Towards the Assessment of Easy-to-Read Guidelines Using Artificial Intelligence Techniques

Mari Carmen Suárez-Figueroa\textsuperscript{1(✉)}, Edna Ruckhaus\textsuperscript{1}, Jorge López-Guerrero\textsuperscript{2}, Isabel Cano\textsuperscript{3}, and Álvaro Cervera\textsuperscript{3}

\textsuperscript{1} Ontology Engineering Group (OEG), Universidad Politécnica de Madrid (UPM), Madrid, Spain
\{mcsuarez,eruckhaus\}@fi.upm.es
\textsuperscript{2} Universidad Politécnica de Madrid (UPM), Madrid, Spain
jorge.liglesias@alumnos.upm.es
\textsuperscript{3} ACCEDES, Madrid, Spain
\{icano,acervera\}@accedes.es

Abstract. The Easy-to-Read (E2R) Methodology was created to improve the daily life of people with cognitive disabilities, who have difficulties in reading comprehension. The main goal of the E2R Methodology is to present clear and easily understood documents. This methodology includes a set of guidelines and recommendations that affect the writing of texts, the supporting images, the design and layout of documents, and the final editing format. Such guidelines are used in the manual processes of (a) adapting existing documents and (b) producing new materials. The process of adapting existing documents is cyclic and implies three activities: analysis, transformation, and validation. All these activities are human resource consuming, due to the need of involving people with cognitive disabilities as well as E2R experts. In order to alleviate such processes, we are currently investigating the development of methods, based on Artificial Intelligence (AI) techniques, to perform the analysis and transformation of documents in a (semi)-automatic fashion. In this paper we present our AI-based method for assessing a particular document with respect to the E2R guidelines as well as an initial implementation of such a method; our research on the transformation of documents is out of the scope of this paper. We carried out a comparative evaluation of the results obtained by our initial implementation against the results of the document analysis performed by people with cognitive disabilities.

Keywords: E2R methodology · Cognitive accessibility · Artificial intelligence

1 Introduction

People with cognitive disabilities have some problems related to reading comprehension, communication, and ability to respond to routine situations as well as to challenging scenarios. These obstacles become a daily barrier in the understanding, interaction, and use of products and services in different environments. To overcome
the aforementioned barriers and to improve the daily life of people with cognitive
disabilities, a methodology called Easy-to-Read (E2R) [1, 2] was created. This
methodology aims to present clear and easy to understand contents to different sectors
of the population that include people with disabilities and people with limited language
or reading proficiency, among others. The E2R Methodology provides guidelines for
content, language, illustrations, as well as graphic layout in documents. The final aim is
to have materials that are compliant with the E2R guidelines and recommendations,
such as to use simple and short phrases, to avoid technicalities, abbreviations, acro-
nyms, among others, to have an adequate typography, to include pictures that reinforce
the message and clarify the content, and to select an editing format that is easy to use
and convenient for the expected use of the material.

When a particular material needs to be adapted to the E2R Methodology, three key
activities are performed as part of a cyclic process: (1) analysis of which E2R guide-
lines are fulfilled in the material, this can be seen as a kind of assessment activity;
(2) transformation of the material by E2R experts based on the unsatis-
fied guidelines discovered during the previous assessment; and (3) validation of the transformed
material by a professional team comprised by people with intellectual disabilities,
called validators. The activities of analysis, transformation (or adaptation), and vali-
dation are complementary activities for achieving good quality and effective final
materials in terms of cognitive accessibility. These activities are human resource
consuming due to the need of involving people with cognitive disabilities as well as
E2R experts.

In order to help in the labour-intensive and costly process of adapting materials to
the E2R guidelines, and thus to facilitate the work of those who manually adapt
documents, our research is currently focused on applying different AI-based methods
and techniques to (semi)-automatically perform both the analysis and the transforma-
tion of documents. To the best of our knowledge, there is no significant research in the
analysis of how E2R compliant is a document, a crucial point for achieving cognitive
accessibility. Nevertheless, there is some work on transforming materials based on E2R
guidelines: a text simplification system for Spanish [3] and a method for extending a
grammar checker with E2R guidelines for English and German [4].

Specifically, in this paper we present our method for assessing how E2R compliant
is a particular document written in Spanish; our research on the transformation of these
documents is out of the scope of this paper. The proposed method is based on the use of
AI techniques such as rule-based techniques, pattern recognition, and Natural Lan-
guage Processing (NLP). This paper also presents an initial implementation of the
proposed method: an E2R compliance tool called Easy-to-Read Advisor. Our method
and the proof of concept are focused on educational presentations written in Spanish.

The rest of the paper is organized as follows: Sect. 2 is devoted to the state of the
art on the E2R Methodology. In Sect. 3 we explain our research methodology; while
Sect. 4 describes our contributions: a method and an initial implementation for
assessing E2R compliance of documents. Section 5 presents the initial evaluation of
our proposal. Finally, we show some conclusions and future work on this research.
2 State of the Art

The need of having understandable content and accessible information for people with learning difficulties has received increased attention in the last decade [5]. In this regard, the E2R Methodology is crucial for having more readable and understandable materials. The main goal of having materials that are easy-to-read is to present clear and easy to understand contents that are appropriate for different groups [1]. To achieve these documents, content, language, illustrations, as well as the graphic layout must be taken into consideration. The E2R Methodology provides guidelines, requirements, and recommendations for different aspects of a document [1]: writing of texts, supporting images, design and layout of documents, and the editing format. However, there is no general acceptance on the level of depth in E2R guidelines [6].

To know whether a document is compliant with the E2R Methodology there are two main manual techniques: (a) using checklists, as for example the one¹ created in the PUZZLE project and (b) involving groups of people who discuss the E2R requirements satisfied in a document. However, manual approaches are labour-intensive and costly. In order to decrease both time and resources spent for assessing E2R guidelines, semi-automatic tools, called validators, can be used. One example is the tool presented in [8], which is based on an approach to empirically evaluate E2R guidelines in German documents. Another example is VisRA [9], whose aim is to support the writer in the task of revising a German text with respect to E2R guidelines. To the best of our knowledge, there are no validators for Spanish documents. Nevertheless, research work on transforming Spanish texts using NLP techniques focused on text simplification has been performed [3]. It is worth mentioning that most of the approaches pays attention only to E2R guidelines related to writing aspects (spelling, grammar, vocabulary, and style), while aspects related to the design and layout of a document have not been considered. Our research work focuses on all E2R aspects as well as on both, analysis and transformation of documents.

3 Our Research Methodology

Our general research question is: Is it possible to use AI-based methods and techniques to facilitate the cognitive accessibility to educational materials? This question can be divided into two specific research questions: (1) Can we use AI-based methods and techniques to analyze whether an educational document conforms to the E2R Methodology? and (2) Can we use AI-based methods and techniques to transform an educational document to one compliant to the E2R Methodology? In this paper we present our initial contributions regarding the answer to the first question; therefore, it should be noted that the second question is out of the scope of this document.

Our research methodology includes the following six activities: the first activity was to acquire the knowledge related to the E2R Methodology. We used the catalogue of guidelines as a basis [7]. This catalogue is organized into the following categories:

¹ http://www.puzzle-project.eu/docs/EN/IO1/IO1_EtRLearningMaterial_Checklist.pdf.
(a) spelling, grammar, vocabulary, and style in the writing field, (b) images, typography, composition of text, and pagination in the design and layout field, (c) paper, binding, and printing in the production field, and (d) others.

The second activity was to select the collection of E2R guidelines to consider in our research. Since we decided to focus on educational materials in the form of sets of slides, the E2R conformance evaluation of the documents included both the writing field and the design and layout field. This implies a total of 69 E2R guidelines related to spelling aspects such as capital letters, punctuation, and dates or to grammar aspects like verbs, parts of the sentence, and complete sentences, among others. Examples of guidelines are ‘to use the dot to divide sentences’ and ‘to avoid the use of the comma to divide phrases’ related to the punctuation aspect; and ‘sentences should always include the subject’ related to parts of the sentence aspect.

Our third activity was to analyze and decide which AI techniques could be the most appropriate for the creation of a method for assessing the easy-to-read compliance in a document written in Spanish. We decided to use the following techniques:

- Since our goal is to create a diagnosis method, a rule-based approach can be applied. The background knowledge that establishes whether a particular document is compliant with a specific E2R guideline can be represented as problem-solving rules. The term rule in AI is defined as an “IF-THEN” structure that relates a set of given information or facts in the “IF part” to some action in the “THEN part”, which means “IF Conditions happens THEN do Actions”.
- Conditions in the “If statements” can be represented in a declarative way and can be expressed as patterns. In particular, we decided to create syntactic patterns taking as an inspiration Hearst patterns [10]. This implies that our proposal includes the use of pattern recognition techniques.
- Due to the fact that our method needs to analyze issues such as spelling, grammar, vocabulary, and style in the text included in the document, we consider also the need to use NLP techniques. In particular, techniques to perform parsing, morphological, and syntactic analysis.

The fourth activity was to create the method for assessing whether a particular material is compliant with the E2R Methodology and to implement such a method.

The fifth activity was to select a corpus of slides in Spanish and to create a gold standard. This gold standard consists of slides annotated with the guidelines fulfilled and not fulfilled by people with cognitive disabilities.

The sixth activity was to validate our proposal using a comparative evaluation using the gold standard and the same set of slides analyzed by the proof of concept.

4 Our Proposal for Assessing E2R Guidelines

This section presents our method for analyzing which E2R guidelines are not fulfilled in a document, and our proof of concept called Easy-to-Read Advisor².

² Available in September 2020 at “easy2read.oeg.fi.upm.es”.

Towards the Assessment of Easy-to-Read Guidelines  77
4.1 E2R Assessment Method

Our method for assessing whether a particular material is compliant with the E2R Methodology is composed of the following activities: (1) gather the document to be analyzed, (2) extract from the document the set of elements related with the writing, and the design and layout fields, (3) parse and analyze morphologically and syntactically the set of elements related with the writing field, (4) recognize patterns in each set of elements, (5) infer knowledge using the collection of rules, and (6) generate the report about the original document.

In more detail, the pattern recognition activity relies on a collection of general syntactic patterns, which aims to identify the relevant elements in a document. This collection has been created using Hearst patterns as inspiration and taking into account that for each of the E2R guidelines, at least one pattern is needed to represent the knowledge. The activity of inferring knowledge related to the original document implies the use of a production system that consists of a collection of rules that formalize the E2R guidelines, and an inference engine that implements the control strategy and applies the rules. The structure of the rules is the following: the “if part” is composed of one or more conditions related via the AND logical connector, while the “then part” is an action to report the compliance or non-compliance of the original document with respect to a particular E2R guideline. An example of a syntactic pattern and a rule associated to an E2R guideline is shown in Fig. 1.

4.2 Proof of Concept: Easy-to-Read Advisor

We have developed an E2R conformance checker, which is based on our E2R assessment method described in Sect. 4.1. This proof of concept is focused on the analysis of educational documents in the form of sets of slides written in Spanish; these slides have been created in HTML. Easy-to-Read Advisor provides as output a report on the guidelines satisfied (marked in green) and the ones not satisfied (marked in red), as shown in Fig. 2.

Easy-to-Read Advisor extracts from each slide the collection of elements, tags and attributes related to the writing and the design and layout fields. Extracted elements, tags and attributes are matched with the catalogue of patterns. This initial prototype implements 22 E2R guidelines as patterns: ten of them are related to the design and layout field and twelve of them are related to the writing field, as shown in Table 1.

---

**E2R GUIDELINE:** G1. To always use a font that is clear and easy to read. The list of suggested fonts are: Arial, Calibri, Candara, Corbel, Gill Sans, Helvetica, Myriad, Segoe, Tahoma, Tiresias, and Verdana.

**GENERAL SYNTACTIC PATTERN:** P1*: Text hasTypeOfFont (ToF,)* (or|and) ToF

**GENERAL RULE:** IF Tb ∋ (Arial | Calibri | Candara | Corbel | Gill Sans | Helvetica | Myriad | Segoe | Tahoma | Tiresias | Verdana) THEN to report that original document is compliant with G1

---

**Fig. 1.** Example of E2R guideline, syntactic pattern and rule ("P1 in E2R Advisor: <TextElement style="font-family: (ToF,)* ToF"> Text </TextElement>“, "T: set of font types extracted from the original document via our general syntactic patterns").

---

3 SlideWiki platform (https://slidewiki.org/) can be used to create slides in HTML.
**Fig. 2.** Easy-to-Read Advisor: Architecture (Easy-to-Read Advisor uses MeaningCloud (https://www.meaningcloud.com/es) to perform the parsing, morphological, and syntactic analysis of the Spanish text in each slide) and Example of Report (Report obtained for “https://slidewiki.org/deck/108653-2/lectura-facil/slide/710658-10/”).

| Design and layout field guidelines                                                                 | Writing field guidelines                                      |
|---------------------------------------------------------------------------------------------------|-------------------------------------------------------------|
| M1. The text font belongs to the accepted styles                                                   | T1. The size of the lines is correct                         |
| M2. The font size must be large enough                                                             | T2. Large numbers are expressed correctly                    |
| M3. Text avoids italics                                                                           | T3. Text avoids the use of special characters                |
| M4. Text includes moderately bold words                                                             | T4. Text avoids the use of ordinal characters                |
| M5. Text includes moderately underlined words                                                        | T5. Text is made up of short sentences                       |
| M6. Text avoids typographic effects                                                                 | T6. Dates are written in full                                |
| M7. Text uses capital letters according to the general rule                                         | T7. The use of pronouns is correct                           |
| M8. The contrast between text and background is correct (focused on text color)                   | T8. Text avoids the use of Roman numbers                     |
| M9. The contrast between text and background is correct (focused on background)                   | T9. Text is written in the second person                     |
| M10. The amount of words in the text is correct                                                     | T10. Text avoids passive voice                               |
|                                                                                                   | T11. Sentences have a subject                                |
|                                                                                                   | T12. Text is made up of simple sentences                     |
5 Evaluation

Our comparative evaluation implied: (1) a manual analysis of a corpus⁴ of slides by a person with cognitive disabilities with experience on the process of adapting documents to the E2R Methodology; as a result we obtained our gold standard; and (2) an automatic assessment using Easy-to-Read Advisor over the same corpus. Each assessment, the gold standard and the output obtained with Easy-to-Read Advisor, is represented as a matrix of E2R guidelines and their compliance (yes/no) for each slide in the corpus; it is worth mentioning that in the case of the gold standard, when a slide is not compliant with a specific guideline, comments and suggestions are provided.

Due to the difficulty of analysing the aforementioned matrices in an automatic fashion, we decided to manually inspect a subset of 10 slides⁵ and their corresponding assessments (the gold standard and the output obtained with Easy-to-Read Advisor). To compare the distributions of the number of coincidences between the gold standard and Easy-to-Read Advisor over the 10 slides, we calculated their descriptive statistics for the Design and Layout field and for the Writing field. No major differences between those fields arise other than the difference in medians, 7 vs. 8.5, and third quartiles, 7.75 vs. 9; being the results for the Writing field slightly higher than the results for the Design and Layout field. With these figures, we could say that Easy-to-Read Advisor (1) is analyzing the slides in a reasonably precise way, and (2) is identifying better the aspects related to the Writing field. Thus, Easy-to-Read Advisor should improve the way in which some guidelines are analyzed. In order to know which are those guidelines, we also analyzed the non-coincidences by computing: (a) the number of “Yes (gold standard) - No (Easy-to-Read Advisor)”, which may indicate that the tool fails because the guideline is satisfied according to the gold standard (false negative) and (b) the number of “No (gold standard) - Yes (Easy-to-Read Advisor)”, which could indicate that the tool fails because the guideline is not satisfied according to the gold standard (false positive). This analysis in the Design and Layout field shows that these non-coincidences between the gold standard and the tool represent 31% of the total number of comparisons. From these non-coincidences, 35.5% are “Yes-No” (false negative) and 64.5% are “No-Yes” (false positive). This same analysis in the Writing field shows that the gold standard and the tool disagree 33.4% of the time. All disagreements are due to false positives (“No-Yes”); there is no “Yes-No” disagreements. In addition, we calculated the number of non-coincidences between the gold standard and Easy-to-Read Advisor for each guideline. It can be observed that guidelines with the largest number of non-coincidences are T9 and M9, with 9 and 8 respectively; T1 and T5, with 7; and T11 and T12, with 6. In all these cases, except in the M9 guideline, the non-coincidence was due to a false positive. These figures imply that Easy-to-Read Advisor should be mainly improved with respect to false positive scenarios. In particular, we should focus our efforts on guidelines related to writing aspects such as using second person in texts and using a correct and simple grammatical structure (“Subject + Verb + Predicate.”).

---

⁴ This corpus is currently composed of 50 slides written in Spanish and created in SlideWiki.

⁵ Slides available at “https://slidewiki.org/deck/91466-1/plan-de-formacion-datos-abiertos-ayto-de-zaragoza/deck/91466-1/”. Slide 8 was not included in the corpus.
6 Conclusions and Future Work

In this paper, we present the initial results of our research on assessing E2R guidelines in a particular document. We propose a method for analyzing how easy-to-read compliant is a particular document. This method is independent on the format in which the material has been created. The method is based on rule-based techniques, pattern recognition, and NLP. We have also implemented a proof of concept of the method as an E2R conformance checker called Easy-to-Read Advisor. This checker is a specific implementation of the method and assesses which ER2 guidelines are satisfied in educational presentations written in Spanish.

We have performed a comparative evaluation between our gold standard and the results obtained by Easy-to-Read Advisor over a set of slides. The results indicate that on average over 70% of the guidelines in the checker are coincident with the gold standard. This evaluation provides us with clues to improve and refine Easy-to-Read Advisor. In addition, E2R experts validated this implementation; they provided positive comments and interesting suggestions for further work as well as remarked that to speeding up the E2R adaptation process is the potential of the tool.

Our plan is to improve Easy-to-Read Advisor (a) in the guidelines in which there was no coincidence with the gold standard, (b) with respect to the comments received, (c) and implementing the complete set of E2R guidelines.

Acknowledgements. This research work has been funded by the project Datos 4.0 (TIN2016-78011-C4-4-R) (Agencia Estatal de Investigación del MINECO y Fondos FEDER). We would like to thank Iván Martínez and Arminda Moreno for their comments, and Sandra Cartas and Álvaro Bermejo for their analysis of E2R guidelines.

References

1. IFLA: Guidelines for Easy-to-Read Materials. International Federation of Library Association and Institutions IFLA Professional Reports 120 (2010)
2. Norma UNE 153101:2018 EX Lectura Fácil. Pautas y recomendaciones para la elaboración de documentos (2018)
3. Bott, S., Saggion, H.: Automatic simplification of Spanish text for e-accessibility. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) ICCHP 2012. LNCS, vol. 7382, pp. 527–534. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31522-0_79
4. Nietzio, A., Naber, D., Bühler, C.: Towards Techniques for Easy-to-Read Web Content. Procedia Comput. Sci. 27, 343–349 (2014)
5. Matausch, K., Nietzio, A.: Easy-to-read and plain language: defining criteria and refining rules. In: Miesenberger, K., Petz, A., Matausch, K. (eds.) W3C WAI Symposium on Easy-to-Read on the Web (2012)
6. Leskela, L.: Guidelines or standards for Easy-to-read? In: Easy-to-Read on the Web. Symposium W3C. WAI 2012
7. García Muñoz, O.: Lectura fácil: Métodos de redacción y evaluación 2012. Real Patronato sobre Discapacidad. M-34791-2012
8. Nietzio, A., Scheer, B., Bühler, C.: How long is a short sentence? – a linguistic approach to definition and validation of rules for easy-to-read material. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) ICCHP 2012. LNCS, vol. 7383, pp. 369–376. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31534-3_55

9. Oelke, D., Spretke, D., Stoffel, A., Keim, D.: Visual readability analysis: how to make your writings easier to read. IEEE Trans. Visual. Comput. Graph. 18(5), 662–674 (2012)

10. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: Conference on Computational Linguistics (COLING 1992) (1992)