First Attempt at an Automatic Adaptation of Explanatory Structures in Spanish to Easy-to-Read

Isam Diab[1], Mari Carmen Suárez-Figueroa[1]

[1]Ontology Engineering Group (OEG), Universidad Politécnica de Madrid (UPM)

Abstract

Explanatory structures in the form of incises (e.g. nominal appositions and adjective clauses) can break the argumentative line of a sentence and lose the focus of the reader's attention. Thus, these structures are considered complex for different groups of the population who present reading comprehension difficulties, including people with cognitive disabilities. The Easy-to-Read (E2R) Methodology was created to provide clear and easily understood contents to people with reading comprehension problems. This methodology recommends avoiding the use of explanations between commas and avoiding the use of appositions that interrupt the natural rhythm of reading. To help people with difficulties in reading comprehension, we have developed a pair of initial Artificial Intelligence (AI)-based methods for adapting in an automatic way explanatory structures in Spanish to E2R. The evaluation of the methods involved unit tests and the calculation of the sentence similarity between the original and the adapted sentences.

Keywords

Easy-to-Read (E2R), Cognitive Accessibility, Automatic Translation, Text Adaptation

1. Introduction

Equal opportunities and universal access to information are fundamental rights that every person should benefit[1]. However, certain groups of society, particularly those with cognitive or intellectual disabilities, present some difficulties related to reading comprehension processes. Therefore, prioritising the so-called cognitive accessibility becomes essential for promoting active participation in diverse social domains, such as politics, education, employment, and culture. For such a reason, a methodology called Easy-to-Read (E2R) [1, 2, 3, 4] was created. The main goal of this methodology is to present clear and easily understood content by providing a set of guidelines on the content and the design and layout of written materials, as, for instance, to use short and simple sentences, to avoid the use of long words, or to divide ideas into paragraphs. This adaptation process is iterative and involves three key activities: analysis, adaptation and validation [4]. Nevertheless, the E2R methodology is currently implemented manually, which is costly and time-consuming, so it would benefit from having a technological support. In this context, our research line is focused on applying different Artificial Intelligence (AI) methods and techniques[2] to automatically perform the analysis and the adaptation of Spanish documents to obtain easy-to-read versions. In particular, this paper concentrates on two of the E2R guidelines that influence the composition of the text [3, 4]: (a) to avoid explanations between commas; and (b) to avoid the use of appositions that interrupt the natural rhythm of reading.

Several studies [5, 6, 7, 8, 9, 10] have shown that this type of explanatory structures present a difficulty in the reading comprehension process, since they break the argumentative line and lead to missing information in the process of understanding the text. In this way, the adaptation of explanatory structures in the form of incises (i.e. nominal appositions and adjective clauses) to easy-to-read versions has a positive impact for people with reading comprehension difficulties.

Within the scope of technological support for addressing E2R guidelines and recommendations in Spanish texts, it is worth mentioning (a) Easy-to-Read Advisor [11], FACILE [12], Comp4Text [13], E2R-Helper [14] and ATECA[3] for an E2R analysis of documents; and (b) Simplext [15], LexSIS [16], DysWebxia [17], EASIER [18], FACILE [12], ATECA[4] and Simple.Text[5] for creating simpler versions of original documents.

However, none of the aforementioned works specifically targets the identification and adaptation[6] of explanatory structures in the form of incises into simpler or easy-to-read forms.

Motivated by the aim of bridging this gap and enhancing the reading comprehension process, our research work focuses on automatically identifying and adapting explanatory structures in the form of nominal appositions and adjective clauses, since they impact linguistic aspects such as sentence length and sentence complexity. Therefore, we propose two methods based on symbolic AI[7] to adapt explanatory structures that are not compliant with the E2R Methodology. We also implemented two proofs-of-concept based on these methods. It is worth mentioning that we have opted to use the term 'adaptation' consistently throughout the paper, as it aligns with the established terminology in E2R disciplines. This adaptation, achieved through our methods, can be viewed as a type of intralinguistic automatic translation tailored specifically for rendering sentences into an E2R version.

The rest of the paper is organised as follows: Section 2 is devoted to (a) how explanatory structures affect reading comprehension and cognitive accessibility, and (b) the automatic approaches for identifying and adapting these structures into simpler ones. In Section 3 we present our first attempts of methods for adapting both structures to the E2R

SEPLN-2024: 40th Conference of the Spanish Society for Natural Language Processing. Valladolid, Spain. 24-27 September 2024.

isam.diab@upm.es (I. Diab); mcsuarez@fi.upm.es (M. C. Suárez-Figueroa)

0000-0002-3967-0672 (I. Diab); 0000-0003-3807-5019 (M. C. Suárez-Figueroa)

[1]Convention on the Rights of Persons with Disabilities (United Nations, 2006). Available at: https://short.upm.es/k34jw

[2]We are investigating both symbolic (e.g. logical rules) and subsymbolic (e.g. neural networks and machine learning) approaches.

[3]https://ateca.linkeddata.es/

[4]https://ateca.linkeddata.es/

[5]https://simpletext.demos.gplsi.es/

[6]It is worth mentioning that 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.

[7]Human knowledge is explicitly represented in a declarative form (e.g. facts and rules). This way of proceeding is part of symbolic AI.

Methodology as well as the versions of proofs-of-concept for the methods. Finally, we present some conclusions and future work.

2. State of the Art

In this work we delve into developing initial methods to automatically adapt explanatory structures in Spanish into easy-to-read and more accessible versions, based on the guidelines provided by the E2R Methodology [3, 4]. Therefore, in this section we (a) highlight some notes about this type of structures and its implication for reading comprehension (Section 2.1), and (b) summarise the automatic approaches carried out on the adaptation of such explanatory structures (Section 2.2).

2.1. Explanatory Structures and Cognitive Accessibility

Explanatory structures are incises that appear between commas within a sentence, interrupting the course of the utterance to add some precision or comment on the nominal element that precedes them [19, 20]. Explanatory structures can occur in two different forms, according to their syntactic nature. On the one hand, in the form of **nominal appositions**, that is, nouns or noun phrases, as in *Julia, our cousin, lives in Canada.* This type of apposition is formally represented as "A, B". The segment B (also called apodosis in linguistic terms) represents in this variety a parenthetical noun phrase which adds some precision or some remark to clarify the reference of A (also called protasis), which is another noun phrase. In this sense, we observe that segment B assumes that the explanation is copulative; that is, the relationship between segments A and B is formed by the verb *to be*, following the pattern "A, B = A is B" (e.g. *Julia, our cousin, lives in Canada > Julia is our cousin. Julia lives in Canada*). On the other hand, explanatory structures can be **non-restrictive**[8] **adjective clauses**. Such clauses can be (a) relative clauses, which are introduced by relative determiners or pronouns (viz. *that, which, who, whom, whose*), such as *The house, which is on the seafront, is very bright*; or (b) participial clauses, introduced by verbs in participle form, as in *The man, tired from work, fell asleep*[9].

Explanatory structures, as a syntactic element that breaks the discourse line [3, 4], have been studied in relation to reading comprehension [6, 8, 10]. For Dillon and colleagues [10], comprehending an explanatory structure in the form of an incise involves the additional step of identifying which of the previous phrases of like type it is coreferential to. Certainly, they mention several studies [5, 7] suggesting that appositive relative subordinate clauses are often forgotten during the reading process, reflecting the well-known phenomenon that sentence details quickly disappear from memory. Furthermore, following this line, different analyses [9, 10] claim that there is increasing evidence that the syntactic form of appositive material that has come and gone, such as appositive relative clauses in medial position, becomes rapidly unavailable in short-term memory. Moreover, in a study [8] conducted to investigate the effects of aspects of contextual meaning on reading comprehension, the author realised that in the examples provided to the study participants, appositive structures in the form of explanatory incises were the most difficult to understand.

Thus, based on the aforementioned studies, there is clear evidence of the complexity of the explanatory structures that we are dealing with in this research work.

2.2. Automatic Approaches Addressing Explanatory Structures

In the context of Natural Language Processing (NLP), Text Simplification (TS) has gained considerable attention over the last decades. Specifically, syntactic simplification, which involves reducing the complexity of embedded sentences, has emerged as a key focus for the research community. Numerous works have been devoted to the simplification of complex sentences into simpler ones applied to different languages. One of the types of complex sentences that have been automatically addressed are relative clause sentences in the form of incises. For the first of these, we have to go back to the 1990s, when Chandrasekar and Srinivas [21] proposed the implementation of an algorithm through which generalised simplification rules are automatically derived from annotated training data in English. The process used a partial parsing technique that integrates constituent structure and dependency information, in order to simplify subordinated sentences that included relative clauses. Along the same line, relative clauses in the form of incises are also addressed in the work done by Siddharthan [22], which introduces a text simplification framework that uses transformation rules applied to a typed dependency representation generated by the Stanford parser[10]. In addition, for English as well, Dornescu and colleagues [23] explored the extraction of relative clauses employing a tagging approach. They manually annotated a dataset encompassing three text genres, enabling the development and comparison of ruled-based and machine learning methods for automatically identifying appositions and non-restrictive relative clauses. They built a supervised tagging model for automatic detection of appositions using the tagged dataset. For languages other than English, in Brazilian Portuguese, Candido and colleagues [24] developed a rule-based system using a parser which provides lexical and syntactic information for the simplification of 22 complex linguistic phenomena, including relative clauses. For the Indonesian language, Haryadi and colleagues [25] replicated the same method of simplification and dataset as in Siddharthan's work, where some relative clauses were handled. Regarding the Basque language, Aranzabe and colleagues [26] presented an architecture for a text simplification system based on hand written rules specific for syntactic simplification. In the case of Spanish, both in the framework of the Simplext project [15], and also in the work by Bott and colleagues [27] some relative clauses are covered in the syntactic simplification approach, making use of a hand-written computational grammar and dependency trees, and focussing on reducing sentence complexity.

However, to the best of our knowledge, explanatory structures in the form of incises, whether appositions or non-restrictive relative clauses, have not been specifically dealt with in any research work in Spanish. Thus, in our work, we delve into analysing these structures in detail with the

[8] A non-restrictive clause adds additional information to a previous noun, called antecedent. It uses commas to show that the information is additional.

[9] Examples provided by the Spanish Royal Academy of Language (RAE): https://short.upm.es/nbmpd

[10] https://short.upm.es/w7gxt

aim of adapting them to easy-to-read versions following the E2R methodology.

3. Initial Methods for an E2R Adaptation of Explanatory Structures

The aim of the proposed methods is (a) to detect explanatory structures in the form of appositions and adjective clauses written in Spanish, and (b) to adapt such structures into easy-to-read versions as the E2R methodology suggests. These initial methods are composed of the following activities: (1) Natural Language Processing (NLP), which includes tokenization, tagging tasks, morphology and dependency detection, (2) Explanatory Structures Identification, and (3) Explanatory Structures Adaptation. Section 3.1 explains the E2R adaptation methods for nominal appositions and Section 3.2 the methods for non-restrictive adjective clauses in Spanish, as well as the proofs-of-concept implemented based on such methods.

3.1. Nominal Appositions

Nominal appositions **identification** relies on the Part-of-Speech (PoS) tagging information, following an IF-THEN ruled-based approach. In particular, we looked for the PoS tag that identifies an apposition, that is, 'appos'[11]. Regardless, we realised in the initial tests that in two cases appositions were not identified by the PoS tag. Therefore, we analysed these cases to extract new apposition identification patterns to complement the identification by the PoS tag. For the first case, illustrated in Listing 1, it was decided to create a rule that whenever a noun phrase between commas functions (i.e. has the same tag) as its antecedent (a noun), it is an apposition. This noun phrase, being equally tagged, is acting as a "symmetrical" structure and both the noun phrase and the antecedent have the same syntactic information.

```
<NounPhrase> ::= <Determiner>? <Adjective>?
 ↪  <Noun>
<IdentificationPattern1> ::= <Antecedent>","
 ↪  <NounPhrase>","
IF IdentificationPattern1 AND Antecedent.POSTag
 ↪  = Noun.POSTag
THEN NominalApposition
```

Listing 1: Identification Pattern 1 for detecting appositional structures.

For the second case, since the apposition can be seen as a complement of the noun, it was decided that in those cases where a noun phrase is enclosed in commas and the noun is treated as a complement of the noun (PoS tag 'amod'), it should be identified as an apposition. Since appositions are noun phrases, it is required that the apposition is marked with the tag 'NOUN'. Listing 2 shows this identification pattern.

[11] https://short.upm.es/zuoji

```
<IdentificationPattern2> ::= ","
 ↪  <NounPhrase>","
IF IdentificationPattern2 AND NounRelation =
 ↪  "amod" AND NounPhrase.POSTag = NOUN
THEN NominalApposition
```

Listing 2: Identification Pattern 2 for detecting appositional structures.

Regarding the **adaptation** of nominal appositions, in general, the transformation of the apposition into a more accessible and easier structure consists of splitting the appositive structure into two simple sentences according to the pattern "A, B = A is B", mentioned in Section 2.1: (a) on the one hand, the main idea, and (b) on the other hand, the explanation: *En el congreso conocí al famoso investigador, quizá la persona que más influyó en mi trabajo. > En el congreso conocí al famoso investigador* (main idea). *El famoso investigador es quizá la persona que más influyó en mi trabajo* (explanation)[12].

Considering this, we have dealt with the adaptation of four cases of appositions occurrences according to their syntactic nature:

- **Case A. The apposition is not marked by a determiner.** As explained in Section 2.1, the apposition is a noun phrase (or nominal syntagm). Typically a nominal syntagm is formed by a determiner (definite article *el/la/los/las/los/las*[13] ('the') or indefinite article *un/una/unos/unas/unas* ('a')) preceding the noun. However, occasionally the apposition lacks a determiner (as it is omitted because it is taken for granted in the discourse). For example: *El búho, ave rapaz, ve bien de noche. > #El búho es ave rapaz*[14]*. El búho ve bien de noche*[15]. The fact that the determiner is assumed by its absence means precisely that the noun it should accompany is not definite, i.e. no previous reference has been made to that noun. When we use definite determiners (*el/la/los/las*) with a noun, we allude to the fact that there has already been a previous reference in the text to that noun. Therefore, we can assume that the absence of a determiner is synonymous with the use of an indefinite determiner (*un/una/unos/unas*). Thus, in this type of apposition, the adaptation is as follows: *El búho, ave rapaz, ve bien de noche > El búho es un ave rapaz. El búho ve bien de noche*[16].

  Then, in the adaptation process, illustrated in Listing 3, the steps to create the two new simple structures are the following:

  1 To remove the commas.
  2 To insert the verb *ser* ('to be') in its concordant form to the subject and to the originally main

[12] Translation (Tr.): *At the congress I met the famous researcher, perhaps the person who most influenced my work. > At the congress I met the famous researcher. The famous researcher is perhaps the person who most influenced my work.*

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

[14] The hash (#) is used in linguistics to express that a structure is unusual, although it makes sense grammatically.

[15] Tr.: *The owl, bird of prey, sees well at night. > #The owl is bird of prey. The owl sees well at night.*

[16] Tr.: *The owl, bird of prey, sees well at night. > The owl is a bird of prey. The owl sees well at night.*

verb. In the example above the inserted form is *es* ('is') because the subject is *búho* ('owl') (3rd person and singular number) and the verb *ve* ('sees') is in the present indicative tense.

3 To insert the indefinite determiner. For this step, similar as before, the subject information is consulted (*búho* is masculine and singular), and the indefinite determiner that meets the same characteristics is added (in this case, *un*).

4 To close the first segment or protasis with a full stop.

5 To create the second segment or apodosis, retrieving the antecedent noun (*El búho* ('the owl')) and placing it preceding the rest of the predicate (*ve bien de noche* ('sees well at night')).

```
<DefiniteDeterminer> ::= el|la|los|las
<IndefiniteDeterminer> ::= un|una|unos|unas
<AdaptationPattern1> ::= Subject
 ↪  ConjugatedVbSer IndefiniteDeterminer
 ↪  NounPhrase"." Subject Predicate"."
IF Determiner NOT IN NounPhrase AND NounPhrase
 ↪  IS NominalApposition AND Sentence = Subject,
 ↪  NounPhrase, Predicate
THEN AdaptationPattern1
```

Listing 3: Adaptation Pattern for Case A appositions.

In the case of proper nouns, which are naturally not accompanied by a determiner, this rule does not apply. It is possible either to eliminate the commas, as in *Mi primo, Juan, vive en Canarias. > Mi primo Juan vive en Canarias*[17]; or to carry out the same process of creating two sentences, as in *Mi primo, Juan, vive en Canarias. > Mi primo es Juan. Mi primo vive en Canarias*[18]. In this situation, we have opted to eliminate commas for these specific cases, since in this case the apposition between commas is a simple element (a proper noun), and not a construction with more elements that can cut the rhythm of the reading. In this way, the deletion of commas is the most direct form of adaptation as it does not interfere with the content.

- **Case B. The apposition is marked by a determiner.** Contrary to Case A, when the apposition contains a determiner, the steps mentioned in Case A are followed, except for the inclusion of a determiner, which is now explicit. Thus, if there is an example of an apposition with a definite determiner, the adaptation is as follows: *Ana, la amiga de Sara, vino a la fiesta > Ana es la amiga de Sara. Ana vino a la fiesta*[19] And, in the case of an apposition with an indefinite determiner, the adaptation is as follows: *Jorge VI, uno de los reyes de Gran Bretaña, tuvo muchas hijas. > Jorge VI fue uno de los reyes de Gran Bretaña. Jorge VI tuvo muchas hijas*[20]. The pattern for this case is illustrated in Listing 4.

```
<AdaptationPattern2> ::= Subject
 ↪  ConjugatedVbSer Determiner NounPhrase"."
 ↪  Subject Predicate"."
IF Determiner IN NounPhrase AND NounPhrase IS
 ↪  NominalApposition AND Sentence = Subject,
 ↪  NounPhrase, Predicate
THEN AdaptationPattern2
```

Listing 4: Adaptation Pattern for Case B appositions.

- **Case C. The apposition is headed by a deictic.** In spoken discourse, or in certain literary contexts, so-called opaque deictics (pronouns, in this case) are sometimes used, which make a non-literal spatio-temporal allusion. For example, in *Alberti, ese poeta políticamente comprometido, llegó el lunes*[21] we see how *ese* ('that') is a deictic pronoun which does not really point to anything, it is opaque, it makes an allusion to the listener's supposed knowledge of the information in the apposition about Alberti. That is, the sender uses it to include the receiver as knowing that Alberti was a politically committed poet, but frames Alberti in a space. For its adaptation pattern (see Listing 5), the same approach is proposed as in Case B, except that the deictic pronoun (*ese/es/es/esa/esas* ('that')) is replaced by an indefinite determiner (*un/una/unos/unas/unas*), matching its gender and number. Thus, the example above is adapted as follows: *Alberti es un poeta políticamente comprometido. Alberti llegó el lunes*[22].

```
<DeicticPronoun> ::= ese|esa|esos|esas|
este|esta|estos|estas|
aquel|aquella|aquellos|aquellas
<AdaptationPattern3> ::= Subject
 ↪  ConjugatedVbSer IndefiniteDeterminer
 ↪  NounPhrase"." Subject Predicate"."
IF DeicticPronoun IN NounPhrase AND NounPhrase
 ↪  IS NominalApposition AND Sentence = Subject,
 ↪  NounPhrase, Predicate
THEN AdaptationPattern3
```

Listing 5: Adaptation Pattern for Case C appositions.

- **Case D. The apposition can be a clause attached to a proper noun.** It is important to mention that in Spanish there are other explanatory structures in the form of incises which are not appositions but can be treated as such. These are parenthetical structures, short and simple syntagms, which give information by the speaker within the discourse (unlike in Case B, these incises lack complements, they are simpler than the noun phrases of appositions). Their function may be explanatory, but may also be due to factors such as emphasis or reiteration of something

[17] Tr.: *My cousin, Juan, lives in the Canary Islands. > My cousin Juan lives in the Canary Islands.*
[18] Tr.: *My cousin, Juan, lives in the Canary Islands. > My cousin is Juan. My cousin lives in the Canary Islands.*
[19] Tr.: *Ana, the friend of Sara, came to the party.. > Ana is the friend of Sara. Ana came to the party.*
[20] Tr.: *George VI, one of the kings of Great Britain, had many daughters. > George VI was one of the kings of Great Britain. George VI had many daughters.*
[21] Tr.: *Alberti, that politically committed poet, arrived on Monday.*
[22] Tr.: *Alberti is a politically committed poet. Alberti arrived on Monday.*

previously said [19]. In these cases, it is proposed that, since they do not express an explanation per se, the sentence is transformed by reorganising the clause, as shown in Listing 6. That is, in *Juan, el pobre, lo perdió todo*, the clause *el pobre* is placed before the proper noun in the adaptation: *El pobre Juan lo perdió todo*[23]. The same applies to *Arturo, mi amigo, no quiere venir* > *Mi amigo Arturo no quiere venir*[24].

```
<AdaptationPattern4> ::= Determiner Noun
 ↪  ProperNoun Verb RestOfSentence"."
IF ProperNoun.POSTag = PROPN AND
 ↪  (Determiner.POSTag = DET AND Noun.POSTag =
 ↪  NOUN) AND Sentence = Subject, ProperNoun,
 ↪  Predicate
THEN AdaptationPattern4
```

Listing 6: Adaptation Pattern for Case D appositions.

Based on the proposed method for identifying and adapting appositions in Spanish, we have developed a proof-of-concept (PoC)[25] implemented in Python 3.9.

In detail, on the one hand, regarding the **identification activity**, we made use of the PoS tags provided by the NLP library spaCy[26] to retrieve the specific tags related to the apposition structure and the different patterns posed in the aforementioned cases. As for the evaluation of this activity, we have manually built a set of unit tests. For this purpose, we collected a sample of 52 sentences[27] (with an average word count of 11.8) extracted from CREA Corpus[28], of which 34 include an apposition. We manually classified the collection of sentences in binary form (true-false classification), and then analysed the classification performance of our system by using a confusion matrix to measure the number of hits and misses the system made when applying the patterns to identify appositions. We observed that the results are apparently favourable, since the reported precision was 0.94 and the recall was 0.94. Analysing these results, we found that the false positives and false negatives were due to PoS tagging errors on the part of spaCy. For example, in the sentence *Julia, mi perra, necesita que la paseen varias veces al día*[29], spaCy labels the noun *perra* ('dog'), which is the apposition, as a continuation of the noun Julia, thus omitting the explanatory apposition.

On the other hand, with respect to the **adaptation activity**, PoS tags provided by spaCy are also used. In addition, in the cases where the adaptation requires the inclusion of the verb *to be* or determiners agreeing with the subject and the original verb, a dictionary has been manually created including the different verb and determiner forms. For the evaluation of the adaptation activity, we opted for a language-model-based approach, since we aimed to measure whether the semantic content of the adaptation maintains that of the original. Specifically, we have used a sentence similarity model[30] for Spanish, (available at Hugging Face repository[31]), to compare the original sentence with apposition and the adapted sentence provided by our PoC according to the E2R methodology. The choice of this language model is based on a previous work[32] we carried out in which we compared the performance of different sentence similarity models in Spanish. By using this model, vector representations of the sentences can be obtained and the semantic similarity between them can be calculated. The model receives as input the original sentence and the transformed sentence, and returns a number indicating the degree of similarity between them, being 0 not similar at all and 1 being completely similar. We obtained an average of 0.94 similarity between all 34 sentences with apposition that were adapted. In more detail, one of the main errors encountered in the adaptation activity has to do with the conjugation of the verb *ser* ('to be'), since an explanation of a subject can be presented in a different tense than the time at which the main action of the sentence occurs. For example, the sentence *Nuestros vecinos, los Pérez, se fueron de vacaciones* [33] is adapted as *#Nuestros vecinos fueron los Pérez. Nuestros vecinos se fueron de vacaciones*[34], because the main verb of the sentence is *fueron* ('they were') (3rd person singular of the preterite perfect simple indicative) and therefore the verb *ser* is in the same verb tense (*fueron*). We can clearly detect that perhaps the conjugation of the verb *ser* is not the right one in this situation, since according to the reader's logic it is assumed that if the Pérez knocked on the door, they are still the speaker's neighbours, so the verb ser should be in the 3rd person plural of the present tense (*son* ('they are')). In addition, as a qualitative analysis, we manually analysed the adapted sentences for grammatical sense. The result is positive, as all sentences generated by our method are grammatically correct.

3.2. Non-restrictive Adjective Clauses

Similarly to the identification of nominal appositions, in the case of the non-restrictive adjective clauses **identification**, we used Part-of-Speech (PoS) tagging information, by means of a rule-based approach. Nevertheless, following the method in Section 3.1, we performed initial tests to analyse those cases in which the adjective clauses were not identified by the specific PoS tag, in order to create patterns to cover these cases. We crafted specific rules for covering the following cases: (a) the relative pronoun consists of two words forming a single semantic unit (e.g. *el que, la que, los que, las que*), where the pattern (See Listing 7) identifies the sentence not only as a relative clause when encountering a comma followed by a relative pronoun, but also when encountering a comma followed by a definite article, and then a relative pronoun. On the other hand, (b) participle sentences with verbal periphrasis in the main sentence, in which the auxiliary verb is treated as the main verb, posing a problem when performing the transformation, since we depend on the main verb to extract the morphological information. The solution by means of this rule is straightforward: whenever the "auxiliary verb + main verb" pattern is detected, the morphological information of the auxiliary

[23] Tr.: *Poor Juan lost everything.*
[24] Tr.: *My friend Arturo does not want to come.* > *My friend Arturo does not want to come.*
[25] The proof-of-concept is not yet available online but we are working to make it available as soon as possible.
[26] https://spacy.io/ We used the trained model for Spanish `es_core_news_lg` for developing all the activities.
[27] Both set of sentences and analysis results are available at: https://doi.org/10.5281/zenodo.11397343
[28] https://short.upm.es/ydq6p
[29] Tr.: *Julia, my dog, needs to be walked several times a day.*
[30] https://short.upm.es/w2slm
[31] https://huggingface.co/
[32] https://oa.upm.es/75516/
[33] Tr.: *Our neighbours, the Pérez, went on holiday.*
[34] Tr.: *#Our neighbours were the Pérez. Our neighbours went on holiday.*

verb is consulted in the adaptation.

```
<RelativePronoun> ::= que
<DefiniteArticle> ::= el|la|los|las
<IdentificationPattern1> ::= ","
 ↪  (<RelativePronoun>|<DefiniteArticle>
 ↪  <RelativePronoun>)"," Predicate
IF IdentificationPattern1
THEN AdjectiveRelativeClause
```

Listing 7: Identification Pattern for Case A adjective clauses.

As for the **adaptation** activity, the easy-to-read adaptation of the adjective clauses into a more accessible and easier structure should consist of splitting the adjective clause into two simple sentences: (a) on the one hand, the explanation and (b) on the other hand, the main idea. For example, the sentence *La enfermera, que tiene 63 años, está a punto de jubilarse*[35] is adapted as follows: the explanation *La enfermera tiene 63 años*[36] and the main idea *La enfermera está a punto de jubilarse*[37]

In more detail, the process of adapting the non-restrictive adjective clauses into more easily understood structures depends on the two types of adjective clause we mentioned in Section 2.1:

Relative Clauses. These type of adjective clauses can be introduced by different elements, and thus be classified as:

(a) Introduced by a relative pronoun. In this case, the procedure is straightforward. The relative pronoun and the commas are removed, and the two new simple structures are reorganised, keeping the verbs and the rest of the complements the same: *Juan, que trabaja mucho, decidió tomarse un descanso*[38] is adapted as *Juan trabaja mucho. Juan decidió tomarse un descanso*[39].

(b) Introduced by a possessive relative determiner. This type of relative (*cuyo, cuya, cuyos, cuyas* ('whose')) presents a relation with the antecedent as a complement to the noun, and this function in Spanish is expressed by means of the prepositional syntagm "de ('of') + noun'. For example, *Esta chica, cuyo padre vive en Malasia, se mudó a las islas*[40] the is adapted as: *Esta chica se mudó a las islas. El padre de esta chica vive en Malasia*[41] In this case, the main idea is placed in the first plance and afterwards, the explanation, in order to avoid problems of correference. Thus, the adaptation process for these cases is based on the following steps:

1 To remove the commas.

2 To remove the possessive.

3 To reorganise the main sentence to the first segment or protasis and close it with a full stop.

4 To create the second segment or apodosis by adding as subject the subject of the subordinate clause and the prepositional phrase "de + the subject of the main clause". For the addition of the subject of the apodosis it has to be marked with a definite article determiner (*el, la, los, las*) according to the gender and number of the noun.

Participial Clauses. Whereas relative clauses presented a nexus (the relative pronoun) linking the antecedent with the subordinate clause, participial clauses follow the same pattern as nominal appositions, i.e. "A, B = A is B". For instance, in the sentence, *El hombre, cansado de trabajar, se durmió*[42], we assume that *El hombre* ('the man') (segment A) "was" tired from working (segment B). Considering that, the adaptation of these type of adjective clauses is as follows:

1 To remove the commas.

2 To insert the verb *estar* ('to be') in its concordant form with the subject and the original main verb. If the main verb appears in the present tense, the verb *estar* also appears in the present tense, and if it occurs in any past tense, we opted to use the imperfect past tense.

3 To close the first segment or protasis with a full stop.

4 To create the second segment or apodosis by retrieving the antecedent preceding the main sentence.

We have developed a proof-of-concept (PoC)[43], based on the proposed method implemented in Python 3.9.

With respect to the **identification activity**, as in the method for nominal appositions, we used the PoS tags provided by spaCy[44] to get the specific tag related to the relative pronoun that introduces the clause. To assess our method of identifying non-restrictive adjective clauses, we assembled a set of 96 sentences[45] from various sources (educational textbooks and literary works), with an average word count of 9.7, of which 62 were adjective sentences of the aforementioned types. We started by manually assigning binary tags to the sentence collection. Next, we evaluated the classification performance of our system by analysing a confusion matrix, which enabled us to quantify both correct classifications and errors made by the system. As results, we obtained 0.95 precision and 0.86 recall. In this case, one of the most frequent errors is the incorrect identification of exhortative (expressing command) or desiderative (expressing desire) subordinate clauses as adjective clauses since spaCy detects the conjunction that acts as a nexus of the main and the subordinate sentence as a relative pronoun, since both nexus and relative pronoun have the same form (*que*): e.g. *Juan, que te portes bien, por favor*[46].

The same methodology has been used to develop the **adaptation activity** as in the case of nominal appositions. This task relies on the PoS tags provided by spaCy. Furthermore, for cases where adaptation involves the addition of the verb *to be* or determiners that match the subject and original verb, we developed a dictionary manually. To evaluate the adaptation process, we employed a language-model-based approach. We have yet again used the Spanish sentence similarity model[47] to compare the vectorial similarity between the original adjective clause to the adapted sentence

[35] Tr.: *The nurse, who is 63 years old, is about to retire.*
[36] Tr.: *The nurse is 63 years old.*
[37] Tr.: *The nurse is about to retire.*
[38] Tr.: *Juan, who works a lot, took a break.*
[39] Tr.: *Juan works a lot. Juan took a break.*
[40] Tr.: *This girl, whose father lives in Malaysia, moved to the islands.*
[41] Tr.: *This girl moved to the islands. The father of this girl lives in Malaysia.*
[42] Tr.: *The man, tired from working, fell asleep.*
[43] The proof-of-concept is not yet available online but we are working to make it available as soon as possible.
[44] https://spacy.io/
[45] Both set of sentences and analysis results are available at: https://doi.org/10.5281/zenodo.11397343
[46] Tr.: *Juan, please behave yourself.*
[47] https://short.upm.es/w2slm

provided by our system. The results showed an average similarity of 0.94 across the 62 adapted sentences in relation with their original versions. As in the previous case, we have manually analysed the adapted sentences and they are all grammatically correct.

4. Conclusions and Future Work

This paper aims to enhance the cognitive accessibility of Spanish texts by proposing two methods for adapting sentences containing explanatory structures such as nominal appositions and adjective clauses, following an E2R approach. While the methods we propose for the identification and adaptation of these structures might seem simple and straightforward, we believe they represent a valuable contribution to the field of cognitive accessibility. These methods have been implemented as proof of concepts to help in the (semi)-automatic adaptation task of improving text accessibility for individuals with reading comprehension difficulties, including those with cognitive disabilities. We have evaluated the methods by unit tests and by using a language model to calculate the similarity between the original sentences and the ones adapted to E2R, obtaining generally satisfactory results. However, since the text adaptation aims to meet the needs of particular groups, a user-based evaluation is essential to complete the assessment of our methods.

As further research, several actions are planned to improve the initial attempt presented in this work: (a) we are going to analyse the possibility of proposing methods based on subsymbolic AI techniques, such as the use of language models or large language models (LLMs); the ultimate goal is to compare the results obtained using the methods proposed in this paper with the ones obtained with the subsymbolic-based methods; (b) we are going to implement a web application in the context of the assistive technologies for adapting the explanatory structures which are not compliant to the E2R methodology; and (c) we have planned to involve people with cognitive disabilities to evaluate the aforementioned web application.

Acknowledgments

This work has been supported by the grant "Ayudas para la contratación de personal investigador predoctoral en formación para el año 2022" funded by Comunidad Autónoma de Madrid (Spain). We would like to thank Miguel Cerezo Durán and Clara Osorio Sanz for their help in the initial methods development.

References

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

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

[3] O. García Muñoz, Lectura fácil: Métodos de redacción y evaluación, Real Patronato sobre Discapacidad, 2012.

[4] AENOR, Lectura Fácil. Pautas y recomendaciones para la elaboración de documentos (UNE 153101:2018 EX), Inclusion Europe, 2018.

[5] J. S. Sachs, Recognition memory for syntactic and semantic aspects of connected discourse, Perception & Psychophysics 2 (1967) 437–442. URL: https://doi.org/10.3758/BF03208784. doi:10.3758/BF03208784.

[6] G. L. Dillon, Language Processing and the Reading of Literature: Toward a Model of Comprehension, Indiana University Press, Bloomington, 1978.

[7] M. C. Potter, L. Lombardi, Regeneration in the short-term recall of sentences, Journal of Memory and Language 29 (1990) 633–654. URL: https://www.sciencedirect.com/science/article/pii/0749596X9090042X. doi:https://doi.org/10.1016/0749-596X(90)90042-X.

[8] A. A. Mirza, The effects of contextual meaning aspects on reading comprehension, Journal of English as a Foreign Language 1 (2011). URL: https://doi.org/10.23971/jefl.v1i2.192. doi:10.23971/jefl.v1i2.192.

[9] M. Kroll, M. Wagers, Is working memory sensitive to discourse status? Experimental evidence from responsive appositives, Poster presented at XPrag 2017, Cologne, Germany, 2017. URL: https://osf.io/view/xprag2017/.

[10] B. Dillon, L. Frazier, C. Clifton, No longer an orphan: evidence for appositive attachment from sentence comprehension, Glossa 3 (2018). URL: https://doi.org/10.5334/gjgl.379. doi:10.5334/gjgl.379.

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

[12] M. Suárez-Figueroa, I. Diab, E. Ruckhaus, I. Cano, First steps in the development of a support application for easy-to-read adaptation, Universal Access in the Information Society (2022). doi:10.1007/s10209-022-00946-z.

[13] A. Iglesias, I. Cobián, A. Campillo, J. Morato, S. Sánchez-Cuadrado, Comp4text checker: An automatic and visual evaluation tool to check the readability of spanish web pages, Springer-Verlag, Berlin, Heidelberg, 2020, pp. 258–265. URL: https://doi.org/10.1007/978-3-030-58796-3_31. doi:10.1007/978-3-030-58796-3_31.

[14] E. Díez, N. Rodríguez, A. Fernández, M. A. Alonso, M. A. Díez-Álamo, M. J. Sánchez, D. Wojcik, E2R-Helper Asistente Lectura Fácil, Instituto Universitario de Integración en la Comunidad - Universidad de Salamanca, 2019. URL: https://eapoyo-inico.usal.es/asistente-lectura-facil/.

[15] H. Saggion, S. Stajner, S. Bott, S. Mille, L. Rello, B. Drndarevic, Making It Simplext: Implementation and Evaluation of a Text Simplification System for Spanish, ACM Transactions on Accessible Computing 6 (2015). doi:10.1145/2738046.

[16] S. Bott, L. Rello, B. Drndarevic, H. Saggion, Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish, in: Proceedings of COLING 2012, The COLING 2012 Organizing Committee, 2012.

[17] L. Rello, R. Baeza-Yates, H. Saggion, DysWebxia: tex-

tos más accesibles para personas con dislexia, Procesamiento del Lenguaje Natural 51 (2013) 205–208.
[18] L. Moreno, R. Alarcon, R. Martínez, EASIER System. Language Resources for Cognitive Accessibility, in: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '20, Association for Computing Machinery, 2020. doi:10.1145/3373625.3418006.
[19] RAE, ASALE, Nueva gramática de la lengua española, Espasa Calpe, Madrid, 2009.
[20] RAE, ASALE, Ortografía de la lengua española, Espasa Calpe, Madrid, 2010.
[21] R. Chandrasekar, B. Srinivas, Automatic induction of rules for text simplification, Knowledge-Based Systems 10 (1997) 183–190. URL: https://www.sciencedirect.com/science/article/pii/S0950705197000294. doi:https://doi.org/10.1016/S0950-7051(97)00029-4.
[22] A. Siddharthan, Text simplification using typed dependencies: A comparision of the robustness of different generation strategies, in: C. Gardent, K. Striegnitz (Eds.), Proceedings of the 13th European Workshop on Natural Language Generation, Association for Computational Linguistics, Nancy, France, 2011, pp. 2–11. URL: https://aclanthology.org/W11-2802.
[23] I. Dornescu, R. Evans, C. Orăsan, Relative clause extraction for syntactic simplification, in: C. Orasan, P. Osenova, C. Vertan (Eds.), Proceedings of the Workshop on Automatic Text Simplification - Methods and Applications in the Multilingual Society (ATSMA 2014), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 1–10. URL: https://aclanthology.org/W14-5601. doi:10.3115/v1/W14-5601.
[24] A. Candido, E. Maziero, L. Specia, C. Gasperin, T. Pardo, S. Aluisio, Supporting the adaptation of texts for poor literacy readers: a text simplification editor for Brazilian Portuguese, in: J. Tetreault, J. Burstein, C. Leacock (Eds.), Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications, Association for Computational Linguistics, Boulder, Colorado, 2009, pp. 34–42. URL: https://aclanthology.org/W09-2105.
[25] M. L. Khodra, A. B. Sasmita, et al., Rule-based text simplification for extractive automatic summarization of news articles in indonesian, in: 2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA), IEEE, 2023, pp. 1–6.
[26] M. Aranzabe, A. Ilarraza, I. Gonzalez-Dios, First approach to automatic text simplification in basque, 2012, pp. 1–8.
[27] S. Bott, H. Saggion, S. Mille, Text simplification tools for Spanish, in: N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA), Istanbul, Turkey, 2012, pp. 1665–1671.