FACILE Team at MER-TRANS 2026: Easy-to-Read Adaptation in Spanish through Two-Stage Fine-Tuning of Large Language Models

Isam Diab-Lozano[1,*], Javier Manobanda-Tutasig[1], Cynthia Torres-Celorio[1], Javier Rubira-Herráez[2] and Mari Carmen Suárez-Figueroa[1]

[1]*Ontology Engineering Group (OEG), Universidad Politécnica de Madrid, Madrid, Spain*
[2]*Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, Madrid, Spain*

Abstract

Easy-to-Read (E2R) adaptation aims to improve the accessibility of written information for people with reading comprehension difficulties. Recent advances in Large Language Models (LLMs) have opened new opportunities for automating this process, although ensuring accessibility while preserving the original meaning remains challenging. This paper presents the FACILE Team submission to the Spanish track of MER-TRANS 2026, the first shared task on multilingual E2R adaptation. Our approach is based on the Qwen3.5-9B model and employs a two-stage fine-tuning strategy. In the first stage, the model is adapted to recognise E2R writing through a binary style-classification task. In the second stage, it is fine-tuned on a parallel corpus of original and E2R sentence pairs to perform automatic adaptation. The system was trained using publicly available E2R resources in Spanish and evaluated using the official MER-TRANS metrics. The proposed approach achieved a SARI score of 41.61 on the Spanish track, outperforming the DoNothing baseline. A qualitative analysis further showed that the model successfully learns common E2R transformations, including sentence segmentation, lexical adaptation, and explicit reformulation of specialised concepts. At the same time, challenges related to factual consistency and meaning preservation remain. These results demonstrate the potential of fine-tuned LLMs for supporting automatic E2R adaptation in Spanish.

Keywords

Easy-to-Read, Cognitive Accessibility, Text Adaptation, Large Language Models, Fine-Tuning,

1. Introduction

Ensuring equal opportunities and universal access to information is a fundamental right for all individuals. However, some groups, including people with cognitive disabilities, low literacy skills, language learners, and older adults, may encounter difficulties when accessing written information. In this context, cognitive accessibility plays a key role in enabling participation in education, employment, public services, culture, and democratic processes on an equal basis with others.

To address these challenges, the Easy-to-Read (E2R) methodology was developed as a set of recommendations aimed at improving text comprehensibility for people with reading difficulties [1, 2, 3]. These recommendations cover aspects such as vocabulary, sentence structure, information organisation, and visual presentation. Although E2R adaptations have proven to be effective in improving accessibility, they are typically produced manually by trained experts, making the process time-consuming and costly in terms of both human effort and resources.

To cover this gap, a growing body of research has explored how Artificial Intelligence (AI) can support the automatic or semi-automatic adaptation of texts into E2R versions. Regarding technological support for E2R in Spanish, it is possible to distinguish between tools aimed at analysing compliance with E2R guidelines and tools aimed at generating adapted texts. Examples of the former include Easy-to-Read Advisor [4], FACILE [5], Comp4Text [6], and E2R-Helper [7]. Examples of the latter include Simplext

IberLEF 2026, September 2026, León, Spain
*Corresponding author.
isam.diab@upm.es (I. Diab-Lozano); javier.mtutasig@upm.es (J. Manobanda-Tutasig); cynthia.tcelorio@upm.es (C. Torres-Celorio); j.rubira@alumnos.upm.es (J. Rubira-Herráez); mcsuarez@fi.upm.es (M. C. Suárez-Figueroa)
0000-0002-3967-0672 (I. Diab-Lozano); 0000-0003-3807-5019 (M. C. Suárez-Figueroa)
© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-09-29

CEUR-WS.org/Vol-4273/MER-TRANS-2026_paper4.pdf

[8], LexSIS [9], DysWebxia [10], EASIER [11], Simple.Text [12], FACILE [5], ATECA [13], and AlicIA [14]. However, automatic E2R adaptation remains a challenging task [15], as systems must balance adaptation, readability, meaning preservation, and compliance with accessibility guidelines.

In recent years, the emergence of Large Language Models (LLMs) has created new opportunities for text adaptation. Their ability to perform complex rewriting operations, including lexical substitution, sentence restructuring, paraphrasing, and information reformulation, makes them attractive for E2R adaptation. At the same time, evaluating and comparing such systems remains difficult due to the limited availability of common benchmarks and evaluation frameworks. MER-TRANS 2026 (Multilingual Easy-to-Read Translation) [16] addresses this limitation by providing the first shared task specifically devoted to multilingual E2R adaptation. Organised within IberLEF 2026 [17], the task evaluates systems on their ability to generate E2R versions of original texts in multiple languages under a common evaluation framework.

This paper presents our proposal as FACILE Team, related to the Spanish track of MER-TRANS 2026. The approach is based on Qwen3.5-9B model [18] and employs a two-stage fine-tuning strategy. First, the model is adapted to recognise E2R writing through a binary style-classification task. Then, it is further fine-tuned on parallel original-E2R sentence pairs to perform automatic adaptation.

The rest of this paper is organised as follows. Section 2 outlines the related work. Section 3 describes the MER-TRANS shared task and the dataset used for evaluation. Section 4 presents the proposed methodology, including the training resources, fine-tuning strategy, and prompt design. Section 5 reports the official evaluation results and a qualitative analysis of the generated adaptations. Finally, Section 6 presents some conclusions and future work.

2. Related Work

Several studies have explored the use of instruction-tuned LLMs for generating E2R texts. For Spanish, Madina and colleagues [19] analysed the use of ChatGPT for E2R adaptation and found that, although the model can produce more accessible texts, it does not always follow E2R recommendations and may generate unnecessarily complex structures. Similar observations have been reported for other languages, where LLMs improve readability but still present challenges related to factual consistency and controllability [20, 21].

Beyond prompting, recent work has investigated the fine-tuning of language models on E2R corpora. [22] showed that task-specific fine-tuning can improve adaptation quality compared with general-purpose prompting, although issues related to meaning preservation and hallucinations remain. These results suggest that specialised training data can help align model outputs with accessibility requirements.

The growing interest in LLM-based adaptation has also led to the creation of shared evaluation tasks. One recent initiative is CLEARS 2025 (Challenge for Plain Language and Easy-to-Read Adaptation for Spanish Texts), organised within IberLEF 2025 [23]. The task evaluated systems for adapting Spanish texts into Plain Language and Easy-to-Read formats and attracted several LLM-based approaches.

3. Task Description

This section summarises the objectives of the MER-TRANS 2026 shared task and describes the dataset and evaluation setting used in the competition.

3.1. The MER-TRANS 2026 Shared Task

MER-TRANS 2026 (Multilingual Easy-to-Read Translation) is a shared task organised within IberLEF 2026 [17, 16]. The task focuses on the automatic generation of E2R versions of texts in multiple languages and constitutes the first shared evaluation campaign specifically devoted to multilingual E2R adaptation.

The shared task includes four languages: Spanish, Catalan, Italian, and Arabic. Given an original text, participating systems must generate an E2R version that preserves the original meaning while improving readability and accessibility for people with reading comprehension difficulties. The task is evaluated using a combination of metrics, including BLEU [24], SARI [25] and BERTScore [26].

MER-TRANS shared task does not provide an official training set. Instead, participants are encouraged to exploit external E2R and text adaptation resources to develop their systems. We participated exclusively in the Spanish track and submitted a single system run.

3.2. Dataset

The MER-TRANS dataset is derived from the iDEM Corpus [27], a multilingual human-annotated resource developed within the Horizon Europe iDEM project[1] to support accessible participation in democratic processes. The corpus contains original texts together with professionally produced E2R adaptations in Spanish, Catalan, and Italian.

The texts cover domains related to civic engagement, political participation, public administration, and legislative communication. E2R adaptations were created by expert translators following language-specific adaptation guidelines. For Spanish, the adaptation process was guided by the principles established in the UNE 153101:2018 EX standard [3].

The corpus is aligned at sentence level and is not parallel across languages, meaning that each language contains its own original texts and corresponding E2R adaptations. For Spanish, the corpus contains 354 original sentences comprising 11,665 words and their corresponding E2R adaptations. The original texts have an average length of 32.95 words per sentence, whereas the E2R versions are characterised by extensive sentence segmentation and shorter textual units.

4. Methodology

This section describes our proposed two-stage fine-tuning strategy for E2R adaptation in Spanish. The training resources, fine-tuning configuration, and prompt design employed in the MER-TRANS submission are presented in the following subsections.

4.1. System Overview

The proposed system employs Qwen3.5-9B [18] as its backbone model and follows a two-stage fine-tuning strategy for E2R adaptation in Spanish. This model was selected because it offers a balance between generative capacity and computational feasibility: its 9B-parameter scale allows task-specific fine-tuning on a single A100 GPU using QLoRA, while still providing strong instruction-following capabilities and broad multilingual coverage.

The first stage performs binary style classification, distinguishing between *Lectura Fácil* and *Lectura Difícil* texts. This stage is not used as a filtering mechanism during inference. Instead, its purpose is to expose the model to the linguistic and stylistic properties that characterise E2R Spanish, including shorter sentences, more explicit information, reduced lexical complexity, and simpler syntactic structures. The resulting LoRA adapters are then merged into the backbone model, producing a style-enriched version of Qwen3.5-9B.

The second stage performs the adaptation task itself. Starting from the style-enriched model obtained in the previous step, a new set of LoRA adapters is trained on parallel original–E2R sentence pairs. This allows the system to learn the transformation from standard Spanish into E2R Spanish while preserving the stylistic knowledge acquired during the classification stage.

[1] https://idemproject.eu/

4.2. Training Data

All training resources used in this work were collected independently from the evaluation corpus. The proposed system relies on two complementary datasets: a labelled corpus for style classification and a parallel corpus for E2R adaptation.

4.2.1. Style Classification Corpus

For the style-classification task, we compiled a corpus of 262,000 Spanish texts labelled as either *Lectura Fácil* or *Lectura Difícil*. The Lectura Fácil subset was collected from publicly available E2R publications in Spanish, including institutional, informative, and accessibility-oriented materials. The Lectura Difícil subset was compiled from non-E2R Spanish texts covering general and institutional domains. The corpus was intended to provide the model with broad exposure to stylistic differences between E2R and non-E2R writing rather than to train the adaptation task directly.

The corpus only contains plain textual content. Metadata related to visual presentation, layout, typography, or document structure was not systematically available and was therefore not used during training. For this reason, the classification stage focuses on textual features associated with E2R writing, such as sentence length, lexical complexity, syntactic simplicity, and explicitness of information.

To reduce training time while maintaining class balance, a stratified subset containing 10,000 training texts and 2,000 validation texts was used during fine-tuning. The subset preserved the binary distribution of the original corpus and was used exclusively for the style-recognition stage.

4.2.2. Adaptation Corpus

For the adaptation task, we assembled a corpus of 3,065 parallel original–E2R sentence pairs from two complementary resources.

The first resource, hereafter *LF-SentPairs*, consists of 665 linguist-reviewed adaptation pairs extracted from the corpus developed by Torrelles Rodríguez [28]. The second resource, hereafter *LF-PubTexts*, contains 2,400 original–E2R pairs compiled from accessibility publications in Spanish by Rubio Pérez [29].

The resulting corpus was randomly divided into training (2,452 pairs), validation (306 pairs), and test (307 pairs) partitions following an 80/10/10 split.

Table 1 summarises the datasets used in the two training tasks.

Table 1
Datasets used for style classification and E2R adaptation.

Task	Dataset	Samples
Style classification	Stratified training subset	12,000
Easy-to-Read adaptation	LF-SentPairs	665
Easy-to-Read adaptation	LF-PubTexts	2,400
Easy-to-Read adaptation	Training split	2,452
Easy-to-Read adaptation	Validation split	306
Easy-to-Read adaptation	Test split	307

For the style-classification task, the table reports the stratified subset used during fine-tuning rather than the complete corpus of 262,000 labelled texts.

4.2.3. Preprocessing

Before training, all texts were subjected to a lightweight preprocessing pipeline. Residual HTML tags, carriage-return encoding artefacts, and formatting inconsistencies were removed. In addition, Unicode normalisation was applied to standardise quotation marks, dashes, and whitespace characters.

For the adaptation corpus, each E2R sentence was explicitly separated by a newline character. This formatting reflects the visual segmentation principles of the UNE 153101:2018 EX standard and enables the model to learn not only lexical and syntactic adaptation patterns but also the sentence-per-line structure characteristic of E2R documents.

4.3. Fine-Tuning Setup

All training and inference experiments were conducted on a single NVIDIA A100 40GB GPU using the Hugging Face ecosystem[2]. The `transformers` library was used to load and train the model, while *Parameter-efficient fine-tuning (PEFT)* was employed to manage the LoRA adapters.

The backbone model, Qwen3.5-9B, was loaded using 4-bit NF4 quantisation with double quantisation enabled and `bfloat16` as the computation data type. This configuration substantially reduces GPU memory requirements while preserving model performance, enabling efficient fine-tuning on a single accelerator.

Both fine-tuning tasks employed QLoRA [30] adapters with rank $r = 16$, scaling factor $\alpha = 32$, and dropout 0.05. Training was performed using the AdamW optimiser, an effective batch size of 32, and a learning rate of 2×10^{-4}.

4.3.1. Style Classification Fine-Tuning

Style recognition was formulated as a binary generative classification task. Given an input text, the model was required to generate exactly one of two labels: *Lectura Fácil* or *Lectura Difícil*.

Model checkpoints were evaluated every 200 training steps. After training, the resulting LoRA adapters were merged into the backbone model, producing a style-enriched version of Qwen3.5-9B.

4.3.2. Easy-to-Read Adaptation Fine-Tuning

The adaptation model was initialised from the style-enriched backbone obtained during the previous fine-tuning task. A new set of LoRA adapters was then trained on the parallel adaptation corpus to learn the transformation from standard Spanish into E2R Spanish.

Training was performed for three epochs. A cosine learning-rate scheduler was applied together with a warm-up phase corresponding to the first 3% of training steps. Given the relatively small size of the adaptation corpus, model checkpoints were evaluated every 50 training steps. The checkpoint achieving the lowest validation loss was selected as the final model used in the shared-task submission. During inference, the model was configured with a temperature of 0.2, a top-p value of 0.9, a maximum generation length of 512 tokens, and a no-repeat 5-gram constraint to reduce repetitive outputs.

4.4. Prompt Design

Both fine-tuning tasks employ dedicated ChatML prompts designed to constrain model outputs through explicit instructions. Rather than relying on open-ended natural-language descriptions, the prompts encode a small set of mandatory rules intended to improve output consistency and compliance with E2R conventions.

The use of ChatML provides a structured separation between the system instruction, the user input, and the expected assistant output. In comparison with a single open-ended prompt, this format makes the training instances more consistent with the dialogue template expected by the instruction-tuned model and reduces the probability of generating explanations, preambles, or task-irrelevant content.

[2] https://huggingface.co/

4.4.1. Classification Prompt

The classification prompt formulates style recognition as a constrained generation task. Given an input text, the model must output exactly one of two labels: *Lectura Fácil* or *Lectura Difícil*. No explanations or additional text are permitted.

The prompt follows the standard ChatML format and explicitly restricts the output space to the two target labels. This design prevents the generation of conversational responses and encourages the model to focus exclusively on stylistic classification. The complete prompt is shown in Listing 1.

```
<|im_start|>system
Eres un sistema de clasificación automática.
TU ÚNICA SALIDA permitida es una de estas dos
etiquetas: "Lectura Fácil" o "Lectura Difícil".
NO des explicaciones. Solo la etiqueta.
<|im_end|>

<|im_start|>user
Clasifica este texto: {input_text}
<|im_end|>

<|im_start|>assistant
```

Listing 1: Prompt used for style classification.

4.4.2. Adaptation Prompt

The adaptation prompt operationalises several recommendations of the UNE 153101:2018 EX standard [3]. Rather than relying on lengthy instructions, it defines a compact set of explicit rules that guide the generation process. The complete prompt is shown in Listing 2.

```
<|im_start|>system
Eres un experto en Lectura Fácil.
Adapta el texto que te dará el usuario.

REGLAS OBLIGATORIAS:

1. Frases cortas: Sujeto + Verbo + Predicado.
2. Sustituye palabras cultas por palabras comunes.
3. Escribe UNA frase por línea.
4. Inserta un salto de línea antes de cada frase.
5. NO des explicaciones ni preámbulo.
<|im_end|>

<|im_start|>user
Adapta este texto: "{input_text}"
<|im_end|>

<|im_start|>assistant
```

Listing 2: Prompt used for Easy-to-Read adaptation.

The mandatory rules in Listing 2 means the following: Rule 1 promotes syntactic adaptation through short sentence structures. Rule 2 encourages lexical adaptation by replacing uncommon vocabulary with more frequent alternatives. Rules 3 and 4 enforce the sentence-per-line formatting characteristic

of E2R documents. Finally, Rule 5 suppresses introductory explanations and other forms of extraneous output frequently produced by instruction-tuned language models.

These constraints are further reinforced during fine-tuning through the structure of the adaptation corpus, where E2R target texts are already formatted according to the sentence-per-line convention. Therefore, the model learns both the linguistic transformations required for adaptation and the formatting conventions expected in E2R documents.

It should be noted, however, that the present system focuses mainly on textual adaptation. Although E2R also includes visual and document-level recommendations, such as typography, spacing, layout, illustrations, and the organisation of information on the page, these aspects were not explicitly modelled in the current approach. Formatting is only partially addressed through the sentence-per-line convention encoded in the prompt and represented in the training corpus.

5. Results and Discussion

This section reports the official MER-TRANS evaluation results obtained by our system and discusses the main characteristics of the generated adaptations through a qualitative analysis.

5.1. Official MER-TRANS Results

Table 2 presents the official results obtained by FACILE Team in the Spanish track of MER-TRANS 2026.

Table 2
Official MER-TRANS results obtained by FACILE Team in the Spanish track.

$BLEU_{orig}$	$BLEU_{gold}$	SARI	BERTScore
3.00	2.86	41.61	0.878

According to the official evaluation results reported by the task organisers [16], the highest SARI score obtained in the Spanish track was 47.02, whereas the DoNothing baseline achieved 13.87. Our proposed system obtained a SARI score of 41.61, outperforming the baseline and demonstrating the ability to perform meaningful E2R adaptations beyond merely copying the source text.

An interesting observation is the combination of this SARI score with very low BLEU values. This pattern suggests that the proposed system performs substantial adaptation of the source text instead of preserving surface-level lexical overlap. The generated outputs frequently involve sentence splitting, lexical substitution, explicit reformulation of concepts, and the introduction of explanatory content. These transformations are generally consistent with E2R objectives, but they may also increase the risk of deviating from the original meaning.

These results should be interpreted in light of the known limitations of current automatic evaluation metrics for text simplification and adaptation. BLEU was originally developed for machine translation and primarily rewards n-gram overlap between system outputs and reference texts. For this reason, valid adaptations involving substantial reformulation, reordering, sentence splitting, or explanatory additions may obtain relatively low BLEU scores despite improving accessibility. Several studies have reported weak or inconsistent correlations between BLEU and human judgments in simplification tasks, particularly when systems perform diverse transformation operations [31].

BERTScore partially addresses this limitation by comparing contextual embeddings instead of exact lexical matches. However, it remains an indirect measure of adaptation quality. High semantic similarity does not necessarily imply improved accessibility, and previous work has shown that meaning preservation represents only one dimension of adaptation quality [31, 32].

For this reason, the relatively low BLEU and BERTScore values obtained by our system should be interpreted together with the SARI score and the qualitative analysis presented in the following subsection.

Table 3
Representative examples of generated Easy-to-Read adaptations.

Original Text	Adapted Text	Observation
El anticomunismo es la oposición política e ideológica al comunismo. El anticomunismo organizado se desarrolló después de la Revolución de Octubre de 1917 en Rusia y alcanzó dimensiones globales durante la Guerra Fría.	*El anticomunismo es la oposición al comunismo. El comunismo es una forma de gobierno en la que todos los ciudadanos son iguales. El anticomunismo se desarrolló después del triunfo del comunismo en Rusia. Durante la Guerra Fría, Estados Unidos y la Unión Soviética se enfrentaron en una intensa rivalidad.*	Sentence splitting and explicit concept elaboration.
Sin embargo, el Ejército Rojo derrotó al movimiento Blanco y se creó la Unión Soviética en 1922.	*El Ejército Rojo derrotó al movimiento Blanco en 1919 y se creó la Unión Soviética.*	Hallucinated date and loss of factual accuracy.
Durante las décadas de 1920 y 1930, conservadores, fascistas, liberales y socialdemócratas promovieron la oposición al comunismo en Europa.	*En los años 20 y 30, los conservadores, fascistas, liberales y socialdemócratas se pusieron en contra del comunismo en Europa.*	Lexical adaptation while preserving the original meaning.

5.2. Qualitative Analysis

To better understand the behaviour of the proposed system, we manually inspected representative outputs generated for the Spanish MER-TRANS test set. The inspected examples illustrate several transformation patterns commonly associated with E2R writing, including sentence splitting, lexical adaptation, and the introduction of explicit explanations for specialised concepts. At the same time, some outputs exhibit typical LLM errors such as factual hallucinations, spelling inconsistencies, and the insertion of information not present in the source text.

Table 3 presents representative examples illustrating both successful adaptations and remaining challenges.

The inspected examples suggest that the proposed approach successfully performs several transformations commonly associated with E2R adaptation, particularly sentence segmentation, lexical adaptation, and explicit reformulation of specialised concepts. These operations are consistent with the objectives of E2R adaptation and help explain the competitive SARI score obtained by the system.

However, the examples also reveal important limitations. In some cases, the model introduces factual inaccuracies, modifies dates, or generates explanatory information that is not present in the source text. Such behaviours are characteristic of generative language models and may negatively affect meaning preservation. These observations may partially explain the combination of a relatively high SARI score and lower BLEU and BERTScore values observed in the official evaluation.

6. Conclusions and Future Work

This paper presented the FACILE Team submission to the Spanish track of MER-TRANS 2026. The proposed approach combines a style-classification fine-tuning task with a following E2R adaptation fine-tuning task using Qwen3.5-9B and QLoRA.

The official evaluation results show that the system achieves a SARI score of 41.61, outperforming the DoNothing baseline and demonstrating the potential of fine-tuned LLMs for E2R adaptation. The qualitative analysis revealed that the model successfully performs transformations commonly associated with E2R writing, including sentence segmentation, lexical adaptation, and explicit reformulation of specialised concepts.

At the same time, the generated outputs occasionally introduce factual hallucinations and explanatory content that is not present in the source text. These limitations suggest that meaning preservation

remains a major challenge for generative approaches to E2R adaptation.

Future work will focus on improving semantic faithfulness, exploring larger adaptation corpora, and incorporating evaluation methodologies that better capture accessibility adaptations beyond surface lexical similarity.

Acknowledgments

This work has been supported by the grant "Ayudas para la contratación de personal investigador predoctoral en formación para el año 2022" (Reference: PIPF-2022/COM-25762) funded by Comunidad Autónoma de Madrid (Spain), and the grant FPI-UPM (PRE2024-003105) under the project MALTA PID2024-159504OB-I00 funded by MICIU/AEI/ 10.13039/501100011033 and by ERDF/EU.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and Copilot to check grammar and spelling.

References

[1] Inclusion Europe, Information for All. European standards for making information easy to read and understand, Inclusion Europe, 2009.

[2] M. Nomura, G. S. Nielsen, International Federation of Library Associations and Institutions, Library Services to People with Special Needs Section, Guidelines for easy-to-read materials, IFLA Headquarters, The Hague, 2010.

[3] AENOR, Easy-to-Read. Guidelines and recommendations for the production of documents (UNE 153101:2018 EX), Asociación Española de Normalización, 2018.

[4] M. C. Suárez-Figueroa, E. Ruckhaus, J. López-Guerrero, I. Cano, Álvaro Cervera, Towards the Assessment of Easy-to-Read Guidelines Using Artificial Intelligence Techniques, in: K. Miesenberger, R. Manduchi, M. C. Rodriguez, P. Peňáz (Eds.), Computers Helping People with Special Needs. ICCHP 2020, volume 12376 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 74–82. doi:10.1007/978-3-030-58796-3_10.

[5] M. C. Suárez-Figueroa, I. Diab, E. Ruckhaus, I. Cano, First steps in the development of a support application for easy-to-read adaptation, Universal Access in the Information Society 23 (2024) 365–377. doi:10.1007/s10209-022-00946-z.

[6] A. Iglesias, I. Cobián, A. Campillo, J. Morato, S. Sánchez-Cuadrado, Comp4text checker: An automatic and visual evaluation tool to check the readability of spanish web pages, in: K. Miesenberger, R. Manduchi, M. Covarrubias Rodriguez, P. Peňáz (Eds.), Computers Helping People with Special Needs, Springer International Publishing, Cham, 2020, pp. 258–265.

[7] E. Díez, N. Rodríguez, A. Fernández, M. A. Alonso, M. A. Díez-Álamo, M. J. Sánchez, D. Wojcik, E2R-Helper (Easy-to-Read Assistant) [Text in Spanish], Instituto Universitario de Integración en la Comunidad - Universidad de Salamanca, 2019. URL: https://eapoyo-inico.usal.es/asistente-lectura-facil/.

[8] H. Saggion, S. Štajner, S. Bott, S. Mille, L. Rello, B. Drndarevic, Making It Simplext: Implementation and Evaluation of a Text Simplification System for Spanish, ACM Trans. Access. Comput. 6 (2015). URL: https://doi.org/10.1145/2738046. doi:10.1145/2738046.

[9] S. Bott, L. Rello, B. Drndarevic, H. Saggion, Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish, in: Proceedings of COLING 2012: Technical Papers, The COLING 2012 Organizing Committee, Mumbai, India, 2012, pp. 357–374.

[10] L. Rello, R. Baeza-Yates, H. Saggion, DysWebxia: textos más accesibles para personas con dislexia, Procesamiento del Lenguaje Natural 51 (2013) 205–208.

[11] L. Moreno, R. Alarcón, P. Martínez, Easier system: Language resources for cognitive accessibility, in: Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20), Association for Computing Machinery, New York, NY, 2020. URL: https://doi.org/10.1145/3373625.3418006. doi:10.1145/3373625.3418006.

[12] B. Botella-Gil, I. Espinosa-Zaragoza, P. Moreda Pozo, M. Palomar, "simple-tool": A tool for the automatic transformation of Spanish texts into easy-to-read, in: G. Angelova, M. Kunilovskaya, M. Escribe, R. Mitkov (Eds.), Proceedings of the 15th International Conference on Recent Advances in Natural Language Processing - Natural Language Processing in the Generative AI Era, INCOMA Ltd., Shoumen, Bulgaria, Varna, Bulgaria, 2025, pp. 204–209. URL: https://aclanthology.org/2025.ranlp-1.24/.

[13] M. C. Suárez De Figueroa Baonza, P. M. Blanco, A. Blanco, I. Diab Lozano, M. García-Agudo, A. Muñoz, C. Torres, ATECA: una aplicación web para agilizar las adaptaciones a lectura fácil, Siglo Cero. Revista Española sobre Discapacidad Intelectual 56 (2025) 115–137. URL: https://revistas.usal.es/tres/index.php/0210-1696/article/view/32147. doi:10.14201/scero.32147.

[14] I. Diab, M. C. Suárez-Figueroa, C. Guerra, In Dialogue with Users: Co-Designing AlicIA, an Application for the Adaptation of Short Stories in Spanish into Easier Versions, in: Proceedings of the XXV International Conference on Human-Computer Interaction (Interacción 2025), volume 4000, CEUR Workshop Proceedings, Valladolid, Spain, 2025, pp. 1–15. URL: https://ceur-ws.org/Vol-4000/paper10.pdf.

[15] M. Madina, I. Gonzalez-Dios, M. Siegel, Towards reliable e2r texts: A proposal for standardized evaluation practices, in: K. Miesenberger, P. Peňáz, M. Kobayashi (Eds.), Computers Helping People with Special Needs, Springer Nature Switzerland, Cham, 2024, pp. 224–231.

[16] H. Saggion, M. Tareh, N. Khallaf, D. Adanza, S. Bott, N. P'erez Rojas, A. Rasc'on, S. Szasz, Overview of MER-TRANS at IberLEF 2026: First Shared Task on Multilingual Easy-to-Read Translation, Procesamiento del Lenguaje Natural 77 (2026).

[17] A. Bonet-Jover, J. Á. González-Barba, L. Chiruzzo, Overview of IberLEF 2026: Natural Language Processing Challenges for Spanish and Other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2026), Co-Located with the 42nd Conference of the Spanish Society for Natural Language Processing (SEPLN 2026), 2026.

[18] Qwen Team, Qwen3.5: Towards native multimodal agents, 2026. URL: https://qwen.ai/blog?id=qwen3.5.

[19] M. Madina, I. Gonzalez-Dios, M. Siegel, A preliminary study of ChatGPT for Spanish E2R text adaptation, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 1422–1434. URL: https://aclanthology.org/2024.lrec-main.126/.

[20] T. Uricchio, S. Ceccacci, I. D'Angelo, N. Del Bianco, C. Giaconi, Investigating OpenAI's ChatGPT Capabilities to Improve Accessibility of Textual Information: An Explorative Study, in: M. Antona, C. Stephanidis (Eds.), Universal Access in Human-Computer Interaction, Springer Nature Switzerland, Cham, 2024, pp. 326–336.

[21] F. Ledoyen, G. Dias, A. Lechervy, J. Pantin, F. Maurel, Y. Chahir, E. Gouzonnat, M. Berthelot, S. Moravac, A. Altinier, A. Khairalla, Inclusive Easy-to-Read Generation for Individuals with Cognitive Impairments, 2025. arXiv:2510.00691.

[22] P. Martínez, A. Ramos, L. Moreno, Exploring Large Language Models to generate Easy to Read content, Frontiers in Computer Science Volume 6 - 2024 (2024). URL: https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2024.1394705. doi:10.3389/fcomp.2024.1394705.

[23] B. Botella-Gil, I. Espinosa-Zaragoza, A. Bonet-Jover, M. Madina, L. M. Pi, P. Moreda, I. Gonzalez-Dios, M. T. Martín-Valdivia, L. A. Ure, et al., Overview of CLEARS at IberLEF 2025: Challenge for Plain Language and Easy-to-Read Adaptation for Spanish texts, Procesamiento del lenguaje natural 75 (2025) 393–400.

[24] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, Bleu: a Method for Automatic Evaluation of Machine

Translation, in: P. Isabelle, E. Charniak, D. Lin (Eds.), Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, 2002, pp. 311–318. URL: https://aclanthology.org/P02-1040/. doi:10.3115/1073083.1073135.

[25] W. Xu, C. Napoles, E. Pavlick, Q. Chen, C. Callison-Burch, Optimizing Statistical Machine Translation for Text Simplification, Transactions of the Association for Computational Linguistics 4 (2016) 401–415. URL: https://aclanthology.org/Q16-1029/. doi:10.1162/tacl_a_00107.

[26] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, Y. Artzi, Bertscore: Evaluating text generation with bert, 2020. URL: https://arxiv.org/abs/1904.09675. arXiv:1904.09675.

[27] S. Bott, V. Riegler, H. Saggion, A. Rascón-Alcaina, N. Khallaf, A Multilingual Human Annotated Corpus of Original and Easy-to-Read Texts to Support Access to Democratic Participatory Processes, in: Proceedings of the Language Resources and Evaluation Conference 2026, Palma, Mallorca, Spain, 2026.

[28] D. R. Torrelles Rodríguez, Creación automática de un corpus paralelo de lectura fácil, Master's thesis, Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, 2024.

[29] F. Rubio Pérez, Creación de un corpus en Lectura Fácil en español, Master's thesis, Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, 2024.

[30] T. Dettmers, A. Pagnoni, A. Holtzman, L. Zettlemoyer, Qlora: efficient finetuning of quantized llms, in: Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23, Curran Associates Inc., Red Hook, NY, USA, 2023.

[31] F. Alva-Manchego, C. Scarton, L. Specia, The (Un)Suitability of Automatic Evaluation Metrics for Text Simplification, Computational Linguistics 47 (2021) 861–889. URL: https://aclanthology.org/2021.cl-4.28/. doi:10.1162/coli_a_00418.

[32] H. Saggion, Artificial Intelligence and Natural Language Processing for Easy-to-Read Texts, Revista de Llengua i Dret (2024) 84–103. URL: https://revistes.eapc.gencat.cat/index.php/rld/article/view/4362. doi:10.58992/rld.i82.2024.4362.