How Artificial Intelligence can help in the Easy-to-Read Adaptation of Numerical Expressions in Spanish

Mari Carmen Suárez-Figueroa and **Alejandro Muñoz-Navarro** and **Isam Diab**
Ontoloy Engineering Group (OEG)
Universidad Politécnica de Madrid
Madrid, Spain
mcsuarez@fi.upm.es, alejandro.mnavarro@upm.es, isam.diab@upm.es

Abstract

Numerical expressions, specifically the use of fractions and percentages in texts, may pose a difficulty in the reading comprehension process for different groups of the population, including persons with cognitive disabilities. As an element that facilitates reading comprehension, the Easy-to-Read (E2R) methodology, created to achieve the cognitive accessibility, recommends avoiding the use of fractions and percentages. If it is necessary to include them, their equivalence or explanation should be described. In order to help people who have difficulties in reading comprehension when they have to deal with fractions and percentages, we have developed an initial method for adapting numerical expressions in an automatic way in Spanish. This method is based on (a) Artificial Intelligence (AI) methods and techniques and (b) the E2R guidelines and recommendations. In addition, the method has been implemented as a web application. With the goal of having our research in the context of the responsible AI, we followed the human-centred design approach called participatory design. In this regard, we involved people with cognitive disabilities in order to (a) reinforce the adaptations provided by E2R experts and included in our method, and (b) evaluate our application to automatically adapt numerical expressions following an E2R approach. Moreover, this method can be integrated into institutional procedures, such as those of university administrations and public organisations, to enhance the accessibility of official documents and educational materials.

1 Introduction

Numerical expressions are defined as expressions that denote quantities, optionally accompanied by a numerical modifier, such as *more than a quarter* or *almost 97 %*, where *more than* and *almost* take the role of numerical modifiers (Bautista et al., 2017). Different groups of the population, including people with cognitive or intellectual disabilities, experience some difficulties regarding the reading comprehension process of such numerical expressions. For such a reason, the Easy-to-Read (E2R) methodology (Inclusion Europe, 2009; Nomura et al., 2010; AENOR, 2018) recommends avoiding the use of fractions and percentages in text materials. The goal of this methodology is to present clear and easily understood content by providing a collection of guidelines concerning both the content of texts and their design and layout. This methodology was created with the aim of improving the cognitive accessibility of those groups of the population who present reading comprehension difficulties.

When a particular document needs to be adapted to Easy-to-Read, the E2R methodology is applied in a manual fashion following three key activities: E2R analysis, E2R adaptation and E2R validation (AENOR, 2018). This manual process is labour-intensive and costly, and it would benefit from having technological support. In this context, our research is focused on applying different Artificial Intelligence (AI) methods and techniques[1] to (semi)-automatically perform both the analysis and the adaptation of documents to obtain easy-to-read versions of original documents. Furthermore, it should be mentioned that our general intention is to conduct research and develop applications in the context of the responsible AI (Akata et al., 2020). Specifically, we understand responsible AI as inclusiveness and explainability. In this paper, we present our approach for including people with cognitive disabilities in our research; while explainability is out of the scope of this paper.

In this paper, we present an initial proposal for

[1] We are investigating symbolic (such as knowledge representation) and subsymbolic (such as machine learning) approaches in AI as well as the combination of both.

Proceedings of the 1st Workshop on Artificial Intelligence and Easy and Plain Language in Institutional Contexts (AI & EL/PL), pages 14–24, June 23, 2025

adapting numerical expressions in an automatic way in texts written in Spanish, in particular expressions that contain fractions and percentages. This adaptation proposal follows the Easy-to-Read (E2R) methodology and is aligned with the preferences expressed by E2R experts in a previous study (Suárez-Figueroa et al., 2022), in which they identified the adaptation of numerical expressions as one of the guidelines they would like to see automated. Based on this feedback, we developed the support tool described in that study.

There are other studies that performed a quite similar process, which is the automatic simplification[2] of numerical expressions in various languages such as English (Bautista et al., 2011; Power and Williams, 2011), German (Suter et al., 2016) and Spanish (Bautista et al., 2013; Bautista and Saggion, 2014b,a; Bautista et al., 2017) using different methods and techniques that are explained in Section 2. Nevertheless, such attempts do not involve target groups with cognitive disabilities in the development of the system. Our approach therefore focusses on covering this gap by involving both E2R experts and people with cognitive disabilities in our work.

Additionally, our development is designed to be used in institutional contexts where accessibility is a priority. For example, the Language Centre[3] of the Universidad Politécnica de Madrid could integrate our system to enhance the accessibility of its learning materials, ensuring that numerical expressions do not become a barrier for students with cognitive disabilities or non-native speakers. Likewise, institutional bodies such as the Rectorate of the Universidad Politécnica de Madrid could leverage this system to adapt official documents, making them more accessible to all members of the university community, including faculty, staff, and students who may face difficulties with complex numerical representations. By applying our approach in such institutional environments, we aim to contribute to fostering inclusiveness and accessibility in academic and administrative communications.

The rest of the paper is organised as follows: Section 2 is devoted to the state-of-the-art on (a) the difficulties that numerical expressions raise regarding cognitive accessibility, and (b) the automatic approaches for identifying and adapting this type of structures in different languages. In Section 3 we present our proposal for performing an automatic E2R adaptation of fractions and percentages. Section 4 summarises the participatory design approach we followed to include people with cognitive disabilities in our AI-based development; this section also shows the results we obtained from this inclusive strategy. Section 5 describes our web application for automatically adapting numerical expressions and shows the results obtained in a preliminar user-based evaluation. Finally, we present some conclusions and future work.

2 State of the Art

Since our work focusses on developing an initial automatic adaptation of numerical expressions (specially fractions and percentages) in text written in Spanish following the E2R guidelines, in this section we address (a) the problematic these structures pose, and (b) the automatic approaches that have been developed to date for the identification and transformation of numerical expressions in different languages.

2.1 Numerical Expressions and Cognitive Accessibility

Disciplines such as experimental psychology and cognitive neuropsychology have dealt with the study of number processing and calculation over the last decades, since mathematical reasoning helps people develop, as it comes from a basic human need to communicate and describe things like quantities and measurements (Piaget and Inhelder, 1969). However, there are groups of people who present some difficulties when reading numerical expressions, i.e. expressions that denote quantities, optionally accompanied by a numerical modifier.

A specific difficulty that involves learning or understanding numeracy in general and numerical expressions in particular is dyscalculia, which includes difficulties in understanding numbers, manipulating, learning math facts, and a number of other symptoms related to counting money, understanding prices or remembering dates (Landerl et al., 2004; Butterworth, 2010). Such a difficulty (also referred to as a cognitive disability) affects the daily life of people with reading comprehension impairments, since everything around us contains numerical information (e.g. daily news or public information). Moreover, it has been evidenced in

[2]Text adaptation always aims to transform texts to meet the needs of a specific audience, while text simplification tends to reduce the complexity of texts and does not always take the final user into account (Saggion, 2022).

[3]https://www.lenguas.upm.es/

some studies (Rello et al., 2013) that the presence of numerical information in a text impacts negatively on its readability and understandability for people with dyslexia.

For these reasons, technological aids can greatly benefit the adaptation of numerical expressions into E2R versions to facilitate cognitive accessibility.

2.2 Automatic Approaches for Adapting Numerical Expressions

In the context of text simplification and E2R adaptation of texts, few works have addressed the automatic adaptation of numerical expressions in different languages. Power and Williams (Power and Williams, 2011) studied how authors present numerical information in English news articles, focusing on variations in mathematical forms (such as fractions and percentages) and in the level of precision used to express the same quantity. They developed a rule-based system to adapt original proportions and evaluated its effectiveness by comparing the model's predictions with the values suggested by survey participants.

Also in English, Bautista and colleagues (Bautista et al., 2011) studied preferences for rounding numerical expressions to common values, as well as different simplification strategies depending on the original proportion. The system they developed was designed specifically for English and was not intended for any particular group of readers. The authors conducted a survey in which experts in numeracy were asked to simplify a range of proportion expressions with three different readerships in mind. The responses were consistent with their intuitions about how common values are considered simpler and how the value of the original expression influences the chosen simplification.

With respect to languages other than English, in the system developed by Suter and colleagues (Suter et al., 2016) for adapting German texts, numbers written as words and special characters are replaced by digits and appropriate word substitutions using manually created dictionaries.

As regards Spanish (the language we are dealing with in our research), Bautista and colleagues have been working on different approaches (Bautista et al., 2013; Bautista and Saggion, 2014b,a; Bautista et al., 2017) to simplify numerical expressions using parallel corpora of original and manually simplified texts. The numerical simplification was implemented by a rule-based system that included a numerical simplification prototype and a syntactic simplification module to preserve simplicity and meaning.

However, none of the aforementioned efforts has been based on responsible AI, as they have not included groups of people with cognitive disabilities in the development of numerical expression simplification systems. Such a gap is indeed an open issue in the text simplification task literature, since such systems are designed without considering the user, which can lead to underestimating the reader's capabilities (Saggion, 2018). For this reason, in this paper we focus on the adaptation of numerical expressions in Spanish, including both E2R experts and people with cognitive disabilities in our research and development processes.

3 Initial Method for an E2R Adaptation of Fractions and Percentages

The final aim of the proposed method is (a) to detect fractions and percentages in texts written in standard Spanish, and (b) to replace such structures by the most appropriate paraphrasing formula with the goal of being E2R compliant. This method is based both on symbolic AI (e.g. production rules and syntactic patterns) [4] and subsymbolic AI (e.g. machine learning-based Natural Language Processing) methods and techniques (Norvig and Russell, 2021).

This initial method consists of the following high-level activities: (1) Natural Language Processing (NLP), which includes a cleanup of the text using regular expressions and a tokenization step, (2) Fractions and Percentages Identification, and (3) Fractions and Percentages Adaptation. Figure 1 shows the low-level steps of our initial method.

The first activity in our proposed method is text preprocessing. The original text is prepared by separating the paragraphs, looking for line breaks, replacing the words *por ciento* ('per cent') with the symbol %, and establishing a space between the number and the symbol[5] to facilitate the tokenization task (e.g. *10 %*). By the same token, double spaces are removed, spaces are included before and

[4] Using symbolic AI makes it relatively straightforward to provide explanations about what has happened and why. Since symbolic systems are based on explicit rules and representations, they allow for transparent reasoning processes. This means that each step taken by the system can be traced and understood, making it easier to explain both the outcomes and the underlying logic that led to them.

[5] Decision based on the rules for writing percentages presented by the Spanish linguistic foundation Fundéu (https://www.fundeu.es/recomendacion/porcentajes-claves-de-redaccion/).

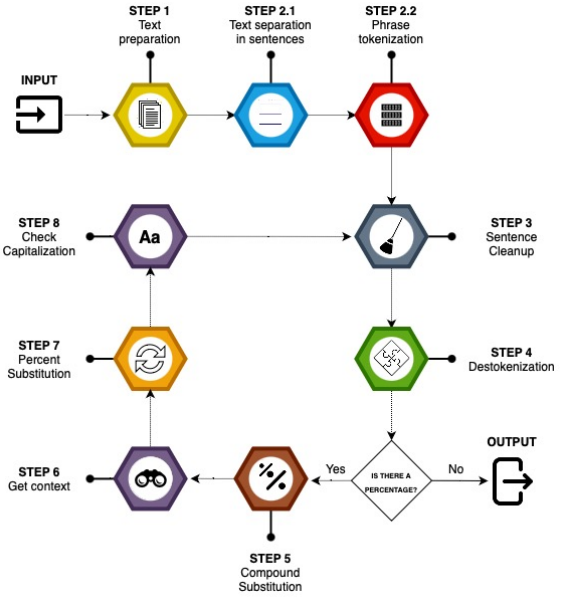

Figure 1: Method Workflow.

after hyphens (e.g. *Madrid-Málaga* to *Madrid - Málaga*) and contractions in Spanish are extended (*al* to *a el* and *del* to *de el*). Figure 1 shows this first activity in Steps 1, 2.1 and 2.2. In addition, a clean-up task is also performed to remove unusual symbols such as parentheses (Step 3 in Figure 1). We decided to use the solution proposed by Drndarevic and colleagues (Drndarevic et al., 2013), thus eliminating both the parentheses and their content; although we are aware of the risk of this action with respect to information loss.

On the other hand, since fractions and percentages are just different ways of showing the same value, we decided to replace fractions whose numerator is less than the denominator with a percentage (e.g. *1/2* is replaced by *50 %*). This action is performed in order to use only one single approach for both situations (fractions and percentages); that is, we generalise the way of treating those numerical structures. After the text preprocessing activity, the identification of fractions and percentages is simplified to detokenize the sentence and look for the symbol % (Step 4 in Figure 1).

The activity of adapting fractions and percentages is based on what is mentioned in the E2R methodology (García Muñoz, 2012; AENOR, 2018), that is, the most appropriate way to adapt fractions or percentages is to use paraphrasing words that preserve the same meaning as the numerical structure. In order to select the most suitable set of words we contacted E2R experts to discuss with them the best options without adding complexity to the text and keeping the original meaning. The result of the discussion with the experts is summarised in Table 1 and Table 2[6]. Such tables show that the percentages were organised by ranges and context, and for each percentage range a set of words is suggested. In addition, during this process, we identified two different structures in sentences with percentages: (a) simple structures, which contain just one percentage (e.g. *un 20 %*) (equivalent to 'just 20 %'); and compound structures, which contain a list of percentages (e.g. *Entre un 10 y un 20 %* ('Between 10 and 20 %'), *Subió desde un 20 a un 50 %* ('It went up from 20 to 50 %'). One way to reduce the complexity of the latter case is to remove the word *entre* or *desde* and replace the two numbers by the mean or the subtraction (e.g. *Un 15 %*, *Subió un 30 %*) (Step 5 in Figure 1). Once substitution is performed, the sentences present a simple structure that can be adapted as a simple percentage (Step 6 in Figure 1).

Table 1: Adaptation proposals (written in Spanish) organised by different established ranges.

Ranges (%)	Adaptation Proposal
0	No + nada / Ninguno / Ninguna
1−9	Muy poco / Muy poca / Muy pocos / Muy pocas
10−24	Un poco / Poca / Pocos / Pocas
25	La minoría
26−40	1 de cada 3
41−49	Casi la mitad
50	La mitad
51−74	Más de la mitad
75	La mayoría
76−90	Mucho / Mucha / Muchos / Muchas
91−99	Casi todo / Casi toda / Casi todos / Casi todas / Casi muy / Casi el total
100	Todo / Toda / Todos / Todas / Muy / El total
101−154	Más del total
155−254	El doble
255−354	El triple
+355	N veces el total

In the following subsections, we explain the details of the proposed E2R suggestions for adapting

[6]Adaptation proposals are written in Spanish. English versions of those proposals are available at: https://zenodo.org/records/15213107.

Table 2: Adaptation proposals (written in Spanish) organised by different established ranges for cases related to units of measurement (e.g. Kg, L, among others).

Ranges (%)	Adaptation Proposal
0	No + nada
1−24	Casi nada
25−49	Menos de la mitad de 1
50	La mitad de 1
51−74	Más de la mitad de 1
75−90	Casi 1
100	1

fractions and percentages in each of the ranges presented in Table 1 and Table 2.

3.1 General Proposal

This proposal is based on the ranges presented in Table 1 to adapt phrases regardless of their context.

Concerning the first case referred to the 0 % interval, it is worth mentioning that in Spanish, double negations occur, that is, a particular scheme of negation in which two negative elements appear (e.g. *no + nada* or *ninguno/ninguna*[7] *('no + nothing or none))*. Double negation does not change the negative meaning of the sentence, but is not always needed. If the verb takes precedence over the percentage, it is necessary to use the adverb *no* in preceding it; otherwise it is not necessary to be denied. In addition, in some cases we have observed that the sentence contained the word *con* ('with') referring to the percentage. In this case it is replaced by *sin* ('without'). To carry out this substitution there are two different paraphrasing suggestions: *nada* or *ninguno/a*. In cases where the percentage is followed by a word in the masculine gender (i.e. *de los*) we use *ninguno*, and for feminine cases (i.e. *de las*) we use *ninguna*. In any other case, we use the word *nada* ('nothing').

For the intervals 1-9 %, 10-24 %, and 76-90 %, our adaptation follows a similar method to the previous scenario. Typically, based on the specific range, we use terms such as *muy poco/muy poca/muy pocos/muy pocas* ('somewhat') for the interval 1-9 %; *un poco/poca/pocos/pocas* ('a little'/'few'), for the interval 10-24 %; and *mucho/mucha/muchos/muchas* ('many') for the interval 76-90 % in the paraphrased adaptations. When the percentage is followed by the preposition *de* ('of'), we examine the subsequent token. If this token is one of *el/la/los/las* ('the'), the percentage is directly substituted by a word matching its number and gender (e.g. *Un 10 % de los libros* ('10 % of the books') to *Pocos de los libros* ('Few of the books'). Otherwise, we use a natural language processor to determine the token's number and gender, and then perform the appropriate percentage substitution. Furthermore, we omit the preposition *de* when the word that follows is not a determiner (e.g. *Un 10 % de estudiantes* ('10 % of the students') to *Pocos estudiantes* ('Few students')), with the exception of masculine singular terms (e.g. *Un 10 % de azúcar* ('10 % of sugar') to *Un poco de azúcar* ('A bit of sugar')). Additionally, when we use the terms *un poca/pocos/pocas* in the substitution, if a preposition precedes the percentage, we replace *poca* by *poco* and introduce a determiner (*un/unos/unas* ('a/an/some/a little/a few')) agreeing in gender and number with the rest of elements of the sentence. For instance, *Con un 10 % de sal* ('With 10 % of salt') is adapted as *Con un poco de sal* ('With a little salt').

As for the ranges from 25 % to 75 % we replace the percentage by the adaptation proposal shown in Table 1. However, in those cases where our new paraphrasing words contain *casi* ('almost') or *más de* ('more than'), we ignore these same words from our original sentence in order not to repeat them (e.g. *Casi un 49 % de los participantes* ('Almost 49 % of the participants') to *Casi la mitad de los participantes* ('Almost half of the participants')).

Additionally, for the range from 26 % to 40 %, we also change the noun to its plural form and the adjectives and verbs referring to the percentage to its singular form (e.g. *El 25 % de los alumnos ha aprobado el examen* ('25 % of the students passed the exam') to *La minoría de los estudiantes ha aprobado el examen* ('The minority of students has passed')). For the rest of the cases, we only replace the adjectives and verbs with their singular feminine form (e.g. *Un 50 % de los estudiantes han sido muy estudiosos* ('50 % of the students have been very studious') to *La mitad de los estudiantes ha sido muy estudiosa* ('Half of the students have been very studious')).

For the 91-100 % range, the procedure remains consistent with the previous intervals. By default, we opt for *casi todo* ('most of it') or *todo* ('all'), depending on the exact percentage. However, if the percentage is succeeded by the preposition *de*, we employ the Part of Speech (PoS) tags provided by spaCy of the subsequent words, simi-

[7]Note that in Spanish we use the slash symbol (/) to indicate gender and number variations of the same word.

lar to the approach for other ranges (e.g. *casi todo/a/os/as* or *todo/a/os/as*). If there is no determiner (*el/la/los/las*) after the preposition (*de*), we add the determiner that agrees in gender and number with the rest of the elements of the sentence (e.g. From *Un 95 % de estudiantes* ('95 % of students') to *Casi todos los estudiantes* ('Almost all students')). Furthermore, if a preposition other than *de* follows the percentage, we revert to the default method (e.g. *Un 95 % en agosto* ('95 % in August') to *Casi todo en agosto* ('Almost all in August')). In the presence of an adjective succeeding the percentage, the structures *casi muy* ('almost totally') or *muy* ('very') replace the percentage (e.g. *Un 95 % rebajado* ('95 % reduced') to *Casi muy rebajado* ('Almost totally discounted')). When a verb precedes the percentage and a punctuation mark follows, we use *casi el total* ('almost the total') or *el total* ('the total') (e.g. *Subió un 95 %* ('It went up by 95 %') to *Subió casi el total* ('It went up almost the total')); otherwise, we use *casi muy* or *muy*. In all other scenarios, the default approach is applied.

For percentages exceeding 100 %, we implement a rounding strategy. If the value is close to 100, we substitute it with *más del total* ('more than the total'). If it approaches 200, we opt for *el doble* ('double'); and for values nearing 300, *el triple* ('triple") is used. For all other cases, the percentage is replaced by the structure *N veces el total* ('N times the total'), where N represents the nearest rounded number divided by 100.

To conclude, it could be possible that we added new words at the beginning of the adapted sentences, and because of these changes, we have to check the capital letters of our text (Step 8 in Figure 1).

3.2 Specific Proposal for Percentages with Units of Measurement

In instances where the percentage is 0 %, we follow the *nada* ('nothing') word adaptation using the double negative *no* + *nada* when a verb precedes the preposition. Additionally, we omit the preposition and the unit of measure (i.e. Kg, L, etc.), as in *El paquete pesa un 0 % de kilo* ('The package weighs 0 % of a kilo'), adapted as *El paquete no pesa nada* ('The package weighs nothing').

For percentages within the range of 1 to 24 %, we use *casi nada* ('almost nothing'), also eliminating the preposition and the unit. For example, *El paquete pesa un 5 % de kilo* ('The package weighs 5 % of a kilo') is adapted as *El paquete pesa casi nada* ('The package weighs almost nothing').

As for the range 25-49 % we use the structure *menos de la mitad de 1* ('less than half of 1'), also removing the preposition before the unit of measurement, if present, and transforming the unit of measurement to singular, e.g. *El paquete pesa 1/4 de kilo* ('The package weighs 1/4 kilo'), is adapted as *El paquete pesa menos de la mitad de 1 kilo* ('The package weighs less than half of 1 kilo').

In the case of 50 % we use *la mitad de 1*, following the same method as above. For instance, *El paquete pesa 1/2 kilo* ('The package weighs 1/2 kilo') is adapted as *El paquete pesa la mitad de 1 kilo* ('The package weighs half of 1 kilo').

For the percentages within the range from 51 to 74 %, we use *más de la mitad de 1* ('more than half of 1'), following the same algorithm as in the previous cases, as we observe in *Usamos 60 % litros de agua* ('We use 60 % litres of water'), adapted as *Usamos más de la mitad de 1 litro de agua* ('We use more than half of 1 litre of water').

With respect to the percentages within 75-99 %, we use *casi 1* ('almost 1') following the same algorithm as in the previous cases. For example, *Usamos 3/4 litros de agua* ('We use 3/4 litres of water') is adapted as *Usamos casi 1 litro de agua* ('We used almost 1 litre of water').

Finally, in instances where the percentages are exactly 100 %, we use *1*, considering the same algorithm as in the previous cases. As an illustration, *Usamos 100 % litros de agua* ('We use 100 % litres of water') is adapted as *Usamos 1 litro de agua* ('We use 1 litre of water').

4 Inclusive AI: Involving People with Cognitive Disabilities

As mentioned in Section 1, our intention is to develop applications in the context of the responsible AI. In this paper we focus on the inclusiveness dimension by means of involving people with cognitive disabilities in the development team. Specifically, their involvement was performed through a human-centred design (Trewin et al., 2019). There are three potential approaches for integrating people with cognitive disabilities in the development processes (Trewin et al., 2019): Inclusive Design, Participatory Design, and Value-Sensitive Design. In our research we would like to design for and with people with cognitive disabilities, so we decided to use the inclusive design approach. As a

first attempt, we used a survey method that allows us to gather feedback about which different ways to express fractions and percentages are closer to E2R structures.

In order to reinforce the collection of linguistic structures for adapting fractions and percentages given by the E2R experts (see Section 3), we decided to use a participatory design approach for gathering feedback about the easier way to express the aforementioned numerical structures from people with cognitive disabilities. For this purpose we conducted a 20-minutes on-line anonymous survey[8]. In this survey[9] we requested (a) opinions of people with cognitive impairments on the use of different ways to express fractions and percentages and (b) several demographic data. We recruited participants by emailing autonomic federations and associations of people with cognitive disabilities in Spain in February 2022.

After analysing the survey responses[10], findings indicate that, overall, participants consider simpler those sentences in which the typical ways of expressing percentages have been adapted, with an E2R approach in mind, by using synonym formulas. In fact, these data confirm the different options proposed by E2R experts (see Table 1).

In more specific detail, data reveal that participants' preferences can be classified into the following scenarios considering the adaptation proposals for each range posed in Table 1:

Scenario A. In this scenario, sentences in which the range of the numerical expression (both percentages and fractions) is between 1 and 49 are considered. In this case, the synonymous adapted option *un(os) poco(s)* ('(a) few') is the most chosen option among the participants in opposition to the most typical ways to represent numerical structures whose meaning is a percentage or a fraction (that is, *30%*, *twenty-five percent*, and *one fifth*).

Scenario B. Likewise, in this scenario sentences in which the numerical expression ranges from 51 to 100 are treated. As a result, participants preferred the adapted version *más de la mitad* ('more than a half') against the rest of non-adapted options (e.g. *60%*, *seventy-five percent*, and *three quarters*).

Scenario C. Finally, this scenario addresses numerical expressions relating to the middle range. Participants similarly chose the synonymous adapted formulas *mitad* and *medio/a* (both have the meaning of 'half').

5 Web Application for Adapting Numerical Expressions

As a proof of concept, we have developed a web application to detect fractions and percentages in texts written in Spanish, and adapt them following the most appropriate E2R translation guidelines. This application is based on the E2R method described in Section 3 and on the feedback gathered about the easier way to express numerical expressions described in Section 4. This application requires a sentence written in Spanish as an input and provides a simpler version of the original sentence as an output. Our application was designed with two options for text input: (a) manual writing or (b) file upload. Furthermore, following the E2R methodology, we adapted the interface of our web application: we limited the number of characters per line (60 characters) and we aligned the text to the left; we used a font size of 14 points or higher, with a line spacing of 1.5 greater than the default; we used font in black with a white background for easy reading; and the different navigation buttons have a yellow background and dark blue lettering to draw the user's attention away from the other elements on the same page. Figure 2 shows a screenshot of the web application in which an example of a text, written in Spanish, with six sentences containing fractions or percentages, is used as input text and the easy-to-read version is the output text.

Our web application has been implemented in Python 3.10, using the Django framework (version 4.0.1). We used the NLTK library[11] for phrase-level tokenization and the spaCy library[12] for word-level tokenization. It is worth mentioning that in Spanish, since words have different forms depending on gender and number, we used spaCy PoS tags to replace the percentage to preserve the consistency of the whole sentence.

Regarding the evaluation carried out with the application, it is worth mentioning that the functional evaluation (Section 5.1) focusses on testing the functionalities and an initial user-based evaluation (Section 5.2) centres on knowing whether the

[8]The questionnaire is implemented as a Google Form and it is available at: https://zenodo.org/records/15213107

[9]Survey design and demographic information about participants are available at: https://zenodo.org/records/15213107.

[10]Responses are available at: https://zenodo.org/records/15213107.

[11]https://www.nltk.org/

[12]https://spacy.io/

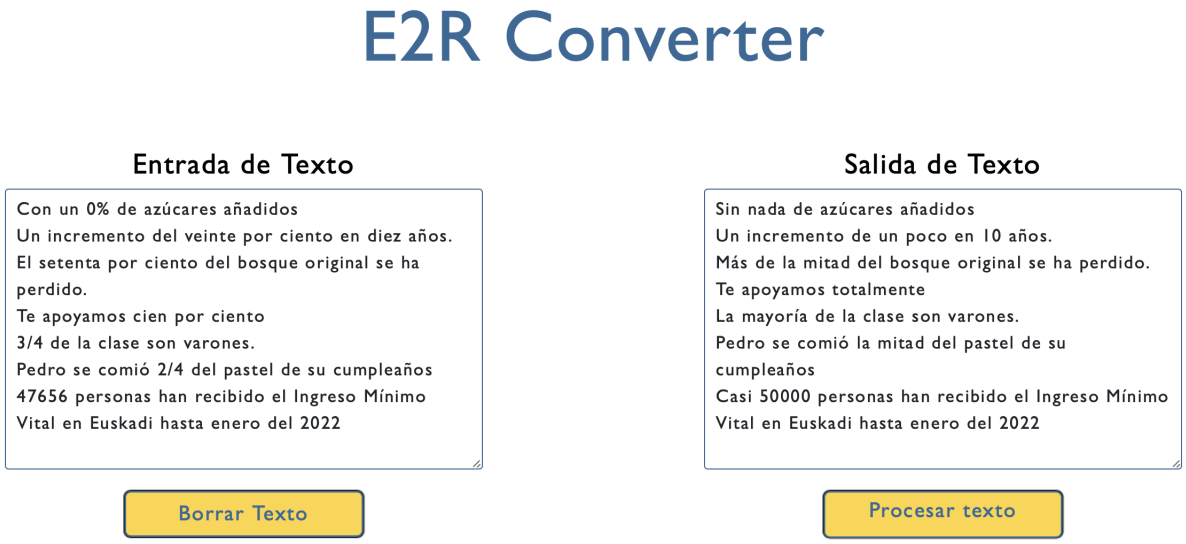

Figure 2: Screenshot of the proof of concept.

E2R adaptations provided by our application are comprehensible for people with cognitive disabilities.

5.1 Functional Evaluation

We tested the identification and adaptation functionalities of our proof of concept with a collection of 360 texts written in Spanish[13]: 180 sentences extracted from a Spanish corpus[14] by searching for the word sequences *por ciento* ('per cent'), *dos quintos* ('two fifths') or *un tercio* ('one third'), and 180 phrases that are quotes attributed to famous people such as Albert Einstein or Leonardo da Vinci among many others, which do not need any substitution. The results have been manually classified into true positives (TP) or true negatives (TN) based on accurate identification and, in the case of TP, on accurate adaptation. In contrast, they were labelled false negatives (FN) if the adaptation was incorrect or false positives (FP) if the adaptation was unwarranted and the original sentence was altered. To measure the effectiveness of our web application, we have tracked two of the metrics used to evaluate classification systems. On the one hand, the so-called sensitivity, which in our context represents the probability of correctly identifying a sentence that needs to be adapted. On the other hand, the metric of specificity, which represents the probability of correctly identifying a sentence that does not need adaptation. Currently, our application has a sensitivity of 91.81% compared to a specificity of 95.23 %. Thus, we could say that it is more specific than sensitive. This is the situation that we are aiming for, as our goal is to avoid false positives.

5.2 Preliminary User-Based Evaluation

This evaluation was carried out using an on-line questionnaire. The questionnaire[15] was divided into two main parts: (1) one part with questions related to participants' demographics, background and experience in E2R validation; and (2) the other part that includes 15 single-answer multiple choice questions to capture participants' opinions about how easy is the adaptation provided by our web application. The format of these 15 questions is always the same. Each question has 2 sentences (one is an original sentence and the other one is the sentence adapted by our service). After reading with calm each question, participants should choose one option among the following ones: (a) Sentence 1 is the one you understand best; (b) Sentence 2 is the one you understand best; (c) I understand both sentences well; and (d) I do not understand either of the two sentences. The questionnaire was validated by an E2R expert.

21 participants responded to the questionnaire, 11 males and 10 females, all from Madrid. 5 of them had a high level of reading comprehension, 13 a medium level, 1 a medium-high level and 2 a low level[16]. Their distribution by age is 4, 12 and 5

[13]The collection of texts is available at: https://zenodo.org/records/15213107.
[14]https://www.wordandphrase.info/span/
[15]The questionnaire is implemented as a Google Form and it is available at: https://zenodo.org/records/15213107.
[16]This information was provided by the support professionals of the organisations.

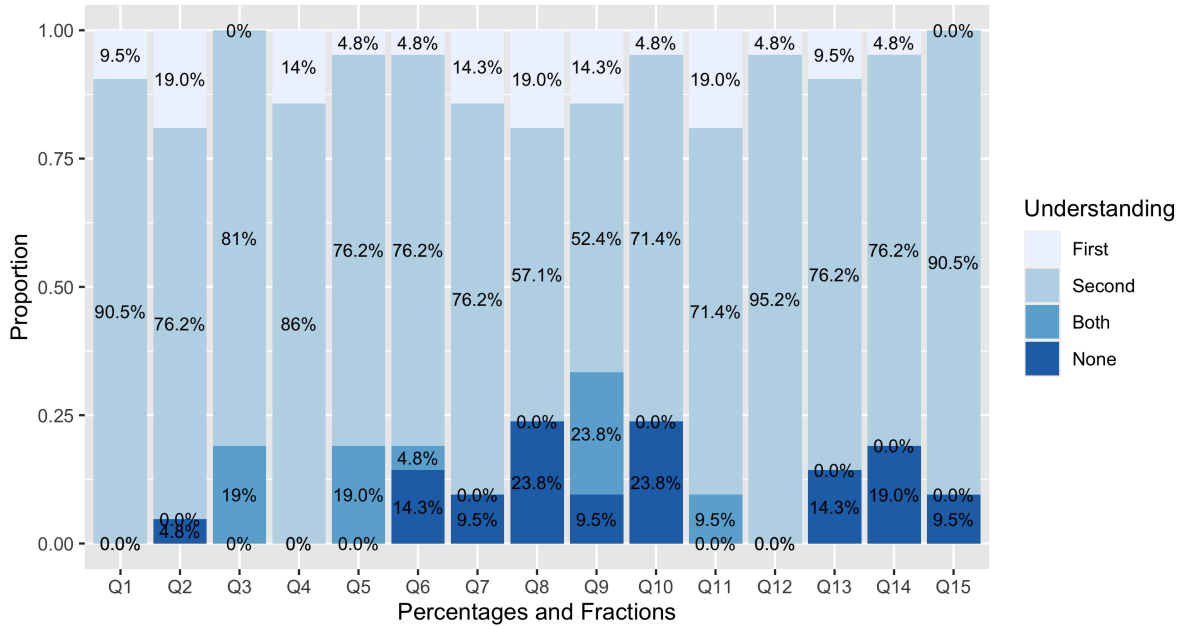

Figure 3: Summary of percentages in the user-based evaluation.

participants in the ranges 18-30, 31-45 and 46-60 respectively. Regarding the impairments, 17 had an intellectual disability, two people had intellectual plus physical disability, one other had intellectual plus mental disability and the last one declared to have a rare disease. As for their occupation, 10 were E2R validators, 7 were users of occupational centres, and the remaining were two retired people, a kit man, and an unemployed.

Figure 3 shows the percentages of responses for each available option and question when assessing the usefulness of the wording our web application gives with respect to percentages and fractions in the text. *First* refers to the original sentence and *Second* refers to the adaptation made by our application. *Both* indicates the participant understands well both sentences and *None* is chosen by a participant who understands neither of them. Through the 15 questions, the proportion of times the second sentence was chosen as the preferred option by the participants is overwhelming. Although there were different levels of understanding throughout the questions, even for the most difficult ones (the ones showing non-zero percentages in the option *None*) the simplification made by our application outnumbers the rest.

6 Conclusions and Future Work

This paper presents an initial method, based on AI, for detecting fractions and percentages in texts written in Spanish, and adapting such numerical structures into easy-to-read versions. A crucial task in the development of this method has been the selection of the most appropriate E2R paraphrasing formulas for such numerical structures. To perform this task we follow a mixed approach: on the one hand, we contacted E2R experts in order to discuss with them the most appropriate adaptation of fractions and percentages; on the other hand, we complement the proposals obtained by the experts with feedback gathered from people with cognitive disabilities. Feedback has been obtained by applying an inclusive design approach. The involvement of people with cognitive disabilities was materialised by participating in one on-line survey. In such a survey, participants were asked about their preferences with respect to the simplicity of parapharsing formulas to express fractions and percentages. Data gathered in this survey reinforced the set of synonym formulas used in our method for adapting fractions and percentages to an E2R version.

This straightforward declarative method has been implemented in a simple-to-use web application; in which the design of the user interface has been developed based on the E2R methodology. Elements such as the number of characters per line, text alignment, font size, line spacing, and colour contrast, among others, have been taken into account. This web application has initially been evaluated by people with cognitive disabilities. This user-based evaluation has been performed using

online questionnaires. This is a first approximation to a deeper evaluation of the web application, since the sample size was not large, but prospects are optimistic. Thus, currently, we could say that adaptations provided by our application seem to be easier than original texts including fractions and percentages written in a standard way.

In addition to its relevance for individual users, our proposal may also be of great value in institutional environments where accessibility is a fundamental requirement. University administrations, government agencies and public institutions frequently produce documents that include numerical expressions, which can pose comprehension problems for some readers. By integrating this method into institutional procedures, organisations such as the Rectorate of the Universidad Politécnica de Madrid or the Language Centre could ensure that official communications, policies and teaching materials are more accessible.

As further research, we are going to analyse in more depth the data gathered in our inclusive co-design process for selecting the most appropriate E2R adaptations for fractions and percentages. In addition, we have planned to design an evaluation activity involving both E2R experts and people with cognitive disabilities. Our plan here is to have a larger sample size of participants. Finally, we would like to explore different ways to explain both the process and the outcomes in our method and application. In this regard, our aim is to cover the explainability dimension in the context of responsible AI.

Acknowledgments

This research has been carried out in the context of the project "Artificial Intelligence for Easy Reading - Cognitive Accessibility (AI&LF)" (Reference: APOYO-JOVENES-21-8AV1UF-119-OEUU22). This research Project has been funded by the Comunidad de Madrid through the call Research Grants for Young Investigators from Universidad Politécnica de Madrid. In addition, the research presented in this paper has been partially financed by Asociación Inserta Innovación (part of Grupo Social Once) through Prosvasi Ciencia y Tecnología Para La Inclusión, A.I.E., within the project ACCESSJOBS. We would like to thank Plena Inclusión España for its help in organising the participatory design involving people with cognitive disabilities, as well as the Federations of Organisations of People with Intellectual or Developmental Disabilities in Madrid, Comunidad Valenciana, and Andalucía for their participation in the study. We would like to thank Isa Cano and María José Sánchez for their help in organising the user-based evaluation of our web application. In addition, we really appreciate the collaboration provided by (a) ACCEDES (Entornos y Servicios Accesibles SL.) and its cognitive accessibility validation team, made up of people with intellectual disabilities, from the "Así Mejor" Program of workshops and activities of the Tres Cantos City Council (Madrid) and (b) the users of the COFOIL "cuarentainueve" of the Association Somos Diferencia (AMP). Finally, we would like to express thanks to Arminda Moreno for her help in the analysis of the data gathered in the user-based evaluation.

References

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

Zeynep Akata, Dan Balliet, Maarten Rijke, Frank Dignum, Virginia Dignum, Guszti Eiben, Antske Fokkens, Davide Grossi, Koen Hindriks, Holger Hoos, Hayley Hung, Catholijn Jonker, Christof Monz, Mark Neerincx, Frans Oliehoek, Henry Prakken, Stefan Schlobach, Linda Gaag, Frank Harmelen, and Max Welling. 2020. A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, 53:18–28.

Susana Bautista, Raquel Hervás, Pablo Gervás, and Javier Rojo. 2017. An approach to treat numerical information in the text simplification process. *Universal Access in Information Society*, 16(1):85–102.

Susana Bautista, Raquel Hervás, Pablo Gervás, Richard Power, and Sandra Williams. 2011. How to make numerical information accessible: Experimental identification of simplification strategies. In *INTERACT 2011. Lecture Notes in Computer Science*, pages 57–64.

Susana Bautista, Raquel Hervás, Pablo Gervás, Richard Power, and Sandra Williams. 2013. A system for the simplification of numerical expressions at different levels of understandability. In *Natural Language Processing for Improving Textual Accessibility (NLP4ITA 2013)*, pages 10–19.

Susana Bautista and Horacio Saggion. 2014a. Can numerical expressions be simpler? implementation and demostration of a numerical simplification system for Spanish. In *Proceedings of the Ninth International*

Conference on Language Resources and Evaluation (LREC'14), pages 956–962, Reykjavik, Iceland. European Language Resources Association (ELRA).

Susana Bautista and Horacio Saggion. 2014b. Making numerical information more accessible: The implementation of a numerical expression simplification system for spanish. *ITL - International Journal of Applied Linguistics*, 165:299–323.

Brian Butterworth. 2010. Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, 14(12):534–541. Special Issue: Space, Time and Number.

Biljana Drndarevic, Sanja Štajner, Stefan Bott, Susana Smith Bautista, and Horacio Saggion. 2013. Automatic text simplification in spanish: A comparative evaluation of complementing modules. In *Conference on Intelligent Text Processing and Computational Linguistics*, pages 488–500.

Óscar García Muñoz. 2012. *Lectura fácil: Métodos de redacción y evaluación*. Real Patronato sobre Discapacidad.

Inclusion Europe. 2009. *Information for All. European standards for making information easy to read and understand*. Inclusion Europe.

Karin Landerl, Anna Bevan, and Brian Butterworth. 2004. Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old students. *Cognition*, 93(2):99–125.

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

Peter Norvig and Stuart J. Russell. 2021. *Artificial Intelligence: A modern approach (4th Edition)*. Pearson Global Editions.

Jean Piaget and Barbel Inhelder. 1969. *Psicología del niño*. Ediciones Morata S.L.

Richard Power and Sandra Williams. 2011. Generating Numerical Approximations. *Computational Linguistics*, 38(1):113–134.

Luz Rello, Susana Bautista, Ricardo Baeza-Yates, Pablo Gervás, Raquel Hervás, and Horacio Saggion. 2013. One half or 50%? an eye-tracking study of number representation readability. In *IFIP Conference on Human-Computer Interaction*, volume 8120, pages 229–245.

Horacio Saggion. 2018. Text Simplification. In *The Oxford Handbook of Computational Linguistics*. Oxford University Press.

Horacio Saggion. 2022. 1114text simplification. In *The Oxford Handbook of Computational Linguistics*. Oxford University Press.

Julia Suter, Sarah Ebling, and Martin Volk. 2016. Rule-based automatic text simplification for german. In *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*, pages 279–287.

Mari Carmen Suárez-Figueroa, Isam Diab, Edna Ruckhaus, and Isabel Cano. 2022. First steps in the development of a support application for easy-to-read adaptation. *Universal Access in the Information Society*, 23(1):365–377.

Shari Trewin, Sara Basson, Michael Muller, Stacy Branham, Jutta Treviranus, Daniel Gruen, Daniel Hebert, Natalia Lyckowski, and Erich Manser. 2019. Considerations for ai fairness for people with disabilities. *AI Matters*, 5(3):40–63.