LanguageTool as a CAT tool for Easy-to-Read in Spanish

Margot Madina[1], Itziar Gonzalez-Dios[2], Melanie Siegel[1]
[1]Darmstadt University of Applied Sciences (Hochschule Darmstadt)
Hochschule Darmstadt, Max-Planck Str.2, 64807 Dieburg
[2]HiTZ Center - Ixa, University of the Basque Country UPV/EHU
Manuel Lardizabal, 1, 20018 Donostia-San Sebastian, Gipuzkoa
margot.madina-gonzalez@h-da.de, itziar.gonzalezd@ehu.eus, melanie.siegel@h-da.de

Abstract

Easy-to-Read (E2R) is an approach to content creation that emphasizes simplicity and clarity in language to make texts more accessible to readers with cognitive challenges or learning disabilities. The Spanish version of E2R is called *Lectura Fácil* (LF). E2R and its variants, such as LF, focus on straightforward language and structure to enhance readability. The manual production of such texts is both time and resource expensive. In this work, we have developed *LFWriteAssist*, an authoring support tool that aligns with the guidelines of LF. It is underpinned by the functionalities of *LanguageTool*, a free and open source grammar, style and spelling checker. Our tool assists in ensuring compliance with LF standard, provides definitions for complex, polysemic, or infrequently used terms, and acronym extensions. The tool is primarily targeted at LF creators, as it serves as an authoring aid, identifying any rule infringements and assisting with language simplifications. However, it can be used by anyone who seek to enhance text readability and inclusivity. The tool's code is made available as open source, thereby contributing to the wider effort of creating inclusive and comprehensible content.

Keywords: Lectura Fácil, readability, cognitive accesibility

1. Introduction

As our world becomes increasingly interconnected, the need for accessible and inclusive communication is more important than ever. This has led to the development of specialized linguistic strategies aimed at bridging communication gaps among diverse audiences. One such strategy is the use of simplified language variants, which enhance understanding and engagement for all. Among these innovations, the Easy-to-Read (E2R) initiative stands out as a key transformation in the way information is conveyed and comprehended in society. E2R refers to a simplified version of a standard language, designed to be less complex and thereby enhance the clarity and understanding of texts. Its purpose is to enhance the clarity and understanding of texts, particularly for individuals with communication challenges. It achieves this by using only the fundamental vocabulary and grammatical structures of the respective natural language. It should not be confused with *Plain Language*, as the Plain Language movement has the entire society as target audience, while E2R focuses on people with communication challenges. Each E2R variant receives a name depending on the standard language it is based on; in the case of Spanish, it is *Lectura Fácil* (LF).

Currently, the availability of E2R or LF texts is limited, as they are created from scratch or adapted from standard texts. The adaptation process is expensive in both time and financial resources, as it involves several steps, including incorporating auditory and/or visual aids, providing additional explanations, simplifying syntax and vocabulary, and summarizing content. It can be especially difficult to keep up with new content in time-sensitive materials such as news articles. On the other hand, translations are often carried out by individuals without formal translation training, resulting in inconsistent and poor quality texts (Hansen-Schirra et al., 2020).

One solution to help overcome these challenges is the use of Computer-Assisted Translation (CAT) tools to convert standard texts to E2R. Such tools would not only accelerate the translation process for professional translators but also provide a means for non-professional translators to verify the accuracy of their work. *LanguageTool* (Naber et al., 2003) is a writing assistant that checks grammar and spelling mistakes, and offers a nuanced analysis of texts, focusing on style, tone, and typography. This approach allows it to provide context-sensitive suggestions, helping users refine their writing[1]. *LanguageTool* finds errors based on rules. Their core technology is available as open-source software, and therefore, users can create their own custom rules and include them in their grammar. This paper explores the extent to which *LanguageTool* can be used as a CAT tool for E2R in Spanish and presents *LFWriteAssist*, an authoring support tool for LF based on *LanguageTool*.[2]

The rest of the paper is organised as follows: in

[1]*LanguageTool*. About us. https://languagetool.org/about?force_language=1

[2]The code and rules are available at https://github.com/margotmg/LFWriteAssist.git

3rd Workshop on Tools and Resources for People with REAding DIfficulties (READI), pages 93–101
20-25 May, 2024May 20th, 2024.

section 2 we present the related work, section 3 introduces *LanguageTool* and its main characteristics, in section 4 we overview the guidelines for LF, in section 5 we present our tool, in section 6 we discuss the main limitations, and in section 7 the conclusions.

2. Related Work

Automatic Text Simplification (ATS) consists of lexical, syntactic, or discourse simplification levels (Chen et al., 2017). Lexical simplification involves Complex Word Identification (CWI) (rare, technical, or abstract words) and substituting them with simpler, more commonly used synonyms or providing their definitions, images, videos, or similar enhancements. Syntactic simplification focuses on simplifying complex sentence structures, such as passive constructions or lengthy sentences; this involves reorganizing, splitting, and adjusting sentence structures, reducing grammatical complexity, and omitting unnecessary information. Discourse simplification addresses coherence and coreference issues to ensure that no important information has been lost during lexical and syntactic simplifications. In E2R, visual or graphic adaptation is also taken into account, that is, the visual design and layout of the text. ATS usually follows three main approaches: rule-based approach, data-driven approach, and hybrid approach. However, there is a lack of tools based on neural approaches (Espinosa-Zaragoza et al., 2023).

Different tools and approaches have been proposed for Spanish text simplification and adaptation [3]:

- *LexSiS* (Bott et al., 2012a) is the first approach to lexical simplification in Spanish. It was created based on the empirical analysis of a sample of data from the *Simplext* corpus, and it relies on freely available resources, such as dictionaries and the web.
- *DysWebxia* (Rello et al., 2013) is the first model for people with dyslexia that presents synonyms for complex words in the text and includes changes in the design of text presentation based on quantitative studies with people with dyslexia.
- *OpenBook* (Barbu et al., 2015) is a rule-based tool that helps the Autistic Spectrum Disorder (ASD) carers and people with ASD to simplify the written documents on a discourse, syntactical and lexical level. It is multilingual, available for Spanish, English and Bulgarian, and it was developed in the scope of the *FIRST* project (Valdivia et al., 2014).
- *Simplext* (Bott et al., 2012b; Saggion et al., 2015a) is a rule-based prototype for syntactic simplification in Spanish, tackling sentence splitting, lexical substitution, and syntactic reordering. This was part of the *Simplext* project (Saggion et al., 2015b).
- *CASSA plug in* (Rello et al., 2015) was created based on the *CASSA* algorithm (Context-Aware Synonym Simplification Algorithm) (Baeza-Yates et al., 2015), a context-aware algorithm for generating simpler synonyms, using resources like Google Books Ngram Corpus and Spanish OpenThesaurus and real web frequencies of the complex word for disambiguation.
- The *Able2Include* project[4] (Saggion et al., 2017) aims at improving the living conditions of people with Intellectual or Developmental Disabilities (IDD) in key areas of society by introducing accessible web-based tools. Some of their tools are also available in languages other than Spanish.
- *MUSST* (Scarton et al., 2017) is a rule-based multilingual syntactic simplification tool, supporting sentence simplifications for Spanish, English, and Italian. It was implemented in the context of the European project SIMPATICO on text simplification for public administration texts.
- *NavegaFácil* (Bautista et al., 2018) is a web application aimed at facilitating the comprehension of text. It allows users to visualize and navigate through the original content of any web page, and provides definitions, synonyms and antonyms, lemmatisations, images, Google search, Wikipedia, translation and text to voice.
- *EASIER* (Alarcon et al., 2021) performs CWI following machine learning techniques and contextual embeddings using Easy Reading and Plain Language resources, and also provides definitions.
- The *ClearText* project[5] (Moreda et al., 2023)

[3] There is also *arText-claro* (http://sistema-artext.com/lenguaje-claro) (da Cunha, 2022), the first Spanish-assisted copywriter that helps to write texts in specialised fields and texts in *Lenguaje Claro* (the Plain Language equivalent of Spanish). It has not been included in this list because it focuses on *Lenguaje Claro* and not LF, and it is therefore beyond the scope of this paper.

[4] *Able2Include* project https://able-to-include.ccl.kuleuven.be/index.html

[5] *ClearText* project https://cleartext.gplsi.es

aims to create a tool that simplifies Spanish texts from the public administration, making them more accessible to people with mild to moderate cognitive impairment.

- *FACILE* (Suárez-Figueroa et al., 2024) is an AI-driven tool to aid, in a semi-automated way, in the E2R adaptation process of documents. It is still under development, but its primary objective is to assist E2R specialists in their routine activities, which include evaluating documents against E2R standards and modifying them in line with E2R principles.

In spite of the existence of these tools and resources, it is worth highlighting that only *arText*, *Simplext* and *EASIER* are operative (the latest being also open source). *MUSST* and *NavegaFácil* offer open source code, but are not operative, to our knowledge, at the writing of this paper. We did not find information on the operativeness of *DysWebxia*, *CASSA plug-in* and *OpenBook* are inoperative, and there is no information on *LexSiS*. In fact, the availability and accessibility of ATS tools is a recurrent problem in various languages (Espinosa-Zaragoza et al., 2023).

To the best of our knowledge, there is no previous study or tool that employs *LanguageTool* to aid in LF text adaptation. However, *LanguageTool* has been used as an authoring support tool for *Leichte Sprache* (the E2R equivalent for German) (Siegel and Lieske, 2015). In this study, they created *Leichte Sprache* rules and implemented them in *LanguageTool*. Our paper follows these steps, but also introduces some important changes (see section 5).

3. LanguageTool as a CAT tool

LanguageTool is an open-sourced, multilingual proofreading tool. As of February 2024, it supports 30 languages and 20 language variants[6] (version 6.3, released October 4th, 2023). It detects spelling, grammatical, and stylistic errors, as well as ambiguities and opportunities for improvement in wording. It can also paraphrase text to improve clarity and fluency. *LanguageTool* integrates seamlessly with a variety of platforms and applications, including web browsers and Office programs, such as Google Docs and OpenOffice. The premium version of LanguageTool offers enhanced capabilities for more thorough and detailed proofreading of texts. *LanguageTool* is known for its focus on open source development, which allows anyone to access and contribute to its code[7], and anyone can set up their own LanguageTool server locally or in the cloud. This approach encourages continuous improvement and adaptability of the software to different linguistic needs and contexts. Additionally, users can create custom rules to adjust the tool to their own writing styles or specific needs, which makes it a versatile and flexible tool suitable for a wide range of users and applications[8].

The integration of *LanguageTool* into applications offers several advantages. Customizability is a standout feature, allowing developers to tailor *LanguageTool*'s rules to specific needs or guidelines, such as adapting it for E2R content. As an open-source tool, *LanguageTool* invites a collaborative approach to development and improvement, offering transparency in its functionality and the flexibility to modify its code to fit different requirements. This open-source aspect ensures that it can evolve continually with contributions from a global community. The multilingual support of *LanguageTool* allows for integration in applications that cater to diverse user groups, ensuring accurate grammar and style checks across many languages. Finally, the consistency that LanguageTool brings to text is crucial for maintaining a coherent narrative in written text. This consistency is particularly important in E2R, as clear and uniform communication is paramount.

3.1. Custom Rules

The rules in *LanguageTool* follow a specific pattern, which has also been followed in *LFWriteAssist* (refer Table 1 for examples). These are the main elements of the rules and their attributes:

- `id`: an internal, unique identifier of the rule.
- `name`: short text displayed in the configuration, describing the rule.
- `antipattern`: complex exception to a rule (optional).
- `pattern`: part of the original text that should be marked as an error.
- `message`: text displayed to the user if the rule matches. Here, we include the sub-element `suggestion` to suggest a replacement to correct the error. If the `suggestion` sub-element is not included in the rule, the text will not be corrected.
- `url`: url to a page explaining the rule in more detail (optional).
- `short`: short description of the rule (optional)

[6] Languages and rules in LanguageTool 6.3 https://dev.languagetool.org/languages

[7] *LangaugeTool* source code on Github https://github.com/languagetool-org/languagetool?tab=readme-ov-file

[8] *LanguageTool* complete development documentation https://dev.languagetool.org

- `example`: example with an incorrect sentence. The position of the error must be marked up with the sub-element `marker`.

At times, it may be necessary to employ multiple rules to identify all instances of an error. All these rules can be combined into a single `rulegroup` element. The `rulegroup id` and `name` attribute are used for all the rules belonging to that group. The rules can also be grouped into categories, depending on their purpose; this allows enabling and disabling those rules at the same time[9].

Custom rules can first be tested in the *LanguageTool* online rule editor[10]; this way, users can check if the rule has any errors and whether it covers all the desired linguistic features. After this, the rules must be included in the `grammar` XML file of the preferred language. The rules are different depending on the language; that means that even if the same language phenomena happen in two different languages (e.g. passive voice), the rules will be different. Therefore, each language has its own `grammar` XML file, and custom rules will only work for the language they were created for.

4. *Lectura Fácil Guidelines*

The Spanish language counts with the standard *Norma UNE 153101:2018 EX de Lectura Fácil. Pautas y recomendaciones para la elaboración de documentos* (UNE 153101 EX of Lectura Fácil. Guidelines and recommendations for the elaboration of documents) (UNE, 2018). This standard aims to guarantee the comprehension of written documents and the entitlement of all individuals to access information. This standard explains the process of adapting texts in LF, as well as the process of adapting standard texts into LF. It also contains guidelines and recommendations for writing text in LF, and guidelines and recommendations for the design of a document in LF. In this paper, we have focused on the former, which contain the following subsets of guidelines and recommendations:

1. Guidelines and recommendations related to orthotypography
2. Guidelines and recommendations related to vocabulary and expressions
3. Guidelines and recommendations related to phrases and sentences
4. Guidelines and recommendations related to text organisation and style

Within these rules, some of them are specific, clearly defined rules that are straightforward to implement. For instance, the rule that claims that "you should not use acronyms". Conversely, we encountered challenges with rules that are inherently vague or overly generic. An example of such a rule is "avoid the use of words that do not add information to the text and make it longer to read". The subjective nature and the broad scope of this rule make its implementation problematic. Consequently, these types of rules have been set aside in our current framework due to the difficulty in quantifying and codifying them into a set of parameters. Furthermore, our system does not incorporate rules that require a deeper understanding of the context and meaning such as "do not use nominal phrases and avoid nominal use of adjectives". These are beyond the scope of our current rule-based approach. The complexity of semantic interpretation presents significant challenges for rule-based systems. In an attempt to get the most out of our tool, we have integrated some other resources that align with our focus on lexicon and syntactic rules. We have utilized *Diccionario Fácil*[11], a dictionary that offers simplified definitions of complex, polysemic, or infrequently used terms. It is designed for individuals with reading comprehension difficulties and is an initiative by *Plena Inclusión Madrid*[12], the Madrid Community Federation supporting people with intellectual or developmental disabilities. We have extracted all dictionary entries and definitions, so that they are provided to our tool users. We have also employed the *EASIER* corpus (Alarcon et al., 2023)[13] to provide easier synonyms for the complex words encountered in the text. Additionally, we also created a list of acronyms and a list of abbreviations, which were extracted from *Wikilengua*[14], an open and participatory site for sharing practical information about the norm, usage and style of Spanish. On the other hand, we have also integrated rules that were created using Python in addition to the standard *LanguageTool* rules. These include the detection of long phrases and long words.

[9] For further details, please refer to *LanguageTool* development overview on custom rules https://dev.languagetool.org/development-overview

[10] *LanguageTool* online rule editor https://community.languagetool.org/ruleEditor2/

[11] *Diccionario Fácil* https://www.diccionariofacil.org/diccionario

[12] *Plena Inclusión Madrid* https://plenainclusionmadrid.org

[13] 260 documents were annotated, from which they gathered 8,100 complex words. A total of 7,892 synonyms were proposed.

[14] *Wikilengua* acronyms https://www.wikilengua.org/index.php/Lista_de_siglas_A and *Wikilengua* abbreviations https://www.wikilengua.org/index.php/Lista_de_abreviaturas_A

Rule with suggestion (figurative meaning)

```
<rule>
  <pattern>
    <token min="0">hasta</token>
    <token>por</token>
    <token>los</token>
    <token>codos</token>
  </pattern>
  <message>Se debe evitar el uso de enunciados con sentido figurado.</message>
  <suggestion>mucho</suggestion>
  <example correction='mucho'>Ella habla <marker>hasta por los codos</marker>.</example>
  <example>Ella habla mucho.</example>
</rule>
```

Rule without suggestion (passive voice)

```
<rule>
  <pattern>
    <token regexp='yes'>asunto|cosa|algo</token>
  </pattern>
  <message>Se debería evitar el uso de palabras de contenido indeterminado como "cosa", "algo" o "asunto".</message>
  <example correction='problema'>Era un <marker>asunto</marker> complicado</example>
  <example>Era un problema complicado.</example>
</rule>
```

Table 1: Examples of a rule with `suggestion`, and a rule without `suggestion`.

5. *LanguageTool* for *Lectura Fácil*

As mentioned in Section 1, to the best of our knowledge, *LanguageTool* has not been previously employed to aid in the LF text adaptation process. There is one study that implements *Leichte Sprache* rules on *LanguageTool*. We follow this work by Siegel and Lieske (2015) and create LF rules to be used with *LanguageTool*.

As mentioned in subsection 3.1, the inclusion or exclusion of a `suggestion` determines whether the text will be automatically changed or not. Some of the UNE rules we have adapted include a `suggestion`, while others do not. Examples of a rule with `suggestion`, and a rule without `suggestion` are provided in Table 1. The rule with a `suggestion` applies for a phrase with figurative meaning. In this case, the rule matches the figurative phrases *hasta por los codos* (even through the elbows) and *por los codos* (through the elbows), which refer to a person that talks a lot. This part of the text is then changed to *mucho* (a lot). Therefore, when having the sentence *él habla hasta por los codos* (he talks even through the elbows), the text will be automatically changed to *él habla mucho* (he talks a lot). The rule without a `suggestion` applies for the rule that claims that "the use of words with indeterminate content such as *thing*, *something* or *issue* should be avoided". The rule matches any of these words, but does not make any changes nor offer any alternative terms, as more information on the context is necessary, and the phrase might need to be rephrased.

The decision to include a `suggestion` that would automatically correct the text depended on various factors:

- **Need for context:** some text may require a deeper understanding of context to make an appropriate correction. The tool flags these areas to prompt the user to review them, as automatic correction could alter the intended meaning or not be feasible without additional context.
- **Limitations of *LanguageTool* rules:** the correction may not fit within the intrinsic nature and structure of *LanguageTool* rules. Sometimes it might not be possible to structure and capture all linguistic nuances or complex structures.
- **Extensive rewriting required:** in cases where a text segment requires significant rephrasing, it may be more efficient to rewrite the entire section rather than attempt to correct it piece by piece. This approach can help maintain the coherence and flow of the text.

In these situations, the tool provides guidance by highlighting the areas of concern in orange, but leaves the decision and manner of revision to the human user, who can take into account the nuance, context, and complexity of the content. Whenever automatic changes have been applied, the areas are highlighted in green. We have named our tool *LFWriteAssist*, and will refer to it as such for now on.

5.1. *LFWriteAssist* Structure

Currently, *LFWriteAssist* functions primarily as an interface, serving as a user-friendly front-end for the more technical aspects of the *LanguageTool* framework and allowing users to interact with the tool's functionalities. The structure of the interface consists of the following parts (see Figure 1):

- Input panel, named *Campo de entrada*, which is a text entry field where users write their source text.
- The second panel, named *Resumido y revisado*, shows the summarized text, the tool's

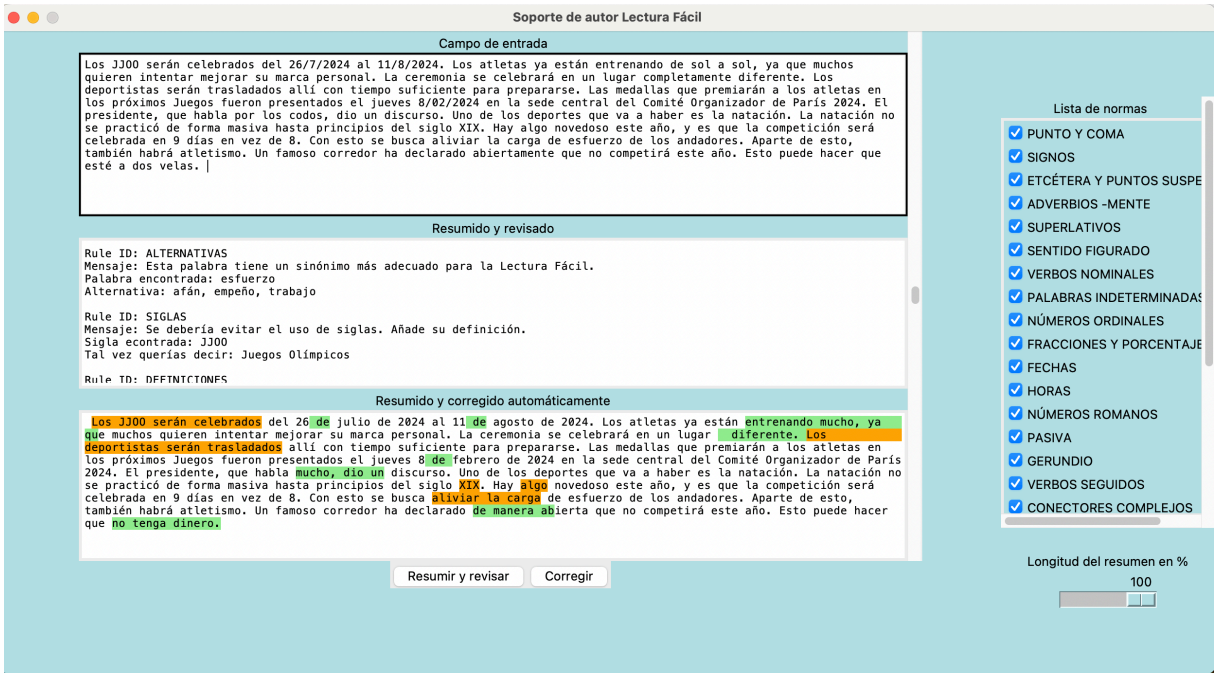

Figure 1: Interface of the *LFWriteAssist* tool displaying three text panels: original text input, automated suggestions for acronyms and complex word definitions, as well as explanations of the broken rules encountered, and the final output. Underlined in green, the changes that have been performed automatically, and underlined in orange, something that should be changed or revised. On the right, a side panel listing various language rules for text simplification. Below the side panel, a slider control to adjust the summary length of the text. This particular example shows a text which has not been summarized, in an aim to show as many error corrections as possible in the output.

suggestions on the original text (both for automatically corrected parts, and areas that need revision), as well as definitions of complex, polysemic, or infrequently used terms and acronym expansions.

- The third panel, named *Resumido y corregido automáticamente*, shows the summarized and automatically corrected version of the source text. Those parts of the text that have been automatically corrected are underlined in green. Those parts underlined in orange are parts of the text that violate some LF rule, but have not been automatically corrected.

- The side panel, named *Lista de normas*, lists the rules for LF that we have created.

- The slider control, named *Longitud del resumen en %*, allows the user to adjust the length of the summary. When choosing 100%, the output text will keep all the information in the source text.

The interface is in Spanish, but if needed, it can be localised to other languages.

6. Limitations

In spite of the strengths our tool offers in the realm of ATS, particularly LF texts, it is important to acknowledge certain limitations and areas for future development. Primarily, our current focus is on the Spanish language, with future research planned for other languages. Notably, a German version exists, but lacks the suggestion feature; therefore, automatic changes are not applied in the final output. Our work is grounded in the guidelines for writing text in LF, but it is important to recognize that LF and E2R texts encompass more than just language simplification. Factors like layout are also vital, which our current tool does not address. All rules have been created manually, which may result in inadvertently missing certain linguistic elements such as figurative phrases, abbreviations, and other language nuances. However, the collaborative nature of *LanguageTool*, on which *LFWriteAssist* is based, allows for the potential addition of more rules by the community, progressively enriching its capabilities. This aspect underscores the tool's evolving nature and the scope for continuous improvement through community involvement. As this is the initial prototype of *LFWriteAssist*, it is important to acknowledge that it may exhibit some errors or limitations. However, it's crucial to note that our pri-

mary audience is LF developers, not the end-users themselves. This distinction is significant because any inaccuracies or shortcomings in the *LFWriteAssist*'s current iteration are less likely to directly impact the target audience. The developers, being more familiar with LF principles and guidelines, can identify and mitigate these issues during the content creation process. Therefore, while the tool aims to aid in producing more accessible texts, its current prototype status implies a phase of testing and refinement primarily within a professional context.

7. Conclusions and Future Work

The traditional process of producing E2R and LF texts is notably resource-intensive, both in terms of time and financial investment. Despite the existence of some ATS tools, including some targeting LF texts, many lack full operational capability. We have proposed *LFWriteAssist*, an authoring support tool based on *LanguageTool*. We perform extractive summarization, cover different language phenomena and provide definitions when needed based on already existing LF resources, such as dictionaries and guidelines. A distinctive feature of *LFWriteAssist* is its ability to perform automatic alterations in the text, which are highlighted in green for ease of recognition. This visual cue assists users in quickly identifying the modifications made for simplicity and clarity. Moreover, the tool also highlights sections that require manual review. The combination of these features makes our tool a comprehensive assistant in the creation of E2R and LF texts. We advocate for the involvement of target users in the creation and evaluation of ATS tools, therefore, future developments include conducting surveys with LF translators to refine the tool according to their needs. Additionally, we aim to enhance accessibility for LF professionals by implementing this tool on a web page, eliminating the current installation requirements. The open-source nature of this tool invites collaboration and continuous improvement, potentially leading to further advancements in this field. It opens up opportunities for other developers and users to contribute to its development, ensuring that the tool remains adaptable and up-to-date with the evolving needs of its user base.

Our tool aims to enhance the overall simplicity of documents, reduce human effort, and ensure adherence to E2R guidelines. Although a specific evaluation method for *LFWriteAssist* has not yet been finalised, a strategic approach is in place. The plan is to involve professional E2R translators in a comprehensive review process. This approach will involve selecting a diverse group of translators, providing them with various texts, and asking them to use the tool in their translation and proofreading tasks. After using the tool, translators will be asked to provide feedback through surveys and interviews. The feedback will focus on the tool's usability, effectiveness in simplifying texts, and integration into their workflow. The feedback will be critically analysed to assess the tool's performance in terms of accuracy, time efficiency, and overall user satisfaction. The evaluation insights will refine the tool, meeting practical needs of professional translators and aiding in creating high-quality E2R content. The expert-driven process enhances functionality and provides valuable research data, demonstrating real-world applicability and impact.

In addition to considering professional feedback, we are exploring the possibility of conducting a false positive/false negative analysis as part of the evaluation for *LFWriteAssist*. This method involves assessing how accurately the tool identifies E2R issues. A false positive occurs when the tool incorrectly flags a piece of text as non-compliant with E2R guidelines when it is compliant, while a false negative is when the tool fails to identify an E2R issue in the text. By analysing these occurrences, we can measure the precision and accuracy of our tool, providing critical insights into its effectiveness and reliability.

8. Acknowledgements

We would like to thank Tatiana González-Ferrero, from *LanguageTool*, for her assistance in the development of some of the rules. This work has been partially supported by the following projects i) Ixa group A type research group (IT-1805-22) funded by the Basque Government ii) DeepKnowledge (PID2021-127777OB-C21) project funded by MCIN/AEI/10.13039/501100011033 and by FEDERe and iii) AWARE Commonsense for a new generation of natural language understanding applications (TED2021-131617B-I00) funded by MCIN/AEI /10.13039/501100011033 by the European Union NextGenerationEU/ PRTR.

9. Ethical Considerations

The primary objective of this study is to foster understanding and inclusivity through our focus on E2R and LF. Our use of these terms is strictly for descriptive purposes. The used terminology carries no judgement on the value of languages, dialects, or linguistic styles. We hold all forms of linguistic expression in high regard and are mindful of the sensitivities surrounding discussions about language. Should any part of our discourse or the terminology we have employed unintentionally imply otherwise, we offer our sincere apologies.

10. Bibliographical References

Rodrigo Alarcon, Lourdes Moreno, and Paloma Martínez. 2021. Lexical Simplification System to Improve Web Accessibility. *IEEE Access*, 9:58755–58767.

Rodrigo Alarcon, Lourdes Moreno, and Paloma Martínez. 2023. EASIER corpus: A Lexical Simplification Resource for People with Cognitive Impairments. *Plos one*, 18(4):e0283622.

Ricardo Baeza-Yates, Luz Rello, and Julia Dembowski. 2015. CASSA: A Context-Aware Synonym Simplification Algorithm. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1380–1385.

Eduard Barbu, M Teresa Martín-Valdivia, Eugenio Martinez-Camara, and L Alfonso Urena-López. 2015. Language Technologies Applied to Document Simplification for Helping Autistic People. *Expert Systems with Applications*, 42(12):5076–5086.

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.

Stefan Bott, Luz Rello, Biljana Drndarevic, and Horacio Saggion. 2012a. Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish. In *COLING*.

Stefan Bott, Horacio Saggion, and Simon Mille. 2012b. Text Simplification Tools for Spanish. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1665–1671.

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.

Iria da Cunha. 2022. Un redactor Asistido para Adaptar Textos Administrativos a Lenguaje Claro. *Procesamiento del Lenguaje Natural*, 69:39–49.

Isabel Espinosa-Zaragoza, José Abreu-Salas, Elena Lloret, Paloma Moreda Pozo, and Manuel Palomar. 2023. A Review of Research-Based Automatic Text Simplification Tools. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 321–330.

Silvia Hansen-Schirra, Jean Nitzke, Silke Gutermuth, Christiane Maaß, and Isabel Rink. 2020. Technologies for Translation of Specialised Texts into Easy Language. *Easy language research: Text and user perspectives*, 2:99.

Marcin Miłkowski. 2012. Translation Quality Checking in LanguageTool. *Corpus Data across Languages and Disciplines. Peter Lang, Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien*, pages 213–223.

Paloma Moreda, Beatriz Botella, Isabel Espinosa-Zaragoza, Elena Lloret, Tania Josephine Martin, Patricio Martínez-Barco, Armando Suárez Cueto, Manuel Palomar, et al. 2023. CLEAR. TEXT Enhancing the Modernization Public Sector Organizations by Deploying Natural Language Processing to Make Their Digital Content CLEARER to Those with Cognitive Disabilities.

Daniel Naber et al. 2003. A Rule-Based Style and Grammar Checker.

Luz Rello, Ricardo Baeza-Yates, and Horacio Saggion. 2013. DysWexia: Textos Más Accesibles Para Personas con Dislexia. *Procesamiento del lenguaje natural*, 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.

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, System Demonstrations*, pages 25–28.

Melanie Siegel, Dorothee Beermann, and Lars Hellan. 2019. Aspects of Linguistic Complexity: A German–Norwegian Approach to the Creation of Resources for Easy-to-Understand Language. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE.

Melanie Siegel and Christian Lieske. 2015. Beitrag der Sprachtechnologie zur Barrierefreiheit: Unterstützung für Leichte Sprache. *Zeitschrift für Translationswissenschaft und Fachkommunikation*, 8(1):40–78.

Mari Carmen Suárez-Figueroa, Isam Diab, Edna Ruckhaus, and Isabel Cano. 2024. First steps in the development of a support application for easy-to-read adaptation. *Universal Access in the Information Society*, 23(1):365–377.

UNE. 2018. UNE 153101:2018 EX Lectura fácil. Pautas y Recomendaciones para la Elaboración de Documentos. Madrid: Asociación Española de Normalización.

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.