje Natural*, 58:155–158.

Raquel Hervás, Susana Bautista, Marta Rodríguez, Teresa de Salas, Ana Vargas, and Pablo Gervás. 2014. Integration of lexical and syntactic simplification capabilities in a text editor. *Procedia Computer Science*, 27:94–103.

Sasikiran Kandula, Dorothy Curtis, and Qing Zeng-Treitler. 2010. A semantic and syntactic text simplification tool for health content. In *AMIA annual symposium proceedings*, volume 2010, page 366. American Medical Informatics Association.

David Kauchak. 2013. Improving text simplification language modeling using unsimplified text data. In *Proceedings of the 51st annual meeting of the association for computational linguistics (volume 1: Long papers)*, pages 1537–1546.

David Kauchak and Gondy Leroy. 2020. A web-based medical text simplification tool. In *53rd Hawaii International Conference on System Sciences, HICSS 2020, Maui, Hawaii, USA, January 7-10, 2020*, pages 1–9. ScholarSpace.

Jumpei Kobayashi and Toshio Kawashima. 2019. Paragraph-based faded text facilitates reading comprehension. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12. ACM.

Silvana Togneri MacMahon, Marco Alfano, Biagio Lenzitti, Giosuè Lo Bosco, Fergal McCaffery, Davide Taibi, and Markus Helfert. 2019. Improving communication in risk management of health information technology systems by means of medical text simplification. In *2019 IEEE Symposium on Computers and Communications (ISCC)*, pages 1135–1140. IEEE, IEEE.

Louis Martin, Benjamin Muller, Pedro Ortiz Suarez, Yoann Dupont, Laurent Romary, Éric Villemonte De La Clergerie, Djamé Seddah, and Benoît Sagot. 2020. Camembert: a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219.

Hans Moen, Laura-Maria Peltonen, Mikko Koivumäki, Henry Suhonen, Tapio Salakoski, Filip Ginter, and Sanna Salanterä. 2018. Improving layman readability of clinical narratives with unsupervised synonym replacement. In *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth - Proceedings of MIE 2018, Medical Informatics Europe, Gothenburg, Sweden, April 24-26, 2018*, volume 247 of *Studies in Health Technology and Informatics*, pages 725–729. IOS Press.

Misako Nomura, Gyda Skat Nielsen, and Bror Tronbacke. 2010. *Guidelines for easy-to-read materials*. International Federation of Library Associations and Institutions (IFLA).

Brian Ondov, Kush Attal, and Dina Demner-Fushman. 2022. A survey of automated methods for biomedical text simplification. *Journal of the American Medical Informatics Association*, 29(11):1976–1988.

Gustavo Paetzold and Lucia Specia. 2016. Anita: An intelligent text adaptation tool. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 79–83.

Jipeng Qiang, Yun Li, Yi Zhu, Yunhao Yuan, Yang Shi, and Xindong Wu. 2021. Lsbert: Lexical simplification based on bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3064–3076.

Luz Rello, Ricardo Baeza-Yates, and Horacio Saggion. 2013. Dyswebxia: Textos más accesibles para personas con dislexia * dyswebxia: Making texts more accessible for people with dyslexia. *Revista nº*, 51:205–208.

Luz Rello, Roberto Carlini, Ricardo Baeza-Yates, and Jeffrey P Bigham. 2015. A plug-in to aid online reading in Spanish. In *Proceedings of the 12th International Web for All Conference*, pages 1–4.

Luz Rello, Martin Pielot, and Mari-Carmen Marcos. 2016. Make it big! the effect of font size and line spacing on online readability. In *Proceedings of the 2016 CHI conference on Human Factors in Computing Systems*, pages 3637–3648.

Evelina Rennes, Marina Santini, and Arne Jönsson. 2022. The Swedish simplification toolkit:–designed with target audiences in mind. In *Proceedings of the*

2nd Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI) within the 13th Language Resources and Evaluation Conference, pages 31–38.

Eva Rolin, Quentin Langlois, Patrick Watrin, and Thomas François. 2021. Frenlys: A tool for the automatic simplification of French general language texts. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1196–1205.

Horacio Saggion, Daniel Ferrés, Leen Sevens, Ineke Schuurman, Marta Ripollés, and Olga Rodríguez. 2017. Able to read my mail: An accessible e-mail client with assistive technology. In *Proceedings of the 14th International Web for All Conference*, pages 1–4.

Horacio Saggion, Montserrat Marimon, and Daniel Ferrés. 2015a. Simplificación automática de textos para la accesibilidad de colectivos con discapacidad: experiencias para el español y el inglés. *IX Jornadas Científicas Internacionales de Inverstigación sobre Personas con Discapacidad*.

Horacio Saggion, Sanja Štajner, Stefan Bott, Simon Mille, Luz Rello, and Biljana Drndarevic. 2015b. Making it simplext: Implementation and evaluation of a text simplification system for Spanish. *ACM Transactions on Accessible Computing (TACCESS)*, 6(4):1–36.

Carolina Scarton, Alessio Palmero Aprosio, Sara Tonelli, Tamara Martín-Wanton, and Lucia Specia. 2017. MUSST: A multilingual syntactic simplification tool. In *Proceedings of the IJCNLP 2017, Tapei, Taiwan, November 27 - December 1, 2017, System Demonstrations*, pages 25–28. Association for Computational Linguistics.

Kim Cheng Sheang and Horacio Saggion. 2023. Multilingual controllable transformer-based lexical simplification. *arXiv preprint arXiv:2307.02120*.

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

Advaith Siddharthan and Angrosh Annayappan Mandya. 2014. Hybrid text simplification using synchronous dependency grammars with hand-written and automatically harvested rules. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*. Association for Computational Linguistics.

Sanja Štajner, Sergiu Nisioi, and Ioana Hulpuș. 2020. Coco: A tool for automatically assessing conceptual complexity of texts. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 7179–7186.

Amalia Todirascu, Rodrigo Wilkens, Eva Rolin, Thomas François, Delphine Bernhard, and Nuria Gala. 2022. Hector: A hybrid text simplification tool for raw texts in French. In *12th International Conference on Language Resources and Evaluation (LREC)*.

María-Teresa Martín Valdivia, Eugenio Martínez Cámara, Eduard Barbu, L Alfonso Ureña López, Paloma Moreda, and Elena Lloret. 2014. Proyecto first (flexible interactive reading support tool): Desarrollo de una herramienta para ayudar a personas con autismo mediante la simplificación de textos. *Procesamiento del Lenguaje Natural*, 53:143–146.

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 international conference on Design of communication*, pages 29–36.

Qiongkai Xu, Chenchen Xu, and Lizhen Qu. 2019. ALTER: auxiliary text rewriting tool for natural language generation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019 - System Demonstrations*, pages 13–18. Association for Computational Linguistics.

Chen-Hsiang Yu and Robert C Miller. 2010. Enhancing web page readability for non-native readers. In *Proceedings of the sIGCHI conference on human factors in computing systems*, pages 2523–2532.

Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. A monolingual tree-based translation model for sentence simplification. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1353–1361.