HULAT1 at MER-TRANS 2026: UNE-Guided Prompting for Multilingual Easy-to-Read Text Simplification

Laura Martín Ybáñez[1], Isabel Segura-Bedmar[1]

[1]Computer Science Department, Universidad Carlos III de Madrid, Spain

Abstract

This paper describes the HULAT1 system submitted to the MER-TRANS 2026 shared task on Multilingual Easy-to-Read Translation at IberLEF 2026. Our approach is based on LLM prompting using Llama 3.3 70B Versatile, accessed via the Groq API. We designed four prompt configurations combining two types of instructions — simple and advanced — with zero-shot and few-shot strategies. The advanced prompts include simplification rules based on the Spanish UNE 153101:2018 EX standard for easy-to-read content. Since no training data was provided, we built an augmented dataset by applying back-translation on the trial data. For both Spanish and Italian, we submitted three runs corresponding to the configurations that performed best in our internal evaluations. In the official evaluation, HULAT1 ranked third in Spanish and achieved both second and third place in Italian. The results show that few-shot prompting consistently outperforms zero-shot prompting in both Spanish and Italian, highlighting the effectiveness of in-context examples for multilingual easy-to-read text simplification in low-resource settings. While the differences between simple and UNE-guided prompts are small, both prompting strategies achieved competitive results across languages, suggesting that further research is needed to better exploit accessibility-oriented simplification guidelines within LLM prompting frameworks.

Keywords

Text Simplification, Large Language Models, Prompt Engineering, Easy-to-read, Multilingual NLP, UNE 153101

1. Introduction

A significant portion of the population struggles with the literacy skills necessary to navigate everyday written texts. According to an adult literacy survey conducted in 24 highly-developed countries [1], 15% to 30% of the adult population can only understand texts written using basic vocabulary [2]. Consequently, people with intellectual disabilities, low literacy levels, or limited language knowledge face severe barriers when trying to understand official documents, health information, and civic content [3]. This situation represents a challenge to the right to information and access to knowledge recognized in the Universal Declaration of Human Rights [4].

Automatic text simplification (ATS) is a Natural Language Processing (NLP) task that aims to reduce the linguistic complexity of a text while preserving its meaning and making information accessible to a wider audience. ATS systems address two complementary dimensions: *lexical simplification*, which replaces complex words with simpler synonyms, and *syntactic simplification*, which restructures complex sentences into shorter and more transparent expressions [5, 6]. Although most of the early work was focused on English, the research community has increasingly addressed other languages, including Spanish [7, 8].

The MER-TRANS 2026 shared task [3] is the first initiative focusing on Multilingual Easy-to-Read Translation, organized as part of IberLEF 2026. It covers four languages — Spanish, Italian, Catalan, and Arabic — and requires participants to generate easy-to-read simplifications of complex documents following accessible language guidelines. A key difficulty of this task lies in the lack of training data, requiring participants to work with only a minimal trial set of parallel examples.

This paper describes the participation of the HULAT1 team in the Spanish and Italian tracks of the competition. Our main contribution is an LLM prompting approach using Llama 3.3 70B Versatile [9] in which we evaluate how prompt complexity and few-shot examples impact simplification quality.

IberLEF 2026, September 2026, León, Spain

lauramartinn1015@gmail.com (L. M. Ybáñez); isegura@inf.uc3m.es (I. Segura-Bedmar)

Specifically, we compare simple prompts that merely instruct the model to simplify the input text with more advanced prompts based on the UNE 153101:2018 EX standard [10], the official Spanish guideline for easy-to-read document production. These advanced prompts are enriched with rules derived from the standard and aim to guide the model towards producing more accessible texts. Our assumption is that incorporating these guidelines into the prompts improves the quality of the generated simplifications compared to using generic instructions alone.

The main contributions of this work are:

- We present a multilingual prompting framework for text simplification in Spanish and Italian using Llama 3.3 70B.
- We investigate the impact of UNE 153101:2018 EX accessibility guidelines when incorporated into prompts.
- We compare zero-shot and few-shot prompting strategies under a low-resource setting.
- We construct an internal development set through back-translation to support model selection and prompt evaluation.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the task and dataset. Section 4 presents our system. Section 5 describes the evaluation setup. Section 6 presents the results and analysis. Finally, Section 7 concludes the paper and discusses future work.

2. Related Work

Automatic text simplification (ATS) has a long research trajectory, with methodologies transitioning from early rule-based systems and statistical approaches to modern deep learning models [5, 11]. While current neural systems achieve strong results, their success relies on large parallel corpora of complex and simplified texts, resources that remain scarce for languages other than English [12, 13]. This data limitation is particularly pronounced for Spanish, where systems have consistently underperformed compared to English benchmarks [6]. A similar situation can be observed for Italian, where the availability of simplification resources and benchmarks remains considerably more limited than for English, motivating the exploration of multilingual approaches. The emergence of Large Language Models (LLMs) offers a promising alternative, as they can perform simplification through prompting without requiring task-specific training data.

For Spanish, the Simplext project [7] was one of the first initiatives, introducing a modular system with syntactic and lexical simplification modules evaluated on news texts adapted for individuals with special needs. More recently, IrekiaLF_es [13] presented an open benchmark consisting of documents published by the Basque Government in both original and easy-to-read format, with baseline experiments evaluated in a zero-shot scenario.

The TSAR-2022 shared task [6], organized at EMNLP 2022, focused on multilingual *lexical* simplification for English, Spanish and Portuguese. A key finding was that LLM-based prompting delivered competitive results. For example, the UniHD system [14] combined GPT with several prompts using zero-shot and few-shot strategies, achieving the best results in the English track by a wide margin. Moreover, the authors demonstrated that this prompting framework could be adapted to Spanish and Portuguese with minimal adjustments, demonstrating that LLM prompting is a viable approach for multilingual simplification without annotated training data.

More recently, the CLEARS 2025 shared task [8], organized at IberLEF 2025, addressed the automatic adaptation of Spanish texts into Plain Language and Easy-to-Read formats. Among the best-performing systems, the CardiffNLP team [15] explored LLM prompting using LLaMA-3.2 and Gemma-3, comparing zero-shot, one-shot, and few-shot strategies. Their findings indicated that few-shot prompting in Spanish with explicit linguistic examples produced better results than zero-shot. Moreover, Gemma-3 showed better instruction-following behavior than LLaMA-3.2, achieving a cosine similarity of 0.71. Similarly, the UNED-INEDA team [16] also explored several prompting configurations with different

Table 1
MER-TRANS 2026 dataset statistics.

Split	Language	Docs	Segments	Avg. words
Trial	ES	1	4	29.2
Trial	IT	1	8	26.5
Trial	CAT	1	7	27.1
Test	ES	16	346	30.1
Test	IT	15	317	27.9
Test	CAT	13	375	25.2
Test	AR	378	378	17.8

LLM models. In particular, they submitted a zero-shot Dolphin Mistral configuration guided by the UNE 153101:2018 EX specifications, obtaining a cosine similarity of 0.68.

Inspired by these previous approaches, we explore the use of LLM prompting for multilingual text simplification in Spanish and Italian. In particular, we investigate the effect of prompt complexity and guideline-based instructions on simplification quality across both languages.

3. Task and Dataset Description

MER-TRANS 2026 is the first shared task on Multilingual Easy-to-Read Translation [3], organized as part of IberLEF 2026 [17]. The task requires participants to automatically generate easy-to-read versions of complex texts, with outputs that should be simplified, readable and meaning-preserving. It covers four languages — Spanish, Italian, Catalan and Arabic — with texts drawn from the iDEM corpus, which focuses on civic and administrative content related to democratic participation [18].

The organizers provided a small trial set for Spanish, Italian and Catalan, containing a single document per language with its corresponding easy-to-read reference simplification. During the evaluation phase, a test set was released for all four languages without reference simplifications. Because no task-specific training data was provided, teams were allowed to use external simplification resources, such as Simplext [7] and FEINA [19] for Spanish, and SIMPITIKI [20] for Italian. Table 1 summarizes the statistics of the provided data.

Trial data were provided only for Spanish, Italian, and Catalan; no trial data were released for Arabic. The trial documents cover diverse text types. The Spanish trial consists of a legal and administrative text regarding employment rights, covering topics such as professional registration and social security. The Italian trial is a journalistic text discussing the relationship between religion and politics. The Catalan trial is also an administrative text about employment and social security procedures. In all cases, each document is segmented at the sentence or paragraph level, with each segment paired with a human-produced easy-to-read simplification.

Although MER-TRANS 2026 covers four languages, we focused our development efforts on Spanish and Italian, as these languages benefit from a broader availability of existing simplification resources and prior shared tasks [7, 13, 8, 20]. Consequently, we did not submit runs for Arabic or Catalan.

4. System Description

Our system is based on prompt-based inference using Llama 3.3 70B Versatile [9], accessed through the Groq API. This cloud inference platform enables efficient execution of LLMs without requiring local GPU infrastructure.

We selected Llama 3.3 70B Versatile after a preliminary comparison with smaller alternatives, such as Gemma 2 9B. Llama 3.3 70B consistently produced higher-quality simplifications with better adherence to the prompt instructions, particularly for complex syntactic transformations. In preliminary experiments, Gemma 2 9B occasionally failed to follow all simplification constraints, for example by omitting

information or generating additional explanations, whereas Llama 3.3 more consistently preserved the original content and produced only the requested simplification.

This aligns with its robust instruction-following capabilities as documented in its model card [9], which reports a score of 92.1 on the IFEval benchmark, outperforming even larger models like Llama 3.1 405B. Furthermore, Spanish and Italian are among the eight languages officially supported by the model [9], making it a suitable choice for our multilingual setup.

To ensure consistency and reproducibility across generations, the model's temperature was fixed at 0.3, reducing output variability while still allowing some flexibility in lexical and syntactic reformulations. No fine-tuning, post-processing, or architectural modifications were applied.

We designed two instruction variants: **Simple** and **Advanced**, which are described below.

Simple prompt. This prompt assigns the model the role of a text simplification expert for people with reading comprehension difficulties, such as individuals with intellectual disabilities or low literacy levels. It then tasks the model to rewrite the input text using plain, clear and direct language in active voice, without adding or removing information. The complete simple prompt template is shown in Appendix A.

Advanced prompt. The advanced prompt extends the simple instruction with a set of simplification rules selected and adapted from the UNE 153101:2018 EX standard [10], the official Spanish standard for the elaboration of easy-to-read documents. We selected the linguistic recommendations most applicable to the text types observed in the trial data, excluding guidelines related to document layout and visual presentation. Although UNE 153101:2018 EX was originally developed for Spanish, the selected rules focus mainly on general simplification principles, such as using shorter sentences, simpler vocabulary, and reducing syntactic complexity. We therefore applied the same rules in both Spanish and Italian. These include:

- Use short sentences with a single idea; separate ideas into distinct sentences. Do not over-fragment: if two related ideas can fit in one sentence, keep them together.
- Remove parenthetical clauses and appositions; convert them into independent sentences.
- Use the present indicative and active voice; avoid passive constructions (including reflexive passives), gerunds, subjunctive, and compound tenses. Preserve past tense when narrating past events.
- Replace complex connectives: *por lo tanto* / *por consiguiente* → *por eso*; *sin embargo* / *no obstante* → *pero* / *aunque*.
- Use simple and commonly used vocabulary; replace technical or abstract terms with everyday language.
- Never explain everyday words or well-known institutions; explain only acronyms and uncommon legal or occupational terms in a separate sentence immediately after first use.
- Present enumerations of three or more elements as bullet lists introduced with a colon, each element on its own line with a dash.
- Avoid double negations; express negations directly or rephrase in the positive.
- Transform nominal constructions into active verb phrases (e.g., *la obtención del permiso* → *para obtener el permiso*).
- Write the subject explicitly when needed for clarity, but do not repeat it in consecutive sentences if it is understood.

The advanced prompt also includes two constraints: the model must not introduce information that is not explicitly present in the source text, and it must simplify questions without answering them. Moreover, the model is required to generate only the simplified text, without notes, comments or alternative versions. This constraint was introduced because, in preliminary experiments, the model occasionally produced additional explanations or multiple simplification variants. The complete advanced prompt template is shown in Appendix A.

Both the simple and advanced prompts were combined with zero-shot and few-shot strategies, yielding four prompt configurations in total: Simple Zero-Shot (SZS), Simple Few-Shot (SFS), Advanced Zero-Shot (AZS), and Advanced Few-Shot (AFS).

For the official submitted runs, the few-shot variants incorporate three examples randomly selected from the MER-TRANS Spanish trial data. The same examples were used for all inputs.

For the internal evaluation on the augmented development set, the three few-shot examples were drawn from the FEINA dataset [19], a Spanish corpus of manually simplified easy-to-read texts. This avoided overlap between the few-shot demonstrations and the evaluation data, since the MER-TRANS trial data had already been used to construct the augmented development set through back-translation.

For the Italian track, we followed the same methodology as for Spanish, with some adaptations. The full prompt instructions, including all advanced simplification rules, were translated into Italian to ensure the model received instructions in the same target language. Similarly, the three few-shot examples were randomly selected from the MER-TRANS Italian trial data. Beyond these adjustments, the same model, temperature and inference configuration were used, and no additional language-specific tuning was performed.

5. Evaluation

5.1. Metrics

The official evaluation of MER-TRANS 2026 uses four metrics [3]:

- **BLEU-Orig** [21]: BLEU score, a metric based on n-gram overlap, computed against the original complex text. Lower values indicate greater divergence from the source, reflecting a higher degree of rewriting, although very low scores may also signal content loss or meaning changes.
- **BLEU-Gold** [21]: BLEU score computed against the reference simplification. Higher values indicate closer similarity with the human reference.
- **SARI** [22]: SARI evaluates simplification quality by measuring which words are correctly added, kept, or deleted relative to both the source text and the reference simplification. Higher values indicate better simplification operations.
- **BERTScore** [23]: BERTScore between the generated and reference simplifications, measuring semantic similarity at the token level. Higher values indicate better meaning preservation.

For our internal evaluation, we also computed semantic similarity (ST), defined as the cosine similarity between sentence embeddings produced by `paraphrase-multilingual-mpnet-base-v2`. We evaluated both ST_OG (original vs. generated simplification) and ST_RG (reference vs. generated simplification), where higher values indicate greater semantic similarity.

We also computed the Fernández-Huerta readability index [24], a Spanish-specific metric that assigns higher scores to easier-to-read texts, reporting both the absolute score of the generated simplification (FH) and the improvement over the original text (FHΔorig).

5.2. Creation of an Internal Development Set

Since the trial set provided by the organizers was very small (4 segments for Spanish, 8 for Italian and 7 for Catalan), we constructed an internal development set to evaluate and compare the different prompt configurations before the official submission.

To this end, we applied data augmentation techniques based on back-translation, which consisted of two steps. First, we translated the Italian and Catalan trial sets into Spanish using Llama 3.3 70B Versatile via the Groq API, and merged them with the original Spanish trial set. Then, we applied back-translation chains to this combined set to generate additional examples. To increase lexical diversity while preserving meaning, we applied three back-translation paths using different intermediate languages:

- Spanish → English → German → English → Spanish
- Spanish → French → Portuguese → Spanish
- Spanish → English → Chinese → English → Spanish

The selected intermediate languages (English, German, French, Portuguese and Chinese) provide varying degrees of linguistic similarity to Spanish, promoting diverse reformulations while maintaining semantic fidelity. This process yielded a total of 76 examples. All translations were performed using Llama 3.3 70B Versatile via the Groq API. The resulting dataset was used to evaluate the four prompt configurations, enabling systematic comparison and selection of the best-performing runs prior to the official submission.

The internal development set was created only in Spanish and was used to select the three runs submitted to MER-TRANS 2026. Due to time constraints, we did not construct an equivalent development set for Italian. Consequently, the same prompting configurations were submitted to both the Spanish and Italian tracks.

5.3. Results on our internal development set

Table 2 presents the internal evaluation results for all four configurations on the augmented Spanish development set. In text simplification, the ideal output should balance two goals: closely matching the human reference while diverging enough from the complex source text to show real simplification.

Advanced Few-Shot prompting (AFS) achieves the highest scores on BLEU, BERTScore-F1 and ST_RG, indicating that its outputs are the most similar to the reference simplifications. Together with its substantial readability improvement (+21.68 FH points), these results suggest that AFS provides the best balance between simplification and meaning preservation among the evaluated configurations.

Advanced Zero-Shot (AZS) ranks second overall, achieving readability gains comparable to Advanced Few-Shot (AFS) while maintaining competitive semantic similarity scores. In contrast, Simple Few-Shot (SFS) obtains lower readability improvements but relatively high ST_OG values, suggesting a more conservative simplification strategy that prioritizes preserving the original meaning over introducing substantial modifications.

Although SZS achieved the highest SARI and readability scores, it showed substantially lower semantic similarity to both the reference simplifications and the original texts. We therefore selected AZS, SFS, and AFS as the official runs, as they provided a better balance between simplification and meaning preservation.

Table 2
Internal evaluation results on the augmented Spanish development set (76 segments). FH Δorig = improvement in Fernández-Huerta readability over the source text. Submitted runs are marked with *.

Configuration	BLEU	SARI	BERTScore-F1	ST_RG	ST_OG	FH	FH Δorig
RUN1 – AZS*	17.30	42.58	0.5390	0.8441	0.9453	72.82	+22.73
RUN2 – SFS*	16.82	40.28	0.5335	0.8449	0.9549	63.54	+13.45
RUN3 – AFS*	**18.85**	41.18	**0.5447**	**0.8503**	**0.9584**	71.77	+21.68
SZS (not submitted)	15.40	**48.07**	0.5326	0.8239	0.9063	**73.02**	+22.93

A potential limitation of this internal evaluation is that the augmented development set was generated through back-translation using the same Llama 3.3 model employed for simplification. This may introduce some bias in the estimated performance and should therefore be interpreted only as a tool for comparing prompt configurations rather than as an absolute measure of system quality.

5.4. Results on the official test set

Table 3 presents the official MER-TRANS 2026 results for all submitted runs across both languages.

Table 3
Official MER-TRANS 2026 results for HULAT1.

Language	Run	BLEU-Orig	BLEU-Gold	SARI	BERTScore-F1
ES	RUN1 (AZS)	37.1581	18.2081	43.3874	0.9276
ES	RUN2 (SFS)	33.1815	17.7135	45.0182	0.9272
ES	RUN3 (AFS)	31.5608	17.1627	44.7465	0.9272
IT	RUN1 (AZS)	34.6460	15.5855	43.9760	0.9229
IT	RUN2 (SFS)	24.7107	15.1086	48.6690	0.9252
IT	RUN3 (AFS)	23.4350	16.0021	48.6830	0.9245

For Spanish, our best-performing configuration was RUN2 (Simple Few-Shot), which achieved a SARI score of 45.02 and ranked third overall among 19 submitted runs from 10 participating teams. This result was only 2.00 SARI points behind the top-ranked system (47.02) and 0.56 points behind the second-ranked submission (45.58). RUN3 (Advanced Few-Shot) ranked fourth overall with a SARI score of 44.75, while RUN1 (Advanced Zero-Shot) ranked seventh with a SARI score of 43.39.

Our two few-shot configurations outperformed the zero-shot variant by approximately 1.5 SARI points, highlighting the importance of in-context examples for effective simplification. Moreover, the two few-shot configurations differed by only 0.27 SARI points, suggesting that the additional UNE-based guidelines provide little benefit beyond the examples themselves. Finally, BERTScore-F1 remained consistently high across all three configurations ($\geq$0.927), indicating strong semantic similarity to the reference texts regardless of the prompting strategy.

For Italian, RUN3 (Advanced Few-Shot) achieved a SARI score of 48.68 and ranked second overall among 10 submitted runs from 6 participating teams. The gap with respect to the top-ranked system was only 1.15 SARI points (49.84). RUN2 (Simple Few-Shot) obtained a nearly identical SARI score (48.67), ranking third overall. RUN1 (Advanced Zero-Shot) ranked sixth with a SARI score of 43.98.

Similarly to the Spanish results, all three configurations achieved high BERTScore-F1 values ($\geq$0.923), suggesting that the generated simplifications remained semantically close to the reference texts. The two few-shot configurations outperformed the zero-shot variant by approximately 4.7 SARI points, further confirming the effectiveness of in-context examples for text simplification. The minimal difference between the two few-shot configurations (0.01) indicates that the examples themselves account for most of the observed performance, with the UNE-based guidelines contributing little additional improvement.

Overall, few-shot prompting consistently provides higher SARI scores than zero-shot prompting in both languages. While the differences between simple and advanced prompts are very small, the combination of advanced simplification guidelines and few-shot examples (RUN3) achieves the best overall results in Italian and remains highly competitive in Spanish. These findings suggest that combining explicit simplification guidelines with in-context examples is an effective strategy for multilingual text simplification. Regarding the comparison with other participating systems, in Spanish, HULAT1 ranked third with a SARI score of 45.02, only 2.00 points below the best system (ClearText, 47.02) and 0.56 points below the second-ranked system (VICOMTECH, 45.58). Similarly, in Italian, HULAT1 achieved second place with a SARI score of 48.68, only 1.16 points below the top-ranked system (ClearText, 49.84). It is worth noting that HULAT1 obtained the highest BERTScore in both languages, indicating stronger semantic preservation than other systems.

5.5. Qualitative Analysis

To better understand the limitations of our system, we performed a qualitative analysis of the generated simplifications with the lowest SARI scores on the official Spanish test set. For each of the four configurations, we selected the outputs falling in the first quartile of SARI scores and examined them manually. We first describe the most critical issue observed for each individual configuration, and then discuss three recurring error patterns that persist across configurations.

SZS (Simple Zero-Shot). Without few-shot examples and UNE-based guidelines to constrain the output format, SZS frequently generates multiple simplification alternatives and meta-commentaries instead of a single output. The model treats the task as open to interpretation, offering several rewrites and explaining its reasoning rather than committing to one simplified version. One problematic behaviour is that SZS sometimes converts questions into full answers, substantially modifying the meaning of the source segment:

Table 4
SZS converting a question into an answer (ES_071, segment 8).

	Text
Original	¿Cómo funciona un panel de ciudadanos?
SZS output	Un panel de ciudadanos es un grupo de personas que se reúnen para hablar sobre temas importantes. Estas personas son elegidas al azar y representan a la comunidad. El objetivo es que den su opinión y ayuden a tomar decisiones. Funciona de la siguiente manera: * Se eligen ciudadanos al azar para formar el panel [...]
Reference	¿Cómo funciona un panel de ciudadanos? El panel de ciudadanos está dividido en sesiones.

AZS (Advanced Zero-Shot). The explicit instruction in the advanced prompt to return only the simplified text resolves the meta-commentary problem. However, AZS shows a tendency towards surface-level simplification: when the source contains technical terminology, legal language or domain-specific proper names, the model makes minimal or no changes, failing to perform the substantial lexical rewrite that the reference requires:

Table 5
AZS surface-level simplification of legal terminology (ES_043, segment 6).

	Text
Original	Para los menores de 5 años, el periodo de residencia se exigirá a quien tenga la guarda y custodia.
AZS output	Para los menores de 5 años, el periodo de residencia se exigirá a quien tenga la guarda y la custodia.
Reference	Para menores de 5 años de edad el tiempo de vivir en España se cuenta a la persona que sea su tutor.

The only change introduced by AZS is the insertion of the article *la* before *custodia*. The reference, in contrast, replaces *periodo de residencia* with *tiempo de vivir en España* and *guarda y custodia* with *tutor*, reflecting a much deeper lexical rewrite.

SFS (Simple Few-Shot). While SFS shares the meta-commentary problem seen in SZS, it mostly restricts this behavior to shorter segments. Because the few-shot examples provided in the prompt are exclusively long and complex, the model assumes that short inputs need no further adjustment. Consequently, rather than providing a simplification, it either dismisses the task entirely or outputs an explanation:

Table 6
SFS refusing to simplify short inputs (ES_078, segment 0; ES_012, segment 10).

	Text
Original	¿Qué es Lábora?
SFS output	No hay texto que simplificar. El texto original solo contiene la pregunta "¿Qué es Lábora?" y no hay información adicional que simplificar. Por lo tanto, no puedo proporcionar un texto simplificado.
Reference	¿Qué es Lábora?
Original	Los gobiernos fascistas se destacaron como principales oponentes del comunismo en la década de 1930.
SFS output	Los gobiernos fascistas se destacaron como principales oponentes del comunismo en la década de 1930. No se requiere simplificación adicional, ya que el texto original es claro y conciso.
Reference	Los principales gobiernos anticomunistas fueron los gobiernos fascistas.

AFS (Advanced Few-Shot). AFS is generally the most consistent configuration, but it struggles with one specific failure mode, where the UNE rule instructing the model to explain acronyms and uncommon terms is over-applied to ordinary vocabulary, generating unnecessary definitions that expand the output and diverge from the reference. Moreover, AFS sometimes splits sentences correctly but fails to simplify complex vocabulary, resulting in an incomplete transformation:

Table 7
AFS over-applying UNE term-explanation rule and insufficient lexical simplification (ES_071, segment 21; ES_078, segment 1).

	Text
Original	De esta forma se puede conseguir un grupo representativo y diverso.
AFS output	De esta forma se puede conseguir un grupo representativo y diverso. Un grupo representativo es un grupo que tiene personas de diferentes características. Un grupo diverso es un grupo que tiene personas diferentes entre sí.
Reference	De esta manera, forman grupos variados.
Original	El programa Làbora trabaja para fomentar el empleo en Barcelona y acerca a las personas en situación de vulnerabilidad al mercado laboral.
AFS output	El programa Làbora trabaja para fomentar el empleo en Barcelona. Este programa acerca a las personas en situación de vulnerabilidad al mercado laboral.
Reference	Lábora es un programa que impulsa el empleo en Barcelona y ayuda a encontrar trabajo a personas con problemas para tener trabajo.

In the first example, AFS defines *representativo* and *diverso* as if they were technical terms, while the reference simplification replaces both with *variados*. In the second example, AFS correctly splits the sentence into two but leaves *situación de vulnerabilidad* and *mercado laboral* unchanged, while the reference replaces them with plain equivalents.

Common error patterns

Beyond the specific configuration behaviors described above, three error patterns appear consistently across all or most configurations.

- **Insufficient simplification of technical and domain-specific content.** Across all configurations, the model struggles to simplify examples containing acronyms, proper names or domain-specific terminology. In these cases, the generated output is often identical to the source, while the reference requires a deep rewrite involving paraphrase, contextual additions and lexical substitution. The segment in Table 8 shows this, all three submitted runs reproduce the source exactly, while the reference replaces the acronym UDEF, the proper names and adds contextual information not present in the original.

Table 8
All submitted runs failing to simplify technical content (ES_083, segment 33).

	Text
Original	Los funcionarios de la UDEF que investigan el caso Caranjuez creen que el caso de Alvarado Ochoa es similar.
AZS	Los funcionarios de la UDEF que investigan el caso Caranjuez creen que el caso de Alvarado Ochoa es similar.
SFS	Los funcionarios de la UDEF que investigan el caso Caranjuez creen que el caso de Alvarado Ochoa es similar.
AFS	Los funcionarios de la UDEF que investigan el caso Caranjuez creen que el caso de Alvarado Ochoa es similar.
Reference	Los investigadores creen que el caso del antiguo ministro venezolano es parecido, y que le dieron la residencia y la nacionalidad española, a cambio de mentir.

- **Meta-commentary and refusal to simplify short inputs.** SZS and SFS both produce outputs where the model explains why the input does not need simplification, instead of producing the requested output. This behaviour is more frequent in SFS because the few-shot examples bias the model towards expecting longer inputs. AZS and AFS avoid this issue thanks to the explicit constraint requiring the model to produce only the simplified text with no notes or alternatives.
- **Unsolicited content additions.** In AFS and occasionally in SZS, the model introduces content not present in the source text. In AFS this is driven by the UNE term-explanation rule; in SZS it results from the lack of output format constraints. Both behaviours penalize SARI heavily through the *add* component, since the added words are not present in the reference.

These limitations motivate future work on adaptive prompting strategies that better handle input length variability and provide more precise constraints on when to expand upon the original text.

6. Conclusions

In this paper, we presented HULAT1, a prompting-based system for multilingual text simplification submitted to the MER-TRANS 2026 shared task. Our system uses Llama 3.3 70B Versatile via the Groq API and explores four prompt configurations combining simple and advanced instructions with zero-shot and few-shot strategies, without any fine-tuning or task-specific training data. The advanced prompts incorporate explicit simplification rules derived from the UNE 153101:2018 EX standard.

In the official MER-TRANS 2026 evaluation, HULAT1 achieved top-tier performance, ranking third in Spanish and occupying both the second and third positions in Italian. Our main findings can be summarized as follows. First, few-shot prompting consistently outperforms zero-shot prompting in both Spanish and Italian, as reflected by higher SARI scores. These results confirm that providing the model with examples of the target simplification style is highly effective. Second, the impact of the advanced UNE-derived instructions appears limited. In Spanish, the difference between the Simple Few-Shot and Advanced Few-Shot configurations is minimal (45.02 vs. 44.75 SARI), suggesting that few-shot examples alone already provide sufficient guidance for the task. Similarly, in Italian, the two few-shot configurations achieve virtually identical results (48.67 vs. 48.68 SARI), indicating that most of the observed gains stem from the examples rather than from the additional simplification guidelines.

Third, semantic similarity, as measured by BERTScore-F1, is high and stable across all configurations and both languages ($\geq$0.922 in the official evaluation), indicating that the generated simplifications remain semantically close to the reference texts.

It is also worth noting that the Simple Zero-Shot configuration, despite achieving the highest SARI score in our internal evaluation (48.07), was not submitted. Its high SARI score was accompanied by substantially lower semantic similarity scores, suggesting that more aggressive simplifications do not necessarily produce better results when meaning preservation is also considered.

Future work should explore alternative mechanisms for incorporating accessibility-oriented guidelines into LLM-based simplification systems, as the current prompting formulation yields only marginal differences with respect to simpler prompts. Other promising directions include adaptive few-shot example selection, multi-step or iterative simplification strategies, and human evaluation involving target user groups to better assess the accessibility of the generated texts.

Acknowledgments

Grant PID2023-148577OB-C21 (Human-Centered AI: User-Driven Adapted Language Models-HUMAN_AI) by MICIU/AEI/ 10.13039/501100011033 and by ERDF/UE.

Declaration on Generative AI

During the preparation of this work, the authors used Llama 3.3 70B Versatile (accessed via the Groq API) in order to: perform automatic text simplification as the core component of the described system, and to perform data translation and back-translation for dataset augmentation. The authors also used Claude (Anthropic) and ChatGPT (OpenAI) for grammar checking, writing assistance, and manuscript revision during the preparation of this document. After using these tools, the authors reviewed and edited all content as needed and take full responsibility for the content of this publication.

References

[1] OECD, OECD Skills Outlook 2013: First Results from the Survey of Adult Skills, Technical Report, OECD Publishing, 2013. doi:10.1787/9789264204256-en.

[2] S. Štajner, D. Ferrés, M. Shardlow, K. North, M. Zampieri, H. Saggion, Lexical simplification benchmarks for English, Portuguese, and Spanish, Frontiers in Artificial Intelligence 5 (2022). doi:10.3389/frai.2022.991242.

[3] H. Saggion, M. Tareh, N. Khallaf, D. Adanza, S. Bott, N. P. Rojas, A. Rascón, S. Szasz, Overview of MER-TRANS at IberLEF 2026: First shared task on multilingual easy-to-read translation, Procesamiento del Lenguaje Natural 77 (2026).

[4] United Nations, Universal declaration of human rights, https://www.un.org/en/about-us/universal-declaration-of-human-rights, 1948.

[5] H. Saggion, Automatic Text Simplification, Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2017.

[6] H. Saggion, S. Štajner, D. Ferrés, K. C. Sheang, M. Shardlow, K. North, M. Zampieri, Findings of the TSAR-2022 shared task on multilingual lexical simplification, in: Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022), Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Virtual), 2022, pp. 271–283. URL: https://aclanthology.org/2022.tsar-1.31.

[7] 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 Transactions on Accessible Computing (TACCESS) 6 (2015) 1–36.

[8] B. Botella-Gil, I. Espinosa-Zaragoza, A. Bonet-Jover, M. Madina, L. M. Piñar, P. Moreda, I. Gonzalez-Dios, M. T. Martín-Valdivia, L. A. Ureña-López, 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.
[9] Meta AI, Llama 3.3 70b instruct model card, https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct, 2024.
[10] UNE, UNE 153101:2018 EX: Lectura Fácil. Pautas y recomendaciones para la elaboración de documentos, Technical Report, Asociación Española de Normalización, 2018.
[11] S. Štajner, Automatic text simplification for social good: Progress and challenges, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, 2021, pp. 2637–2652.
[12] F. Alva-Manchego, C. Scarton, L. Specia, Data-driven sentence simplification: Survey and benchmark, Computational Linguistics 46 (2020) 135–187.
[13] I. Gonzalez-Dios, I. G.-F. no, O. M. Cumbicus-Pineda, A. Soroa, IrekiaLF_es: 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), Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Virtual), 2022, pp. 86–97. URL: https://aclanthology.org/2022.tsar-1.8.
[14] D. Aumiller, M. Gertz, UniHD at TSAR-2022 shared task: Is compute all we need for lexical simplification?, in: Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022), Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Virtual), 2022, pp. 251–258. URL: https://aclanthology.org/2022.tsar-1.28.
[15] M. Ayesh, N. Gutiérrez-Rolón, F. Alva-Manchego, CardiffNLP at CLEARS-2025: Prompting large language models for plain language and easy-to-read text rewriting, in: Proceedings of IberLEF 2025, 2025.
[16] L. Plaza, L. Araujo, F. López-Ostenero, J. Martínez-Romo, Prompting large language models for Spanish easy-to-read text generation: UNED-INEDA at CLEARS 2025, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), CEUR-WS.org, 2025. URL: https://ceur-ws.org/Vol-4098/.
[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), CEUR-WS.org, 2026.
[18] 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.
[19] S. Calvo, FEINA: a Spanish easy-to-read dataset, https://huggingface.co/datasets/saul1917/FEINA, 2023.
[20] S. Tonelli, A. P. Aprosio, F. Saltori, SIMPITIKI: a simplification corpus for Italian, in: Proceedings of the Third Italian Conference on Computational Linguistics (CLiC-it 2016), Naples, Italy, 2016. URL: https://ceur-ws.org/Vol-1749/paper52.pdf.
[21] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, BLEU: a method for automatic evaluation of machine translation, in: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, 2002, pp. 311–318.
[22] W. Xu, C. Napoles, E. Pavlick, Q. Chen, C. Callison-Burch, Optimizing statistical machine translation for text simplification, in: Transactions of the Association for Computational Linguistics, volume 4, 2016, pp. 401–415.
[23] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, Y. Artzi, BERTScore: Evaluating text generation with BERT, in: International Conference on Learning Representations, 2020.
[24] J. Fernández-Huerta, Medidas sencillas de lecturabilidad, Consigna 214 (1959) 29–32.

A. Prompt Templates

This appendix contains the four prompt templates used in our system. All prompts use a `{texto}` placeholder that is replaced with the input segment at inference time.

A.1. Simple Zero-Shot (SZS)

```
Eres un experto en simplificación de textos en español para personas
con dificultades de comprensión lectora, como personas con discapacidad
intelectual o bajo nivel de alfabetización. Reescribe el siguiente texto
para que sea más fácil de entender. Usa frases simples, claras y directas
en voz activa. No añadas, elimines ni cambies información importante.

Texto a simplificar:
{texto}

Texto simplificado:
```

A.2. Advanced Zero-Shot (AZS)

```
Eres un experto en simplificación de textos en español para personas
con dificultades de comprensión lectora, como personas con discapacidad
intelectual, bajo nivel de alfabetización o conocimiento limitado del
idioma. Tu objetivo es transformar textos complejos en versiones claras,
sencillas y accesibles, preservando siempre el significado original
completo.

REGLA ABSOLUTA: Tu única tarea es reescribir el texto de entrada con
palabras más simples. NUNCA añadas contenido que no esté literalmente en
el texto de entrada. Si el texto es una pregunta, simplifica la pregunta,
no la respondas.

Sigue estas instrucciones:
- Usa frases cortas con una sola idea cada una. Separa las ideas en oraciones
distintas. No fragmentes en exceso: si dos ideas están relacionadas, puedes
mantenerlas en una sola frase.
- Elimina los incisos y aposiciones. Conviértelas en oraciones
independientes que vayan justo después.
- Usa el presente de indicativo y la voz activa. Evita la voz pasiva, los
gerundios, el subjuntivo y los tiempos compuestos. Evita también la pasiva
refleja (se vende, se garantiza).
- Respeta el tiempo verbal del texto original cuando narre hechos pasados.
- Elimina los conectores complejos. Sustituye "por lo tanto" y "por
consiguiente" por "por eso"; "sin embargo" y "no obstante" por "pero"
o "aunque".
- Usa palabras sencillas y de uso frecuente. Sustituye términos técnicos,
abstractos o poco comunes.
- NUNCA expliques palabras cotidianas ni organismos conocidos. Solo explica
siglas y tecnicismos legales poco conocidos en una oración independiente
justo después, sin repetirlos.
- Evita la doble negación.
- Transforma construcciones nominales en oraciones con verbo activo.
- Cuando enumeres tres o más elementos, usa lista con guiones (-).
- Conserva toda la información del texto original.
- Entrega únicamente el texto simplificado. No escribas notas ni ofrezcas
alternativas.

RECUERDA: NUNCA añadas contenido que no esté en el texto de entrada.
```

```
Texto a simplificar:
{texto}

Texto simplificado:
```

A.3. Simple Few-Shot (SFS)

The Simple Few-Shot prompt uses the same instruction as SZS, followed by three examples and a final instruction before the input segment:

> [Instrucción simple como en SZS]
>
> — EJEMPLO 1 —
> Original: En España los abogados han de ser licenciados en Derecho y estar inscritos en el correspondiente Colegio Profesional.
> Simplificado: Los abogados deben ser licenciados en Derecho y estar inscritos en un colegio profesional. Un colegio profesional es un registro público en el que los profesionales se pueden inscribir.
>
> — EJEMPLO 2 —
> Original: Si has de trabajar por cuenta propia, eres tú quien ha de solicitar el alta en el régimen especial de trabajadores autónomos (RETA) de la Seguridad Social.
> Simplificado: Si trabajas por tu cuenta, tú debes pedir el alta en el régimen especial de trabajadores autónomos de la Seguridad Social. El RETA es el sistema especial de los autónomos de la Seguridad Social.
>
> — EJEMPLO 3 —
> Original: Para comprender cuánto cuenta todavía hoy el mundo religioso, hemos realizado una investigación en dos artículos: este que están leyendo está dedicado a los partidos en el gobierno, mientras que aquí encontrará el otro artículo, dedicado a los partidos en la oposición.
> Simplificado: Queremos entender cuánto la religión es importante en la política hoy. Por esto, hemos escrito 2 artículos. En este artículo hablamos de los partidos en el gobierno. El otro artículo es sobre los partidos que no están en el gobierno.
>
> Ahora simplifica el siguiente texto. Conserva toda la información. No añadas nada nuevo.
>
> Texto a simplificar:
> {texto}
>
> Texto simplificado:

A.4. Advanced Few-Shot (AFS)

The Advanced Few-Shot prompt combines the full advanced instruction from AZS with the same three examples used in SFS, placed after the rules and before the input segment.

A.5. Italian Few-Shot Examples

The following three examples were used as few-shot demonstrations for the Italian track, randomly selected from the MER-TRANS Italian trial data:

— ESEMPIO 1 —
Original: Per capire quanto conti ancora oggi il mondo religioso, abbiamo realizzato un’inchiesta in due articoli: questo che state leggendo è dedicato ai partiti al governo, mentre qui trovate l’altro articolo,

dedicato ai partiti all'opposizione.
Semplificato: Vogliamo capire quanto la religione è importante in politica oggi. Per questo, abbiamo scritto 2 articoli. In questo articolo parliamo dei partiti al governo. L'altro articolo è sui partiti che non sono al governo. Lo trovate qui.

— ESEMPIO 2 —
Original: Nell'ottobre 2019, Giorgia Meloni saliva su un palco a Roma e pronunciava una frase diventata il manifesto della sua identità politica: «Io sono Giorgia, sono una donna, sono una madre, sono italiana, sono cristiana».
Semplificato: Giorgia Meloni è dal partito Fratelli d'Italia. Nell'ottobre 2019, ha fatto un discorso a Roma. Ha detto: "Io sono Giorgia, sono una donna, sono una madre, sono italiana, sono cristiana". Questa frase mostra come Giorgia Meloni si vede in politica.

— ESEMPIO 3 —
Original: E ora che governa il Paese, la destra ha trasformato quei segnali con atti politici, nomine e alleanze.
Semplificato: Ora i partiti di destra sono al governo italiano. Hanno realizzato il loro messaggio:
* con i loro decisioni
* con le loro scelte di persone per ruoli importanti
* con nuovi accordi