Automatic Text Simplification for People with Cognitive Disabilities: Resource Creation within the ClearText Project

Isabel Espinosa-Zaragoza[1], **José Abreu-Salas**[2], **Paloma Moreda**[3] **and Manuel Palomar**[3]
[1] Centre 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
{paloma,mpalomar}@dlsi.ua.es

Abstract

This paper presents the ongoing work conducted within the ClearText project, specifically focusing on the resource creation for the simplification of Spanish for people with cognitive disabilities. These resources include the CLEARSIM corpus and the Simple.Text tool. On the one hand, a description of the corpus compilation process with the help of APSA is detailed along with information regarding whether these texts are bronze, silver or gold standard simplification versions from the original text. The goal to reach is 18,000 texts in total by the end of the project. On the other hand, we aim to explore Large Language Models (LLMs) in a sequence-to-sequence setup for text simplification at the document level. Therefore, the tool's objectives, technical aspects, and the preliminary results derived from early experimentation are also presented. The initial results are subject to improvement, given that experimentation is in a very preliminary stage. Despite showcasing flaws inherent to generative models (e.g. hallucinations, repetitive text), we examine the resolutions (or lack thereof) of complex linguistic phenomena that can be learned from the corpus. These issues will be addressed throughout the remainder of this project. The expected positive results from this project that will impact society are threefold in nature: scientific-technical, social, and economic.

1 Introduction

People with cognitive disabilities have significant limitations in their intellectual functioning and/or may also lack the ability to adapt to everyday situations. In fact, they have spoken and written word comprehension deficits that may include misinterpretation of literal meanings and difficulty understanding complex instructions, to name a few. Among the different language phenomena they struggle with, there are idioms, figures of speech, abstractions, uncommon words, lack of precision, and complex syntax, among other aspects.

Currently, Natural Language Processing (NLP) technologies are developed and mature enough to provide a sound basis for (1) developing components to automatically detect and remove obstacles to reading comprehension and (2) generate additional content to facilitate reading comprehension. Thus, we begin this project with the main hypothesis that the research, development, and deployment of NLP technology can support the authoring of accessible content in Spanish for people with cognitive disabilities with the aim of increasing both their inclusion and empowerment in Europe.

With this hypothesis in mind, the ClearText project[1], funded by the Spanish Government and the European Union (grant reference TED2021-130707B-I00) and developed by the GPLSI research group[2] of the University of Alicante, focuses on researching, implementing, deploying, evaluating, and ultimately providing robust technologies for NLP to support the authoring of accessible Spanish content for public sector organisations —at local, regional, and national levels— that is intelligible to people with cognitive disabilities, thereby widening their inclusion and empowerment in Europe. This, in turn, will improve the ability to access written information for everyone, thereby reducing the risk of exclusion for those with cognitive disabilities. The project is expected to positively impact the quality of life of people with cognitive disabilities, by facilitating their access to educational, vocational, cultural, and social opportunities in public sector organisations.

This paper is structured as follows: Section 2 includes a literature review covering automatic text simplification and focusing on related work

[1] https://cleartext.gplsi.es/
[2] https://gplsi.dlsi.ua.es/

Proceedings of the Second Workshop on Text Simplification, Accessibility and Readability associated with RANLP-2023, pages 68–77,
held in Varna Bulgaria, Sept 7, 2023.
https://doi.org/10.26615/978-954-452-086-1_007

tackling corpora and tools; Section 3 presents the scientific and technological objectives of the project; Section 4 delves into the composition of the project's members; Section 5 describes the different resources created in this project, namely the CLEARSIM corpus and the Simple.Text tool; Section 6 details the expected scientific-technical and socio-economic impact of this project, while Section 7 concludes with the future work ahead.

2 Automatic Text Simplification: Review of Literature

Automatic Text Simplification (henceforth ATS) can be defined as "the process of reducing the linguistic complexity of a text to improve its understandability and readability, while still maintaining its original information content and meaning" (Al-Thanyyan and Azmi, 2022).

ATS can be achieved by following different approaches, namely rule-based, data-driven, or hybrid approaches, and by simplifying some or all language levels (i.e. lexical, syntactic, semantic, and stylistic). The more language levels and language phenomena tackled in the simplification, the more refined the simplified text will be. Specific domains are also an aspect to take into account when simplifying texts.

The audience for which these simplifications are created is diverse (e.g. children, non-native speakers, poor readers, and cognitively impaired individuals) although it is sometimes left unspecified. Hence, the importance of the current focus on customisation, as a given simplification solution may not work for all audiences (Alva-Manchego et al., 2020; Scarton et al., 2018).

Sections 2.1 and 2.2 cover a brief review of the main corpora and tools for ATS in Spanish in more detail.

2.1 ATS Corpora

According to Martin et al. (2023), there are 10 corpora for simplification in Spanish: FIRST (Štajner and Saggion, 2013), Automatic Noticias Fácil (Štajner et al., 2014), IrekiaLF (Gonzalez-Dios et al., 2022), CLARAMED (Campillos-Llanos et al., 2022) and some others which remain unnamed by Bott and Saggion (2011; 2014), Mitkov and Štajner (2014) and Štajner et al. (2019). Additionally, there are two English-Spanish: Newsela (Xu et al., 2015) and SIMPLETICO (Shardlow and Alva-Manchego, 2022).

The corpora review carried out by Martin et al. (2023) presented the following conclusions: (1) the majority of the corpora produced for ATS is in English, and only 7 out of the 24 official languages of the European Union are present, namely, Danish, English, French, German, Italian, Portuguese, and Spanish; (2) there is a scarcity of resources that address ATS aimed at domains that are important for social inclusion, such as health and public administration; (3) there is a lack of parallel corpora whose target is people with mild-to-moderate cognitive impairment; (4) there is a lack of experiments where the target audience was directly involved in the development of the corpus; and, lastly, (5) more than half of these corpora lack adequate documentation of how the simplification was performed, or at the very least, fail to identify the linguistic phenomenon tackled by the simplification.

Additionally, the domain is usually general information, like Wikipedia or news media, with the exception of CLARA-MED (Campillos-Llanos et al., 2022) and SIMPLETICO (Shardlow and Alva-Manchego, 2022) belonging to the health domain and IrekiaLF (Gonzalez-Dios et al., 2022) for the public administration. For more detailed information regarding aspects like language, domain, audience, alignment, size, and metadata, consult Martin et al.'s (2023) work.

2.2 ATS Tools

Regarding simplification tools, Espinosa-Zaragoza et al. (2023) concludes that (1) many languages are still not represented in ATS tools; (2) all language levels should be borne in mind; (3) multiplicity of options or, in other words, NLP solutions, should be presented to the user, as well as (4) customised simplifications to fulfil the need(s) of the varied targets users and, lastly, (5) the need for these tools to be fully accessible and operational for the public.

According to Espinosa-Zaragoza et al. (2023), there are 7 ATS tools for Spanish: arText (da Cunha Fanego et al., 2017), Simplext (Saggion et al., 2015), DysWebxia (Rello et al., 2013), EASIER (Alarcón et al., 2021), LexSIS (Bott et al., 2012), NavegaFácil (Bautista et al., 2018) and Open Book (Barbu et al., 2015). From those, only three are operational at the moment (i.e. accessible for people to simplify text): arText[3], EASIER[4], and Simplext[5]. The first one helps in the identification

[3] http://sistema-artext.com/
[4] http://163.117.129.208:8080/
[5] http://simplext.taln.upf.edu/

of complex language phenomena in a given text; the second one highlights complex vocabulary and provides a simpler substitute; and the last one allows for the simplification of sentences, as it has a character limitation.

3 Scientific and Technological Objectives

The main objective of the ClearText project can be divided into the following specific objectives:

- O1. To analyse the main comprehension obstacles posed by the language used in the web content of Spanish public sector organisations, such as ministries and other government agencies, for people with cognitive disabilities.
- O2. To analyse the needs of people with cognitive disabilities.
- O3. To research and adapt the COMPENDIUM System developed by Lloret et al. (2013) to the needs of public sector documentation.
- O4. To research, implement, deploy, and ultimately provide robust technologies to support the processing of structural complexity.
- O5. To research, implement, deploy, and ultimately provide robust technologies to support the processing of ambiguity in meaning.
- O6. To research, implement, deploy, and ultimately provide a robust text simplification system oriented toward public administration documentation.
- O7. To evaluate the simplification system.
- O8. To promote and disseminate the research results obtained from the project through different national and international media including well-indexed journals, conferences, seminars, etc., as well as exploit the potential for transferring this technology to society.

4 Human Resources

A multidisciplinary research team consisting of five computer science experts and three linguists, with seven of them holding doctorate degrees and one serving as a technician, is in charge of the project. All members belong to the GPLSI research group. The composition of the team reflects a slight positive gender imbalance with five women and three men. The team has extensive experience in technological research and development in NLP, spanning more than 30 years, and, more specifically, in relation to the requested project in word sense disambiguation, anaphora resolution, coreference, named entity, lexical and syntactic analysis, text summarisation and text simplification.

5 Tool and Corpus: Work in Progress

We are currently contemplating and developing both a traditional or conventional approach and also one with the training of a language model. For both of them, this project's aim includes the creation of two resources: (1) the CLEARSIM corpus of simplified texts in Spanish and (2) the Simple.Text tool.

5.1 CLEARSIM Corpus

The compilation process is determined to take place during the first year of the project, that is, 2023. As previously mentioned, the language used is Spanish and the domain pertains to public administration texts. Our target audience consist of people with cognitive disabilities and our alignment approach is document to document. Regarding the size of the corpus, the estimation of texts compiled by the end of the project is 18,000 texts, including 15,000 silver standard texts and 3,000 golden standard texts. The different compilation stages are described below:

- **Stage 1. Original text compilation:** The texts selected are published articles dealing with sports, leisure activities, and culture. These articles are sourced from the websites of town halls within the Alicante province, more specifically, from the following cities: Elche, Benidorm, Alicante, Alcoy, Elda, Torrevieja, and Orihuela.
- **Stage 2. Automatic text summarisation and simplification with ChatGPT:** Since the automatic summarisation task deals with the reduction of content to maintain the most important ideas and text simplification involves removing unnecessary information, this common ground led us to summarise the original text using ChatGPT, which yielded the RGPT texts (i.e. resumen GPT). Additionally, we also prompted a simplification from ChatGPT to compare the results from the summarisation and the simplification process and iden-

tify which processes this generative AI employs depending on the provided instructions. This ultimately generates the simplified version from ChatGPT (SGPT, i.e. simplificación GPT). Both summarisation and simplification processes were applied to the original text.

- **Stage 3. ChatGPT revised versions:** Subsequently, a human revision is manually carried out to ensure that the simplifications are properly performed. Due to time constraints, a set of easy-to-read guidelines is being considered and applied to the ChatGPT texts generated in the previous stage. Additionally, the summarisation helps the reviser check that no crucial information is deleted (e.g. dates, locations, and other pieces of information) in the simplified version. This stage was crucial for refining the prompt, which iterated and began with a simpler version (e.g. *Can you simplify this text?*) which, however, lacked conciseness. Although we still obtained simplified texts, manual revision was time-consuming due to the subjective nature of simplification. Nevertheless, a more refined prompt that explicitly indicated which language phenomena we consider difficult and required replacements generated better simplified outputs. This, in turn, accelerated the manual revision stage.

- **Stage 4. Easy-to-read and facilitated versions:** This stage is conducted by our collaborators, APSA[6], a Non-Governmental Organisation (NGO) which comprises a group of individuals with cognitive disabilities who possess expertise in the adaptation of texts in adherence to the easy-to-read guidelines. In this stage, our collaborators are provided with the original text and our revised simplified version (i.e. SUA, see Table 1) to create both an easy-to-read version (LF, i.e. lectura fácil) following all the easy-to-read guidelines (AENOR, 2018) and a "facilitated" version (FAC, i.e. facilitada), which yields a simplified version according to the legislation but disregarding outlay aspects (e.g. font type, size, color, and others). We utilise Google Drive for text interchange and provision, and APSA creates 50 texts weekly in both versions.

[6] https://www.asociacionapsa.com/

This compilation of different texts provides a bronze, silver, and gold standard in simplification, respectively (see Table 1). As can be observed, 7 different text types are included: the original text; a summary and a simplification created with ChatGPT; a revision for each of those versions created by ChatGPT made by our institution; and two manually simplified texts by our collaborator, one following all the easy-to-read guidelines and another disregarding some presentation guidelines. To date, we have compiled approximately 2,000 texts classified as silver standard and 400 texts classified as gold standard.

5.2 Simple.Text Tool

This section describes the tool's objectives, some technical aspects, and the preliminary results derived from early experimentation.

As commented before, text simplification implies solving different language phenomena such as co-references, complex words, or sentence structure. An automatic simplification system may address all or only a subset of these problems. Also, it may work at the sentence or document level. The first setup expects a sentence as input and outputs the simplified version. As Cripwell et al. (2023) noted, sentence-level systems may be leveraged for document-level simplification by iteratively processing the sentences. However, this approach may present problems such as failing to preserve the discourse structure.

Pure document-level simplification may pose challenges, such as the scarcity of datasets aligned at document level (Sun et al., 2021). In addition, the approaches to teaching the system to simplify full documents by simultaneously solving the different linguistic phenomena seem to be at an early stage (Sun et al., 2021; Cripwell et al., 2023). These issues are particularly relevant in the context of data-driven neural generative models.

We aim to explore LLMs in sequence-to-sequence setup for text simplification at the document level. In Sections 5.2.1 and 5.2.2 we cover the technical details of the implementation of Simple.Text tool. Section 5.2.3 discusses early findings from ongoing experiments we are carrying out at the moment.

5.2.1 Technical Details

Simple.Text Tool core is a T5 (small) model (Raffel et al., 2020) fine-tuned using the current version of SUA (see Table 1). The data comprises 925

Code	Description	Stage	Standard
TXT	Original texts	1	
RGPT	Summaries created with ChatGPT	2	Bronze
SGPT	Simplifications created with ChatGPT	2	Bronze
RUA	Summarised texts validated by our institution	3	Silver
SUA	Simplified texts validated by our institution	3	Silver
LF	Easy-to-read documents created by APSA	4	Gold
FAC	Facilitated texts created by APSA	4	Gold

Table 1: Summary of the Texts Created for the CLEARSIM Corpus

instances, which were split into 749 for training, 83 for validation and hyper-parameter tuning, and 93 for testing.

As the base model, we utilise *flax-community/spanish-t5-small*[7]. This model was trained on the large Spanish corpus provided by Cañete et al. (2020). Hyper-parameters were set taking into consideration *oskrmiguel/mt5-simplification-spanish*[8], which is a model for text simplification, although at sentence-level. We employed a learning rate of $2e-5$, weight decay of 0.01, and per device batch size of 8. The other hyper-parameters were set to defaults, training up to 10 epochs.

Evaluation over the validation set using a default generation strategy yielded SARI of 30.45, and BERT score F1 (average) of 0.66 for the best model (9th epoch).

When using the models for generation, we set beam search with 10 beams, with a repetition penalty of 1.2 as in Keskar et al. (2019) to generate from 0.8 to 1.1 the original text length.

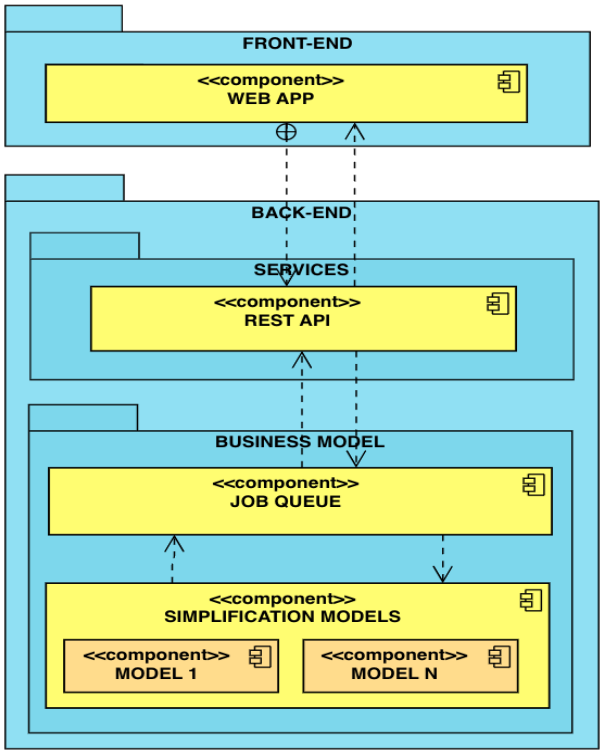

Figure 1: Architecture of Simple.Text Tool

5.2.2 Implementation Details

The tool implements a server-client architecture with the primary objective of providing text simplification services that can be queried from different front-ends or other applications. Figure 1 depicts the main components of the architecture.

The Services component is implemented using Flask[9] as well as Celery[10] for the Job Queue. The Simplification Models component is backed by Hugging Face Transformers Library[11]. Currently, the Web App is a prototype allowing users to select the simplification model and subsequently submit the text for simplification. Figure 2 shows the interface.

5.2.3 Preliminary Tool Testing and Assessment

Experimentation is in a very preliminary stage. Despite presenting different flaws inherent to generative models, such as hallucinations or repetitive text, that need to be addressed, we are currently examining the resolutions (or lack thereof) of the different linguistic phenomena that can be learned from the corpus. It must be pointed out that the language phenomena identified do not have a single NLP solution but several simplification options.

Figure 2 presents the results from imputing

[7] https://huggingface.co/flax-community/spanish-t5-small
[8] https://huggingface.co/oskrmiguel/mt5-simplification-spanish
[9] https://flask.palletsprojects.com
[10] https://docs.celeryq.dev
[11] https://huggingface.co

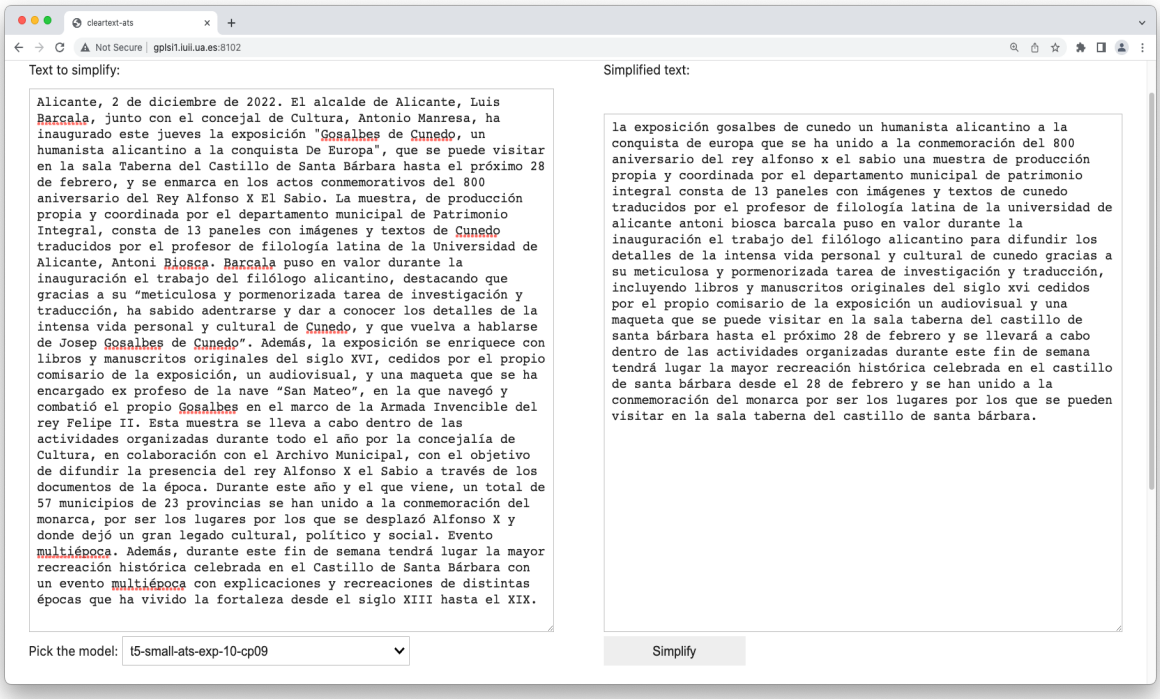

Figure 2: Simple.Text Tool Front End

text 65 from the corpus —which was randomly selected— and the output created by the system. The original text contains several complicated language phenomena identified in the easy-to-read guidelines. The examples presented below showcase some of them, but by no means are those the only complex language phenomena that could be found in that text:

- **Complex sentences:** sentence complexity is often derived from its length, due to the inclusion of appositions, relative clauses, coordinations and other constructions. This complexity can be remedied by splitting the sentence and, as a result, having less information per sentence. An example of a complex sentence from text 65 is the following: *El alcalde de Alicante, Luis Barcala, junto con el concejal de Cultura, Antonio Manresa, ha inaugurado este jueves la exposición "Gosalbes de Cunedo, un humanista alicantino a la conquista De Europa", que se puede visitar en la sala Taberna del Castillo de Santa Bárbara hasta el próximo 28 de febrero, y se enmarca en los actos conmemorativos del 800 aniversario del Rey Alfonso X El Sabio.*

- **Complex enumerations:** In the following example, even though there are not many elements enumerated (i.e. not more than three), additional information is included per each element enumerated. This aspect also increases the complexity level of the enumeration. *[...] se enriquece con libros y manuscritos originales del siglo XVI, cedidos por el propio comisario de la exposición, un audiovisual, y una maqueta que se ha encargado ex profeso de la nave "San Mateo", en la que navegó y combatió el propio Gosalbes en el marco de la Armada Invencible del rey Felipe II.* A potential solution could involve including that additional information in separate sentences, thereby simplifying the enumeration and enhancing the overall readability of the text.

- **Complex vocabulary:** Expressions such as *poner en valor* or *llevar a cabo* could be substituted by more direct verbs such as *valorar/reivindicar* and *realizarse/hacerse*, respectively.

At the moment, as illustrated in Figure 2, the output is far from perfect and presents several issues. Even though there is a reorganisation of the information (e.g. the information about when the event is going to take place appears at the end),

there is a total failure in maintaining punctuation and capitalisation in the output text (e.g. the entire text has no full stops and only one comma is present. Additionally, no capitalisations are maintained for entities). Furthermore, there is a loss of information (e.g. entities) and a patent repetition of source sentences without undergoing any simplification operation. Some sort of simplification has occurred in the first sentence: *la exposición gosalbes de cunedo un humanista alicantino a la conquista de europa que se ha unido a la conmemoración del 800 aniversario del rey alfonso x el sabio*. The subject and the additional information from the appositions, rather than being presented in an independent sentence, are elided.

This preliminary evaluation of the output was performed by the linguists in the group. Nonetheless, a comprehensive evaluation campaign, involving both human validators with cognitive disabilities and a control group of laypeople, will be conducted to assess the effectiveness of the Simple.Text tool once the project is more advanced. It is apparent that a significant amount of work remains to be done, given the preliminary nature of this test. However, this presents an opportunity for substantial improvements to be attained throughout the remaining duration of this project.

6 Expected future impact

The expected positive results from this project that will impact society are three-fold in nature: (1) scientific-technical, (2) social, and (3) economic.

6.1 Scientific-Technical Impact

Language technologies are at the cornerstone of Artificial Intelligence (AI) and are among those tools for which there will be the greatest demand in the next decade. Concerning the scientific and technical impact, our project focuses on researching and developing technologies for NLP to support the authoring of accessible Spanish content for public sector organisations that is intelligible to people with cognitive disabilities. Among the resources developed, which will pique the interest of NLP and AI research communities, are the following:

- Text summarisation, text simplification, lexical analysis, syntactic analysis, anaphora resolution, word sense disambiguation, and summarisation reports.
- The methods, models, resources, and systems that will be researched, developed, and deployed in the project.

6.2 Social Impact

The following positive social impacts for people with cognitive disabilities can be attributed to the fulfilment of this project:

- Facilitation of access to digital information to promote social, and educational inclusion.
- Reduction of the digital divide by identifying barriers that prevent people with disabilities from accessing information on equal terms.
- Promotion of cooperation between the technological and social fields, fostering the design of technological solutions that consider the needs of people with disabilities.
- Facilitation of the daily actions of people with disabilities and widening of inclusion and empowerment in Europe.
- Improvement in the quality of life of those with cognitive disabilities, more specifically, their access to educational, vocational, cultural, and social opportunities in Europe.
- Promotion of an independent life and the capability to realise personal goals.
- Facilitation of equitable access to a meaningful education.
- Promotion of active engagement of individuals in all decisions that have an impact on their future.
- Encouragement of participation in the benefits offered by cultural, recreational, and sporting activities.

6.3 Economic Impact

The development of this project yields the following positive economic impacts for individuals with cognitive disabilities:

- Facilitation of access to digital information to promote economic and political inclusion.
- Promotion of full labor inclusion within the 2040 goal by access to employment for those with cognitive disabilities, and improving productivity via facilitating the performance of work-related functions for those with cognitive disabilities.

- Inclusion and empowerment increase for people with cognitive disabilities in Europe
- Enabling participation in the services provided for promoting effective management of personal finances.
- Provision of equal access to and use of all facilities, services, and activities in the public sector organisations at local, regional, and national levels, such as filing tax returns, paying fines, and managing community charges, among others.

7 Conclusion and Future Work

As a preliminary conclusion, we are currently developing a simplification system in Spanish for people with cognitive disabilities. We are collaborating with APSA, an NGO comprising a group of experts in text simplification, in the creation of a corpus of simplified texts in Spanish. The outcomes of our project will not only contribute to the development of resources for public administration but also facilitate the simplification process for our collaborator, by enabling automated workflows, thereby eliminating the need for manual simplification in the first stage of the simplification-validation process. The resources created by this project will be available on Huggingface[12] and the group's GitHub[13].

Future work is planned in several directions. On the one hand, by improving the corpus. This can be done by increasing its size, the domains as well as the universe of linguistic phenomena covered in it. This may benefit the development of data-driven solutions for ATS at the document level. Also, more research is needed to either validate or reject our hypothesis. Building an automatic document-level text simplification system based on large language models appears to be a challenging task given the scarce number of antecedents. Besides the corpus, other strategies need to be explored such as pre-training the model for specific simplification operations or reinforcement learning from human feedback. On the other hand, concerning the tool, a more advanced user interface needs to be developed so as to provide the user with automatic-to-fine-grained control of the simplification process. For instance, allowing the user to adjust the level of simplification. Additionally, the tool also needs to comply with accessibility recommendations.

[12]https://huggingface.co/gplsi

[13]https://github.com/gplsi

Lay Summary

People with cognitive disabilities face challenges in understanding written language, such as grasping the real meanings of words, phrases, and expressions, as well as retaining information in lengthy sentences, to mention a few. In order to improve their situation, promote their autonomy, and offer unrestricted access to information, Natural Language Processing (NLP) technologies provide ways to automatically simplify texts for these individuals.

In the ClearText project, we are undertaking two important actions to help with the understanding of Spanish texts from the Spanish administration. Specifically, we are creating two resources: the CLEARSIM corpus and the Clear.Text tool.

Firstly, we are putting together a collection of texts —a corpus—, called CLEARSIM, and transforming them into simpler versions with the help of a non-governmental organisation called APSA. These simplified versions have simpler vocabulary and syntax than the original. We are aiming to have approximately 18,000 of these simplified texts by the time the project concludes.

Secondly, we are using a Large Language Models (LLM), a resource that identifies complicated language aspects and automatically simplifies them to create the Clear.Text tool. This model is trained and assessed using the CLEARSIM corpus we are compiling.

We are currently in the process of learning and testing to fine-tune the tool's performance. However, like many LLMs, it frequently makes mistakes such as repeating sentences from the original text without simplifying the complex aspects that we want. Additionally, it may even invent information that was not present in the original text, a phenomenon known as 'hallucination'. We plan to solve these issues and perfect the output text as the the project evolves.

This project has three main goals: first, advancing our understanding of language and technology; second, helping people with cognitive disabilities be more included in society; and third, making the simplification of texts more efficient economically.

Acknowledgments

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

References

AENOR. 2018. Norma Española Experimental UNE 153101 ex. Lectura Fácil: Pautas y recomendaciones para la elaboración de documentos.

Suha Al-Thanyyan and Aqil M. Azmi. 2022. Automated Text Simplification: A survey. *ACM Computing Surveys*, 54(2):43:1–43:36.

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

Fernando Alva-Manchego, Louis Martin, Antoine Bordes, Carolina Scarton, Benoît Sagot, and Lucia Specia. 2020. ASSET: A dataset for tuning and evaluation of sentence simplification models with multiple rewriting transformations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4668–4679, Online. Association for Computational Linguistics.

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.

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. 2012. 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. 2011. An unsupervised alignment algorithm for text simplification corpus construction. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pages 20–26.

Stefan Bott and Horacio Saggion. 2014. Text simplification resources for Spanish. *Language Resources and Evaluation*, 48(1):93–120.

Leonardo Campillos-Llanos, Ana R Terroba Reinares, Sofía Zakhir Puig, Ana Valverde-Mateos, and Adrián Capllonch-Carrión. 2022. Building a comparable corpus and a benchmark for Spanish medical text simplification. *Procesamiento del Lenguaje Natural*, 69:189–196.

José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, and Jorge Pérez. 2020. Spanish pre-trained BERT model and evaluation data. In *PML4DC at ICLR 2020*.

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 Computatinal Linguistics. Proceedings of the Software Demonstrations; 2017 Apr 3-7; Valencia, Spain. Stroudsburg (PA): ACL; 2017. p. 57-60.*

Isabel Espinosa-Zaragoza, José Ignacio Abreu Salas, Elena Lloret, Paloma Moreda, and Manuel Palomar. 2023. A review of research-based automatic text simplification tools. In *Proceedings of the International Conference RANLP-2023*. Accepted for publication.

Itziar Gonzalez-Dios, Iker Gutiérrez-Fandiño, Oscar M Cumbicus-Pineda, and Aitor Soroa. 2022. IrekiaLFes: A new open benchmark and baseline systems for Spanish automatic text simplification. In *Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022)*, pages 86–97.

Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. 2019. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*.

Elena Lloret, María Teresa Romá-Ferri, and Manuel Palomar. 2013. Compendium: A text summarization system for generating abstracts of research papers. *Data & Knowledge Engineering*, 88:164–175.

Tania Josephine Martin, José Ignacio Abreu Salas, and Paloma Moreda Pozo. 2023. A review of parallel corpora for automatic text simplification. key challenges moving forward. In *International Conference on Applications of Natural Language to Information Systems*, pages 62–78. Springer.

Ruslan Mitkov and Sanja Štajner. 2014. The fewer, the better? A contrastive study about ways to simplify. In *Proceedings of the Workshop on Automatic Text Simplification-Methods and Applications in the Multilingual Society (ATS-MA 2014)*, pages 30–40.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

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. *Procesamiento del Lenguaje Natural*, 51:205–208.

Horacio Saggion, Montserrat Marimon, and Daniel Ferrés. 2015. 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*.

Carolina Scarton, Gustavo Paetzold, and Lucia Specia. 2018. SimPA: A sentence-level simplification corpus for the public administration domain. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA).

Matthew Shardlow and Fernando Alva-Manchego. 2022. Simple TICO-19: A dataset for joint translation and simplification of COVID-19 texts. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3093–3102.

Sanja Štajner, Ruslan Mitkov, and Gloria Corpas Pastor. 2014. Simple or not simple? A readability question. *Language Production, Cognition, and the Lexicon*, 48:379.

Sanja Štajner and Horacio Saggion. 2013. Adapting text simplification decisions to different text genres and target users. *Procesamiento del Lenguaje Natural*, 51:135–142.

Sanja Štajner, Horacio Saggion, and Simone Paolo Ponzetto. 2019. Improving lexical coverage of text simplification systems for Spanish. *Expert Systems with Applications*, 118:80–91.

Renliang Sun, Hanqi Jin, and Xiaojun Wan. 2021. Document-Level Text Simplification: Dataset, criteria and baseline. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7997–8013.

Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3:283–297.