Procesamiento del Lenguaje Natural, Revista nº 71, septiembre de 2023, pp. 191-203 recibido 31-03-2023 revisado 14-05-2023 aceptado 17-05-2023

Automatic Easy-to-Read Translation of Morphological Structures in Spanish Texts

Traducción Automática a Lectura Fácil de Estructuras Morfológicas en Español

Mari Carmen Suárez-Figueroa,[1] **Isam Diab,**[1] **Álvaro González Sanz,**[1] **Jesica Rivero-Espinosa**[2]
[1]Ontology Engineering Group (OEG), Universidad Politécnica de Madrid (UPM), Madrid, Spain
[2]Inserta Innovación, Madrid, Spain
mcsuarez@fi.upm.es, isam.diab@upm.es, a.gsanz@alumnos.upm.es, jrivero@fundacionONCE.es

Abstract: The Easy-to-Read (E2R) Methodology was created to improve the daily life of people with cognitive disabilities. This methodology aims to present clear and easily understood documents by means of providing a set of guidelines related to the writing and layout of texts. Some of these guidelines focus on morphological aspects that may cause difficulties in reading comprehension. Examples of those guidelines are: (a) to avoid the use of adverbs ending in -mente (-ly in English), and (b) to avoid the use of superlative forms. Currently, the E2R guidelines are applied manually to create easy-to-read text materials. To help in such a manual process, our research line is focused on applying the E2R Methodology in an automatic fashion to improve cognitive accessibility. Specifically, in this paper we present (a) the inclusive design approach for the development of E2R translation methods for avoiding both adverbs ending in -mente and superlative forms, and (b) the initial rule-based methods for adapting those linguistic structures into an Easy-to-Read form.
Keywords: Easy-to-Read, Cognitive Accessibility, Automatic Translation.

Resumen: La Metodología de Lectura Fácil (LF) tiene como objetivo presentar documentos claros y de fácil comprensión para personas con discapacidad cognitiva. Para ello, incluye una serie de pautas relacionadas con la redacción y disposición de textos. Algunas de estas pautas se centran en aspectos morfológicos que pueden causar dificultades en la comprensión lectora, como (a) evitar el uso de adverbios en -mente y (b) evitar el uso de formas superlativas. Las pautas de LF se aplican manualmente, por lo que, para ayudar en este proceso, nuestra línea de investigación se centra en aplicar la Metodología de LF de forma automática con el fin de mejorar la accesibilidad cognitiva en textos en español. Así, en este trabajo presentamos (a) un enfoque de diseño inclusivo en el desarrollo de métodos de adaptación de LF para evitar adverbios en -mente y superlativos, y (b) dos métodos iniciales basados en reglas para adaptar dichas estructuras lingüísticas a versiones en lectura fácil.
Palabras clave: Lectura fácil, Accesibilidad cognitiva, Traducción automática.

1 Introduction

The Easy-to-Read (E2R) Methodology (Inclusion Europe, 2009; AENOR, 2018; Nomura et al., 2010) was created to improve the daily life of people with cognitive disabilities and different sectors of the population who present some difficulties related to reading comprehension processes. The goal of this methodology is to present clear and easily understood by providing a collection of guidelines on both 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. At the present time, the E2R methodology is applied in a manual fashion in the processes of (a) adapting existing documents and (b) producing new materials. The manual adaptation of documents is an iterative process and involves

ISSN 1135-5948 DOI 10.26342/2023-71-15

three key activities: E2R analysis, E2R adaptation and E2R validation. This manual process is labour-intensive, so it would benefit from having support applications.

In this context, this paper presents our efforts for the (semi)-automatic identification and adaptation of two specific E2R guidelines that affect the writing of texts: (a) to avoid the use of adverbs ending in -mente (-ly in English) and (b) to avoid the use of the superlative form of adjectives and adverbs. These two guidelines are considered useful for the daily work of E2R experts by one in every four experts as reported in our previous research (Suárez-Figueroa et al., 2022).

In linguistic terms, these two structures are the result of the so-called process of word formation. Among other word classes, this process gives rise to derived words, those formed from another word by the addition of morphemes (e.g. *teach: teacher*). Both the adverbs ending in -mente and the superlatives are examples of derived words, as they are made up of a base attached to a morpheme (i.e. a suffix). It is important to point out that poor comprehenders' difficulties have to do with derivational morphology, since the process of derivation can change the word class of the base (for instance, from an adjective to an adverb, e.g. *correcta* ('correct'): *correctamente* ('correctly')) (Spencer et al., 2015). Therefore, these structures are normally long words and can clutter the text if used too often, since the more stimuli a text contains, the more cognitive processing the reader requires in the reading comprehension task (Difalcis, Ferreres, and Abusamra, 2020). Further along this line, the reduction of morphological complexity of these structures is related to another E2R guideline that recommends to avoid the use of long words, as they negatively affect language comprehension (Barton et al., 2014).

To improve the E2R adaptation process, tools and applications could be developed. In this regard, research work on an automatic E2R analysis for Spanish is quite scarce. We can only mention an E2R conformance checker called Easy-to-Read Advisor (Suárez-Figueroa et al., 2020). Regarding E2R transformations for Spanish texts in an automatic way, it is worth mentioning Simplext (Saggion et al., 2015), LexSIS (Bott et al., 2012), DysWebxia (Rello, Baeza-Yates, and Saggion, 2013), and easier (Moreno, Alarcon, and Martínez, 2020), which are based on simplification techniques. However, none of the aforementioned works specifically addresses the E2R adaptation of either adverbs ending in -mente nor superlative forms.

To cover this gap, we pose the following research question: "Is it possible to develop an automatic method for adapting sentences written in Spanish that include both adverbs in -mente (-ly in English) and superlative forms into a simpler version, that is, an E2R version?" To answer our research question, we decided to follow an inclusive design approach. This means that first we conducted a user study to involve people with cognitive disabilities in the activity of selecting the best proposal for paraphrasing sentences with such linguistic constructions (adverbs in -mente and superlatives). Based on these results, we have created rule-based methods for identifying and adapting these types of morphological structures to the E2R Methodology and implemented a pair of proofs of concept based on such methods. Since most of E2R experts prefer an application providing suggestions of E2R adaptations (Suárez-Figueroa et al., 2022), when possible we create our methods based on different possible E2R adaptations. The E2R adaptation made by our method can be considered an intralinguistic automatic translation for the specific purpose of having sentences in an E2R version. However, we decided to use the word *adaptation* along the paper since this is the most appropriate terminology in the E2R area.

The rest of the paper is organised as follows: Section 2 is devoted to the state of the art on (a) the linguistic features of both adverbs in -mente and superlatives in Spanish, and (b) the automatic approaches for identifying and adapting these structures. In Section 3 we explain the user study we conducted with the aim of selecting the best proposal for paraphrasing sentences with adverbs ending in -mente and/or superlatives. Section 4 presents our first attempts of methods for adapting both structures to the E2R Methodology as well as the first versions of proofs of concept for those methods. Finally, we present conclusions and future work.

2 State of the Art

As mentioned in Section 1, in this work we concentrate our efforts on developing

initial methods (a) to automatically detect morphologically-derived structures such as adverbs ending in -mente and superlative forms in Spanish, and (b) to adapt them into easy-to-read versions, in the light of the guidelines provided by the E2R Methodology (AENOR, 2018). Thus, this section is devoted to (a) highlight some notes on both linguistic structures to better understand the problematic they pose, and (b) summarize the automatic approaches on the identification and adaptation of such structures.

2.1 Linguistic Features of Adverbs Ending in -mente

Adverbs ending in -mente belong to the so-called adverbs of manner. Formally, they are composed of feminine and singular adjectives (*lenta*: *lentamente*, 'slow: slowly'), or of single-ending adjectives (*feliz*: *felizmente*, 'happy: happily').

These adverbs conform a paradox, not always noticed in classical grammatical studies, but recognised directly or indirectly in modern ones (RAE, 2009). Such a paradox is a consequence of the fact that the noun *manera* ('manner') acquire a very abstract meaning in the paraphrase "de manera + adjective". Hence, when we say *Los votantes acudieron masivamente a las urnas*[1], we literally express a certain 'manera de acudir' ('manner of attendance'), but, at the same time, it is not evident that the property of 'ser masivo' ('to be massive') can express any manner itself (RAE, 2009). As an indirect consequence of this paradox, the classification of adverbs of manner depends on the meaning of the adjectives from which they derive, as well as on the syntactic segment they modify.

Thus, since there are numerous classifications of adverbs and some overlap with others, in accordance with the Spanish Royal Academy (RAE, 2009), we distinguish among the following types of adverbs in -mente: (a) **Subject-oriented adverbs**, which refer to a certain property of an action, but also of the person or thing designated in the concrete situation being described. Some examples of these adverbs are the following[2]: *deliberadamente, descuidadamente, or (in)conscientemente*. Moreover, (b) **Object-oriented adverbs** indicate the way in which the action affects the complement of some predicate. For instance, in *Cortó el pastel profundamente*[3], the adverb *profundamente* ('deeply') conveys indeed a way of slicing, but it mainly reports a certain change of state in the object which receives the cut, and not a situation of the subject who caused the deep cut. Following, (c) **Action-oriented adverbs** refer to certain obstacles that the action has to overcome, or to other circumstances that intrinsically characterise the action (e.g. *arduamente, dificultosamente, dolorosamente, or fácilmente*[4]). Finally, (d) **Adverbs of point of view or relational adverbs** are derived from relational adjectives. Since such adjectives establish a connection with a certain field or domain represented by the modified noun instead of denoting a property or quality, relational adverbs are hardly commutable by the paraphrase "de manera/forma + adjective". For instance, in *Tomaron decisiones políticamente justificadas*[5] the adverb *políticamente* ('politically') does not manifest the manner in which the action is performed.

2.2 Automatic Approaches for the Identification and Adaptation of Adverbs Ending in -mente

In Spanish the attempts on identifying and/or transforming this type of adverbs are way scarce. The automatic identification has been covered in some studies dealing with adverbs in general (Rodrigo, 2011); however, in the case of transformation, the way forward is still open. In contrast, in other languages such as French (Tolone and Voyatzi, 2011) or Portuguese (Baptista, 2018), progress has been made along this way. On the one hand, the main goal of Tolone and Voyatzi (2011) is to extend the adverbial entries of LGLex, a Natural Language Processing (NLP) oriented syntactic resource for French. To do that, they first identify the adverbs ending in -mente, and then, for each type of adverb, they include the different paraphrasing alternatives in the lexical tables of LGLex. On the other hand, Baptista (2018) aimed to provide a comprehensive set of paraphrasing strategies, which can be used in several natural language applications, such as text simplification or even machine translation. In this case, an annotated corpus was used to

[1]Tr.: *Voters went massively to the polls.*
[2]Tr.: *deliberately, negligently, (un)consciously.*
[3]Tr.: *He sliced the cake deeply.*
[4]Tr.: *arduously, difficultly, painfully, easily.*
[5]Tr.: *They made politically justified decisions.*

propose different paraphrases for each type of adverb, taking into account the lexical-semantic information. Then, they formalised such paraphrases through a finite state automata to transform the equivalent adverb.

2.3 Some Notes on Superlative Structures

Superlative is understood as the ponderation in maximum or minimum degree of quantity or quality (Olmos, 2001). The superlative degree can be denoted by grammatical categories such as adjectives and adverbs. The expression of the superlative can be performed by means of numerous formulas (Calvo, 1984): (a) **Morphemic expression**, also called synthetic form, which is expressed through the process of affixation on adjective and adverbial bases, i.e. the use of prefixes such as "super- + base" (e.g. *supergrande* ('superlarge') and suffixes such as "base + -ísimo/a[6] or -érrimo/a[7]" (e.g. *grandísimo* ('mighty') or *libérrimo* ('very free')). Another formula is the so-called (b) **Lexical expression**. In that case, the expression of the superlative occurs when a word carries itself the feature 'superlative', such as some quantifier adverbs (e.g. *demasiado* ('too much')), absolute adjectives (e.g. *excelente* ('excellent')), or adverbs ending in -mente, e.g. *completamente* ('completely')). For its part, the (c) **Syntagmatic expression**, also called analytic form, it is denoted by the adverb mark *muy* ('very') preceding an adjective or adverb (e.g. *muy triste* ('very sad')). Such a formula is the most common in Spanish (Olmos, 2001). Finally, superlative structures can be made up by some expressions as *un montón, miles de, una barbaridad* (all of them mean 'a lot' and belong to a colloquial register of language), called (d) **Locutions or superlative expressions.**

Since the superlative structure affects the degree of the word it quantifies, it can only occur with words that are susceptible to gradation, which means that it must be possible to place them on a scale of comparison, at a position higher or lower than the one indicated by the adjective or adverb alone. In the case of adjectives, most of the so-called qualifying adjectives accept gradation, with some exceptions such as the adjectives of extreme degree (also called *elatives*) which correspond to the aforementioned lexical expression of superlative, since they express themselves the maximum degree of gradation, so they do not admit any type of affixation (e.g. *fabuloso: *fabulosísimo; enorme: *enormísimo*[8]). Due to their affective connotations, adjectives derived through this suffix (in this case we only refer to the form -ísimo/a) are very rare in scientific and technical language, but very frequent in colloquial language (RAE, 2009).

At last but not least, following Olmos (2001) the morphemic expression using the suffixes -ísimo/a and -érrimo/a represent an exceptional paradigm in the word derivation process, due to the excessive length of the result (the suffixes add three syllables to the base) and the resulting proparoxytone[9] schema of the superlative, which is not common in the Spanish phonetics and can be a hard-to-pronounce word.

2.4 Automatic Approaches Addressing Superlatives

Although superlative forms received considerable attention in formal linguistics (Bos and Nissim, 2006), this interest is not mirrored in computational linguistics and NLP. For such a reason, the study carried out by Bos and Nissim (2006) is seen as the first automated approach to the interpretation of superlatives for open-domain texts in English. In such a work, they present a corpus annotated for superlatives and propose an interpretation algorithm that uses a wide-coverage parser. The system they implemented is able to recognise a superlative expression and its comparison set.

In addition, Jindal and Liu (2006) studied the identification of superlatives in the framework of identifying comparative sentences in evaluative texts, and extracting comparative relations from them. To achieve such aims,

[6]The suffixes -ísimo and -érrimo refer to the masculine gender, while -ísima and -érrima to the feminine gender. (Henceforth, we will use the slash symbol (/) to include both genders (e.g. -ísimo/a)).

[7]Both suffixes express the same superlative feature. However, superlatives ending in -ísimo/a are widespread used in comparison to those in -érrimo/a.

[8]In the colloquial register it is common to use the suffix -ísimo/a in elative adjectives to focus on the meaning of the adjective, so we can find this type of structures in spoken language, even though they are non-normative. (Henceforth, the asterisk symbol (*) will be used to indicate ungrammatical or non-normative structures).

[9]Linguistic term for a word with stress on the antepenultimate (third last) syllable such as the words in English *cinema* and *operational*.

they proposed two techniques to perform the tasks, based on class sequential rules and label sequential rules. Furthermore, it is worth mentioning the work led by Scheible (2007), who proposed a computational treatment of superlatives aimed to automatically extract useful information from superlatives occurring in free text. Further on, the author extended this work (Scheible, 2009).

In the case of Spanish language, to the best of our knowledge, there are no automatic approaches on the treatment of superlatives both in identification and/or transformation into simpler paraphrases.

3 E2R Adaptation Design via a User Study

As mentioned in Section 1, we decided to apply an inclusive design approach including people with cognitive disabilities in the team in charge of designing the most appropriate E2R adaptation for both adverbs ending in -mente and morphemic superlatives. For this purpose, we developed a pair of questionnaires written in Spanish and implemented as a Google Form[10]. The main goal of these questionnaires is to gather opinions of people with cognitive impairments on the use of adverbs ending in -mente and superlatives. Such questionnaires were launched in February 2022 through mailing lists of autonomic federations and associations of people with cognitive impairments in Spain.

3.1 User Study Design

Both questionnaires are divided into two main parts: (1) a section that includes single-answer multiple choice questions to capture data about which linguistic structures (adverb paraphrasing formulas in the case of adverbs ending in -mente and superlative forms in the other case) are easier or better understood; and (2) a part with questions related to the participants' demographics, knowledge, background, and experience. In the first part of the questionnaires, questions[11] are of two different types. Such types of questions were validated by an E2R expert in a pilot survey before the user study begins. On the one hand, the questionnaire poses questions where the participant has to choose an answer. In this case, participants are asked about their preferences or about the simplest answer. The possible answers to those questions consist of (a) sentences including original linguistic structures (i.e. adverbs ending in -mente or a morphemic superlative form that used a suffix (e.g. -ísimo/a)) and (b) one or more sentences that are the result of adapting the original linguistic structures with an E2R approach in mind, that is, trying to find a synonym formula which is easier to understand. On the other hand, we can find questions where participants have to complete a sentence by selecting one of the possible answers. The set of answers includes the original linguistic structure and one or more synonym formulas.

The collection of original sentences used in the questionnaire was built using two oral corpus: COSER[12] and C-Or-DiAL[13]. Synonym formulas were manually built by consulting several dictionaries such as WordReference[14] and Reverso[15]. A linguistic expert validated the collection of synonym formulas.

In the case of the questionnaire involving adverbs ending in -mente, the first part is composed of 16 questions of which four include sentences with subject-oriented adverbs, six include sentences with action-oriented adverbs, and six include sentences with adverbs of point of view or relational adverbs, based on the classification presented in Section 2.1. While in the case of the questionnaire about superlatives, the first part is composed of 11 questions including morphemic superlative expressions (see Section 2.3). Out of these questions, nine contain sentences with superlative adjectives and two with superlative adverbs.

3.2 Participants in User Studies

On the one hand, for the survey on adverbs in -mente, 139 people responded the questionnaire (72 male, 60 female, and seven participants who preferred not to provide gender information). The participants include representatives from five different autonomous communities in Spain: Andalusia (73), Madrid (32), the Community of Valencia (31), one participant from Galicia and an-

[10] Both questionnaires in PDF format are available at https://doi.org/10.5281/zenodo.7990932.

[11] See Annex A for examples of questions posed in the questionnaires.

[12] https://hispanismo.cervantes.es/recursos/coser-corpus-oral-sonoro-del-espanol-rural

[13] http://lablita.it/app/cordial/corpus.php

[14] https://www.wordreference.com/

[15] https://synonyms.reverso.net/sinonimo/

other one from Catalonia. Most of the participants (51.1%) had a medium level[16] of reading comprehension, whereas 35.3% had a high level, 0.7% a medium-high level, 5.8% a low level, 4.3% of the participants declined to provide their level, while 2.9% did not know about it. Regarding the age range, half of the participants ranged from 31 to 45 years old, 25.9% from 18 to 30, 18.7% were from 46 to 60, two participants were over 60 years old, and 3.6% of the participants declined to provide their age. With respect to their impairments, most of the participants (80.3%) had an intellectual disability, followed by those (8%) who had intellectual and physical disabilities. On the most frequent occupation of the participants, from the 139 participants, 68 were users of occupational centres, 10 were public examination candidates, 7 were unemployed, while 7 were E2R validators.

On the other hand, 121 participants (62 male, 54 female, and 5 participants who preferred not to provide gender information) took part in the survey on superlative forms. In this case, these participants came from seven different autonomous communities in Spain: Andalusia (68), the Community of Valencia (29), Madrid (20), one participant from Galicia, one from Catalonia, one from the Community of Castilla y León, and another one from Extremadura. Most of the participants (80.2%) answered the questionnaire alone, while the rest needed the support of another person. On the occupation of the participants, 58.7% were occupational center users. With respect to participant's level of reading comprehension, 48.8% had a medium level of reading comprehension, while 0.8% had a very high lever, 34.7% had a high level, 0.8% a medium-high level, 9.9% a low level, 1.7% of the participants declined to provide their level, while 0.8% did not know about their level and 2.5% did not know how to read. On the age range, around half of the participants (52.1%) ranged from 31 to 45 years old, 21.5% from 18 to 30, 20.7% were from 46 to 60, 2.5% were over 60 years old, 2.5% of the participants declined to provide their age and 0.8 provides a non-valid age. With respect to their impairments, most of the participants (86%) had an intellectual disability, followed by those (6.6%) who had intellectual and physical disabilities.

[16]The level of comprehension is based on a self-assessment question.

3.3 User Study Outcomes: E2R Adaptation Proposals

Findings indicate that, overall, participants consider simpler those sentences that have been manually simplified during the creation of the sentence collection by means of substituting the original linguistic structure (adverbs ending in -mente and superlatives) by other synonym formulas. In the questionnaire on adverbs ending in -mente, only in 3 out of 16 questions the preferred option was the one with the adverb ending in -mente. While in the case of questionnaire involving superlatives, in the 11 questions the preferred option was the one with a synonym paraphrasing of the original superlative form. The detail analysis of the data gathered[17] is provided in Sections 3.3.1 and 3.3.2 for the case of adverbs ending in -mente and the case of superlatives, respectively.

3.3.1 Case 1: Adverbs ending in -mente

The situation in which the preferred option was the original sentence with an adverb ending in -mente includes, in particular, the following adverbs *directamente* ('directly'), *normalmente* ('normally') and *seriamente* ('seriously'). Our first hypothesis for explaining this outcome was that these three words would be very frequent in Spanish. In order to confirm such an hypothesis we analyzed the frequencies of these three adverbs in the list of most frequent words in CORPES (Corpus del Español del Siglo XXI)[18]. However, these adverbs were not in the top of frequencies; indeed, other adverbs in our questionnaire have highest frequencies (e.g. *actualmente* ('currently') and *solamente* ('only')). Nevertheless, in very broad terms, we could interpret this situation from a morphological point of view, since those three preferred adverbs (*directamente* ('directly'), *normalmente* ('normally') and *seriamente* ('seriously')) belong to the group of relational adverbs, and thus they are made up by relational adjectives, as we mentioned in Section 2.1. This type of adjectives do not express a single property or quality of the noun they accompany, but denote a set of properties and link them to those of the modified noun, thus they establish different types of

[17]Data in CSV files are available at https://doi.org/10.5281/zenodo.7990932.

[18]https://www.rae.es/banco-de-datos/corpes-xxi

more complex semantic relations (Demonte, 1999). In addition, the constituent -mente neither alters the semantics of the adjective nor changes its category, hence the adjective still exhibits all the formal properties that it exhibited as an independent adjective (Alfaro, 2007). This means that, for example, in the paraphrase "de manera + adjective" the adjectives *seria*, *normal* and *directa* do not qualify the property of the noun *manera*, but express a set of properties that affect the whole sentence. For such a reason, the commutation of those three adverbs (*directamente, normalmente* and *seriamente*) by the paraphrase "de manera + relational adjective" may seem unusual for persons with reading comprehension difficulties.

A more in depth analysis[19] of the survey data reveals that participants' preferred selections can be classified into the following scenarios:

Scenario A. Sentences that are adapted by using the pattern "de forma + adjective", e.g. original sentence: *El problema de las humedades hay que analizarlo seriamente*[20]; adapted sentence: *El problema de las humedades hay que analizarlo de forma seria*[21]. Two responses fall into this category.

Scenario B. Sentences that are adapted by replacing the adverb ending in -mente by synonym paraphrasing structure, e.g. original sentence: *Pau Gasol es conocido internacionalmente*[22]; adapted sentence: *Pau Gasol es conocido en todo el mundo*[23]. Six survey responses fall into this grouping.

Scenario C. Sentences that are adapted by eliminating the adverb ending in -mente, e.g. original sentence: *María asiste a clase de español solamente un día por semana*[24]; adapted sentence: *María asiste a clase de español Ø un día por semana*[25]. Two survey responses fall into this category.

Scenario D. Sentences that are adapted by replacing the adverb ending in -mente by a synonym word, e.g. original sentence: *Actualmente, la plaza mayor de mi pueblo es más pequeña que antes.*[26]; adapted sentence: *Ahora la plaza mayor de mi pueblo es más pequeña que antes.*[27]. Three survey responses fall into this group. In this case, it is worth mentioning that the sentence with the adverb ending in -mente was the third preferred option in all the questions. In addition, in Question 6 the first preferred sentences were those that have the same direct synonym but in a different position in the sentence.

Figure 1 in Annex B shows, as an illustration, which percentage of participants selected the most preferred option and what percentage the sentence with adverb ending in -mente in the four identified scenarios. As we can see in the figure, there is a clear difference between the percentage of participants who selected the most preferred option and those who selected the sentence with the adverb in -mente. In fact, the percentages for the most preferred options are in almost all cases more than double the percentages for the sentences with adverbs in -mente.

3.3.2 Case 2: Superlatives

A deeper analysis of the survey data reveals that participants' preferred selections can be classified into the following scenarios:

Scenario A. Sentences that are adapted by directly using the original adjective or adverb, without including any mark for its gradation, e.g. the original sentence: *Había muchísimos alumnos en la clase del profesor Martín*[28] and the preferred adapted sentence: *Había muchos alumnos en la clase del profesor Martín*[29]. Five survey responses fall into this category. In this scenario an interesting point is the fact that we identified two behaviour patterns with respect to the second preferred option choose by participants.

- **Pattern A.1.** Synonym of the original adjective or adverb is used as an E2R paraphrasing. For instance, *feliz* ('happy') is a common synonym for *contento* ('pleased'). Three responses fall into this category.
- **Pattern A.2.** The original adjective or adverb, including the gradation mark *muy*

[19] The aggregation of ratings was performed manually grouping the different percentages of response types (no statistical analysis tool was used).

[20] Tr.: *The problem of dampness needs to be seriously considered.*

[21] Tr.: *The problem of dampness needs to be considered in a serious way.*

[22] Tr.: *Pau Gasol is internationally known.*

[23] Tr.: *Pau Gasol is known all over the world.*

[24] Tr.: *María attends classes only one day a week.*

[25] Tr.: *María attends classes one day a week.*

[26] Tr.: *Currently, the main square in my village is smaller than before.*

[27] Tr.: *Nowadays, the main square in my village is smaller than before.*

[28] Tr.: *There were a lot of students in Professor Martin's class.*

[29] Tr.: *There were many students in Professor Martin's class.*

('very') is used as E2R paraphrasing; e.g. *rarísimo* ('strange') is substituted by *muy raro* ('very strange'). Two survey responses fall into this category. Given that, this Scenario A indicates the participants' preference in selecting the lexical expression of superlative when (a) an extreme-degree adjective is proposed as synonym of the morphemic expression ending in -ísimo/a (e.g. most participants prefer the extreme-degree synonym adjective *feliz* ('happy') instead of the morphemic form *contentísimo* ('very pleased')); and (b) the morphemic expression with the suffix -ísimo/a is added to a extreme-degree adjective or adverbial base (e.g. most participants prefer the original adjective *enojada* ('angry') instead of its morphemic version of superlative *enojadísima* ('very angry'), since *enojada* carries itself the maximum degree of superlative). Hence, we can find at a glance that participants consider easier the lexical expression of the superlative rather than its morphemic version with the suffix -ísimo/a.

Scenario B. Sentences that are adapted by replacing the original superlative form by the original adjective or adverb, including the gradation mark *muy* ('very'), e.g. the original sentence: *El pan de la panadería de Mario está riquísimo*[30]; and the preferred adapted sentence: *El pan de la panadería de Mario está muy rico*[31]. Six questions fall in this grouping. We observed that those six cases correspond to the so-called restrictive qualifying adjectives, which point out characteristics that distinguish nouns among their peers, i.e. for instance, in the previous example *El pan de la panadería de Mario está muy rico*, the adjective *rico* is "differentiating" the bread from the rest. Some other examples of restrictive qualifying adjectives in this Scenario B are *majo* ('nice') or *simpático* ('friendly').

Although both the morphemic expression (suffix -ísimo/a and -érrimo/a) and the syntagmatic expression (formula "*muy* + adjective or adverb") are considered synonymous structures for expressing superlative, there are semantic differences in their use. According to (Olmos, 2001), the synthetic form with -ísimo/a expresses a higher gradation than the analytic formula "*muy* + adjective or adverbial base". However, thanks to the findings extracted we see that people with reading comprehension difficulties lean towards the latter form using *muy*, either due to the phonetic and morphological difficulties posed by the suffix -ísimo mentioned in 2.3, or due to the common use of the formula "*muy* + adjective or adverb" in Spanish.

Figure 2 in Annex B presents which percentage of participants selected the most preferred option for those responses that fall into Scenario A (five responses) and Scenario B (six responses), respectively. As can be seen in the figure, there is a clear difference between the percentage of participants who selected as the most preferred option the superlative paraphrasing (original adjective or adverb, with or without the gradation mark *muy* ('very')) and those who selected the sentence with the superlative form using a suffix. In fact, the percentages for the most preferred options are in almost all the cases more than double the percentages for the sentences with superlative structures. The only case in which this statement is not valid is the case of Question 6 (Scenario B). For such a question the most preferred option was *Ana es muy guapa*[32] and the option with the superlative was *Ana es guapísima*[33].

4 Initial Methods for an E2R Adaptation of Morphological Structures

The final aim of the proposed methods is (a) to detect a couple of morphological structures in texts written in Spanish, and (b) to replace such structures by the most appropriate paraphrasing formula. The selected paraphrasing formulas are based on the data gathered from people with cognitive disabilities during the inclusive design action explained in Section 3. The initial methods are composed of the following activities: (1) NLP, which includes a cleanup of the text using regular expressions and a tokenization step, (2) Morphological Structure Identification, and (3) Morphological Structure Adaptation. The following sections explain the E2R adaptation method for adverbs ending in -mente (Section 4.1) and the method for adapting morphemic superlatives ending in -ísimo/a or -érrimo/a (Section 4.2).

[30] Tr.: *The bread from Mario's bakery is delicious.*
[31] Tr.: *The bread from Mario's bakery is very tasty.*
[32] Tr.: *Ana is very pretty.*
[33] Tr.: *Ana is gorgeous.*

4.1 Case 1: Adverbs ending in -mente

Adverbs identification activity relies first on a pre-filter of the words in the text to obtain only those words ending in -mente, and second on a selection of the words that are adverbs from this pre-filtered set. The activity of adapting adverbs was conceived as an activity to provide first the most appropriate substitution and, second, an ordered list with other possible substitutions manually crafted. This design decision was based on the E2R experts preferences about having adaptation suggestions in a support application (Suárez-Figueroa et al., 2022). All possible substitutions are based on the data gathered from people with cognitive disabilities described in Section 3. Specifically, the method proposes as first adaptation the option described in Scenario B, i.e. a synonym paraphrasing structure. This was the first option because it is the most recurrent situation as shown in Section 3. Currently, the method uses a declarative mapping catalogue with adverbs ending in -mente and possible synonym paraphrasing structures. The ordered list of possible substitutions is created with (1) the option presented in Scenario D, (2) the option described in Scenario A, and (3) the option shown in Scenario C.

We have developed a proof of concept, as a RESTful web service[34], to detect adverbs ending in -mente in Spanish texts and replacing them by the most appropriate E2R translation formula, based on the aforementioned method. The service requires as input a sentence in Spanish and provides as output an easier version of the original sentence in which adverbs in -mente have been adapted with an E2R approach in mind. Both the input and the output are JSON (JavaScript Object Notation) objects.

The developed service has been implemented in Python 3.9, using the development framework Flask. This services uses LibrAIry[35] (Badenes-Olmedo, García, and Corcho, 2017) for detecting adverbs ending in -mente. In particular, our service uses the following LibrAIry functionalities: Part-of-Speech tagging and stemming. We tested the identification functionality of our service with a collection of 2502 sentences written in Spanish extracted from minutes of sessions held in municipalities[36]. All sentences included adverbs in -mente. The tests used these 2502 texts and 2433 adverbs were correctly identified. Thus, such a functionality has a 97% of success identifying adverbs ending in -mente. An on-going work is the design of the tests for evaluating the adaptation functionality.

4.2 Case 2: Superlatives

Superlatives identification activity relies first on the PoS tagging information, in which EAGLE tags[37] are used.

As in Case 1, the activity of adapting superlatives was conceived as an activity for providing first the most appropriate substitution and, second, an ordered list with other possible substitutions manually crafted. The idea is to have adaptation suggestions as preferred by E2R experts (Suárez-Figueroa et al., 2022). All possible substitutions are based on the data gathered from people with cognitive disabilities described in Section 3. As a first approximation, our initial method proposes as main adaptation the most recurrent situation shown in Section 3; that is, the option described in Scenario B (to use "muy + adjective or adverb". as a synonym paraphrasing structure). The ordered list of possible substitutions is created with (1) the option presented in Scenario A, that is, using the original adjective or adverb, without including any mark for its gradation; and (2) a synonym of the original adjective or adverb (Pattern A.1). After selecting the most appropriate substitution, a post-processing is performed to maintain the syntactic coherence in the sentence.

We have developed a proof of concept, as a RESTful web service[38], for superlatives in Spanish texts and replacing them by the most appropriate E2R translation formula, based on the aforementioned method. Such a service requires as input a sentence written in Spanish and provides as output an easier version of the original sentence in which superlatives have been adapted with an E2R

[34]The service is not currently available due to privacy constrains in the context of the project in which such a service has been developed.

[35]http://librairy.linkeddata.es/nlp/api.html

[36]https://ayuntamientoboadilladelmonte.org/tu-ayuntamiento/gobierno-municipal

[37]https://www.cs.upc.edu/~nlp/tools/parole-sp.html

[38]The service is not currently available due to privacy constrains in the context of the project in which such a service has been developed.

approach in mind. Both the input and the output are JSON objects.

The developed service has been implemented in Python 3.9, using the development framework Flask. This services uses spaCy[39] for sentence tokenization and TextServer (Padró and Turmo, 2015) for tagging the obtained tokens with EAGLE PoS tags. The superlatives detection is based on a filtered process using the following PoS tags: category A (adjective) and degree S (superlative). When the superlatives have been identified, TextServer is used again in order to get the gender and number of such superlatives. This information is needed to maintain the syntactic coherence in the adapted sentence. We tested the service with a collection of 567 sentences including superlative forms randomly extracted from the CREA corpus[40]. Currently, the proof of concept has a 95% success rate in identifying superlatives, with 540 out of 567 tests passed correctly. At present we are designing the tests for evaluating the adaptation functionality.

5 Conclusions and Future Work

With the general aim of improving cognitive accessibility of texts written in Spanish, in this paper, we present a pair of rule-based methods for adapting sentences which include morphologically-derived structures, such as adverbs ending in -mente and superlative forms, following an E2R approach. Thus, we can respond in a positive manner to the research question presented in Section 1 ("Is it possible to develop automatic methods for adapting sentences written in Spanish that include both adverbs ending in -mente (-ly in English) and superlative forms into a simpler version, that is, an E2R version?").

A crucial task in the development of those methods was the selection of the most appropriate paraphrasing formula for such morphologically-derived structures. To perform this task we applied an inclusive design approach involving people with cognitive disabilities. Their involvement was materialize by participating in two surveys. In those surveys participants were asked about their preferences with respect to the simplicity of a collection of short sentences written in Spanish. Thanks to the data collected in such questionnaires, we discovered the patterns to be used in the E2R adaptation of adverbs ending in -mente and superlatives. As further research, we have planned two specific directions. On the one hand, we are going to analyse in more depth the data gathered in our inclusive design activity with the aim of finding linguistic patterns to explain the different decisions taken by participants with respect to the types of such structures and the identified scenarios. On the other hand, several technical actions are planned: (a) we are going to create the test collections for evaluating the adaptation functionality of our services. The plan is to involve people with cognitive disabilities and E2R experts in such an evaluation; (b) we are going to investigate different ways to obtain the equivalences between adverbs in -mente and possible synonym paraphrasing structures; and (c) we consider to integrate our services in a web user interface. After these technical improvements, we have planned to perform a user study to test both the user interface and the services.

[39] https://spacy.io/

[40] https://www.rae.es/banco-de-datos/crea/crea-version-anotada

Acknowledgements

This research has been 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 managing the participation of different organizations in the inclusive design activity. We also thank the Federations of Organizations of people with intellectual or developmental disabilities in Madrid, Valencia and Andalusia for their participation in such activity.

References

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

Alfaro, A. F. 2007. Adverbios en –mente y la estructura del adjetivo en español. *Revista de Lengua Española y Lingüística en General*, 21.

Badenes-Olmedo, C., J. L. R. García, and Ó. Corcho. 2017. Distributing text mining tasks with *librAIry.* In K. P. Camilleri and A. Bonnici, editors, *Proceedings of the 2017 ACM Symposium on Document En-*

gineering, DocEng 2017, Valletta, Malta, September 4-7, 2017, pages 63–66. ACM.

Baptista, J. 2018. Paráfrase de advérbios terminados em –mente em Português. *Linguamática*, 10(2):21–30.

Barton, J., H. Hanif, L. Björnström, and C. Hills. 2014. The word-length effect in reading: A review. *Cognitive neuropsychology*, 31.

Bos, J. and M. Nissim. 2006. An Empirical Approach to the Interpretation of Superlatives. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, pages 9–17, Sydney. Association for Computational Linguistics.

Bott, S., L. Rello, B. Drndarevic, and H. Saggion. 2012. Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish. In *Proceedings of COLING 2012*. The COLING 2012 Organizing Committee.

Calvo, J. M. G. 1984. Sobre la expresión de lo "superlativo" en español (I). *Anuario de estudios filológicos*, 7:173–205.

Demonte, V. 1999. El adjetivo: clases y usos. La posición del adjetivo en el sintagma nominal. In Ignacio Bosque and Violeta Demonte, editor, *Gramática descriptiva de la lengua española*, volume 1, pages 129–216.

Difalcis, M., A. Ferreres, and V. Abusamra. 2020. Syllabic length effect in spanish - evidence from adult readers. *Ocnos Revista de Estudios sobre Lectura*, 19:19–28, 11.

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

Moreno, L., R. Alarcon, and R. Martínez. 2020. EASIER System. Language Resources for Cognitive Accessibility. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20. Association for Computing Machinery.

Nomura, M., 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.

Olmos, B. P. 2001. El afijo -ísimo en el español actual. *Verba: Anuario galego de filoloxia*, 28:159–185.

Padró, L. and J. Turmo. 2015. Textserver: Cloud-based multilingual natural language processing. pages 1636–1639, 11.

RAE. 2009. *Nueva gramática de la lengua española*. Espasa Calpe.

Rello, L., R. Baeza-Yates, and H. Saggion. 2013. DysWebxia: Textos más Accesibles para Personas con Dislexia. *Procesamiento del Lenguaje Natural*, 51:205–208.

Rodrigo, A. 2011. *Tratamiento automático de textos: el sintagma adverbial núcleo*. Ph.D. thesis, Universidad Nacional de Rosario.

Saggion, H., S. Stajner, S. Bott, S. Mille, L. Rello, and B. Drndarevic. 2015. Making It Simplext: Implementation and Evaluation of a Text Simplification System for Spanish. *ACM Transactions on Accessible Computing*, 6(4).

Scheible, S. 2009. *A Computational Treatment of Superlatives*. Ph.D. thesis, University of Edinburgh, Edinburgh, UK.

Spencer, M., A. Muse, R. Wagner, B. Foorman, Y. Petscher, C. Schatschneider, E. Tighe, and D. Bishop. 2015. Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and Writing*, 28.

Suárez-Figueroa, M. C., I. Diab, E. Ruckhaus, and I. Cano. 2022. First steps in the development of a support application for easy-to-read adaptation. *Universal Access in the Information Society*.

Suárez-Figueroa, M. C., E. Ruckhaus, J. López-Guerrero, I. Cano, and Á. Cervera. 2020. Towards the Assessment of Easy-to-Read Guidelines Using Artificial Intelligence Techniques. In K. Miesenberger, R. Manduchi, M. C. Rodriguez, and P. Peňáz, editors, *Computers Helping People with Special Needs. ICCHP 2020*, volume 12376 of *Lecture Notes in Computer Science*, pages 74–82. Springer.

Tolone, E. and S. Voyatzi. 2011. Extending the adverbial coverage of a NLP oriented resource for French. In *Proceedings*

of 5th International Joint Conference on Natural Language Processing, pages 1225–1233. Asian Federation of Natural Language Processing.

A Annex 1

Tables 1 and 2 show two examples of the type of questions from the questionnaires on adverbs in -mente and superlatives, respectively. On the one hand, in Table 1, the user is asked to complete the beginning of the sentence[41] with one of the options provided. These options include (1) the adverb in -mente *Definitivamente* ('definitely'), (2) the expression *Sin lugar a dudas* ('without doubt'), (3) the expression *Está claro que* ('it is clear that'), (4) the structure *En definitiva ('in the end')*, the (5) the option "I don't know", and, finally, (6) the users can freely write their own proposal. On the other hand, the question in Table 2 asks the user to select the simplest sentence out of the following options: (1) the sentence with the morphemic superlative (*lejísimos*, 'far away'), (2) the sentence with the superlative expression (*muy lejos*, 'very far'), (3) the choice that neither sentence is simple, (4) the choice that both sentences are simple, (5) the option "I don't know", and, finally, (6) the user can freely paraphrase the sentence.

Question
Por favor completa el principio de esta frase con una de las opciones: "..... Ana necesita más vocabulario técnico."
Options
1. Definitivamente
2. Sin lugar a dudas
3. Está claro que
4. En definitiva
5. No lo sé
6. Otro: (texto libre)

Table 1: A sample question from the questionnaire related to adverbs in -mente.

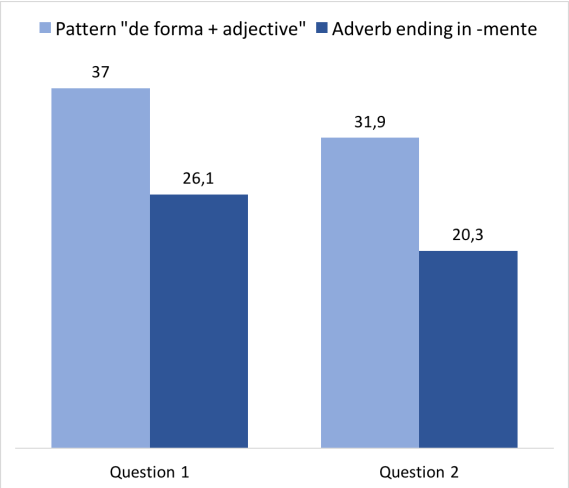
Question
Por favor, lee las siguientes 2 frases. ¿Qué frase te parece más sencilla?
Options
1. David vive lejísimos de su prima Anabel.
2. David vive muy lejos de su prima Anabel.
3. Niguna de las frases es sencilla.
4. Todas las frases son sencillas.
5. No lo sé.
6. Otro: (texto libre)

Table 2: A sample question from the questionnaire related to superlatives.

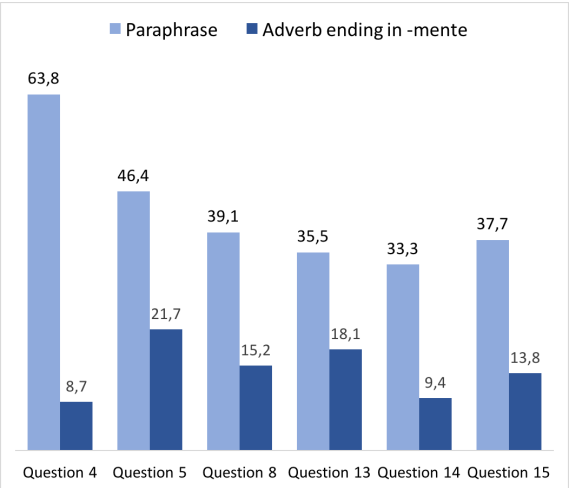
B Annex 2

Both Figure 1 and Figure 2 show the percentages of selections for the most preferred options in the corresponding scenarios related to the questionnaires.

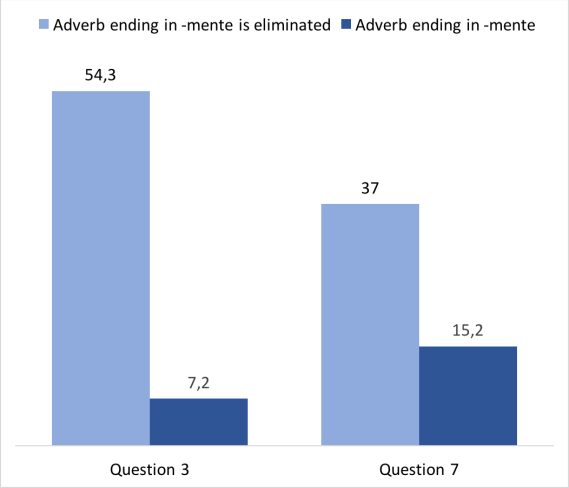
[41] The translation of the sentence that the users had to complete is the following: *[Option] Ana needs more technical vocabulary.*



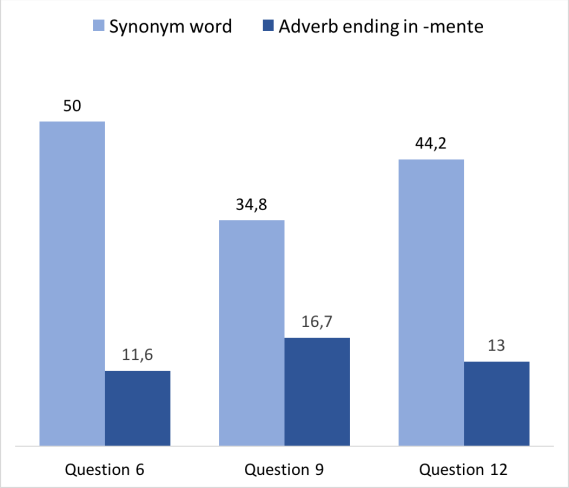
(a) Scenario A



(b) Scenario B

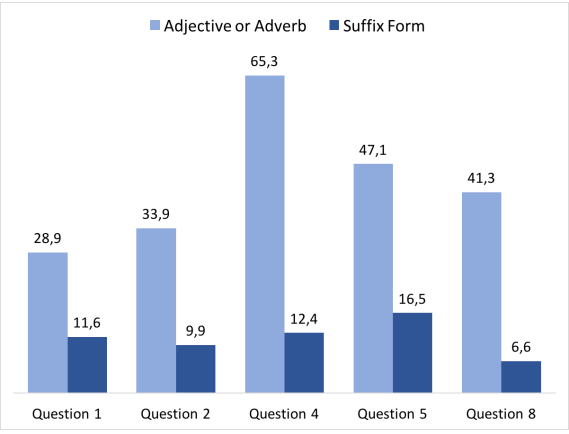


(c) Scenario C

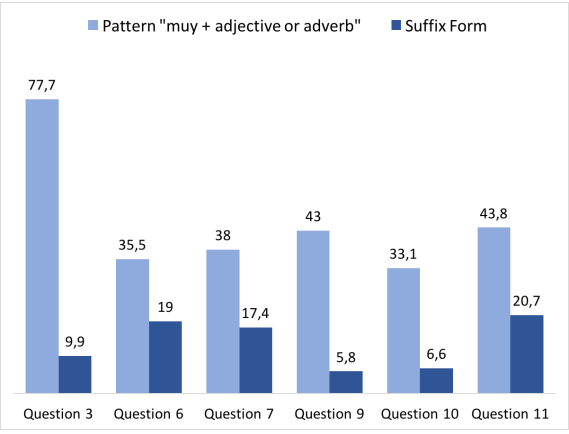


(d) Scenario D

Figure 1: Percentage of selections for the most preferred option and for the sentence with adverb ending in -mente in the four identified scenarios.



(a) Scenario A



(b) Scenario B

Figure 2: Percentage of selections for the most preferred paraphrase and for the sentence with the superlative (using the suffix form) in the two identified scenarios.