IREKIER: An Easy Read Corpus for Basque and Spanish

Jesús Calleja[*,1,2] , **Thierry Etchegoyhen** [*,†,1]
[1] Fundación Vicomtech, Basque Research and Technology Alliance (BRTA)
[2] University of the Basque Country UPV/EHU
{jcalleja, tetchegoyhen}@vicomtech.org

Abstract

Easy Read (ER) text adaptation is one of the main means to provide accessible content for people with reading difficulties. ER text features aspects of text simplification, along with specific characteristics such as the need for short sentences, clearly structured content, and explanations for complex concepts. Support for ER text generation is still lacking overall, with few available resources to build automated systems upon. In this work, we describe the IREKIER corpus, based on ER news in Basque and Spanish from the Irekia transparency portal of the Basque Government. This corpus is currently one of the largest publicly shared resource to support training and evaluation of ER text adaptation models in these two languages, and the first of its kind for Basque. We describe our methodology to create the resource, along with the specific challenges raised by ER text. We also provide both intrinsic and extrinsic evaluations of the corpus, which is shared with the scientific community under a CC-BY-NC-ND 4.0 license.

Keywords: Easy Read, Corpora, Basque, Spanish

1. Introduction

Easy Read (ER), also known as Easy-to-read, is a text adaptation method designed to make information accessible to people with reading difficulties, including cognitive disabilities, low literacy, or other challenges in understanding standard texts. ER material adheres to strict guidelines, such as the Inclusion Europe guidelines[1] or the UNE 153101:2018 EX norm for Spanish, which specify expected simplification standards, structure, and text presentation. These guidelines emphasise using short sentences, simple vocabulary, and including explanations of complex concepts where necessary. Additionally, ER guidelines address stylistic elements, such as providing the line spacing and adding images to improve the understanding of the text. Experts in ER adhere to these standards, and their work is validated by the intended audience to ensure its effectiveness.

The growing need to create accessible ER content, supported by specific laws and regulations, requires the development of supporting technology to adapt texts in multiple domains and languages, as current processes are manual and costly. High-quality ER datasets are needed to train and evaluate ER adaptation models, but are scarce overall. Whereas ER corpora have been developed for Swedish (Språkbanken Text, 2017; Mühlenbock, 2008) or Dutch (Vandeghinste et al., 2019), for Spanish only limited resources are publicly shared at present; for Basque no ER corpus is currently available.

In this work, we describe the IREKIER corpus, a novel resource for Basque and Spanish created from original and ER-adapted news from the Basque Government's transparency portal Irekia.[2] Due to ER text characteristics, where information is distributed in specific ways, without systematic 1-1 information correspondence with complex texts, and with additional information in terms of titles or explanations, ER text alignment features specific challenges. We describe both automatic and manual alignment results for our intrinsic evaluation of the corpus, and provide an extrinsic evaluation based on generative models trained on, or exploiting, the prepared datasets. IREKIER is the largest publicly available ER corpus for Spanish and the first of its kind for Basque, shared with the scientific community under a CC-BY-NC-ND 4.0 license.[3]

2. Related Work

The field of ER adaptation is still relatively unexplored in the scientific literature, with a noted lack of empirical research overall (González-Sordé and Matamala, 2024). There is however a growing trend towards approaching ER in terms of natural language processing, from the development of new applications (Suárez-Figueroa et al., 2024; Madina et al., 2024; Diab et al., 2024) and the evaluation of Large Language Models for the task (Martínez et al., 2024; Freyer et al., 2024), to dedicated studies on specific aspects of automated ER adaptation such as text segmentation (Calleja et al., 2024) or explanatory structures (Diab and Suárez-Figueroa, 2024).

[*] Equal contribution.
[†] Corresponding author.
[1] https://www.inclusion-europe.eu/easy-to-read/
[2] https://www.irekia.euskadi.eus
[3] https://huggingface.co/Vicomtech/irekier-corpus.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 1633–1649
11–16 May 2026.

Corpus resources are also scarce in the field, which hinders the development of new approaches and tools to automate ER text generation. Among ER corpora, a notable example is the Swedish sentence-aligned dataset described in Rennes and Jönsson (2016), based on ER and standard texts sourced from public administration websites and containing approximately 30 million tokens. Other Swedish ER resources include the SUC corpus (Språkbanken Text, 2017), which consists of 500 text files covering a range of topics, and the LäsBarT corpus (Mühlenbock, 2008), which comprises 104,058 sentences from easy-to-read books for children and other simplified texts. For Dutch, the Wablieft corpus contains over two million words from a Belgian newspaper written in ER (Vandeghinste et al., 2019).

For the languages covered by the IREKIER corpus, namely Spanish and Basque, ER corpora are also fairly limited. For Spanish, Martínez et al. (2024) describe a corpus of 1,941 aligned sentences, extracted from 13 documents of the Amas Fácil Foundation covering sports guides, literature, competitive examinations, and exhibitions. The corpus is not publicly available although the raw data can be made available by the authors upon request. The ClearSim corpus (Espinosa-Zaragoza et al., 2023) contains 15,000 original texts, 10,000 of which have been automatically adapted to Easy Read and Plain Language using ChatGPT and subsequently validated by humans, with the remaining texts having been manually adapted by professionals. As of this writing, the corpus is not yet publicly available in full. A small subset of ER-adapted data from the ClearText project is publicly shared though, containing three and nine original and ER-adapted documents, respectively.[4]

For evaluation purposes, the IrekiaLFes corpus has been publicly shared, containing 35 manually aligned news documents from the Irekia transparency portal of the Basque government, for a total of 705 sentences (Gonzalez-Dios et al., 2022). Finally, within the CLEARS challenge (Botella-Gil et al., 2025), a dataset comprising 2,400 documents adapted to Easy Read was made available to the participants of the shared task, although the corpus is not publicly shared as of this writing. For Basque, there are currently no available ER corpora we are aware of.

In addition to ER datasets, simplification corpora can provide a basis for the text simplification aspects of ER adaptation. The largest such corpora have been mainly prepared for English, with several corpora derived from aligning Wikipedia content with their simplified counterpart in the Simple Wikipedia pages (Zhu et al., 2010; Coster and Kauchak, 2011; Hwang et al., 2015; Laban et al., 2023).

For Spanish, the Newsela corpus consists of 1,130 professionally simplified news articles, with approximately 60,000 sentences per level (Xu et al., 2016). Another notable resource is CLARA-MeD, which contains 24,298 pairs of medical documents and their corresponding simplified texts, as well as a subset of 3,800 parallel sentences for benchmarking (Campillos-Llanos et al., 2022). For Basque, the CBST corpus includes 227 science-related sentences simplified by a court translator and a teacher (Gonzalez-Dios et al., 2018).

3. Data Collection & Processing

We collected data by first crawling the portal of the Basque Government's transparency portal (Irekia)[5] in both Basque and Spanish. We identified all original articles for which there was a corresponding ER-adapted version, collecting all pairs of documents published until 9 July 2024.

The HTML files were processed with in-house tools to remove boilerplate and extract textual content in the form of titles, paragraphs, and bullet point lists, the latter being treated uniformly as paragraphs. Since the structure of the HTML files was inconsistent across articles, we analysed various HTML tag combinations to correctly discriminate between paragraph boundaries and sentence segmentation breaks in ER articles.

We performed filtering and normalisation over the extracted text with in-house scripts. This process included correcting misplaced line breaks inside sentences, handling unusual characters, and removing redundant line breaks. Afterwards, the text was segmented into paragraphs, which were in turn split into sentences with the Moses scripts (Koehn et al., 2007). We evaluated both types of granularity in our experiments.

Statistics from the extracted texts are presented in Table 1. The number of collected document pairs is larger in Spanish, as more news were adapted to ER in this language, although Basque still counts with 449 document pairs. Several key differences can be pinpointed from the raw data statistics. First, whereas the number of original paragraphs is similar in both languages, the Basque ER version features almost twice the number of original paragraphs. This is mainly due to differences in adaptation styles, a prevalent feature of ER adaptation which may vary depending on the expert or institution in charge of the adaptation task. The differences in the number of sentences indicate that the adaptation style featured paragraphs with very few sentences in Basque, often including only one sentence.

[4] https://github.com/gplsi/corpus-cleartext-cas-v1.0/tree/main

[5] https://www.irekia.euskadi.eus/lf

	ES		EU	
	Original	Easy Read	Original	Easy Read
Documents	548	548	449	449
Paragraphs	6,534	7,002	6,357	12,686
Avg. #chars	300.3	162.41	281.51	140.63
Avg. #words	47.78	27.54	33.36	16.73
Sentences	12,300	13,893	6,961	13,709
Avg. #chars	158.97	81.39	153.18	77.32
Avg. #words	25.39	13.88	19.42	9.86

Table 1: IREKIER corpus statistics for Spanish (ES) and Basque (EU).

In both languages, the average number of characters and words showcases the compactness of ER paragraphs and sentences, as expected per other adaptation guidelines. The large differences in terms of number of words between the two languages, in both the original and the ER versions, is mainly due to Basque being an agglutinative language; differences in number of characters are more indicative of text density at the paragraph or sentence level.

As a final processing step, we split the dataset into training, development, and test sets. We randomly selected 20 documents for development and another 20 for testing in each language, while the remaining documents were used for training. In the next section, we describe the alignment processes we opted for to create datasets suitable for model training and evaluation.

4. Alignment

After collecting and preprocessing the news articles from the Irekia corpus, 40 articles were manually aligned at the paragraph level: 20 for the development partition and 20 for the test partition. The remaining articles were aligned automatically and used for the training partition. This approach ensured a reasonable balance between manual precision for evaluation sets and viability for the larger training dataset.

4.1. Manual alignment

Aligning complex and ER-adapted data is a challenging task, due to ER characteristics such as the non-guaranteed preservation of information between the original and adapted texts, contrary to standard text simplification. Thus, ER adaptations may ignore some of the information in the complex text, for instance if the concepts are too complex or could be confusing for the intended audience. Additionally, information may be present in both texts but adapted in the ER variant, e.g., complex numbers such as *1,150,000* may be adapted to *more than 1 million*. ER adaptation also typically introduces new explicit information in the adapted text, in the form of explanations for complex concepts or section titles. Finally, the information in ER text is typically provided via grouping related information that may be dispersed throughout the original text (see examples in Appendix C).

All these aspects present significant challenges for automatic alignment and we opted to perform manual alignment for the development and test sets. We developed an in-house alignment tool for this task, as described in Appendix A, providing a web interface showing blocks of text from the original and adapted documents, at the paragraph level, which the user could select and validate as alignments. Any number of blocks could be selected from either text, to support many-to-many alignments, which are needed for the task.

There were three annotators in total, all proficient in the corresponding language. One performed alignments in both languages, whereas the other two focused on one of the languages only. There were thus two annotators for the development and test alignments in each language.

The manual alignment process was performed under specific guidelines, as it is not always obvious how specific text blocks should be aligned. The main choices we established for this process are summarised below, using C to denote the original complex text and E for the Easy Read version:

- One-to-one, one-to-many, many-to-one and many-to-many alignments are all permitted.
- Cross-alignments and non-consecutive alignments are all permitted.
- Titles with similar meaning should be aligned individually if they occur in both C and E.
- Titles that only occur in E (respectively, C) should be grouped with the block(s) in E (respectively, C) whose content relates to the title.
- Information in E that does not occur in C (e.g., explanations) should be grouped with related

information in E.

- Information in C that does not occur in E should be left unaligned.
- The lead paragraphs in C that summarise key points in the text should be ignored, unless they contain unique information not mentioned elsewhere in C.

Examples are provided in Appendix B. The guidelines were refined after an initial training phase, during which the annotators were presented with examples and had the opportunity to discuss and clarify any ambiguities. The alignments were exported as JSON files, one for each pair of documents. In case more than one paragraph was chosen in either C or E, the paragraphs were grouped following their order in the text.

Inter-annotator agreement for each language on the dev and test sets is indicated in Table 2. Despite the challenges of ER alignment, the agreement was notably strong between annotators, and the manually aligned datasets thus provide a strong basis for model development and evaluation. However, it is worth noting that an alignment inter-annotator agreement of at most 0.86 indicates that the task is non-trivial even for human annotators operating under specific guidelines.

	ES	EU
Dev set	0.8301	0.8351
Test set	0.8119	0.8656

Table 2: Kappa inter-annotator agreement

The distribution of manual alignments is indicated in Table 3, for alignment types with at least two occurrences. Although the most represented alignment was of type 1-1, a significant number of other types of alignments occurred in both the development and test datasets, more than half in both languages, illustrating the variety of cases posed by ER alignment.

4.2. Automatic alignment

Due to the large number of documents in the corpus, manually aligning all the data would be prohibitively expensive and time-consuming. We thus performed automatic alignment for the training partition, to streamline the process while maintaining reasonable alignment quality. Due to the challenges of ER alignment, as discussed in the previous sections, we evaluated three different alignment tools with different alignment strategies, strengths and weaknesses, namely: CATS (Štajner et al., 2018), Hunalign (Varga et al., 2008), and Vecalign (Thompson and Koehn, 2019).

	ES		EU	
Alignment	Dev	Test	Dev	Test
1-1	48.5%	40.2%	42.1%	41.3%
1-2	16.2%	29.5%	24.0%	19.8%
1-3	13.1%	15.2%	13.2%	16.7%
1-4	6.9%	2.3%	6.6%	12.7%
1-5	1.5%	1.5%	5.8%	-
2-1	6.9%	4.5%	2.5%	3.2%
2-2	2.3%	1.5%	-	0.8%
2-3	2.3%	0.8%	-	-
3-1	0.8%	1.5%	-	2.4%
4-1	-	0.8%	1.7%	0.8%
Others	1.5%	2.3%	4.1%	2.4%

Table 3: Distribution of manual n-m alignments (paragraph level). Alignments with at most 1 occurrence are grouped under *Others*.

CATS is a standard alignment tool for text simplification, which supports 1-to-n alignments and provides three similarity measures. We selected the C3G measure for similarity, which calculates character trigram similarity with log TF-IDF weighting, to capture alignments based on key surface elements such as named entities. As for the alignment strategy, we opted for the Most Similar Text option, which aligns source text fragments with their most similar target counterparts, as it does not assume that information is presented in the same order in the target text.

Hunalign is a standard aligner for parallel corpora, which supports many-to-many alignments via a statistical approach, selecting the optimal alignment path using dynamic programming, while penalising omitted and merged sentences. It uses a hybrid scoring algorithm that combines token-based and length-based features and iteratively merges alignments close to the best 1-1 alignment pairs. Additionally, a reward term is applied if the proportion of shared numeric tokens is sufficiently high, which is particularly useful to align segments containing numerical named entities. We used the *realign* hyperparameter and did not provide a specific dictionary for the alignment process.

Finally, Vecalign is a modern aligner for parallel corpora, which provides an efficient alignment mechanism that can be used with any vector-based similarity scoring function. It is designed to identify minimal similar groups of sentences, assuming non-crossing alignments and allowing local sentence reordering. We used LASER (Artetxe and Schwenk, 2019) as the default scoring metric, and set the maximum number of alignable elements in a given alignment to 8 and 10 for paragraphs and

	ES			EU		
Alignment	**CATS**	**Hunalign**	**Vecalign**	**CATS**	**Hunalign**	**Vecalign**
1-1	46.6%	59.4%	56.2%	46.5%	57.9%	56.7%
1-2	26.0%	19.2%	10.7%	27.3%	20.3%	10.2%
1-3	14.4%	7.0%	6.8%	13.4%	7.1%	6.5%
1-4	7.3%	2.4%	5.6%	6.9%	2.7%	4.9%
1-5	2.9%	0.9%	2.4%	3.0%	1.0%	2.7%
1-6	1.6%	0.3%	1.2%	1.7%	0.4%	1.4%
1-7	0.7%	0.2%	1.5%	0.6%	0.1%	1.6%
2-1	-	7.0%	6.5%	-	7.0%	7.1%
3-1	-	1.2%	3.0%	-	1.3%	3.0%
4-1	-	0.3%	1.8%	-	0.3%	1.7%
5-1	-	0.2%	0.9%	-	0.1%	0.8%
7-1	-	0.1%	2.2%	-	0.1%	2.1%
Others	0.5%	1.7%	1.3%	0.4%	1.7%	1.1%

Table 4: Distribution of n-m alignments on the train set with different aligners (CATS, Hunalign, Vecalign). Alignments with at most 1 occurrence are grouped under *Others*.

		Precision		Recall		F1	
		Dev	Test	Dev	Test	Dev	Test
ES	CATS	**0.7311**	**0.6923**	0.6517	0.7186	0.6891	**0.7052**
	Hunalign	0.5154	0.5281	0.5019	0.6084	0.5085	0.5654
	Vecalign	0.6603	0.5510	**0.7715**	**0.7605**	**0.7116**	0.6390
EU	CATS	**0.6283**	**0.6818**	0.7100	0.7040	**0.6667**	**0.6927**
	Hunalign	0.5138	0.5034	0.6245	0.5379	0.5638	0.5201
	Vecalign	0.5347	0.5610	**0.7732**	**0.7473**	0.6322	0.6409

Table 5: Precision, recall and F1 results on human dev and test references at the paragraph level for all three alignment tools.

sentences, respectively.

Although these tools rely on specific assumptions that might not be met when aligning ER text, such as 1-to-many alignments for CATS or non-crossing alignments for Vecalign, they have demonstrated their efficiency in parallel or simplified text alignment and provide a reasonable basis for automated alignment in the absence of aligners specifically designed for ER text.

The type of alignments resulting from each alignment tool over the IREKIER training sets are indicated in Table 4. CATS clearly diverges from the other two approaches by producing almost half fewer 1-1 alignments that Hunalign or Vecalign in both languages. This is mainly due to CATS not covering many-to-1 alignments, whereas the other two tools assigned 300 such alignments.

To measure the quality of the automatic alignment hypotheses, we first computed precision, recall, and F1 scores against the development and test reference alignments, with the results shown in Table 5. Overall, CATS markedly outperformed the other two aligners in terms of precision and F1, indicating the significant impact of text simplification aspects to compute alignments between complex and ER-adapted text. Vecalign performed better than the alternatives in terms of recall, at the cost of lower precision, which is likely due to CATS ignoring relevant alignments outside its 1-many scope.

To further measure the quality of the alignment hypotheses in terms of similarity, we computed BERTScore (Zhang et al., 2020) and Levenshtein Distance (Levenshtein et al., 1966) over the alignments provided by each tool on the test partition. The results are shown in Table 6. In both languages, CATS outperformed both Hunalign and Vecalign in terms of BertScore similarity, though Hunalign aligned content with lesser distance between complex and ER paragraphs overall. To contrast these results with human alignments, we also include

		ES		EU	
		BERTScore	Lev.	BERTScore	Lev.
Sentence-level	CATS	**0.7546**	155.89	**0.7546**	140.99
	Hunalign	0.7264	**143.43**	0.7080	**113.26**
	Vecalign	0.7305	171.48	0.7078	156.78
Paragraph-level	CATS	0.7653	313.95	**0.7418**	298.00
	Hunalign	0.7190	**270.92**	0.7021	**251.76**
	Vecalign	0.7273	326.89	0.7053	310.20
	Human1	0.7667	355.50	0.7319	348.02
	Human2	**0.7670**	366.39	0.7361	311.79
	Human Common	0.7662	346.41	0.7321	313.77

Table 6: Alignment similarity results at the sentence and paragraph level on the test set, in terms of BERTScore and Levenshtein Distance (Lev).

the results at the paragraph level achieved by the two annotators in each language, and the common alignments for the annotators in each case.

Interestingly, the automated aligners provide alignments that feature shorter Levenshtein distance, compared to human alignments, at the paragraph level. This is a direct by-product of alignment methods, which tend to group source and target elements with similar material, as opposed to human alignments, which may align content that are semantically related without strictly abiding by surface similarity.

These results should be taken with all due caveats, as the scores are relatively low for all methods, automated and human alike, with BERTScore results around 0.75. This would seem to indicate that these metrics might not be fully accurate to measure ER text alignment quality. There is a clear need for new metrics that could take into account the characteristics of the ER alignment task.

Considering these results, we selected the alignments provided by CATS for the training set. Along with the results from the common human alignments, the final partition statistics are provided in Table 7 in terms of aligned paragraphs (Par.) and sentences (Sent.).

	ES		EU	
	Par.	Sent.	Par.	Sent.
train	5,854	20,090	5,278	19,018
dev	391	953	387	1,015
test	393	998	403	1,004

Table 7: Data partition statistics

4.3. Challenging Alignments

ER text alignment presents several challenges, as previously noted. One issue is that complex ideas in the original text may be represented across multiple ER paragraphs, sometimes with significant overlap. This can create redundancy and require arbitrary selection when aligning content.

Additionally, some ER paragraphs condense information heavily, reflecting a key difference between ER adaptation and simplification. This phenomenon can make alignment more difficult, as it may require merging information across ER paragraphs to maintain a balanced distribution of content relative to the original complex text.

Alignment is further complicated by repetition and crossed information, where multiple paragraphs in both the original and ER texts convey overlapping content or refer to information located elsewhere. Such cases increase the risk of inconsistent or incomplete alignment.

Another consideration specific to the Irekia data involves lead paragraphs, which typically summarize key points in the original complex text, and should be ignored in most cases due to their redundancy. However, in some cases, lead paragraphs contain unique information, requiring careful attention during the alignment process.

Finally, structural elements such as titles introduce additional challenges, as they may or may not appear in both versions, potentially leading to intra- or inter-paragraph alignment choices. Together, these factors illustrate the multifaceted nature of ER alignment and the care required to map content accurately between the original and adapted texts. Examples of challenging alignment are provided in Appendix C.

		SARI	BERTScore
ES	Llama3 8B Instruct ZS	41.90	0.7557
	Llama3 8B Instruct FS-Random	44.52	0.7567
	Llama3 8B Instruct FS-Similarity	44.62	0.7633
	Llama3 8B Instruct FS-BM25	45.68	0.7710
	Llama2 7B Base FT	48.87	0.7791
	Llama3 8B Instruct FT	**51.44**	**0.7861**
EU	Latxa3 8B Instruct ZS	32.20	0.7281
	Latxa3 8B Instruct FS-Random	37.92	0.7224
	Latxa3 8B Instruct FS-Similarity	30.71	0.7355
	Latxa3 8B Instruct FS-BM25	39.13	0.7360
	Latxa2 7B Base FT	**39.59**	0.7348
	Latxa3 8B Instruct FT	39.34	**0.7418**

Table 8: Results for models adapted via zero-shot (ZS), few-shot (FS), and fine-tuning (FT)

5. Extrinsic Evaluation

To evaluate the usefulness of the aligned ER corpora, we performed extrinsic evaluations along two lines: (i) fine-tuning LLMs on the training data and (ii) performing both zero-shot and few-shot in-context learning adaptation. To establish comparative results between languages, we selected the following models:

- Llama2 base models with 7B billion parameters, using the Latxa version (hereafter, Latxa2) adapted to Basque (Etxaniz et al., 2024) for that language. These models were only trained via fine-tuning (FT) over the training data, as they do not feature instruction following capability.
- Llama3.1 8B models pre-trained for instruction following, using the Latxa version adapted to Basque (Latxa3) for that language (Sainz et al., 2025). For these models, we experimented with FT over the training data, zero-shot (ZS) with a specific prompt for the task, and the same prompt complemented with five few-shot (FS) examples. For the few-shot variant, we experimented with three different approaches: random sampling, lexical similarity with BM25 (Trotman et al., 2014),[6] and embedding similarity.[7] In the last two cases, five examples were selected for each input, randomly sampled from the development set, with a minimal length of 15 words to discard sampling titles or other short material.

For all models, we used the versions available in HugginFace.[8] For all variants, we used the following instruction in its Spanish and Basque versions: *Adapt the following text, maintaining the information and using simpler words and short sentences*. Although several other instruction variants could be conceived of, adding additional ER-related instructions for example, in our preliminary experiments this instruction provided either similar or better results than all other variants. Additional training details can be found in Appendix D.

For inference, we used a greedy decoding approach with the following hyper-parameters from the HugginFace generation pipeline: $num\_beams = 1$ and $do\_sample = False$. This was mainly meant to support reproducible results, and other values may provide better results for a given model variant.

All models were evaluated in terms of SARI (Xu et al., 2016) and BERTScore on the test sets. Although neither metric is specifically geared towards evaluating ER text, SARI can help measure aspects of simplification, whereas BERTScore helps assess meaning preservation between the source and the ER system hypotheses.

The results are shown in Table 8. Overall, fine-tuned models in both languages outperformed all other variants on both metrics, to a significant degree in Spanish. The IREKIER corpus can thus be seen as solid resource to improve automated ER adaptation in both languages. It is worth noting that even fine-tuned Llama2 models with 1B fewer parameters outperformed Llama3 models queried with any of the ICL variants.

Zero-shot variants were outperformed by the

[6] https://pypi.org/project/rank-bm25/

[7] We used the sentence-transformers *paraphrase-multilingual-mpnet-base-v2* embedding model and the *langchain* library for retrieval (https://www.langchain.com/).

[8] Specifically: meta-llama/Llama-2-7b-hf, meta-llama/Meta-Llama-3.1-8b-Instruct, HiTZ/latxa-7b-v1.2 and HiTZ/Latxa-Llama-3.1-8B-Instruct.

Source
Ligado a la educación, el Gobierno ha acordado en su reunión de hoy conceder una subvención directa a la UNED Universidad Nacional de Educación a Distancia, por un importe de 500.000 euros . La distribución de esta subvención será entre los tres centros de la UNED en Euskadi (Vitoria-Gasteiz, Bergara y Bizkaia) y se reparte en función del número de estudiantes matriculados en cada uno de los centros asociados en Euskadi, en virtud del convenio suscrito entre el Departamento de Educación del Gobierno Vasco y la UNED en abril de 2021. *In relation to education, the Government agreed during its meeting today to grant a direct subsidy to the UNED (National University of Remote Education) for an amount of 500,000 euros. This subsidy will be distributed among the three UNED centers in the Basque Country (Vitoria-Gasteiz, Bergara, and Bizkaia) and will be divided based on the number of students enrolled at each of the associated centers in the Basque Country, pursuant to the agreement signed between the Basque Government's Department of Education and the UNED in April 2021.*
Llama3 8B Instruct FT
El Gobierno Vasco dará 500 mil euros a la Universidad Nacional de Educación a Distancia (UNED). Estos 500 mil euros serán para los 3 centros de la UNED que hay en Euskadi: • Vitoria-Gasteiz • Bergara • Bizkaia. La subvención se repartirá entre los centros en función del número de estudiantes que están en cada centro. *The Basque Government will grant €500,000 to the National University of Remote Education (UNED). These €500,000 will be allocated to the three UNED centers in the Basque Country: Vitoria-Gasteiz, Bergara, and Bizkaia. The grant will be distributed among the centers based on the number of students enrolled in each center.*
Llama3 8B Instruct FS-BM25
El Gobierno Vasco ha aprobado dar 500.000 euros a la UNED (Universidad Nacional de Educación a Distancia) para ayudar a sus centros en Euskadi. Estos centros están en Vitoria-Gasteiz, Bergara y Bizkaia. El dinero se repartirá según el número de estudiantes que están matriculados en cada centro. *The Basque Government has approved a €500,000 grant to the UNED (National University of Remote Education) to support its centers in the Basque Country. These centers are located in Vitoria-Gasteiz, Bergara, and Bizkaia. The money will be distributed according to the number of students enrolled at each center.*

Table 9: Examples of automated ER adaptation with fine-tuned and few-shot models.

other variants for the most part, unsurprisingly as they lack relevant information for ER adaptation. The sole exception was the few-shot model based on embedding similarity in Basque, whose comparatively poor results may be attributed to the limitations of the retrieval model for that language.

Among few-shot variants, the lexically-oriented BM25 proved to be the optimal choice in both languages, which is indicative of the importance of lexical anchors in ER adapation. This aspect was noted by the annotators during the human alignment phase, as they often had to rely on key lexical cues to be able to determine specific alignments.

Although the tendencies were similar in both languages, the differences in scores between models were significantly more salient in Spanish than in Basque. This could be attributed to two main factors. First, the pretrained models provide a wider coverage in Spanish, even compared to the Basque-adapted Latxa models, leading to stronger adaptation capability on downstream tasks. Secondly, the evaluation metrics themselves are typically more precise for higher-resource languages, leading to smaller differences between variants in low-resource languages.

In future work, we will perform dedicated human evaluations of automatically adapted ER text to gain further understanding of the strengths and limitations of adaptation methods and current automatic evaluation metrics. Due to the need to involve experts in ER content creation and/or people with reading difficulties, this type of evaluation was beyond the scope of this work. Table 9 presents illustrative examples of automated adaptation with the two best variants, namely Llama3-8B in its fine-tuned and few-shot variants. More examples are provided in Appendix E.

6. Conclusion

In this work, we presented IREKIER, a corpus to train and evaluate Easy Read adaptation models in Basque and Spanish. The corpus was built from news published by Irekia, the transparency portal of the Basque Government, exploiting the available texts professionally adapted to Easy Read. Alignment between complex and adapted texts was performed manually for the development and test

datasets, to ensure alignment quality, and automatically for the larger training dataset. We contrasted three different automatic aligners, selecting CATS alignments for the final datasets, although the specific challenges of ER text alignment call for dedicated tools better suited for the task.

Beyond alignment and similarity metrics to measure the quality of the corpus, we performed extrinsic evaluations by adapting large language models via fine-tuning and in-context learning under zero-shot and different few-shot variants. Models fine-tuned over the IREKIER corpus achieved significantly better results in terms of SARI and BERTScore, particularly in Spanish. Few-shot with lexical retrieval was a second-best approach in our experiments, showcasing the usefulness of the datasets for automated ER adaptation.

IREKIER is the first aligned ER corpus suitable to train and evaluate ER adaptations in Basque, and currently one of the largest publicly available ER corpus for Spanish. One of its defining characteristics is its reliance on professionally adapted texts, for both languages. The corpus is shared with the community under a CC-BY-NC-ND 4.0 license.[9]

7. Limitations

The main limitation of the IREKIER corpus relates to the topics it covers, which are mainly related to socio-economic and governmental news in the Basque Country, as published by Irekia. Thus, the data feature a preponderance of specific named entities (e.g., the Basque Government and its main representatives) and topics related to governmental activities. The corpus does however cover a wide range of sub-topics, including public initiatives, activities of diverse associations, health campaigns, and economic news, among others.

As another potential limitation, the manual alignment process was not performed by experts in Easy Read content creation. However, the annotators were very familiar with the topic and characteristics of ER adaptation, and dedicated practice sessions were performed on sampled training data to consolidate their contributions to the alignment tasks. Additionally, the alignment task is not among the typical activities of ER experts, as it is a rather specific task for the creation of parallel corpora.

Finally, our work centred on two languages, Basque and Spanish, which are the only two languages of the Irekia portal. Future work will be needed to provide similar publicly shareable resources in other languages.

[9]We contacted Irekia representatives to confirm that the creation and distribution of IREKIER was in line with their open data policy.

8. Acknowledgements

We wish to thank the annotators for their time and the anonymous LREC reviewers for their comments and suggestions. This work was partially supported by the Department of Economic Competitiveness of the Basque Government (Spri), via projects IRAZ (ZL-2022/00788) and ERAI (ZL-2026/00434).

9. Bibliographical References

Mikel Artetxe and Holger Schwenk. 2019. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610.

Beatriz Botella-Gil, Isabel Espinosa-Zaragoza, Alba Bonet-Jover, Margot Madina, Lucas Molino Pi, Paloma Moreda, Itziar Gonzalez-Dios, M Teresa Martín-Valdivia, L Alfonso Ure, et al. 2025. Overview of clears at iberlef 2025: Challenge for plain language and easy-to-read adaptation for spanish texts. *Procesamiento del Lenguaje Natural*, 75:393–400.

Jesús Calleja, Thierry Etchegoyhen, and David Ponce. 2024. Automating easy read text segmentation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 11876–11894, Miami, Florida, USA. Association for Computational Linguistics.

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(0):189–196.

William Coster and David Kauchak. 2011. Simple English Wikipedia: A new text simplification task. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 665–669, Portland, Oregon, USA. Association for Computational Linguistics.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. Qlora: efficient finetuning of quantized llms. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA. Curran Associates Inc.

Isam Diab, Mari Carmen Suárez-Figueroa, and Roberto Peris. 2024. Towards an automatic easy-to-read adaptation of dialogues in narrative texts

in spanish. In *Computers Helping People with Special Needs*, pages 208–216, Cham. Springer Nature Switzerland.

Isam Diab and Mari Carmen Suárez-Figueroa. 2024. First attempt at an automatic adaptation of explanatory structures in spanish to easy-to-read. In *CEUR Workshop Proceedings*, volume 3846, page 182 – 189.

Isabel Espinosa-Zaragoza, José Abreu-Salas, Paloma Moreda, and Manuel Palomar. 2023. Automatic text simplification for people with cognitive disabilities: Resource creation within the ClearText project. In *Proceedings of the Second Workshop on Text Simplification, Accessibility and Readability*, pages 68–77, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Julen Etxaniz, Oscar Sainz, Naiara Miguel, Itziar Aldabe, German Rigau, Eneko Agirre, Aitor Ormazabal, Mikel Artetxe, and Aitor Soroa. 2024. Latxa: An open language model and evaluation suite for Basque. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14952–14972, Bangkok, Thailand. Association for Computational Linguistics.

Nils Freyer, Hendrik Kempt, and Lars Klöser. 2024. Easy-read and large language models: on the ethical dimensions of llm-based text simplification. *Ethics and Information Technology*, 26(3):50.

Itziar Gonzalez-Dios, María Jesús Aranzabe, and Arantza Díaz de Ilarraza. 2018. The corpus of basque simplified texts (CBST). *Lang Resources & Evaluation*, 52(1):217–247.

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, Abu Dhabi, United Arab Emirates (Virtual). Association for Computational Linguistics.

Mariona González-Sordé and Anna Matamala. 2024. Empirical evaluation of easy language recommendations: a systematic literature review from journal research in catalan, english, and spanish. *Universal Access in the Information Society*, 23(3):1369–1387.

William Hwang, Hannaneh Hajishirzi, Mari Ostendorf, and Wei Wu. 2015. Aligning sentences from standard wikipedia to simple wikipedia. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 211–217.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions*, pages 177–180. Association for Computational Linguistics.

Philippe Laban, Jesse Vig, Wojciech Kryscinski, Shafiq Joty, Caiming Xiong, and Chien-Sheng Wu. 2023. SWiPE: A dataset for document-level simplification of Wikipedia pages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10674–10695, Toronto, Canada. Association for Computational Linguistics.

Vladimir I Levenshtein et al. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710. Soviet Union.

Margot Madina, Itziar Gonzalez-Dios, and Melanie Siegel. 2024. Languagetool as a cat tool for easy-to-read in spanish. In *Proceedings of the 3rd Workshop on Tools and Resources for People with REAding DIfficulties (READI)@ LREC-COLING 2024*, pages 93–101.

Paloma Martínez, Alberto Ramos, and Lourdes Moreno. 2024. Exploring large language models to generate easy to read content. *Frontiers in Computer Science*, Volume 6 - 2024.

Katarina Mühlenbock. 2008. Readable, legible or plain words–presentation of an easy-to-read swedish corpus. In *Multilingualism: Proceedings of the 23rd Scandinavian Conference of Linguistics*, volume 8, pages 327–329. Acta Universitatis Upsaliensis Uppsala, Sweden.

Evelina Rennes and Arne Jönsson. 2016. Towards a corpus of easy to read authority web texts. In *Proceedings of the Sixth Swedish Language Technology Conference (SLTC2016), Umeå, Sweden*.

Oscar Sainz, Naiara Perez, Julen Etxaniz, Joseba Fernandez de Landa, Itziar Aldabe, Iker García-Ferrero, Aimar Zabala, Ekhi Azurmendi, German Rigau, Eneko Agirre, et al. 2025. Instructing large language models for low-resource languages: A systematic study for basque. In

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 29124–29148.

Sanja Štajner, Marc Franco-Salvador, Paolo Rosso, and Simone Paolo Ponzetto. 2018. CATS: A tool for customized alignment of text simplification corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

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

Brian Thompson and Philipp Koehn. 2019. Vecalign: Improved sentence alignment in linear time and space. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1342–1348, Hong Kong, China. Association for Computational Linguistics.

Andrew Trotman, Antti Puurula, and Blake Burgess. 2014. Improvements to bm25 and language models examined. In *Proceedings of the 19th Australasian Document Computing Symposium*, pages 58–65.

Vincent Vandeghinste, Bram Bulté, and Liesbeth Augustinus. 2019. Wablieft: An easy-to-read newspaper corpus for dutch. In *Proceedings of CLARIN Annual Conference*, pages 188–191.

Dániel Varga, Péter Halácsy, András Kornai, Viktor Nagy, László Németh, and Viktor Trón. 2008. Parallel corpora for medium density languages. In *Recent advances in natural language processing IV: selected papers from RANLP 2005*, pages 247–258. John Benjamins Publishing Company.

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

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

10. Language Resource References

Språkbanken Text. 2017. *SUCX 2.0*. PID https://doi.org/10.23695/n91v-yy47.

A. Manual Alignment Tool

To facilitate the manual alignment process, we built a Web-based tool, with a simple interface as illustrated in Figure 1. Users can upload a json file containing the source and target text, split at the paragraph level in this case. Each paragraph is represented as an autonomous text block, with the original text blocks on the left and the ER adapted blocks on the right. Users can select multiple text blocks, which will then be visualised in blue. Selected blocks can be unselected as needed by simply clicking specific blocks, or by clicking on the *Unselect* button to clear all selected blocks.

Once the user clicks on the *Validate* button, all selected aligned blocks appear in green and the alignments are listed at the bottom of the screen, as shown in Figure 2. The user can remove specific alignments upon revision or correction of the alignments. Once all alignments have been performed, the alignments are exported as JSON files.

B. Manual Alignment Guidelines Examples

Table 10 provides the guidelines on the main difficult cases for the manual alignment process.

C. Challenging Alignment Examples

ER text alignment presents several challenges in terms of alignment, as illustrated with some typical examples in Table 11. In the first example, the main idea provided in the complex paragraph is represented in two separate ER paragraphs with large overlaps, which can lead to either the selection of both redundant paragraphs or the arbitrary selection of one of the two.

The second example illustrates summarised information, highlighting one of the key differences between ER adaptation and simplification. In this example, both paragraphs might be aligned, although there could be cases where ER information is so condensed that merging it with other sections of the ER text might provide for a more complete distribution of information between the original and adapted texts.

Figure 1: Text block selection.

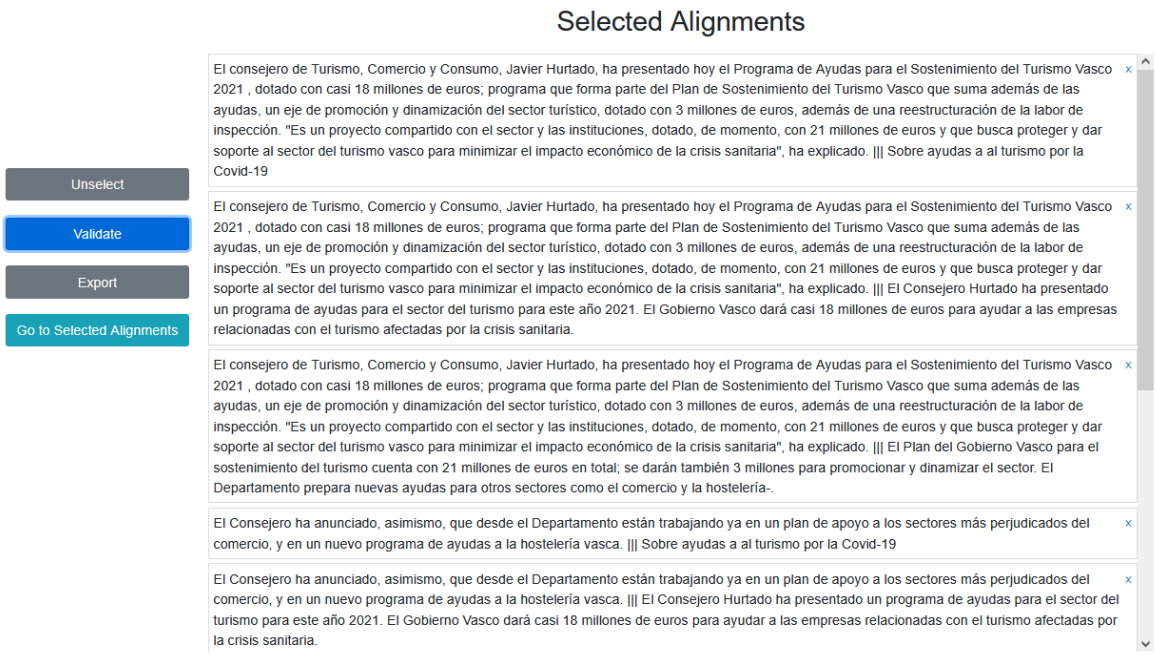

Figure 2: Validation, removal and data export.

In the third example, similar ideas are repeated in both versions of the text, with crossed information. In this case, both include redundant information (e.g., *117 km* road size) distributed in 2 separate paragraphs, while the two ER paragraph also contain redundant information originating from the first complex paragraph (e.g. *Alavese* origin) or the second (e.g., *familia*). Note that in this case, the ER paragraphs also refer to additional information elsewhere in the original text (e.g., *Llanada*).

Finally, the fourth example illustrates a specific challenge of alignment in Irekia data: whereas in

Guideline	Example
If, in either the complex or ER side, there is a title and the following paragraph matches the content of the title, both elements should be selected together, unless the following case applies.	**Title:** La campaña **Paragraph:** La campaña Emakumeak gora! durará desde el 27 de febrero hasta el 14 de marzo. *(Gloss: The campaign // The Emakumeak gora! campaign will last from February 27 to March 14.)*
If there is an equivalent title on both sides, align only those two together.	**Source title:** DOBLE APUÑALAMIENTO SANTURTZI **Target title:** Apuñalan a dos personas en Santurtzi *(Gloss: Double stabbing in Santurtzi / Two people stabbed in Santurtzi.)*
Content appearing in bullet points in the lead of the original text should not be selected.	**Source example:** • El departamento de Igualdad, de la mano de Emakunde, impulsa la campaña junto a las tres diputaciones forales, EUDEL y las capitales. • Nos recuerda que el hecho de que solo el 6% de las gerencias estén ocupadas por mujeres no es por falta de preparación, sino por la persistencia de los roles sexistas. Gasteiz, 2023/02/27 *(Gloss: The Department of Equality, together with Emakunde, is promoting the campaign alongside the three provincial councils, EUDEL, and the capitals. It reminds us that the fact that only 6% of management positions are held by women is not due to a lack of qualifications, but to the persistence of sexist roles.)*
If a given paragraph contains information that only occurs in another paragraph with additional information, the two paragraphs should be aligned.	**Source:** El Consejero ha señalado que el reto en esta ocasión es que sean unos presupuestos que ofrezcan una "visión más allá del PIB" y en los que toma una especial relevancia el enfoque de género. **Target:** El Consejero Aspiazu ha dicho que estos presupuestos: • Están dentro de su programa de gobierno 2016-2020 y de la Estrategia de Desarrollo Humano. • Son necesarios para poder hacer políticas públicas que son competencias del Gobierno Vasco. • Van a invertir mucho dinero sobre todo en salud; después en educación, empleo, políticas sociales y desarrollo económico. • Dan también mucha importancia a los temas de género, es tener en cuenta la situación de las mujeres con respecto a los hombres. *(Gloss: The Councillor has pointed out that the challenge this time is to create budgets that offer a "vision beyond GDP" and in which the gender perspective is particularly relevant. // Councillor Aspiazu has said that these budgets: • Are part of his 2016-2020 government program and the Human Development Strategy. • Are necessary to implement public policies that fall under the jurisdiction of the Basque Government. • Will allocate a large amount of money, mainly to healthcare, followed by education, employment, social policies, and economic development. • Also place great importance on gender issues, considering the situation of women in relation to men.)*

Table 10: Main alignment guidelines and examples.

the vast majority of cases the lead information in the original text only summarises the key points and is present in the rest of the text, thus to be ignored per our guidelines, there are isolated cases where the lead contains information not found anywhere else in the original text and linked to parts of the ER text.

These are only a few illustrative examples of ER alignment challenges. Other cases include titles, which may or may not occur in both versions of the text, leading to inter- or intra-alignment depending on the case. Approximate translations for the examples of challenging alignments are shown in Table 12.

Original	Easy Read
Paragraph Selection Conflict	
“Por eso, esta campaña quiere enseñar las dificultades que hoy en día siguen teniendo las mujeres para tener los mismos puestos que los hombres con las mismas condiciones."	“...las mujeres siguen encontrándose con barreras que les impiden ascender en igualdad a posiciones..." “...en esta campaña mostramos cuáles son los principales obstáculos que aún impiden a las mujeres llegar en igualdad a estos puestos de liderazgo..."
Summarised Information	
“Consciente del delicado momento que vivimos, el Lehendakari ha manifestado la necesidad de superar la etapa de confrontación política que se está dando en el ámbito del Estado y propiciar un escenario de diálogo y acuerdo que permita dar respuesta a los problemas globales que afectan a ambos Gobiernos. “Compartimos las prioridades de responder con eficacia a la crisis sanitaria y superar la crisis económica y de empleo derivada de la misma", ha añadido."	“...El Lehendakari cree que no debe haber enfrentamientos en la política del país, sino dialogar y llegar a acuerdos. El Gobierno Vasco y el español comparten la idea de que hay que actuar para responder a la crisis sanitaria y superar la crisis económica y de empleo."
Repeated Ideas & Cross Information	
“...Un total de 117 kilómetros para el disfrute de visitantes, y de alaveses y alavesas que, según ha augurado, “será todo un éxito"." “...se ha diseñado para que pueda ser disfrutada por el público general y, en especial, por familias. La ruta está ideada para ser completada en dos o tres días, su distancia global es de 117 km..."	"Esta ruta es para recorrer la Llanada Alavesa en bicicleta, disfrutar de la naturaleza y hacer turismo en familia. La Gran Ruta de la Llanada Alavesa tiene 117 kilómetros." “...con esta ruta se quieren dar a conocer los lugares que tiene Álava y la comarca de la Llanada para: • las personas que disfrutan de la naturaleza • las personas que hacen deporte al aire libre • el turismo en familia Durante los 117 kilómetros de la Gran Ruta..."
Information only in the Lead	
"• La misión de las y los trabajadores que lo conforman será divulgar el proyecto y favorecer la activación de las y los hablantes."	"...Este grupo dará a conocer la iniciativa Euskaraldia a los trabajadores y trabajadoras de su Departamento."

Table 11: Examples of challenging alignments in Spanish.

D. Training Parameters

Fine-tuning was performed with QLoRA (Dettmers et al., 2023) and Hugging Face’s SFT Trainer,[10] configured with the following parameters: *dropout=0.05*, *alpha=8*, and *r=16*. The tuning targeted the query and value components, with a learning rate of *3e-4* and a 4-bit model. All other settings remained as per their default values. Training continued until convergence, with early stopping applied after 5 non-improving steps.

[10] https://huggingface.co/docs/trl/sft_trainer

E. Output Examples

Tables 13 and 14 provide illustrative examples of ER adaptation with the selected models in Spanish and Basque, respectively.

Original	Easy Read
Paragraph Selection Conflict	
“Therefore, this campaign aims to show the difficulties that women still have today in obtaining the same positions as men under the same conditions."	“...women continue to face barriers that prevent them from rising to equal positions..."
	“...in this campaign we show the main obstacles that still prevent women from reaching these leadership positions on an equal footing..."
Summarised Information	
“Aware of the delicate moment we are experiencing, the Lehendakari has expressed the need to overcome the current period of political confrontation within the State and foster a framework for dialogue and agreement that will allow for a response to the global problems affecting both governments. “We share the priorities of responding effectively to the health crisis and overcoming the resulting economic and employment crisis," he added."	“...The Lehendakari believes there should be no confrontation in the country’s politics, but rather dialogue and agreement. The Basque and Spanish governments share the belief that action is needed to respond to the health crisis and overcome the economic and employment crisis."
Repeated Ideas & Cross Information	
“...A total of 117 kilometers for the enjoyment of visitors and the people of Alava, which, he predicted, “will be a complete success."."	"This route is for exploring the Llanada Alavesa plains by bicycle, enjoying nature, and going on a family outing. The Great Route of the Llanada Alavesa is 117 kilometers long."
“...this route is perfect for exploring the Llanada Alavesa by bike, enjoying nature, and family sightseeing. The Great Route of the Llanada Alavesa is 117 kilometers long...."	“...with this route we want to make known the places that Álava and the Llanada region have for: • people who enjoy nature • people who do outdoor sports • family tourism During the 117 kilometers of the Great Route..."
Information only in the Lead	
"• The mission of the workers who make it up will be to disseminate the project and promote the activation of the speakers."	"...This group will present the Euskaraldia initiative to the workers of its Department."

Table 12: Examples of challenging alignments translated from Spanish to English.

Source
Otro de los temas que se ha tratado en la rueda de prensa de hoy ha estado relacionado con el mundo del deporte. Y es que, tal y como se ha podido conocer esta mañana, Vitoria-Gasteiz va a ser en 2019 la sede de la Final Four de Basket. El Consejero de Cultura y Política Lingüística Bingen Zupiria, que se encontraba en la sala presentando un estudio Sociómetro sobre cultura, ha señalado que se trata de una noticia muy importante para la capital alavesa así como para el resto Euskadi.
Target
Bingen Zupiria es el Consejero de Cultura y Política Lingüística. El consejero ha dicho que Vitoria-Gasteiz será en 2019 la sede donde se celebrará la Final Four de la Euroliga de baloncesto.
Llama2 7B Base FT
Bingen Zupiria es el Consejero de Cultura y Política Lingüística del Gobierno Vasco. En la rueda de prensa, ha hablado sobre cultura y ha dado muchos datos sobre el estudio que han hecho sobre la cultura vasca. También ha hablado de otro tema relacionado con la cultura: Vitoria se va a hacer la Final Four de Baloncesto en 2019.
Llama3 8B Instruct FT
El Gobierno Vasco ha presentado un estudio sobre cultura. El estudio se llama Sociómetro y tiene que ver con la cultura en Euskadi. En la presentación del estudio también se habló de deporte, ya que se ha anunciado que Vitoria-Gasteiz albergará la Final Four de Baloncesto en 2019.
Llama3 8B Instruct ZS
Hoy en la rueda de prensa se habló sobre deporte. Se ha anunciado que Vitoria-Gasteiz será la sede de la Final Four de Basket en 2019. El consejero de Cultura y Política Lingüística, Bingen Zupiria, dijo que esto es una noticia muy importante para Vitoria y para Euskadi.
Llama3 8B Instruct FS-Random
La ciudad de Vitoria-Gasteiz va a ser la sede de la Final Four de Baloncesto en 2019. Esto es una noticia muy importante para la ciudad y para Euskadi.
Llama3 8B Instruct FS-BM25
Vitoria-Gasteiz va a ser la sede de la Final Four de Basket en 2019. Esto es una noticia importante para Vitoria-Gasteiz y para Euskadi.
Llama3 8B Instruct FS-Semantic
Hoy se ha hablado de deporte en una rueda de prensa. Se ha anunciado que Vitoria-Gasteiz será la sede de la Final Four de baloncesto en 2019. Esto es una noticia importante para la ciudad y para Euskadi.

Table 13: Example of ER adaptation with the selected models in Spanish.

Source
Azken urtean irakurritako liburuen batez besteko kopurua 6 edo 7 dira. Eta irakurri duten azkenean jartzen badugu arreta ikusten dugu egilearen sexua gizonezkoa dela kasu gehienetan (% 56) eta emakumezkoa % 27tan. Gainontzekoetan gazteek ez dakite egilearen sexua edo egile batzuen artean idatzita dago eta gizonezkoak eta emakumezkoak dira egile horiek.
Target
Azken urtean, zenbat liburu irakurri dituzte? Nork idatziak ziren? • 6 eta 7 liburu artean irakurri dituzte. • Irakurri duten azken liburuagatik galdetuta, erdiak baino gehiagok erantzun du autorea gizonezkoa zela; erdiak baino nabarmen gutxiagok, aldiz, erantzun du autorea emakumezkoa zela.
Latxa2 7B Base FT
Azken urtean irakurritako liburuen batez besteko kopurua 6 edo 7 liburu dira. Irakurri duten azken liburua zein den galdetzen zaienean, gehienek esaten dute gizonezko batek idatzitako liburu bat irakurri dutela.
Latxa3 8B Instruct FT
Gazteek irakurtzen dituzten liburuen inguruan, honako datu hauek aipatu ditu: • Gazteek irakurtzen dituzten liburuen batez besteko kopurua 6 edo 7 liburu da. • Gehien irakurtzen dituzten liburuak gizonek idatzitakoak dira. • Irakurri duten azken liburua emakume batek idatzitakoa da. • Gazteek ez dakite egilea gizonezkoa edo emakumezkoa den.
Latxa 8B Instruct ZS
Azken urtean irakurri ditudan liburuen batez bestekoa 6 edo 7 da. Arreta jartzen badugu, ikusten dugu egilearen sexua gizonezkoa dela kasu gehienetan (% 56) eta emakumezkoa % 27tan. Gainontzekoetan, ez dakit egilearen sexua edo egile batzuen artean idatzita dago eta gizonezkoak eta emakumezkoak dira egile horiek.
Latxa3 8B Instruct FS-Random
Azken urtean, euskal gazteek 6 edo 7 liburu irakurri dituzte batez beste. Liburu horien egileei begiratzen badiegu, ikusten dugu gehienetan gizonezkoak direla (% 56). Emakumezko egileak % 27 dira. Beste kasu batzuetan, gazteek ez dakite egilea gizonezkoa edo emakumezkoa den, edo liburua gizonezko eta emakumezko egile batzuek idatzi dutela.
Latxa3 8B Instruct FS-BM25
Azken urtean, gazteek 6 edo 7 liburu irakurri dituzte batez beste.
Latxa3 8B Instruct FS-Semantic
Azken urtean irakurritako liburuen batez besteko kopurua 6 edo 7 da. Irakurri duten liburuetan arreta jartzen badugu, ikusten dugu egilearen sexua gizonezkoa dela kasu gehienetan (% 56) eta emakumezkoa % 27tan. Gainontzekoetan, gazteek ez dakite egilearen sexua edo egile batzuen artean idatzita dago eta gizonezkoak eta emakumezkoak dira egile horiek.

Table 14: Example of ER adaptation with the selected models in Basque.