Democracy Made Easy: Simplifying Complex Topics to Enable Democratic Participation

Nouran Khallaf[1], Stefan Bott[2], Carlo Eugeni[1], John O'Flaherty[3], Serge Sharoff[1], Horacio Saggion[2]

[1]School of Languages, Cultures and Societies, University of Leeds, United Kingdom
[2]Department of Engineering, Universitat Pompeu Fabra, Spain
[3]The National Microelectronics Applications Centre (MAC) Ltd, Ireland

Abstract

Several groups of people are excluded from democratic deliberation because the language used in this context may be too difficult for them to understand. Our iDEM project aims to reduce existing linguistic barriers in deliberative processes by developing technology to facilitate the translation of complicated text into Easy-to-Read formats that are more suitable for many people. In this paper, we describe classification experiments for detecting different types of difficulties which should be amended in order to make texts easier to understand. We focus on a lexical simplification system that can achieve state-of-the-art results with the use of a free and open-weight large language model for the Romance Languages in our project. Moreover, a sentence segmentation system is introduced to produce text segmentation for long sentences based on training data. Finally, we describe the iDEM mobile app, which will make our technology available as a service for end-users of our target populations.

1 Introduction

Representative democracy is based on delegating policy matters to elected representatives, while the deliberative democratic process aims at involving the stakeholders directly (Bächtiger et al., 2018). Modern democratic institutions aim at a greater focus on stakeholders' involvement. However, this has the requirement of clearer language, which is accessible to the stakeholders, especially in cases where the stakeholders face challenges in understanding, for example, in such cases as people with intellectual disabilities or non-native speakers. The demand for better communication is also reflected in the international treaties, in particular, the Universal Declaration of Human Rights, in its Article 19, affirms everyone's right to seek and receive information.

Moreover, particularly important in this context is the United Nations Convention on the Rights of Persons with Disabilities (CRPD), which includes *accessibility* as one key enabler for a more inclusive society. That is, the ability of any product, service, content, environment, etc., to be used by people with the widest range of abilities (including linguistic and cognitive abilities). The CRPD also considers accessibility as, for example, an enabler for democratic participation rights such as freedom of expression and self-determination. Consequently, a lack of accessibility can be linked to a risk of exclusion for persons who cannot participate equally due to linguistic barriers.

The focus of our paper is on providing an introduction into technologies developed in the context of our project, iDEM[1] [2]: in the area of intersectionality and equality in deliberative and participatory democratic spaces, iDEM aims at making information more accessible and inclusive in the context of democracy and in particular in deliberative and participatory processes. More specifically, in this paper, we will discuss:

1. A tool for assessing sentence-level complexity and predicting appropriate simplification strategies;
2. Applications of these tools to real-world corpora, including the United Nations Parallel Corpus and Europarl;
3. A text simplification pipeline powered by Large Language Models (LLMs), focusing on lexical simplification and Easy-to-Read (E2R) sentence segmentation.

The rest of the paper is structured as follows:

[1]Innovative and Inclusive Democratic Spaces for Deliberation and Participation
[2]https://idemproject.eu/

Section 2 provides an overview of the iDEM project. Section 3 reviews related work on complexity assessment and text simplification. Section 4 details our sentence complexity classifier and simplification approach, including the evaluation results. Section 5 outlines the mobile application in development, while Section 6 discusses current limitations. Finally, Section 7 offers concluding remarks.

2 Project Overview

The iDEM Project in the area of intersectionality and equality in deliberative and participatory democratic spaces aims at making information more accessible and inclusive in the context of democracy and, in particular, in deliberative and participatory processes. In the first phase of the project we have investigated, using a theoretical approach, current marginalization from deliberative processes of diverse underrepresented groups due to language skills in order to understand what are the linguistic barriers which hamper their participation. By working with different associations, iDEM adopts a user-centered approach in use case design and data collection and annotation to ensure maximum impact in the community, thus contributing to making democracy more accessible and inclusive. An innovative iDEM service (e.g., mobile app) is being implemented to host the developed language technologies to support on-demand simplification in Catalan, English, Italian, and Spanish. In the current phase of the project, we are developing the underlying natural language processing technology as well as fine-tuning the use cases to test and evaluate our proposed approach to a more inclusive deliberative democracy. The interested reader is referred to (Saggion et al., 2024b) for an overview of the project.

3 Related Work

3.1 Easy-to-Read

Since the late nineties, many organisations have raised awareness about fundamental information being written in a way that is too difficult to understand for many people. Initiatives to palliate this deficit include "Plain Language" (U.S. Government, 2011) and "Easy-to-read" (Inclusion Europe, 2009). They are two different methods whose overall objective is to make information more accessible. They proposed guidelines for how to write more accessible texts; however, applying them to produce accessible material heavily depends on well-trained human editors and, therefore, considerably limits the production of easy-to-read or plain language texts.

Compared to standard language, easy-to-read language is a simplified version for the sake of readability for specific audiences (Caro, 2017). In this paper, we adopt E2R over Plain Language because its structured guidelines form the foundation of diverse and adaptable translation strategies designed to make information accessible to people with reading difficulties, including people with intellectual disabilities. They have little command of the language and poor literacy.

Empirical research in the field is uncommon (González-Sordé and Matamala, 2024), especially when compared to fields such as automatic text simplification. Although the topic has gained greater scholars' attention in recent years, sometimes research reports on apparently contradictory findings (Fajardo et al., 2014) between guidelines and actual text understanding by target E2R populations; moreover, even guidelines appear to take on different aspects with little agreement between them (Maaß, 2020). With the advances that natural language processing has achieved in recent years, interest in the automatic adaptation of texts to plain language or E2R has intensified (Alarcon et al., 2021; Da Cunha Fanego, 2021; Saggion, 2024).

3.2 Complexity Identification

The first focus of our research within iDEM is on theoretically understanding the factors that contribute to the complexity of a text or the sentences within this text. The guidelines described in the previous section are directed to human editors and often leave much room for interpretation and are hard to operationalise, for example the instruction to avoid difficult words. We are interested in combining theoretical insights with data-driven analysis and classification.

In previous work, computational studies typically overlook insights from translation studies, particularly the various strategies proposed (Vinay and Darbelnet, 1971; Newmark, 1988; Chesterman, 1997; Zabalbeascoa, 2000; Molina and Hurtado Albir, 2002; Gambier, 2006), focusing on the systematic processes involved in translating a source text into a target text across languages. Translation studies provide a complementary approach in examining strategies used in intra-lingual translation, where a source text is translated into a target text in the same language. Eugeni and Gambier (2023) argue that such transfers habitually achieve a complete correspondence between source and target texts. One key task in order to transform sentences into E2R is lexical simplification, i.e., simplifying individual words or short phrases independent of the effect of such simplifications on the overall sentence coherence. For instance, Paetzold and Specia (2016) developed methods that specifically targeted complex word identification (CWI), which detects difficult words and suggests simpler alternatives. These techniques usually ignore how such simpler words would fit the general sentence structure.

Datasets developed to evaluate lexical simplification, e.g., SemEval-2012 Task 1 (Specia et al., 2012), ALEXSIS (Ferrés and Saggion, 2022), TSAR 2022 (Saggion et al., 2022) or MLSP 2024 (Shardlow et al., 2024) have aided a focus on single word-level replacements. Though helpful, these datasets primarily cover single word substitutions in isolation rather than more general context-sensitive simplifications. As a result, simplifications generated with the assistance of these tools sometimes sound unnatural, which needed a post generative model to refine sentence coherence. This issue was also highlighted by Shardlow (2014), who reviewed various lexical simplification approaches and noted that, while effective for readability, they frequently ignore sentence coherence and grammatical correctness.

Corpora for sentence simplification includes ASSET (Alva-Manchego et al., 2020) that provides multiple quality simplifications per sentence. However, ASSET still focuses to some extent on fine-grained lexical or phrase-level modifications and lacks annotations for deeper grammatical or discourse-level modifications. Similarly, WikiLarge (Zhang and Lapata, 2017) provides large parallel sentence pairs for training simplification models but does not explicitly annotate the simplification strategies, making it difficult to study in detail exactly how sentences are simplified. The Simplext corpus (Saggion et al., 2015) provides full document simplifications following E2R guidelines for the Spanish language without indication of transformation type while PorSimples (Aluísio and Gasperin, 2010) provides document simplification in Portuguese covering several operations.

3.3 Text Simplification

Our focus for this paper is on lexical simplification; for an overview of full text simplification approaches and methods, the reader is referred to (Saggion, 2017). Several past approaches to lexical simplification used traditional count-based word-vectors and available dictionaries for modelling word semantics and to select simple word replacements for complex words (Biran et al., 2011; Bott et al., 2012); in later work, word embedding were used, which is learned from huge text collection (Glavaš and Štajner, 2015). More recently, large-scale language models such as BERT and its variations have been applied to predict substitution candidates for complex words. For example, LS-BERT (Qiang et al., 2020) uses the masked language model (MLM) of BERT to predict a set of candidate substitution words and their associated probability. In this context, the MLM predicts substitute words which are ranked for simplicity using: probabilities, a language model, a paraphrase database, word frequency and word semantic similarity with the target word. Very recent work presented in the TSAR 2022 (Saggion et al., 2022) and MLSP 2024 (Shardlow et al., 2024) evaluation frameworks have demonstrated that Large Language Models (LLMs) are in fact the best performing models for the lexical simplification. Techniques such as "prompting" are used to condition the LLMs to produce a simplification or to suggest alternative words. Note, however, that these models underperform when simplifying low-resourced languages. We define 'low-resourced languages' as those with limited digital text corpora (e.g., Catalan vs. English), impacting LLM performance as noted in Section 4.4. Despite advances in lexical simplification (e.g., TSAR 2022, MLSP 2024), key gaps remain: (1) How can simplification strategies be systematically categorised beyond lexical substitution? (2) What taxonomies exist for intra lingual translation, and how do they apply to automation? Section 4.2 addresses these by proposing a strategy taxonomy, testing it on institutional corpora, and leveraging LLMs without prompt engineering—a less-explored approach due to its complexity (Shardlow et al., 2024).

4 Natural Language Processing for Easy-to-Read Translation

4.1 Datasets

We use a range of datasets across different components of our system[3]. The primary dataset used

[3]Where applicable, datasets are available on request from the authors or are publicly accessible through the cited sources.

for complexity assessment and simplification strategy classification comprises 76 parallel texts collected from Scottish care services, UK political manifestos (2024), and Disability Equality Scotland newsletters. These cover diverse topics such as healthcare, environmental policies, disability advocacy, and accessibility. The texts were manually aligned at the sentence level, resulting in 4,166 words in 206 original ("complex") sentences and 3,259 words in 210 simplified counterparts. Despite the reduction in word count, the number of sentences increased slightly, reflecting a key simplification strategy that is splitting longer sentences to improve readability. We additionally use a French dataset of 370 manually aligned sentence pairs. The original texts were retrieved from the Réfugiés.info website and were anonymised to remove any personally identifiable information (PII) (Team, 2025). These parallel sentence pairs provide training data for our simplification strategy classifier (Section 4.2).[4]

For evaluating our system on larger, multilingual corpora, we use the European Parliament (Koehn, 2005) and the United Nations Parallel Corpus (Ziemski et al., 2016). These are publicly available and provide high-quality sentence-aligned translations in English, Spanish, and Italian. We applied our multilingual classifier to these datasets to analyse simplification needs across languages (Section 4.3).

For lexical simplification, we use few-shot prompting on pre-trained Salamandra models with trial data from the MLSP 2024 shared task (Shardlow et al., 2024), covering English, Spanish, Italian, and Catalan (Section 4.4).

Finally, for sentence segmentation in Spanish according to E2R standards, we accessed a private annotated dataset provided by Calleja et al. (2024). This dataset includes 3,826 training, 484 validation, and 1,452 development sentences, each annotated with E2R-compatible cut points (Section 4.6).

4.2 Complexity Assessment

The simplification strategy prediction task aims to determine the types of transformations needed to make a sentence more accessible. Table 1 provides examples of these transformations.

Our taxonomy is informed by Inclusion Europe's guidelines (Inclusion Europe, 2009), intralingual translation practices into E2R (Hansen-Schirra et al., 2020), and a qualitative analysis of our dataset. While previous taxonomies in Translation Studies have offered valuable models for interlingual and diamesic translation, they lack the granularity needed to describe all strategies observed in E2R practice. On the other hand, typologies in Automatic Text Simplification (ATS) are based on corpus analysis (Bott and Saggion, 2014) or on edit operations that mainly deal with adding, deleting, replacing, and moving words Cardon et al. (2022). However, texts translated in E2R language clearly show that professionals in the field apply many more operations that pertain to the field of pragmatics and semiotics, focused on how concepts are distributed and or explained to help the user understand them.

To address this gap, our framework adapts insights from both domains. Based on Inclusion Europe's three levels of simplification—lexical, syntactic, and semantic—we define eight macro-strategies that range along a continuum from additive operations (e.g., *Explanation*) to reductive ones (e.g., *Omission*). These are outlined in Table 2 comprises 8 macro-strategies (excluding transcript since it is a non-simplification operation), 8 strategies, and 30 micro-strategies. For the full set of strategies, see Table 10 (Appendix E).

Cross-linguistic differences in simplification strategies are also relevant. In our multilingual experiments, we observed variations in dominant strategies across English, Spanish, and Italian, which suggests that language-specific features influence how simplification is operationalised. This will be further explored in Section 4.3.

The classifier is built by application of pre-trained transformer-based models (such as multilingual BERT (Devlin et al., 2019)) for multiclass text classification, focusing on the prediction of simplification strategies need to simplify the respective sentences. We employed Stratified 5-fold Cross-Validation for rigorous evaluation and generalisation. We took the average of the validation scores across all the folds to determine the final scores. Early stopping was also employed, wherein the training was halted if the validation loss did not see an improvement for the patience period.

Class imbalance in the data, with certain strategies being underrepresented, was a problem during training. To counter this, we used a *weighted*

[4]English dataset was annotated by a linguist with expertise in translation and text simplification, using the same predefined set of simplification strategy categories described in Appendix B; the French dataset was labelled by the Réfugiés.info editorial team following the same guidelines and category definitions.

In 2018-20 From 2018 to 2020	life expectancy at birth in Scotland was babies born in Scotland were expected to live	76.8 years for males 77 years if they were boys	and 81.0 years for females. and 81 years if they were girls.
`Modulation`	`Explanation`	`Synonymy,Syntax`	`Synonymy,Syntax`

Table 1: Segment alignment for the original (top) and simplified (bottom) sentences

Strategy	Description	Example
Omission	Removing unnecessary rhetorical or diamesic constructs.	"Sir Keir Rodney Starmer KCB KC" → "Starmer"
Compression	Simplifying grammatical/semantic constructs.	"to guide the group" → "to the group"
Syntactic Change	Adjustments between syntactic levels.	"citizens" → "people in Scotland"
Transcript	No changes made to the text.	"I love music"
Transposition	Word class change.	"our aim is" → "we want"
Synonymy	Simplifying technical or abstract words.	"conversation" → "talk"
Modulation	Redistributing information linearly.	"joins in activities... supported by family" → "He joins activities. His family helps."
Explanation	Making hidden content or terms explicit.	"co-design services..." → "co-design means sharing your ideas"
Illocutionary Change	Making implied meaning explicit.	"know your body's library" → "know your body"

Table 2: Simplification strategies required for a sentence, with examples

cross-entropy loss function. Class weights were calculated as the inverse frequency of each class:

$$w_c = \frac{1}{\text{freq}_c} \cdot \frac{N}{2} \tag{1}$$

where w_c is the weight assigned to class c, freq_c is the frequency of class c, and N is the number of samples. This approach ensured that underrepresented classes contributed more to the overall loss, so the model became more capable of predicting the minority classes.

Additionally, gradient clipping was applied during training to stabilise the optimisation. Gradient clipping limits the maximum value of gradients during backpropagation to prevent extremely large updates of model parameters that could destabilise training or lead to divergence. Mathematically, gradient clipping can be expressed as:

$$g_{\text{clipped}} = \min\left(g, \frac{g_{\text{threshold}}}{\|g\|}\right), \tag{2}$$

where g represents the original gradient vector, $g_{\text{threshold}}$ is the clipping threshold, and $\|g\|$ is the norm of the gradient vector. Gradient clipping ensures consistent updates to model parameters, improving training stability.

See the summary of hyper-parameters in Table 8 (Appendix B). The use of medium-sized PLMs (such as multilingual BERT) instead of LLMs helps with the possibility of applying the classifiers to large institutional datasets (such as the entirety of Europarl or the United Nations Corpus), as well as with the possibility of deploying the classifiers to guide the corrections.

We used standard precision, recall, and F1-score metrics (Manning et al., 2008) to evaluate model performance. Given the class imbalance, we report the weighted macro F1-score (Sokolova and Lapalme, 2009), which better reflects the classifier's ability to handle both frequent and rare simplification strategies. The fine-tuned classifier model achieved a weighted macro F1-score of 0.8089, demonstrating its ability to generalize across majority and minority classes. In particular, it outperformed the baseline majority-class strategy, which corresponds to the weighted macro F1-score of 0.096.

The F1 score of the multilingual model (trained on English, tested on French) is 0.6339, thus reflecting the need to improve its ability to generalize across languages. However, given that its errors are balanced, i.e., the model is confused with predicting Synonymy for Explanation and vice versa, see the confusion matrix in Figure 2 (Appendix C). Omission and Compression categories tend to be confused with one another, with Omission commonly predicted as Explanation or Transcript, mirroring the need to enhance the separation between removal and rewriting strategies. Modulation is also commonly confused with Synonymy, mirroring the need to strengthen sentence restructuring cues in training.

Category	English # Sent.	%	Spanish # Sent.	%	Italian # Sent.	%
			Europarl			
Total Sentences	2,005,688	100	1,788,913	100	1,928,874	100
Complex	1,932,492	96.3	1,660,631	92.8	1,868,714	96.8
Omission	59,065	3.1	23	0.001	57	0.003
Syntactic Change	254,483	13.2	11,777	0.7	21,321	1.1
Transposition	13,075	0.7	35,053	2.1	40,633	2.2
Synonymy	1,104,564	57.2	37,259	2.2	81,468	4.4
Modulation	41,802	2.2	724,469	43.6	1,004,438	54.2
Explanation	459,503	23.8	852,050	51.3	702,526	37.9
			UN Corpus			
Total Sentences	10,600,000	100	10,665,709	100		
Complex	9,628,533	96.2	9,987,750	93.6		
Omission	75,217	0.7	62	0.0006		
Syntactic Change	181,228	1.8	503,047	5.0		
Transposition	39,356	0.4	68,878	0.7		
Synonymy	4,587,340	45.0	198,479	1.9		
Modulation	445,095	4.3	5,345,515	53.5		
Explanation	4,878,679	47.7	3,871,769	38.7		

Table 3: Sentence counts and proportions of simplification strategies in institutional datasets

4.3 Experiments with assessing institutional repositories

We experimented with two institutional repositories, which include English, Italian and Spanish, some of the languages of our project, the corpus of the European Parliament (Koehn, 2005) and the United Nations Parallel Corpus (Ziemski et al., 2016). Both resources include high-quality translations, so the content of each sentence is the same. However, we can expect that the three languages differ in their traditions for maintaining linguistic complexity in such formal texts as the parliamentary proceedings. Total sentences row in Table 3 presents the amount of data in each dataset. We used sentence-aligned versions from the respective repositories and applied the multilingual classifiers described in the previous section to make predictions. If the complexity classifier detected the need to simplify the sentence, i.e., it was predicted as "Complex", we estimated the likely strategy needed for this task. As the classification model is limited to the one-label setup, out of several edit operations required for a sentence (see the example in Table 1), our current version of the model predicts the single most likely operation (*Explanation* in this example).

Table 3 shows that the majority of sentences in both datasets and across all the languages considered (English, Spanish, and Italian) require some form of simplification. For English sentences, the most common simplification operations found are (1) lexical substitution (synonymy), primarily through the choice of simpler synonyms, and (2) Explanation which provides more explanation to facilitate reading.

Conversely, for both datasets of Spanish and Italian sentences, the predominant simplification strategy is modulation, with a particular emphasis on sentence restructuring for the purpose of achieving a more linear and straightforward reading experience.

4.4 Simplifying Complex Words

As reported in recent lexical simplification challenges (i.e. TSAR 2022 (Saggion et al., 2022) and MLSP 2024 (Shardlow et al., 2024)), most recent state-of-the-art lexical simplification systems rely on decoder-only autoregressive LLMs like GPT-4 (Enomoto et al., 2024). These systems seem to systematically outperform other systems, like encoder-only language models (e.g. BERT), also because recent developments of LLMs have mostly concentrated on decoder models. Decoders are generally more flexible and have strong zero-shot or few-shot abilities. Commercial closed-weight models like GPT-4, however, carry concerns for the purpose of our project since they lack guarantees of privacy protection and generate costs by using the API. In addition, their closed nature does not usually allow us to fine-tune them.

In preliminary experiments, we found out that the LLMs of the Salamandra family [5](Gonzalez-Agirre et al., 2025) perform very well on European Languages, especially on Romance languages, and within the last group, they especially excel at the performance of Catalan. This can be explained because Salamandra models are part of the Alia initiative (Government of Spain) funded by the Spanish government with a strong focus on languages spoken in Spain. Salamandra models were trained as decoder-only, and they are also provided as instruction-tuned versions. With this, we decided to use a simple few-shot system as our first approach.

Few-shot prediction from a pre-trained model refers to the process where a model that has already been trained on a large dataset (a pre-trained model) is used to make predictions or perform tasks with no or only a few labeled examples for a new task. The *shots* are examples provided in the prompt, as opposed to being used as training data for fine-tuning. Zero-shot prediction does not provide any example. The pre-trained model is typically a decoder-only model, which produces output based on an input prompt that conditions

[5] https://huggingface.co/collections/BSC-LT/salamandra-66fc171485944df79469043a

the output. In essence, few-shot prediction from a pretrained model means leveraging a model's prior knowledge from a large dataset to perform well on a new task or dataset, even with very few labeled examples. As our pre-trained models, we used the 2 billion parameters (2B) and the 7 billion parameter (7B) versions of Salamandra.

We used the following prompt without doing refinement through prompt engineering:

> *Given the context and the specified target word in {LANGUAGE}, answer 10 simpler alternative words. Do not give less than 10 alternative words. Give different words as alternatives. {SHOT_EXAMPLES} Context: {CONTEXT} Target Word: {TARGET} Alternatives Words:*

Here LANGUAGE is a variable which is set according to the language (Catalan, English, Italian, Spanish) in which we want to produce predicted solutions. For few-shot prediction we used examples from the trial section of the MLSP data. The shot examples were selected randomly, but we made sure that unique contexts were selected. An instance of a SHOT_EXAMPLE is given here:

> *´Given the context....* Context: A continue statement will skip the remainder of the block and start at the controlling conditional statement again. Target Word: remainder Alternative Words: rest, restrictive, remaining, remainder, balance

For a 2-shot or 4-shot prompt, 2 or 4 of these different examples would be included in the prompt given to the system. The CONTEXT and TARGET variables have the same form as in the provided shot examples.

As evaluation measures, we used the same as in the MLSP shared task (see Section 3.3). Accuracy (ACC) expresses the percentage of right solutions given out of all given solutions. Here we use Accuracy@1@top1 which is defined as the percentage of instances where the first top-ranked substitute matches the most frequently suggested synonym in the gold data (*top1*). MAP@k (Mean Average Precision) uses a ranked list of generated substitutes, which can either be matched or not matched against the set of the gold-standard substitutes. The first *k* solutions of the ranked list are considered.

The results can be seen in Table 4. We use the same baseline here as was used in the MLSP shared task. It has to be noted that the baseline used there was very strong, since it used zero-shot prompting with the use of the chat-fine-tuned version of Llama-2-70B. This is a version with 70 Billion Parameters and thus ten times larger than the Salamandra-7B model we use here. In fact, many participating systems in the MLSP shared task could not outperform this baseline. In the tables, we mark those results with an asterisk that are higher than this baseline. As a further reference we also list the performance of the different winning systems of the shared task. These winners, however, use GPT-3 for Catalan (Dutilleul et al., 2024) and GPT-4 (Enomoto et al., 2024) for the rest of the languages, and for reasons we describe above, we cannot use them for the iDEM project.

As expected, the 2B version of Salamandra could not outperform the baseline (Table 9 in Appendix D). We attribute this to the fact that this model is too small to produce reliable results in a task that requires quite a large amount of general knowledge about language, such as synonymy and simplicity. The results from this table are still interesting because we want to use fine-tuning on Salamandra-2B in future work. The 7B version of Salamandra, on the other hand, could outperform the baseline nearly systematically in few-shot settings. Interestingly, the difference between 2-shot and 4-shot predictions is not very large. In some cases, the 4-shot predictions perform even worse than 2-shot predictions. Another observation that can be made is that Salamandra mostly excels at the three Romance languages Spanish, Catalan and Italian, while for English, it performs very close to the baseline. In this case, it means that the baseline is higher and harder to beat for English than for the other languages because of the multilingual capabilities of the baseline system or the lack thereof. These observations confirm our assumption that Salamandra is a good choice for the set of languages that we have to treat in iDEM.

4.5 Integration of Complexity Assessment and Lexical Simplification

This section presents ongoing work toward integrating two core modules of our system: complexity assessment (Section 4.2) and lexical simplification (Section 4.4). The classifier first detects complex lexical items in a sentence, and the simplification module then proposes easier alternatives. While the full pipeline has not yet been formally evaluated, we have implemented a proof-of-concept

	0-Shot		2-Shot		4-Shot		MLSP Baseline		MLSP Winner	
	ACC	MAP@3	ACC	MAP@3	ACC	MAP@3	ACC	MAP@3	ACC	MAP@3
English	0.1280	0.1912	0.4017*	0.3868	0.4035*	0.4242*	0.3877	0.4241	0.5245	0.5762
Spanish	0.0286	0.1213	0.3541*	0.5148*	0.3608*	0.3644	0.3254	0.4157	0.4536	0.6763
Catalan	0.0426	0.1390	0.2292*	0.3742*	0.2022*	0.3357*	0.1977	0.3024	0.2719	0.5003
Italian	0.035	0.1419	0.3596*	0.4108*	0.3315*	0.3868*	0.2964	0.3310	0.4762	0.5661

Table 4: Results of Zero and Few Shot Lexical Simplification Performance for a big model (Salamandra-7B). Results are compared to the state of the art as reported in the recent MLSP 2024 lexical simplification shared task. Asterisks (*) indicate the model outperformed the strong baseline of the competition.

Sentence	Easy to Read Segmentation
The way this sentence is cut is easy to read.	The way this sentence is cut is easy to read.
Validar es comprobar si un documento es fácil de comprender.	Validar es comprobar si un documento es fácil de comprender.

Table 5: Examples of segmented sentences in English and Spanish taken from Easy-to-Read guidelines.

to illustrate its feasibility. Table 6 provides multilingual examples where the complexity classifier flags difficult words, which are then simplified by the Salamandra-7B lexical simplifier. For instance, in the English sentence “The reason why hypothalamic lesions affect body fat…,” the words ‘hypothalamic’ and ‘lesions’ are identified as complex and replaced with ‘brain’ and ‘damage,’ respectively—substitutions that significantly enhance readability.

In the context of the iDEM project, this integration is intended for deployment within the mobile application currently under development (see Section 5), where users with cognitive or linguistic barriers can receive real-time support in understanding complex information. Future work will involve formal evaluation, expansion to full sentences, and deeper cross-linguistic adaptation.

4.6 Segmenting Sentences for Easy-to-Read

According to E2R standards (Inclusion Europe, 2009; 153101, 2018), sentences in E2R are recommended to be short and should fit on one line on the printed page (or screen). Since this is not always possible, guidelines recommend cutting the sentence where people would pause when reading out loud. Research on sentence segmentation is somehow related to the prediction of prosodic markers in text-to-speech systems, where syntactic structure and word/token information is used (Fitzpatrick and Bachenko, 1989). Examples of how sentences should be segmented in E2R in English and Spanish are presented in Table 5.

Although datasets for sentence and lexical simplification exist (as reported above), there is a lack of publicly available datasets of E2R segmentation. We have gained private access to a dataset of segmented E2R texts in Spanish (Calleja et al., 2024). This dataset is organized into three files corresponding to train (3,826 sentences), validation (484 sentences), and development (1,452 sentences). Each sentence is explicitly marked to indicate where it should be segmented following E2R standards. We adopt a machine learning approach to sentence segmentation, developing a classifier based on linguistic information and other features such as the position of the token in the sequence (first, second, etc.) or the distance to the previous cut. We process the dataset in order to convert the original sentences into instances for learning. The instances for learning are based on the tokens (words, punctuation, numbers, etc.) in each sentence; our aim is to classify all tokens as cut-point or not. In order to create the learning instances, we linguistically analyze each sentence using a Spanish model from the SpaCy library (Honnibal et al., 2020), which produces information on parts of speech, syntactic dependencies, and named entities. We extract several features including the Parts Of Speech (POS) tag of the token, the case of the token (lower cased, upper cased), whether the token is a punctuation, whether the token is part or a named entity (begin, inside, outside), the position of the token in the sentence, the distance to the previous cut point (or -1), and the distance to the end of the sentence. The learning instances (one per token) are stored in a CSV file for use by a machine learning algorithm. We report results using a Decision Tree algorithm (Steinberg, 2009) due to its simplicity and explanatory power (i.e. set of rules). Other algorithms were less successful on our data. The learning algorithm is an instance from the De-

Lang	Context (Sentence)	Complex word (by CA)	Substitute (by LS)
Eng	The reason why hypothalamic lesions affect body fat and feeding behavior has in fact much to do with leptin signaling.	hypothalamic	brain
		lesions	damage
		affect	influence
Spa	Si este indicador baja de 1, implicaría que la empresa no está en capacidad de cubrir sus obligaciones de corto plazo con los activos líquidos que posee. (If this indicator is below 1, this implies that the enterprise is not in conditions to cover its obligations in the long run with the liquid assets it possesses.)	implicaría	significaría
		indicador	medida
		plazo	tiempo
		activos	bienes
Cat	La formació sosté que "els posicionaments excloents en vers a altres realitats educatives fonamentades amb idees polítiques distorsionen la realitat del model" català. (The formation maintains that "exclusionary positions towards other educational realities based on political ideas distort the reality of the Catalan model".)	sosté	defensa
		posicionaments	posicións
		vers	contra

Table 6: Examples of cases where the Complexity Assessment (CA) system identifies a word that needs simplification and the Lexical Simplification (LS) system simplifies it.

cision Tree implementation provided by the Scikit Learn library[6] (Pedregosa et al., 2011).

Table 7 reports segmentation results for the decision tree classifier and two baselines. The baselines are based on (i) the Parts of Speech (POS) tag, which on training data is the best predictor of the token where the sentence should be segmented, and (ii) on the most common length of the segment. As for the decision tree, two methods are applied: the oracle configuration knows about the *true* previous cuts, while the blind configuration has only access to the *predicted* previous cuts. The difference between oracle and blind configurations are expected. The difference in performance between the decision tree and the baselines is an indication that the features are contributing to the classification performance. Future work should look at analyzing feature contribution and improving the models, and providing segmentation support for Catalan, English and Italian.

Algorithm	F1 (Cut)	F1 (No Cut)	Avg. F1
Decision Tree (Oracle)	0.43	0.89	0.66
Decision Tree (Blind)	0.26	0.91	0.58
POS Tag Baseline	0.17	0.95	0.56
Seg. Length Baseline	0.12	0.91	0.52

Table 7: Segmentation results (based on F1 measure) into Easy to Read (Spanish data). Comparison of a Decision Tree with baselines.

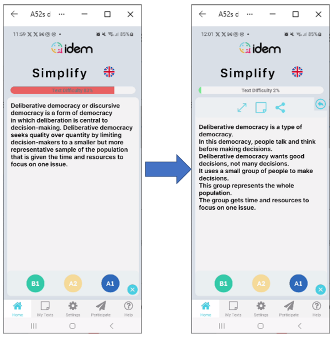
5 Accessing Simplification Technology through the iDEM App

The iDEM project implements and deploys a cloud-based, open-API iDEM platform to deliver text-simplification services, integrating components for complex language detection (Section 4.2) phenomena and adaptation through text simplification (Section 4.4). It supports diverse audiences, languages, and domains, and solutions are made available for deliberative participatory spaces as open-source products. The current version of the app supports iteration via typed text, speech, OCR or PDF. A participation functionality allows the user to check proposals currently being discussed and simplify them for better understanding. For example, the *Decidim* platform (Aragón et al., 2017) can be directly accessed from the app to translate, or simplify active participatory processes. Examples of the APP in action can be seen in Figures 1a and Figures 1b. Note that the current simplification technology supported by the app is not yet the one described in the paper; it still serves as a demonstrator of what it will look like in the coming months.

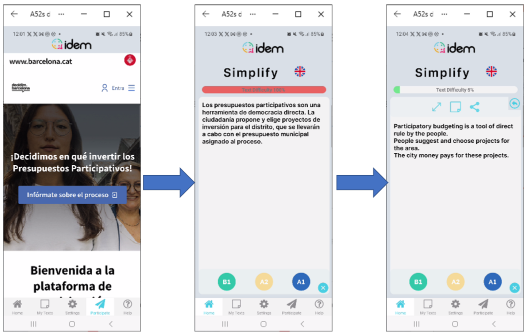
6 Limitations and Ethical Considerations

The studies on Complexity Assessment in section 4.2 and 4.3 argue for an analysis and simplification of a large array of factors, one of which is lexical simplification. We are aware that this is a current limitation, but future versions of the iDEM simplification tools will include a full treatment of sentence simplification. Our current simplification model, although achieving good performance in comparison with a strong baseline, does not do so with respect to the state of the art. This can be explained by our aim to keep models open and accessible to a broader community of stakeholders, i.e. lighter, open models could be afforded by more disadvantaged communities in the spirit of our project. Since our project deals with pro-

[6] https://scikit-learn.org/stable/modules/tree.html



(a) English Wikipedia excerpt on deliberative democracy.



(b) Decidim platform: input in Spanish, output simplified in English.

Figure 1: Examples of cross-lingual text simplification.

viding accessible information to people who need language support, special attention has to be put in the assessment of the underlying models used as backbones for our technology as well as on the data we trained or fine-tuned our models with. An assessment of data quality and ethics has already been carried out (Saggion et al., 2024a). As for the involvement of human subjects in our case studies, we are following strict norms for data protection and ethical principles.

7 Conclusions

First, our intralingual translation-borrowed framework facilitates comparison between source and target texts more easily when the texts are simplified. Second, we proposed a taxonomy of simplification strategies inspired by intralingual translation and E2R principles, consisting of 8 macro-strategies, illustrating the cognitive complexity of intralingual translation. Such challenges underscore current automation tool limitations, as computational analyses illustrate the subtle competencies that are engaged in transcription and alteration strategies.

We applied our classifier to a parallel dataset from institutional sources and observed that Explanation and Modulation were among the most frequently predicted strategies, especially in English texts. While the classifier demonstrates promising results, a limitation of this study is that the observations were not verified through systematic manual analysis but rather were generated automatically. Therefore, further systematic validation and error analysis should be included in future work.

This first study on simplification strategies and complexity assessment underlines the importance to carry out lexical simplification. In our second study we explored lexical simplification using few-shot prompting with open-source LLMs from the Salamandra family. The most important finding is that for Romance Languages LLMs of the Salamandra family show very promising results because, in contrast to most other LLMs, they are trained on much larger amounts of data in these languages. It was important to note, that our system can obtain results similar to those obtained with commerical closed-weights LLMs without having the same disadvantage of those of being only available over APIs that generate costs and being hosted on servers for which we cannot control the protection of sensitive data. The last point is potentially important especially in a project which handles data of vulnerable populations. Further on, commercial models usually do not allow fine-tuning, since their weights are not public. Even though our current experiments do not outperform the state of the art reached by GPT-3 and GPT-4 based models, we have not experimented with fine-tuning of Salamandra models and we are confident that such an approach will give room for improvement.

Finally, we presented a proof-of-concept integration of complexity assessment and lexical simplification, demonstrating its potential for real-world applications such as accessible mobile interfaces. While formal evaluation of the full pipeline remains future work, our preliminary results suggest that strategy-aware simplification can meaningfully support inclusive democratic participation.

Acknowledgments

This document is part of a project that has received funding from the European Union's Horizon Europe research and innovation program under Grant Agreement No. 101132431 (iDEM Project). The views and opinions expressed in this document are solely those of the author(s) and do not necessarily reflect the views of the European Union. Neither the European Union nor the granting authority can be held responsible for them. The University of Leeds (UOL) was funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (Grant Agreement No. 10103529).

References

UNE 153101. 2018. Lectura Fácil. Pautas y recomendaciones para la elaboración de documentos.

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

Sandra Aluísio and Caroline Gasperin. 2010. Fostering digital inclusion and accessibility: The PorSimples project for simplification of Portuguese texts. In *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, pages 46–53, Los Angeles, California. Association for Computational Linguistics.

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

Pablo Aragón, Andreas Kaltenbrunner, Antonio Calleja-López, Andrés Pereira, Arnau Monterde, Xabier E. Barandiaran, and Vicenç Gómez. 2017. Deliberative platform design: The case study of the online discussions in decidim barcelona. In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part II*, pages 277–287, Cham. Springer International Publishing.

André Bächtiger, John S Dryzek, Jane Mansbridge, and Mark E Warren. 2018. *The Oxford handbook of deliberative democracy*. Oxford University Press.

Or Biran, Samuel Brody, and Noémie Elhadad. 2011. Putting it Simply: A Context-Aware Approach to Lexical Simplification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, ACL, pages 496–501, Portland, Oregon, USA.

Stefan Bott, Luz Rello, Biljana Drndarevic, and Horacio Saggion. 2012. Can spanish be simpler? lexsis: Lexical simplification for spanish. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 357–374, Mumbai, India.

Stefan Bott and Horacio Saggion. 2014. Text simplification resources for spanish. *Lang. Resour. Evaluation*, 48(1):93–120.

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.

Rémi Cardon, Adrien Bibal, Rodrigo Wilkens, David Alfter, Magali Norré, Adeline Müller, Patrick Watrin, and Thomas François. 2022. Linguistic corpus annotation for automatic text simplification evaluation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1842–1866, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Rocío Bernabé Caro. 2017. Propuesta metodológica para el desarrollo de la lectura fácil según el diseño centrado en el usuario. *Revista Española de Discapacidad (REDIS)*, 5(2):19–51.

Andrew Chesterman. 1997. *Memes of Translation: The Spread of Ideas in Translation Theory*. John Benjamins Publishing Company, Amsterdam.

Iria Da Cunha Fanego. 2021. El sistema ARTEXT CLARO: un auxiliar per a la redacció de textos administratius en llenguatge planer. *Terminàlia*, 2(24):78–79.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Benjamin Dutilleul, Mathis Debaillon, and Sandeep Mathias. 2024. Isep_presidency_university at mlsp 2024 shared task: Using gpt-3.5 to generate substitutes for lexical simplification. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, pages 605–609.

Taisei Enomoto, Hwichan Kim, Tosho Hirasawa, Yoshinari Nagai, Ayako Sato, Kyotaro Nakajima, and Mamoru Komachi. 2024. TMU-HIT at MLSP 2024: How Well Can GPT-4 Tackle Multilingual Lexical Simplification? In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, pages 590–598.

Carlo Eugeni and Yves Gambier. 2023. La traduction intralinguistique: les défis de la diamésie. *Editura Politehnica, Timisoara*.

Inmaculada Fajardo, Vicenta Ávila, Antonio Ferrer, Gema Tavares, Marcos Gómez, and Ana Hernández.

2014. Easy-to-read texts for students with intellectual disability: linguistic factors affecting comprehension. *Journal of applied research in intellectual disabilities*, 27(3):212–225. Publisher: Wiley Online Library.

Daniel Ferrés and Horacio Saggion. 2022. Alexsis: A dataset for lexical simplification in spanish. In *Proceedings of TSAR-2022*, pages 39–49.

Eileen Fitzpatrick and Joan Bachenko. 1989. Parsing for prosody: what a text-to-speech system needs from syntax. *[1989] Proceedings. The Annual AI Systems in Government Conference*, pages 188–194.

Yves Gambier. 2006. La traduction audiovisuelle : une traduction sélective. In Jorma Tommola and Yves Gambier, editors, *Translation and Interpreting – Training and Research*, pages 21–37. University of Turku, Department of English Translation Studies, Turku.

Goran Glavaš and Sanja Štajner. 2015. Simplifying Lexical Simplification: Do We Need Simplified Corpora? In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 63–68, Beijing, China. Association for Computational Linguistics.

Aitor Gonzalez-Agirre, Marc Pàmies, Joan Llop, Irene Baucells, Severino Da Dalt, Daniel Tamayo, José Javier Saiz, Ferran Espuña, Jaume Prats, Javier Aula-Blasco, Mario Mina, Adrián Rubio, Alexander Shvets, Anna Sallés, Iñaki Lacunza, Iñigo Pikabea, Jorge Palomar, Júlia Falcão, Lucía Tormo, Luis Vasquez-Reina, Montserrat Marimon, Valle Ruíz-Fernández, and Marta Villegas. 2025. Salamandra technical report. *Preprint*, arXiv:2502.08489.

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. Publisher: Springer.

Government of Spain. ALIA: The public ai infrastructure in spanish and co-official languages. `https://alia.gob.es/eng/`. Accessed: March 2025.

Silvia Hansen-Schirra, Walter Bisang, Arne Nagels, Silke Gutermuth, Julia Fuchs, Liv Borghardt, Silvana Deilen, Anne-Kathrin Gros, Laura Schiffl, and Johanna Sommer. 2020. Intralingual translation into easy language – or how to reduce cognitive processing costs. In Silvia Hansen-Schirra and Christiane Maaß, editors, *Easy Language Research: Text and User Perspectives, Easy – Plain – Accessible*, volume Volume 2, pages 197–226. Frank & Timme, Berlin.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spacy: Industrial-strength natural language processing in python.

Inclusion Europe. 2009. *Information for All: European Guidelines for the Production of Easy-to-Read Information*. Inclusion Europe, Brussels, Belgium. Available online at `http://www.easy-to-read.eu`.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the 10th Machine Translation Summit*, pages 79–86, Phuket, Thailand. Accessed: April 2025.

Christiane Maaß. 2020. *Easy language - plain language - easy language plus: balancing comprehensibility and acceptability*. Number Vol. 3 in Easy - plain - accessible. Frank & Timme, Verlag für wissenschaftliche Literatur, Berlin.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press.

Lucía Molina and Amparo Hurtado Albir. 2002. Translating techniques revisited: A dynamic and functionalist approach. *Meta*, 47(4):498–512.

Peter Newmark. 1988. *A textbook of translation*, volume 66. Prentice hall New York.

Gustavo Henrique Paetzold and Lucia Specia. 2016. Semeval 2016 task 11: Complex word identification. In *Proceedings of SemEval-2016*, pages 560–569.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake VanderPlas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830.

Jipeng Qiang, Yun Li, Yi Zhu, Yunhao Yuan, and Xindong Wu. 2020. LSBert: Lexical Simplification Based on BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pages 3064–3076.

Horacio Saggion. 2017. *Automatic Text Simplification*, volume 10 of *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers.

Horacio Saggion. 2024. Artificial intelligence and natural language processing for easy-to-read texts. *Revista de Llengua i Dret*, (82):84–103.

Horacio Saggion, Stefan Bott, Sandra Szasz, Nelson Pérez, Saúl Calderón, and Martín Solís. 2024a. Lexical complexity prediction and lexical simplification for Catalan and Spanish: Resource creation, quality assessment, and ethical considerations. In *Proceedings of the Third Workshop on Text Simplification, Accessibility and Readability (TSAR 2024)*, pages 82–94, Miami, Florida, USA. Association for Computational Linguistics.

Horacio Saggion, John O'Flaherty, Thomas Blanchet, Serge Sharoff, Silvia Sanfilippo, Lian Muñoz, Martin Gollegger, Almudena Rascón, José L. Martí, Sandra

Szasz, Stefan Bott, and Volkan Sayman. 2024b. Making democratic deliberation and participation more accessible: The idem project. In *SEPLN (Projects and Demonstrations)*, volume 3729 of *CEUR Workshop Proceedings*, pages 71–76. CEUR-WS.org.

Horacio Saggion, Sanja Štajner, Stefan Bott, Simon Mille, Luz Rello, and Biljana Drndarevic. 2015. Making It Simplext: Implementation and Evaluation of a Text Simplification System for Spanish. *ACM Transactions on Accessible Computing*, 6(4):1–36.

Horacio Saggion, Sanja Štajner, Daniel Ferrés, Kim Cheng Sheang, Matthew Shardlow, Kai North, and Marcos Zampieri. 2022. Findings of the TSAR-2022 shared task on multilingual lexical simplification. In *Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022)*, pages 271–283, Abu Dhabi, United Arab Emirates (Virtual). Association for Computational Linguistics.

Matthew Shardlow. 2014. A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications*, 4:58–70.

Matthew Shardlow, Fernando Alva-Manchego, Riza Batista-Navarro, Stefan Bott, Saul Calderon Ramirez, Rémi Cardon, Thomas François, Akio Hayakawa, Andrea Horbach, Anna Hülsing, Yusuke Ide, Joseph Marvin Imperial, Adam Nohejl, Kai North, Laura Occhipinti, Nelson Peréz Rojas, Nishat Raihan, Tharindu Ranasinghe, Martin Solis Salazar, Sanja Štajner, Marcos Zampieri, and Horacio Saggion. 2024. The BEA 2024 shared task on the multilingual lexical simplification pipeline. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, pages 571–589, Mexico City, Mexico. Association for Computational Linguistics.

Marina Sokolova and Guy Lapalme. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.

Lucia Specia, Shashi Narayan, and Carolina Gasperin. 2012. Semeval-2012 task 1: English lexical simplification. In *Proceedings of SemEval-2012*, pages 347–355.

Dan Steinberg. 2009. Cart: Classification and regression trees. Technical report.

Réfugiés.info Team. 2025. Réfugiés.info genai for public good hackathon submission. Accessed April 2025.

U.S. Government. 2011. Federal plain language guidelines. Accessed: March 2025.

Jean-Paul Vinay and Jean Darbelnet. 1971. *Stylistique comparée du français et de l'anglais*. Didier. Translated into English as *Comparative Stylistics of French and English*, 1995.

Patrick Zabalbeascoa. 2000. From techniques to types of solutions. In Allison Beeby, Doris Ensinger, and Marisa Presas, editors, *Investigating Translation*, pages 117–127. John Benjamins, Amsterdam and Philadelphia.

Wei Zhang and Mirella Lapata. 2017. Sentence simplification with deep reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594.

Michał Ziemski, Marcin Junczys-Dowmunt, and Bruno Pouliquen. 2016. The United Nations parallel corpus v1.0. In *Proc. LREC*, Portorož, Slovenia.

A Plain Language Summary

This paper describes how to make difficult text easier to read. It is part of the iDEM project, which aims to help more people participate in democratic processes. Some people find official or complex documents hard to read. Because of that, they might not be able to vote or take part in important discussions.

The iDEM project studies ways to make public information easier to read. By removing or replacing hard words, adding helpful explanations, and splitting text into shorter sentences. To do this, it uses computer programs. One program can detect which words or sentences are difficult. Another program can suggest simpler words to replace the difficult ones. There is also another program to split long sentences into shorter ones.

The project uses freely available language models. These models are trained to understand many languages, such as English, Spanish, Italian, and Catalan.

B Classifier Configuration

Parameter	**Value**
Pre-trained Model	`bert-base-multilingual`
Max Sequence Length	512 tokens
Tokenisation	Pre-trained tokenizer
Loss Function	Weighted Cross-Entropy Loss
Class Weights	Inverse frequency of labels
Gradient Clipping Threshold	1.0
Learning Rate	5×10^{-6}
Batch Size	8
Weight Decay	0.01
Number of Epochs	Up to 20 (early stopping)
Cross-Validation	Stratified 5-Fold
Early Stopping Patience	3 epochs
GPU	NVIDIA Tesla T4 (Google Colab). Occasionally V100 (our HPC cluster).

Table 8: Hyperparameters and Training Configuration for experiments in Section 4.2

C Confusion Matrix

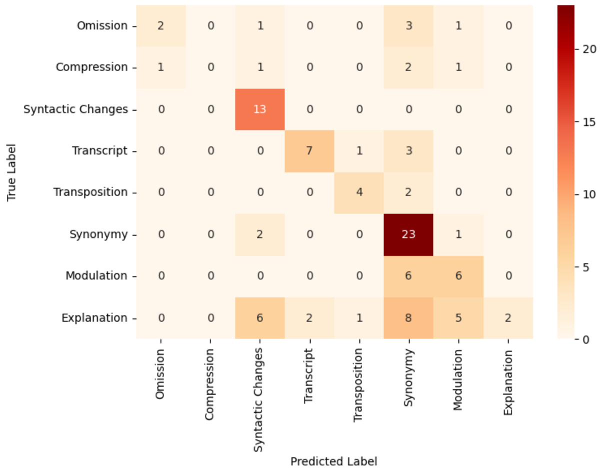


Figure 2: Confusion Matrix of the multilingual model.

D Zero and Few Shot Lexical Simplification Performance Smaller Model

	0-Shot		2-Shot		4-Shot		MLSP Baseline		MLSP Winner	
	ACC	MAP@3	ACC	MAP@3	ACC	MAP@3	ACC	MAP@3	ACC	MAP@3
English	0.2877	0.261	0.3017	0.2601	0.3315	0.2765	0.3877	0.4241	0.5245	0.5762
Spanish	0.2192	0.2608	0.2596	0.3011	0.2681	0.3356	0.3254	0.4157	0.4536	0.6763
Catalan	0.1438	0.1817	0.1438	0.1961	0.1348	0.1710	0.1977	0.3024	0.2719	0.5003
Italian	0.228	0.1983	0.2736	0.2251	0.1684	0.2006	0.2964	0.3310	0.4762	0.5661

Table 9: Results of Zero and Few Shot Lexical Simplification Performance for a small model (Salamandra-2B). Results are compared to the state of the art, as reported in the recent MLSP 2024 lexical simplification shared task.

E Classification Strategies

Strategy	**MacroStrategy**	**Explanation and Examples**	**Total**
WorExp	Explanation	Explanation of a word. e.g. “co-design services...” → “co-design services with people who use or work in them...”	
ExpExp	Explanation	Explanation of an expression. e.g. “Accessible Transport...” → “Accessible travel means making buses, trains, ferries and taxis easier to use...”	
HidGra	Explanation	Making hidden grammar explicit. e.g. “this is the music I love” → “This is the music that I love”	4
HidCon	Explanation	Making hidden content explicit. e.g. “COVID-19” → “Covid pandemic”	
ModInf	Modulation	Splitting sentence based on number of ideas. e.g. “He joins in community activities. . . ” → “He likes to take part. . . He gets support. . . ”	

(continued on next page)

(continued from previous page)

Strategy	**MacroStrategy**	**Explanation and Examples**	**Total**
ModLin	Modulation	Redistribution of sentence components: - **ModWord**: "...collaboration and information sharing..." → "...working together and sharing information..." - **ModGrou**: "Accessible Museums is a topic..." → "Our members think it is important to talk about Accessible Museums" - **ModClau**: "To improve community health... the Government works..." → "The Government works... to improve..."	2
PraSyn	Synonymy	Pragmatic synonyms: - **PraProp**: UN → United Nations, Nutella → chocolate cream - **PraCont**: "Sir Keir Starmer" → "the new Prime Minister"	
SemSyn	Synonymy	Semantic synonyms: - **SemStere**: ponder → think - **SemHype**: lecturers → teachers - **SemHypo**: flora → trees and flowers	3
GraSyn	Synonymy	Grammatical synonyms: - **GraPron**: "you don't see it" → "you don't see the mistake" - **GraTens**: "we have been doing" → "we have done" - **GraPass**: passive → active - **GraNega**: "not an obstacle" → "facilitated"	
TraNou	Transposition	Noun transposition. e.g. "our aim" → "we want"	
TraVer	Transposition	Verb transposition. e.g. "listening to music" → "music"	
TrAdje	Transposition	Adjective transposition. e.g. "mountainous landscapes" → "mountains"	4
TrAdve	Transposition	Adverb transposition. e.g. "behaving happily" → "was happy"	
Transcript	Transcript	A sentence is left unchanged.	
SynW2G/S/C	Syntactic Change	Word to group/clause/sentence	
SynG2W/C/S	Syntactic Change	Group to word/clause/sentence	12
SynC2W/G/S	Syntactic Change	Clause to word/group/sentence	
SynS2W/G/C	Syntactic Change	Sentence to word/group/clause	
Illocutionary Change	Illocutionary Change	Making implied meaning explicit.	1
GraSim	Compression	Grammatical simplification. e.g. "so as to" → "to"	
SemSim	Compression	Semantic simplification. e.g. condensing explanations	2
OmiEle	Omission	Omission of elements: - **OmiSubj**: "Sir Keir Rodney Starmer..." → "Starmer is..." - **OmiVerb**, **OmiComp**, **OmiClau**, **OmiSent** (e.g. full sentence removed)	

(continued on next page)

(continued from previous page)

Strategy	MacroStrategy	Explanation and Examples	Total
OmiDia	Omission	Omission of discourse elements: - **OmiFil**: “you know” → removed - **OmiRef**: “I was tight… right when...” → “I was right when...” - **OmiRhe**: “wasn’t I?” → removed	2
		Total	**30**

Table 10: Macro-strategies, Strategies, Micro-strategies, and Examples with Annotated Totals