Lexical simplification for the systematic support of cognitive accessibility guidelines

by

Rodrigo Alarcón García

A dissertation submitted by in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Computer Science and Technology

Universidad Carlos III de Madrid

Advisor(s):

Lourdes Moreno López

January 2022

To my parents, brother and my girlfriend for their constant support and understanding. For always being there for me, whether it was a good or bad experience.

To my advisor, for having been my guide both inside and outside the doctoral program. I thank her for her great patience with me.

And finally, to my friends, who, living so far away from home, never lost contact with me, supporting me at every moment and wishing me the best.

Published and submitted content

The research publications realized by the author are partially or fully included for the methods proposed in the present document and are listed as follows:

1. **Journals:**

   - **Rodrigo Alarcón**, Lourdes Moreno, Isabel Segura Bedmar, Paloma Martínez Fernández. (2019). Lexical simplification approach using easy-to-read resources. Procesamiento del Lenguaje Natural. 63, 95-102. Sociedad Española para el Procesamiento del Lenguaje Natural, 1989-7553. 2019, Septiembre. [SCImago Journal Rankings (SJR), Computer Science Applications, 0,270, 3]. Partially included in Chapters 3 and 4.
   - **Rodrigo Alarcón**, Lourdes Moreno, Paloma Martínez. (2021). Lexical Simplification System to Improve Web Accessibility. IEEE Access. 9, 58755-58767. 2169-3536. 2021, Abril. 10.1109/ACCESS.2021.3072697. [Computer Science, Information Systems, 3.745, Q1]. Partially included in Chapters 1, 2, 3 and 4.

2. **Conferences:**

   - Lourdes Moreno, **Rodrigo Alarcón**, Isabel Segura Bedmar, Paloma Martínez Fernández. (2019). Lexical simplification approach to support the accessibility guidelines. Proceedings of the XX International Conference on Human Computer Interaction, Interacción 2019. Donostia, Gipuzkoa, Spain. 2019, Junio. ACM, 978-1-4503-7176-6. 14:1-14:4. 10.1145/3335595.3335651. Partially included in Chapters 2, 3 and 4.
   - **Rodrigo Alarcón**, Lourdes Moreno, Paloma Martínez Fernández. (2020). Hulat - ALexS CWI task - CWI for Language and Learning Disabilities applied to University Educational Texts. ALexS 2020: Lexicon Analysis Task. IberLEF (Iberian Languages Evaluation Forum) co-located with SEPLN 2020. Málaga, Spain. 2020, Septiembre. CEUR-WS.org, 1613-0073. 2664, 24-30. Partially included in Chapters 2, 3 and 4.
   - Lourdes Moreno, **Rodrigo Alarcón**, Paloma Martínez Fernández. (2020). EASIER system. Language resources for cognitive accessibility. 22nd International ACM SIGACCESS Conference on Computers and Accessibility.

ASSETS 2020. (Virtual). 2020, Octubre. ACM Digital Library, [GII-GRIN-SCIE, GGS Class 3]. Partially included in Chapters 2, 3, 4 and 5.

- **Rodrigo Alarcón**, Lourdes Moreno, Paloma Martínez Fernández. (2020). Word-Sense disambiguation system for text readability. DSAI 2020 (9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion). 2020, Diciembre. ACM Digital Library. 147–152. https://doi.org/10.1145/3439231.3439257. Partially included in Chapters 2, 3 and 4.
- **Rodrigo Alarcón**. (2020). Simplificación Léxica para la Mejora de la Accesibilidad Cognitiva (Lexical Simplification to Improve Cognitive Accessibility). PLNnet-DS-2020 (Proceedings of the Doctoral Symposium on Natural Language Processing). Jaén, Spain. 2020, Diciembre. CEUR-WS.org, 1613-0073. 2802, 1-7. Partially included in Chapters 2, 3 and 4.
- Lourdes Moreno, **Rodrigo Alarcon**,Paloma Martnez. (2021). Designing and Evaluating a User Interface for People with Cognitive Disabilities. Interacción '21.XXI International Conference on Human Computer Interaction. September 22–24, 2021. Málaga, Spain. Partially included in Chapters 2 and 5.
- **Rodrigo Alarcón**, Lourdes Moreno, Paloma Martínez. (2021). Exploration of Spanish Word Embeddings for Lexical Simplification. First Workshop on Current Trends in Text Simplification (CTTS-2021) (SEPLN 2021). Málaga, Spain. 2021, Septiembre. CEUR. Vol-2944. Partially included in Chapters 2, 3 and 4.

Abstract

The Internet has come a long way in recent years, contributing to the proliferation of large volumes of digitally available information. Through user interfaces we can access these contents, however, they are not accessible to everyone. The main users affected are people with disabilities, who are already a considerable number, but accessibility barriers affect a wide range of user groups and contexts of use in accessing digital information. Some of these barriers are caused by language inaccessibility when texts contain long sentences, unusual words and complex linguistic structures. These accessibility barriers directly affect people with cognitive disabilities.

For the purpose of making textual content more accessible, there are initiatives such as the Easy Reading guidelines, the Plain Language guidelines and some of the language-specific Web Content Accessibility Guidelines (WCAG). These guidelines provide documentation, but do not specify methods for meeting the requirements implicit in these guidelines in a systematic way. To obtain a solution, methods from the Natural Language Processing (NLP) discipline can provide support for achieving compliance with the cognitive accessibility guidelines for the language.

The task of text simplification aims at reducing the linguistic complexity of a text from a syntactic and lexical perspective, the latter being the main focus of this Thesis. In this sense, one solution space is to identify in a text which words are complex or uncommon, and in the case that there were, to provide a more usual and simpler synonym, together with a simple definition, all oriented to people with cognitive disabilities.

With this goal in mind, this Thesis presents the study, analysis, design and development of an architecture, NLP methods, resources and tools for the lexical simplification of texts for the Spanish language in a generic domain in the field of cognitive accessibility. To achieve this, each of the steps present in the lexical simplification processes is studied, together with methods for word sense disambiguation. As a contribution, different types of word embedding are explored and created, supported by traditional and dynamic embedding methods, such as transfer learning methods. In addition, since most of the NLP methods require data for their operation, a resource in the framework of cognitive accessibility is presented as a contribution.

Resumen

Internet ha avanzado mucho en los últimos años contribuyendo a la proliferación de grandes volúmenes de información disponible digitalmente. A través de interfaces de usuario podemos acceder a estos contenidos, sin embargo, estos no son accesibles a todas las personas. Los usuarios afectados principalmente son las personas con discapacidad siendo ya un número considerable, pero las barreras de accesibilidad afectan a un gran rango de grupos de usuarios y contextos de uso en el acceso a la información digital. Algunas de estas barreras son causadas por la inaccesibilidad al lenguaje cuando los textos contienen oraciones largas, palabras inusuales y estructuras lingüísticas complejas. Estas barreras de accesibilidad afectan directamente a las personas con discapacidad cognitiva.

Con el fin de hacer el contenido textual más accesible, existen iniciativas como las pautas de Lectura Fácil, las pautas de Lenguaje Claro y algunas de las pautas de Accesibilidad al Contenido en la Web (WCAG) específicas para el lenguaje. Estas pautas proporcionan documentación, pero no especifican métodos para cumplir con los requisitos implícitos en estas pautas de manera sistemática. Para obtener una solución, los métodos de la disciplina del Procesamiento del Lenguaje Natural (PLN) pueden dar un soporte para alcanzar la conformidad con las pautas de accesibilidad cognitiva relativas al lenguaje

La tarea de la simplificación de textos del PLN tiene como objetivo reducir la complejidad lingüística de un texto desde una perspectiva sintáctica y léxica, siendo esta última el enfoque principal de esta Tesis. En este sentido, un espacio de solución es identificar en un texto qué palabras son complejas o poco comunes, y en el caso de que sí hubiera, proporcionar un sinónimo más usual y sencillo, junto con una definición sencilla, todo ello orientado a las personas con discapacidad cognitiva.

Con tal meta, en esta Tesis, se presenta el estudio, análisis, diseño y desarrollo de una arquitectura, métodos PLN, recursos y herramientas para la simplificación léxica de textos para el idioma español en un dominio genérico en el ámbito de la accesibilidad cognitiva. Para lograr esto, se estudia cada uno de los pasos presentes en los procesos de simplificación léxica, junto con métodos para la desambiguación del sentido de las palabras. Como contribución, diferentes tipos de word embedding son explorados y creados, apoyados por métodos embedding tradicionales y dinámicos, como son los métodos de transfer learning. Además, debido a que gran parte de los métodos PLN requieren datos para su funcionamiento, se presenta como contribución un recurso en el marco de la accesibilidad cognitiva.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This Chapter shows the context and the motivation that has led to the definition of this Thesis. The objectives and goals to achieve with this Thesis are described. Next, the research hypotheses formulated which drive the research are presented. And finally, the methodology followed during the research process is described.

1.1. Background

We have access to an overwhelming amount of information, but this information is not accessible to everyone. Many texts we come across in our everyday lives are lexically and syntactically very complex. Some people face accessibility barriers when reading texts that contain long sentences, unusual words, complex linguistic structures and others. Although people with cognitive, language and learning disabilities are directly affected, cognitive accessibility barriers affect other user groups such as the deaf, deaf-blind, elderly, illiterate and immigrants with a different native language.

Currently, 25% of the population cannot read documents that contain a large amount of information that needs to be simplified. Moreover, this need for more simplified texts is becoming increasingly important as the number of people with disabilities is growing due to the ageing of the population [1]. Specifically in Spain, the majority of the adult population has difficulty understanding dense texts [2] and 1.7% of the population is functionally illiterate and there are 277,472 people with an intellectual disability [3].

In order to provide universal access to information and make texts more accessible, there are initiatives that work to improve cognitive accessibility to language, such as the Easy Reading guidelines [4] which propose guidelines for adapting texts to make them easier to read and understand. In addition, there is the Plain Language initiative [5] that promotes plain language in the content of the information society . Also, there are the Web Content Accessibility Guidelines (WCAG) [6], which support web developers in providing accessible content. These include criteria that involve offering resources that provide text simplification, which is difficult to fulfil. Few tools exist that provide systematic support for simplification processes. Usually, the websites that offer simplified versions of their main sites are manually created. Manual simplification of written documents is quite costly, mostly because the information is continually being produced.

As part of the solution, automatic text simplification methods, which are found in the Natural Language Processing (NLP) field, provide systematic support to promote compliance with these cognitive accessibility guidelines. This task consists of the process of reducing the linguistic complexity of a text while preserving the original information and meaning [7]. Several approaches to this task have been presented over the years, of which include sentence-level simplifications and lexical simplifications, the latter being the main focus in this Thesis, which consists in replacing words in a given sentence in order to make it simple, without applying any modifications to its syntactic structure [8].

While by definition it may appear to be a simple task, lexical simplification has proven to be a non-trivial task, especially because there are research challenges to address to provide new methods and resources. On the other hand, focusing on the field of disability, the target audience will have different needs, specifically the elderly or people with some type of cognitive disability; in this scenario, resources specifically created oriented to people with disabilities should be used in the processes simplification.

The context of this Thesis is within the NLP discipline. This Thesis is focused on text simplification for Spanish texts in a generic domain, more specifically, on the lexical simplification task, to support audiences with cognitive disabilities.

1.2. Motivation

Recently, the task of text simplification has motivated more and more researchers to advance new approaches. This can offer different research points to address cognitive accessibility barriers in texts to support audiences with cognitive disabilities, which are described below:

- **Systematic support for accessibility requirements:**A large amount of information is produced every day, however, this information is not accessible to the public at large, and specifically, to people with disabilities. Also, there are standards such as WCAG and initiatives such as easy reading and plain language guidelines that indicate accessibility requirements that must be met. In this sense, there is a need to provide systematic methods to simplify text to support developers of websites and applications that include textual content.

- **Basis for Text Simplification:** Lexical simplification aims to perform text simplification by using lexical information from a target text. Many works have shown that lexical simplification can benefit various groups of people, including people with autism [9], aphasia [10], low vision [11], dyslexia [12] or people with intellectual disabilities [13]. In this sense, the production of research related to lexical simplification is always advantageous for many people's text comprehension.

- **Lack of Spanish linguistic resources:** Although the Spanish language is receiving more and more interest in the research and development of NLP resources, it is far from what the English language possesses. Although English is the language that receives the most research, the more time passes, the more interest in the Spanish language increases worldwide, as it is a language that currently has more than 480 million native speakers. In this sense, the need for the production of Spanish language resources is high and always advantageous for research.

- **Lack of linguistic resources in the scope of cognitive accessibility:** The resources produced are written by people who do not fully understand the needs of the target audience such as people with cognitive disabilities and older people. Therefore, it is necessary to increase the quality of existing resources by having texts annotated by people specialized in manual text simplification, with knowledge in plain language and easy reading.

- **Possibilities for new approaches:** It is no surprise that the Spanish language is an ambiguous language, where many words can have different meanings depending on the context in which they are found. This can pose a problem in many aspects of text comprehension, either in determining whether a word is complex or in understanding the meaning of the word. Taking into account these disadvantages, word embedding methods have recently been introduced, which takes large amounts of information and vectorizes it in order to provide semantic and contextualized information of the textual content. Currently, this technology is constantly being improved, so there is a great variety of embedding types, as it has shown good results in disambiguating textual content [14]. In this sense, the exploration of possible applications of word embeddings for text comprehension is valuable for the research community.

- **Support for other NLP tasks:** The simplification of the text serves as a basis for many other crucial areas of information management. It initially started as a preprocessing stage for tasks such as parsing, question generation, information extraction, fact retrieval and semantic role labeling [15]. In these tasks, different methods of text simplification are applied, either at the lexical level, syntactic level or sentence level simplification. In this sense, the production of useful technologies for text simplification represents a wide benefit for other areas in the NLP community.

1.3. Objectives

The main objective of the Thesis focuses on the study, analysis, design and development of architecture, NLP methods, resources and tools for the lexical simplification task in a generic domain for the Spanish language and scope of cognitive accessibility.

This general objective has been defined through the following specific objectives:

- Study and analysis of the state of the art of cognitive accessibility requirements in combination with lexical simplification methods and resources.
- Design of an architecture, methods and resources to support lexical simplification processes in the scope of cognitive accessibility.
- Experimentation and analysis of the results in comparison with state of the art.
- Identification of open research questions through discussion of data and conclusions, as a step towards proposals for future lines of research.

1.4. Research Hypothesis

In this Thesis, the problem of Spanish lexical simplification in a generic domain through the use of cognitive accessibility resources and word embedding combinations supported by NLP techniques is addressed. Based on the objectives set out above, the general Hypothesis can be summarized as follows:

Is it possible to improve the accessibility of Spanish texts in a generic domain using NLP techniques to support audiences with cognitive disabilities?

In order to confirm this research Hypothesis, a breakdown of specific hypotheses taking into account the objectives of this Thesis is performed. These specific hypotheses are described below:

- Hypothesis 1. *Accessibility resources can offer support for lexical simplification steps aimed at people with cognitive disabilities.*

  There is no doubt that obtaining annotated data is very important for the evaluation of NLP methods and the creation of annotated data is always appreciated by researchers, especially for the Spanish language which does not have the amount of resources as the English language. However, many of these data are annotated by volunteers or by a team briefly trained for the task, consequently, obtaining results with a lower quality. Therefore, this Thesis believes that the creation of data annotated by specialists who understand the needs of the target user can offer better support in the evaluation of lexical simplification steps.

  In the framework of this Thesis, an annotated corpus in the field of accessibility is proposed. Consequently, this corpus suggests that this corpus can offer support in the training and validation of methods for the lexical simplification procedures, because the corpus possesses a fair amount of instances that offer detected complex words and simple substitutes.

  Additionally, this Thesis believes that the resources offered by the accessibility area can offer support in other procedures, such as the incorporation of this information as a feature when representing a word in the discernment between a complex and a simple word.

- Hypothesis 2. *Combining different types of word embeddings can improve results in the steps in the lexical simplification process.*

  Word embedding technology has gained popularity in the NLP community and has led to significant advances in a wide variety of tasks. These embeddings can be described as a way of mapping words into an n-dimensional space.

  Models created by this method attempt to capture as much context information as possible in a word, and may even contain semantic and syntactic information. Currently, these models can be classified as static representations (or classical embedding) and contextual representations. The former focuses on obtaining autonomous representations that do not depend on the context (words surrounding the target word). The second represents an improvement on the former, since it incorporates the context in its representations [16].

  Therefore, this Thesis takes advantage of the versatility of this technology by exploring and applying different methods that use information from a variety of embedding models, with the objective of providing solutions to the different stages of the lexical simplification task and the word sense disambiguation task.

- Hypothesis 3. *Transfer learning approaches can improve results in the steps in the lexical simplification process.*

  Transfer learning methods have gained popularity in recent years. These methods consist of optimizing a model trained for one task in order to use it for a different task and aim to improve learning in the target task by leveraging knowledge from the source task [17].

  This Thesis takes the objective described above and believes that an embedding model trained with generic content can be fine-tuned to support lexical simplification tasks in a more efficient way.

1.5. Methodology

In order to achieve Thesis objectives, the following methodology has been followed:

- **Study and review of state of the art.** Study of the current state of the literature, reviewing important previous works related to cognitive accessibility, NLP methods and text simplification to focus on the task of lexical simplification as a next step.

- **Design.** This step consists of the design of an architecture, NLP methods, accessibility resources and tools to support the simplification procedures and experiments to be carried out to meet the objectives and aim at better results than those of state of the art in lexical simplification.

- **Implementation and experimentation.** After the designing step, the implementation and experimentation step is carried out. In this step, resources and methods investigated in the design stage are created, adapted, integrated and then tested in different scenarios, depending on which stage of lexical simplification is being dealt with. In addition, an experimentation with target users is performed on the proposed architecture.

- **Analysis of results.** After experimentation, the procedures for each stage of lexical simplification are evaluated and the results are compared with those of the state of the art. This is an important part of the methodology because it allows us to see errors and points for improvement at each stage, which allows us to go back to previous stages and aim for improved procedures.

1.6. Document Outline

This document is organized into five Chapters. This first Chapter contains an introduction explaining the motivation and objectives that led to carrying out the study. Furthermore, the hypotheses with the research questions that the Thesis aims to solve and the methodology have been presented. The remainder of this document is divided into different Chapters and is organized as follows.

Chapter 2 summarizes language accessibility background, which leads the Section to previous work on NLP tasks in text simplification and shows a literature summary on the task of lexical simplification and word sense disambiguation. In addition, existing resources for the assessment of these tasks are briefly described.

Chapter 3 shows the proposed architecture to support language accessibility guidelines. After an extensive review of previously applied methodologies, different resources and approaches are proposed for lexical simplification.

Chapter 4 presents the experimentation carried out using the architecture proposed previously. For each scenario, this Chapter contains a description of the problem addressed, the dataset used, the methodology employed and the results obtained. In addition, a study with the target users is presented, evaluating the proposal.

Chapter 5 shows a proof of concept which presents the design of an accessible user interface that brings together the proposal of this Thesis.

Chapter 6 contains conclusions where findings and main contributions are summarized. Moreover, this Chapter provides an outlook into future works.

Chapter 2

Related Work

This Chapter introduces important aspects about cognitive accessibility requirements to textual content, text simplification methods and language resources. However, as a first Section, basic knowledge is provided to facilitate reading and to establish the basis for the following Sections. Finally, conclusions are offered to summarize everything seen in the Section.

2.1. Background

Nowadays, the development of artificial intelligence (AI) has become more popular thanks to the increase in data volume, improvements in computer performance, increased storage capacity and consequently allowed the support of advanced algorithms. According to MacCarthy [18], AI can be defined as "the science and engineering of making intelligent machines, especially intelligent computer programs, task-related to make computers understand human intelligence". As a subfield of this discipline, NLP is introduced, which seeks to make computers understand utterances or words written in human languages [19]. And because much information on the Internet is presented in an unstructured form, NLP plays an essential role in the extraction of valuable information that can benefit the structuring and classification of information, decision making, entity recognition, among others.

Different methods to achieve these tasks have been proposed over the years, reaching more complex methods based on machine learning (ML). This field can be defined as [20]: "study of computer algorithms that allow computer programs to automatically improve through experience". These approaches can be organized into different categories, for example, considering whether the algorithm is trained with labeled data or not, these algorithms can be classified into supervised learning, semi-supervised learning and unsupervised learning. In supervised approaches, rules are automatically induced from the annotated training data. The semi-supervised machine learning technique involves a small degree of supervision. In unsupervised machine learning no model training is represented as they arise due to lack of annotated data in some tasks. The task is performed by finding the intra-similarity and inter-similarity between objects.

In the next subsection, some machine learning algorithms are described for the contextualization of this Thesis (as shown in figure 2.1).

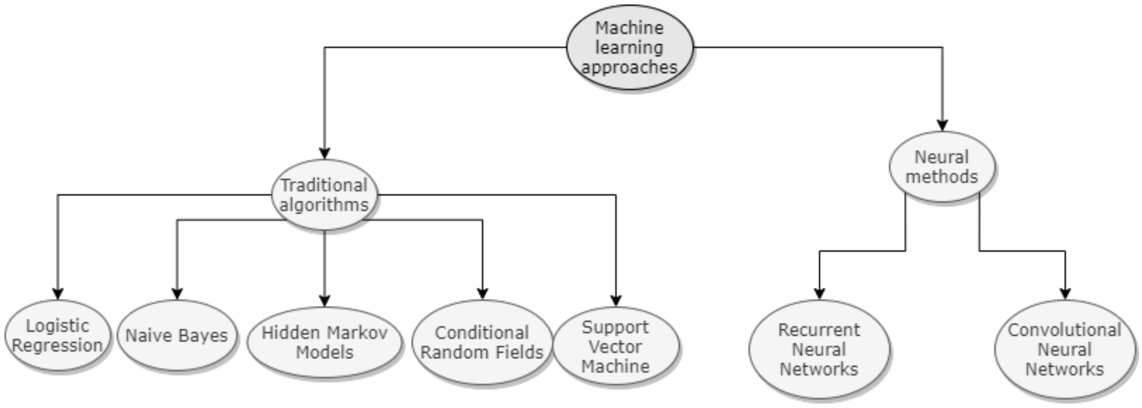

Figure 2.1: Machine learning approaches described in this Thesis.

2.1.1. Traditional classifiers

This Section describes traditional algorithms (also known as classical algorithms) that have served as the basis for today's more advanced approaches. Likewise, the following algorithms are still relevant in NLP tasks such as text classification or NER.

2.1.1.1 Logistic Regression

Logistic regression is a linear classifier [21] that is used to predict the probability of an event as a function of the independent variables. This classifier is a linear method, but the predictions are transformed using the following logistic function (equation 2.1).

$$P(y|X) = \frac{\exp(\sum_{m=1}^{M} \lambda_m f_m(y, X))}{\sum_{y'} \exp(\sum_{m=1}^{M} \lambda_m f_m(y', X))} \tag{2.1}$$

This function describes the weight $\lambda_m f_m$ of the features f_m defined with respect to y and X to generate a class prediction. In addition, pairs of state observations $f_m(y, X)$ are defined as features [22].

One of its main advantages is that the results are easy to interpret compared to other classifiers, while on the other hand, it usually requires a large amount of data to obtain acceptable results. It is widely used in various fields, one of them being the medical sciences.

2.1.1.2 Naive Bayes

Naive Bayes (NB) is a probabilistic classifier that assumes that the value of a particular feature is independent of the value of any other feature, given the class variable. This classifier describes the joint distribution $p(y, x)$ by the prior probability $p(y)$ and the likehood function $p(x|y)$ as shown in equation 2.2.

$$p(y, x) = p(y) \prod_{m=1}^{M} p(X_m|y) \tag{2.2}$$

2.1.1.3 Hidden Markov Models

Hidden Markov Models (HMM) is a statistical model in which the system to be modeled is assumed to be a Markov process of unknown parameters [23]. The objective is to determine the unknown or "hidden" parameters of such a chain from the observable parameters [24]. This model is the sequential version of the Naive Bayes that represents the join distribution $p(y, X)$ as shown in equation 2.3.

$$p(y, X) = \prod_{n=1}^{N} p(y_n|y_{n-1})p(x_n|y_n) \tag{2.3}$$

2.1.1.4 Conditional Random Fields

The Conditional Random Fields (CRFs) operates by the principle of the Maximum Entropy Markov Model (MEMMs) [25] and are used for sequence labeling. The basic principle of CRF is to define the conditional probability distribution over the label sequences in a given observation [26]. Take for example equation 2.4, where the conditional probability $p(y|X)$ is described. This classifier uses the sigmoid function with the weights λ_m of features f_m defined with respect to y_n,y_{n-1} and x_n to generate a class prediction like a Markov Chain. At least one feature for each transition $f_m(y_n, y_{n-1})$ is needed to be defined, and one for each observation pair $f_m(y_n, x_n)$

$$p(y|X) = \frac{\exp(\sum_{m=1}^{M} \lambda_m f_m(y_n, y_{n-1}, x_n))}{\sum_{y'} \exp(\sum_{m=1}^{M} \lambda_m f_m(y'_n, y'_{n-1}, x_n))} \tag{2.4}$$

2.1.1.5 Support Vector Machines

A non-probabilistic classifier that focuses on finding the hyperplane that best separates classes, maximizing the margin between them and minimizing the number of classification errors [27]. The main reason for its success is that most text classification problems have been shown to be linearly separable [27].

Another advantage is that SVM classifiers are able to learn independently of the dimensionality of the feature space, since they are based on maximizing the margin and not on the number of features [27]. Thanks to the latter, this classifier was taken in this Thesis (see Section 3), since the resulting model is able to provide good results in those problems where classes are widely separated from each other, even with a large number of features.

Figure 2.2 shows data points with variables $x = [x1, x2]$ of two classes $y = -1, 1$, where the maximum margin is defined by the margins $wx - b \leq 1$ and $wx - b \geq -1$ which are boundaries of the classes $y_i = 1$ and $y_i = -1$, respectively. The margin distance is represented as $\frac{2}{||w||}$ and hyperplane offset as $\frac{b}{||w||}$.

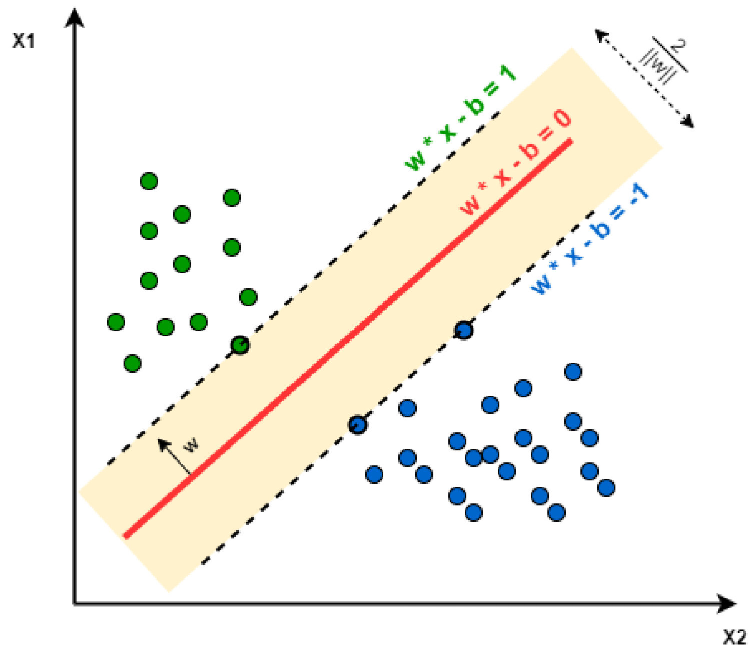

Figure 2.2: An SVM separating two classes by an hyperplane $wx - b = 0$ [a]
[a] Image retrieved from https://www.baeldung.com/cs/svm-multiclass-classification

This classifier has a variety of kernel tricks to deal with nonlinear problems [28]. Kernels used in the scope of this Thesis are shown below:

- **Gaussian radial basis function (RBF):** $K(X_i, X_j) = e^{-\gamma||x_i - x_j||^2}$ for $\gamma > 0$
- **Linear:** $K(X_i, X_j) = X_i^T X_j$

2.1.2. Neural Methods

Most recently, deep neural networks have gained popularity in ML approaches. Deep neural networks are part of a larger family of machine learning methods based on artificial neural networks (ANNs). An ANN employs a hierarchy of layers in which each layer considers information from a previous layer and then passes its output to other layers [29]. While the algorithms described above are typically linear, deep learning approaches are made up of a hierarchy of increasing complexity and abstraction. Figure 2.3 shows an overview of an ANN structure, where input, hidden and output layers can be seen. In addition, the neurons in one layer can be seen to be connected only to neurons belonging to adjacent layers.

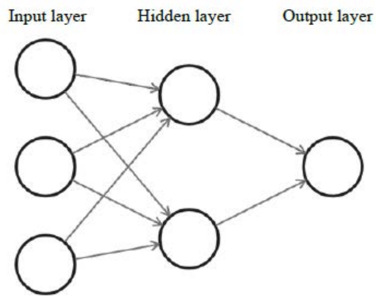

Figure 2.3: Basic ANN Architecture

The main two architectures built for classification are the Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). The first architecture is prepared to classify spatial signals and the last one is prepared to classify temporal signals. In this Thesis, a brief explanation of these architectures are given:

2.1.2.1 Recurrent Neural Networks

RNNs are a class of ANNs in which the connections between nodes form a directed graph along a temporal sequence. RNNs are commonly used for ordinal or temporal problems, such as speech recognition, image captioning, and NLP. While traditional deep neural networks assume that inputs and outputs are independent of each other, the output of RNNs depends on the above elements within the sequence [29]. Figure 2.4 illustrates how the RNN has a recurrent connection in the hidden state, in contrast to the ANN shown in figure 2.3. This loop constraint ensures that sequential information is captured in the input data.

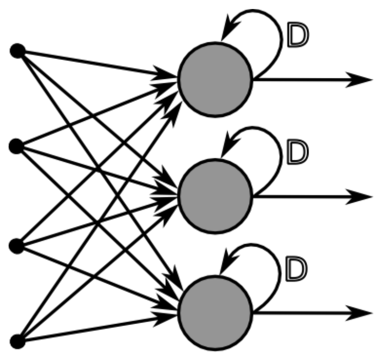

Figure 2.4: Example of recurrent connections

2.1.2.2 Convolutional Neural Networks

CNN´s are a specialized type of neural network model designed for working with two-dimensional image data, although they can be used with one-dimensional and three-dimensional data. CNN uses layers with convolution filters that are applied to local features in order to represent local information [30]. In this context, a convolution can be understood as a linear operation that involves the multiplication of a set of weights with the input, much like a traditional neural network. In this type of neural network, the connections between nodes do not form a loop like RNNs. Figure 2.5 shows a basic CNN architecture, where the first layer is the convolutional layer that is in charge of extracting features from the input, then it goes through a pooling layer, where the goal is to decrease the size of the convolutional feature map to reduce computational costs, to finally go through a fully connected layer, which is formed by the weights and biases together with the neurons to be used to connect neurons between two different layers. These layers are usually placed before the output layer and form the last layers of the CNN architecture.

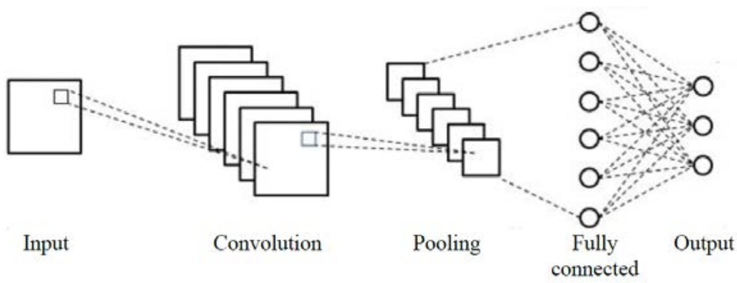

Figure 2.5: Example of a basic CNN architecture

Although this architecture was originally designed to solve computer vision problems, over the years it has been demonstrated that this architecture offers solutions to NLP problems, achieving good results in text classification [31]. CNNs can also be applied to NLP tasks with textual data, since the inputs are the vector representation of each word in a sentence. However, the shape of the filters has to be the dimension of the word vector and a predefined context window to compute the convolution operations. The same operation is then applied by sliding the filter through the context window for each word in the sentence.

2.1.3. Word Embeddings

Moving on to the next important concept in this Thesis, word embeddings are described. In general terms, a word embedding is a form of learned representation of a text, where words that have the same meaning have a similar representation. Once Mikolov's [32] paper was published and Word2Vec became public, a new era in NLP was unleashed, where word embeddings represent a fundamental part of many papers. This form of representation can be considered as one of the major breakthroughs of Deep Learning in natural language processing problems. Figure 2.6 shows some of the most popular word embedding methods.

Described in a more technical way, embeddings are a class of techniques in which individual words are represented as real-valued vectors in a predefined vector space. Each word is mapped to a vector and the vector values are learned in a way that resembles those of a vector [32].

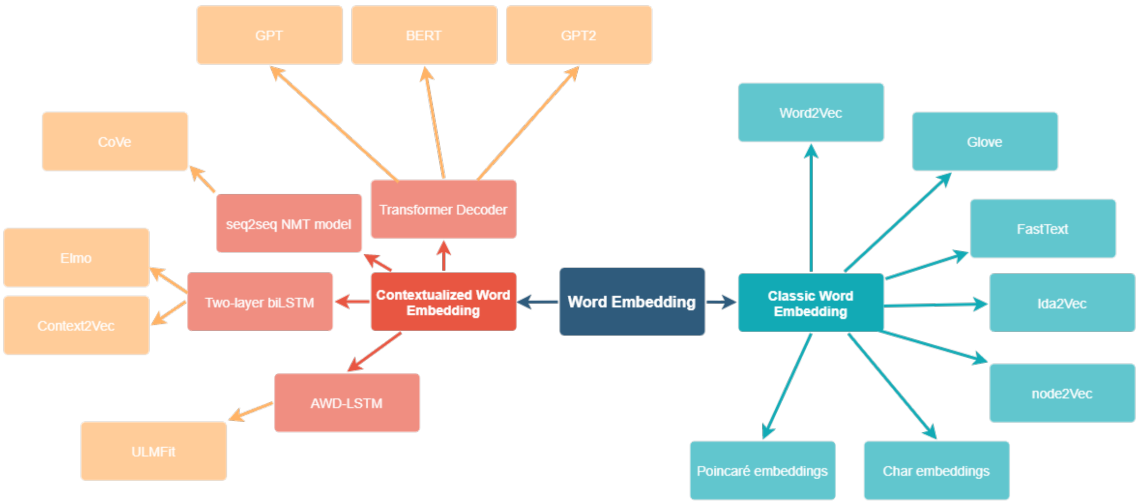

Figure 2.6: Brief list of different word embedding methods

2.1.3.1 Classic Word Embeddings

Classic word embedding techniques (also known as static word embedding) refer to the fact that the same word will always have the same representation, regardless of the context in which it appears.

This concept was introduced in the work of Mikolov [32] where they essentially model a neural network with a single hidden layer and the embeddings represent the weights of the hidden layer in the neural network (see Figure 2.7).

Below, some methods for efficiently learning classic word embeddings from the text are described:

1. **Word2Vec:** Word2Vec is a statistical method to efficiently learn an embedded independent word from a text corpus. As told before, this method was proposed in the work of Mikolov [32] to make neural network-based training of embedding more efficient and since then this method has become a starting point for the development of pre-trained word embedding (see Figure 2.8).

2. **Glove:** This method was published shortly after Word2vec (Skip-Gram) [33]. This method, unlike Word2vec, takes advantage of the co-occurrence statistics of the corpus and does not ignore a large amount of repetitions in the data.

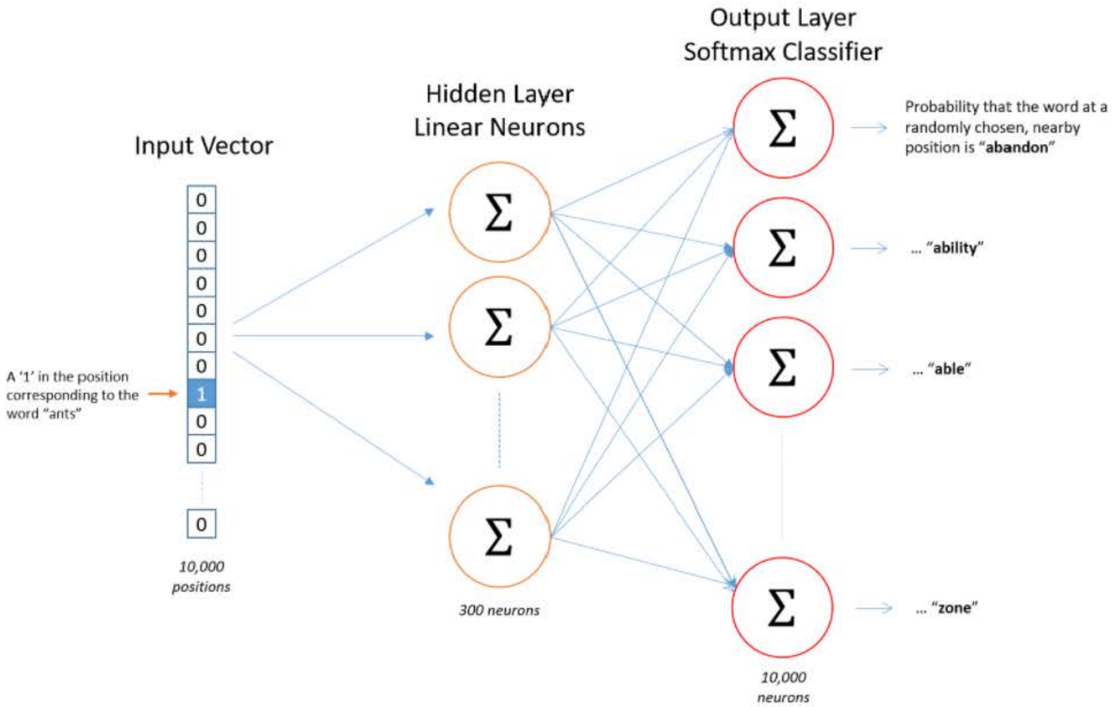

Figure 2.7: Neural Architecture example [a]

[a] Image retrieved from https://www.davidsbatista.net/blog/2018/12/06/Word_Embeddings/

Essentially these models learn the vectors essentially by doing some sort of dimensionality reduction on the co-occurrence counts matrix.

3. **FastText:** This method emerged from a common disadvantage among the previous methods, as it presented problems with out-of-vocabulary word (OOV) situations. In this embedding research work [34], a method is proposed where each word is presented as a bag of characters, where a vector is associated with each character (char-gram) and the word is represented as the sum of these vectors. As an example, one can have the word "vector" with length 3 of ngram, where as a result one would have the sequence "<ve, vec, ect, cto, tor, or>" and the sequence <vector>. This allows us to have an additional detail to previous methods, providing not only semantic and syntactic information, but also morphological information of the words.

At this point, the fact that embeddings were a valuable contribution to the research community was indisputable. However, in matters of polysemy, the previous models presented common problems by generating the same embedding for the same word in different contexts. For example, the polysemous word "solution" where it can have the following contexts:

- Work out the **solution** in your head.

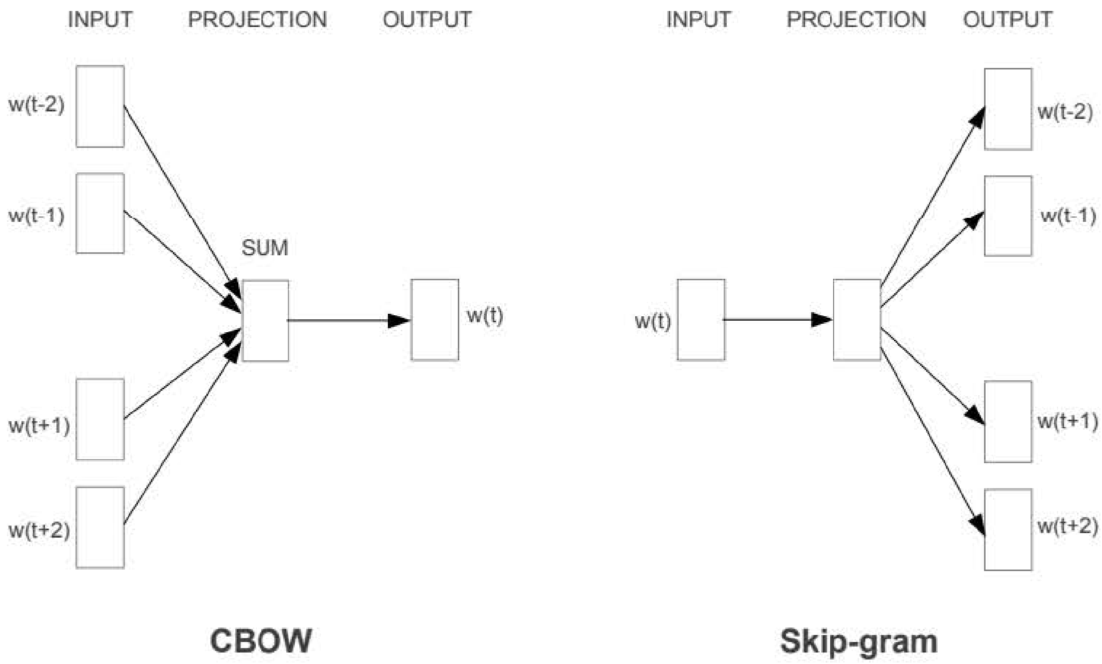

Figure 2.8: Word2Vec Training Models [a]

[a] Image retrieved from the research work of Mikolov et al. [32]

- Heat the **solution** to 75° Celsius.

In the previous methods, the representation of the word "solution" would always be the same, regardless of whether it appears in context and despite the fact that they obviously have different meanings. For this reason, the classical methods evolved to the contextual methods presented in the following Section.

2.1.3.2 Dynamical Word Embeddings

Dynamic word embedding techniques (also known as contextual embeddings) can be considered as an evolution of earlier techniques. In these techniques, the context of the word is taken into account and in most of these techniques, language models to help model the representation of a word are used.

Language Models In this type of embeddings one of the important concepts to explain is "language models". Language models are in charge of understanding the probability distribution over a sequence of words, in other words, these models calculate the probability distribution of the next word in a sequence given the sequence of previous words. the novelty of language model pre-training is the fact that they are fine-tuned for a particular downstream task and not as widely generalized like embeddings are for texts [35]. An example of one of the technologies used to accomplish this task is Long short-term memory (LTSM), which receives sequences of words and uses the internal state of the LSTM along with the previous word in the sequence to predict [36].

Transformer-based Models Another important concept to describe is "transformers". Transformers are designed to handle sequential data such as natural language for tasks relevant to this Thesis, such as text classification and NER. The most important part of the Transformer is the attention mechanism. The attention mechanism represents the importance of other tokens in an input for the encoding of a given token. In other words, the attention mechanism allows the Transformer to focus on certain words both on the left and on the right to deal with the current word according to the NLP task that is addressed [35]. Another advantage of the Transformer architecture is that learning in one task can be transferred to other tasks through transfer learning. This method has the principle of taking the knowledge gained from performing one task and applying it to a different task. This is the big difference between the traditional ML methods described above with the transfer methods, since having a previously trained model, the knowledge can be extended, creating a new task specific model.

Some important dynamic word embedding methods are described below:

1. **ELMo (Embeddings from Language Models):** As a first method, embeddings that had good results in several tasks are introduced such as NER or question-answer. ELMo models [36] train a multi-layer, bidirectional, LSTM-based linguistic model and extract the hidden state of each layer for the input word sequence. As a next step, a weighted sum of those hidden states to obtain an embedding for each word is obtained. The weights of each state are adjusted according to the task and learned during the training of the final task.

2. **BERT:** This embedding method introduced novelties to previous approaches by including bidirectional contextual representations [37], as opposed to other unidirectional methods. To understand the term "bidirectional", take for example the sentence "The cat sits on the mat", where the target word is "sits". In unidirectional approaches the representation would be based on "The cat" but not on "on the mat". However, in these bidirectional approaches both contexts of the target word are represented. This is done by masking words in the input text and then passing the sequence through a multi-layer bidirectional Transformer encoder, and then predicting only the masked words. These representations can also be used to predict sentences [37] (as shown in Figure 2.9).

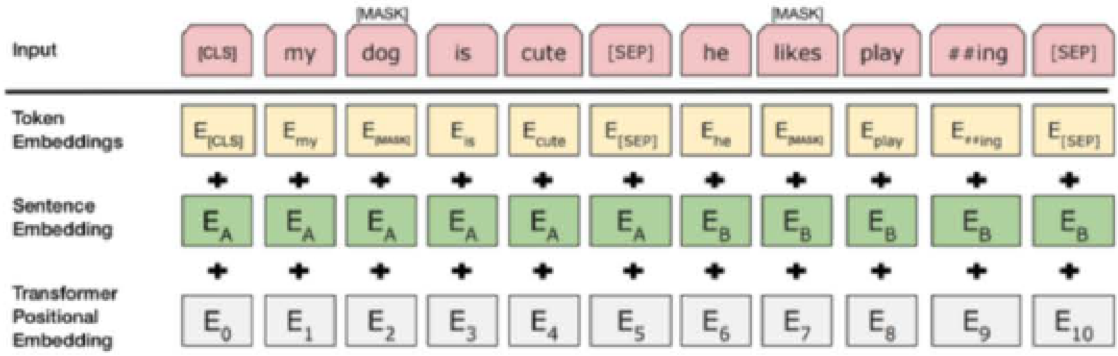

Figure 2.9: BERT: Next Sentence Prediction (NSP) [a]

[a] Image retrieved from https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270

2.2. Cognitive accessibility

This Section describes fundamental information about cognitive accessibility requirements, legislation and standards to aid people dealing with different cognitive barriers.

2.2.1. Background

People continually use information and communications technology (ICT) to perform daily tasks. Through the use of the Web, people can access an increasing number of services. While these barriers affect users with disabilities to a greater degree, there are other groups of users who are also at risk of exclusion.

People access ICTs in very different ways depending on their own characteristics in access, contexts of use, technological characteristics, among others. Thus, the presentation of information in an accessible way implies guaranteeing access to it regardless of: accessibility features of people (people with disabilities, functional diversity), equipment (hardware), applications (software) and user agents (graphical browsers, voice browsers, text browsers, media players, etc).

There is a wide range of people with disabilities, including blindness and low vision, deafness and hearing loss, learning disabilities, cognitive limitations, limited movement, speech disabilities and combinations of these. In addition, some people may have disabilities due to illness, or may develop impairments with age.

In addition to these disabilities, the disability due to ageing must be taken into account, since disability increases with age. Sometimes disability is seen as something external by those who do not suffer from it. The increase in quality of life and life expectancy together with the low birth rate is causing progressive ageing of the population. The UN predicts that the number of people over 60 years of age will triple in the world by 2050 [38].

According to their access characteristics and special needs, people with disabilities can access digital content in different ways [39]. The needs vary from one disability to another. For example, some people with hearing disabilities and people with intellectual disabilities, they may not read written language fluently, so using simple and understandable text in a digital document is beneficial.

2.2.2. Regulatory frameworks

Legislation is therefore vital for all of us, even for those without disabilities. Some directives provide guidelines for making text content more accessible for individuals with intellectual, language, and learning disabilities, which are detailed below:

The United Nations (UN) Convention on the Rights of Persons with Disabilities [40] ratified by the European Union and its member states in 2007 details in Article 9 that states must ensure the accessibility of, among others, information systems and technology and electronic services. The convention recognizes disability as the result of the interaction between an individual and physical, technological and economic barriers, among others, that prevent full participation in society. For the different countries, the convention has been the main mechanism for legislating on accessibility, and since its establishment there have been several laws, as well as derogations.

In Spain, as well as in the rest of the European Union countries, the European Directive (EU) 2016/2102 has meant a great boost to existing accessibility policies. Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of websites and mobile applications of public sector bodies [41] aims to create a barrier-free EU in which all citizens have access to public websites regardless of whether they have any kind of disability. The directive establishes obligations of compliance with accessibility regulations in the public sector, both state or territorial bodies and those managed mainly with public funds, with some exceptions.

The directive set deadlines for the different member states to incorporate it into national legislation. In the case of Spain, the transposition of the directive had to be done before September 2018. Therefore, on September 20 of the same year, most of the provisions of Royal Decree 1112/2018 (RD 1112/2018), September 7, on the accessibility of public sector websites and applications for mobile devices [42] came into force.

RD 1112/2018 compels all websites and native mobile applications of the public sector, or receiving public funding, to be accessible in accordance with the European standard EN 301 549 [43]. This European standard, in turn, was adopted as a standard as UNE-EN 301 549 by AENOR [44]. The standard is aligned with the WCAG 2.1 standard for web content. It applies to all content on public sector websites and native mobile apps and although it does not specify exactly the types of content that must be accessible, like the Directive, it does not expressly exclude any type.

In this sense, there is another directive that was signed on April 17, 2019, the European Directive on accessibility requirements for products and services, known as Directive 2019/882, which extends the regulatory framework for accessibility to the private sector. This new directive, which aims to improve the trade of accessible products and services in the EU in order to increase the quality and accessibility of products for people with disabilities, will require member states to adopt them into law by June 28, 2022.

2.2.3. Standards

In relation to standardization, the role of the W3C with the Web Accessibility Initiative (WAI) founded in 1997 should be highlighted [45]. The WAI working group, in coordination with organizations around the world, pursues the accessibility of the Web and its contents through five main areas of work: technology, guidelines, tools, training and dissemination, and research and development. Of the work developed by the WAI, the Web Content Accessibility Guidelines (WCAG), cited as a mandatory reference in most ICT legislation around the world, stand out.

The WCAG provide recommendations to make web content more accessible to a wider range of people with disabilities. In addition, following these guidelines often make web content easier to use. The official recommendation is the 2018 WCAG 2.1 [46], although there is a very advanced draft of a new version, WCAG 2.2 [47] that is expected to become the new recommendation before the end of 2021, which has included aspects of taking plain language into account.

Another benchmark in standardization is the European standard EN 301 549, which has been introduced previously, sets out the accessibility requirements that must be met by any product and service: web pages, mobile applications, documents, hardware, etc [43]. This standard also includes guidelines based on the WCAG.

The WCAG aim to address guidelines to avoid web accessibility barriers for people with all types of disabilities. Although this is too ambitious, there are ongoing projects within the W3C to cover accessibility guidelines for some disabilities that are weakly addressed, such as the "Cognitive and Learning Disabilities Accessibility Task Force (Coga TF)" project, which has resulted in an interesting document on cognitive accessibility to be taken into account when creating textual content in digital documents, entitled "Making Content Usable for People with Cognitive and Learning Disabilities" [48]. This document presents guidelines to make the textual content more understandable and easy-to-read, highlighting among them some related to the use of simple language and the use of simple grammatical structures. This approach is aligned with the plain language initiative [49], aimed at the general public, which promotes the use of simple language in all administrative and governmental electronic content and information to provide better service to all citizens. At present, governments must favor the provision of accessible information in plain language in their eAdministration, eHealth, etc. services. For this reason, they are developing guidelines and adapting many of their public communications

[50] such as easy reading guidelines [4], aimed at meeting the text accessibility needs of people with intellectual disabilities, as well as other affected user groups. All these standards and documentation have common cognitive accessibility guidelines in relation to the development of accessible digital content, in how to make texts more readable and understandable.

Guidelines	**Description**
WCAG 2.1 [45]	· Unusual Words (Success Criteria (SC) 3.1.3) A mechanism is available for identifying specific definitions of words or phrases used in an unusual way. · Abbreviations (SC 3.1.4) A mechanism for identifying the expanded form or meaning of abbreviations is available. · Reading Level (SC 3.1.5) When the text requires advanced reading ability, a reading version is available that does not require advanced reading ability.
Easy-to-Read [51]	· Do not use difficult words. · Speak to people directly. · Use positive sentences rather than negative ones where possible. · Use active language rather than passive language where possible. · Use personal pronouns.
Plain Language [49] [52]	· Write in clear and easily understandable language. · Use short, concise sentences. · The content can be repeated. Explain difficult words. Give examples. · Do not write in metaphors. Do not use abbreviations or explain to them if you do.
COGA documents [48]	· 4.4.1 Pattern: Use Clear Words. · 4.4.2 Pattern: Use a Simple Tense and Voice. · 4.4.3 Pattern: Avoid Double Negatives or Nested Clauses. · 4.4.4 Pattern: Use Literal Language. · 4.4.5 Pattern: Keep Text Succinct. · 4.4.6 Pattern: Use Clear, Unambiguous Text Formatting and Punctuation. · 4.4.8 Pattern: Provide Summary of Long Documents and Media.

Table 2.1: Readability and Understandability Guidelines.

Table 2.1 shows an analysis of these guidelines. Although there are some differences between these initiatives, some of them are completely related and it can be seen that the use of a simple lexicon is an element that is repeated in all the guidelines. People with language problems often have a small vocabulary and learning new terms is a slow and difficult process. In this Thesis, WCAG 2.1 and COGA documentation have been taken into account to provide solutions and comply with this guideline. The first success criterion indicates that a mechanism should be in place to identify specific definitions of words or phrases used in unusual or restricted ways, including idiomatic expressions and slang (WCAG 3.1.3 (Unusual Words)). This guideline is also part of Principle 3 (Understandable), which states that the information and operation information and operation of the user interface should be understandable. To achieve this, Table 2.2 describes success criterion 3.1.3, which requires that the definition of a word be provided when it is used in an unusual or restricted way on a web page. To follow the techniques and provide definitions for unusual words, it is necessary to differentiate between the following two situations: if a word has just one meaning within the webpage or if different meanings for the same word appear within the same webpage.

Furthermore, design pattern 4.4.1 of the COGA documentation indicates that common and clear words must be used in all content. Some techniques add a simple language term and provide a definition if complex words are used.

2.2.4. Accessibility requirements

At this point in the Thesis, fundamental information concerning the context of the Thesis has been described. Starting with concepts referring to the NLP discipline and ending with the importance of this discipline in providing systematic support to comply with standards and laws regarding cognitive web accessibility. In addition, by conducting this study, this Thesis concludes that in order to satisfy the requirements of accessibility to textual content, there must be procedures that (1) detect which words are unusual or complex; (2) offer simpler synonyms; (3) offer definitions; and (4) contextualize the meaning of the unusual word in the text to offer the correct synonym or definition. These accessibility requirements drive the focus of the Thesis. In the following Sections, concepts, related works are introduced and the contributions of this Thesis are briefly exposed to fulfill these points.

Situation A: If the word or phrase has a unique meaning within the web page:	
G101: Providing the definition of a word or phrase used in an unusual or restricted way	G55: Linking to definitions · H40: Using description lists · H60: Using the link element to link to a glossary G112: Using inline definitions · H54: Using the dfn element to identify the defining instance of a word G62: Providing a glossary G70: Providing a function to search an online dictionary
Situation B: If the word or phrase has different meanings within the same web page:	
G101: Providing the definition of a word or phrase used in an unusual or restricted way	G55: Linking to definitions · H40: Using description lists · H60: Using the link element to link to a glossary G112: Using inline definitions · H54: Using the dfn element to identify the defining instance of a word

Table 2.2: Success criterion 3.1.3 techniques (WCAG 2.1)

2.3. Text Simplification

Automatic text simplification is the process of reducing the linguistic complexity of a text to improve its understandability and readability, while still maintaining its original information content and meaning [15]. In recent years, research on text simplification has been applied with the aim of transforming a text into an equivalent text with the difference that it is more accessible to people with cognitive disabilities. Interest in automatic text simplification has increased in recent years and, despite the many approaches and techniques proposed, automatic text simplification is, so far, far from perfect. This interest in text simplification is evidenced by the number of languages being addressed in research around the world [13].

First works on text simplification began more than 20 years ago by performing a superficial analysis of texts to identify verbs and nouns in complex phrases [53]. Most of the early work in the field involved methods of summarization and extraction of sentences from a document that conveyed the most meaning [54] and although there are many features of a text that can be modified to make it more understandable, including the way the text is presented, recent automatic text simplification has generally concentrated on two different tasks: syntactic simplification and lexical simplification. Each addresses different sub-problems [13]. Syntactic simplification identifies syntactic phenomena in sentences that may hinder comprehension, in an effort to transform the sentence into more understandable and readable equivalents. This involves replacing particular syntactic constructors (such as relative clauses, apposition and conjunction) in sentences to make the text easier for some target group to read or easier for some other algorithm to process [55]. On the other hand, lexical simplification attempts to modify the vocabulary of the text by choosing words that are considered more appropriate for a specific reader [13]. In addition to these methods, there are papers that present approaches that combine lexical and syntactic methods in order to optimize each other [56] [57]. This Thesis focuses on the lexical simplification task, which is detailed in the following subsection.

2.3.1. Lexical Simplification

Lexical simplification aims at replacing difficult words for a certain audience with understandable expressions while preserving the meaning of the original text segments. The best substitution should be simpler, keeping the grammatical sentence and preserving its meaning as much as possible [8].

Lexical simplification directly supports the accessibility guidelines described in Section 2.2, by offering a mechanism for detecting unusual words and proposing simple replacements for a specific audience. These mechanisms can provide support for understanding textual content for people with low literacy levels, such as children and non-native speakers [58]. Also, there is research showing that this task supports people with autism [59] [60] [9], aphasia [61] [10], low vision [11], dyslexia [12] [62] or people with intellectual disabilities [63] [64] [13] [65]. These works offer a wide variety of approaches, where their systems include, for example, rule-based approaches, traditional supervised, unsupervised machine learning approaches or recent more sophisticated approaches based on deep learning.

In a recent research [66], experiments were conducted to help people with dyslexia with text adaptations by combining lexical simplification and visual support with a group of people with and without dyslexia. These users were tasked with reading 60 sentences of different structures, some with pictures and some with high-frequency words (see Figure 2.10). These tests were performed with the aid of eye tracking, including metrics such as target gaze duration, target reading time, regression path duration and probability of regression. The results suggested that pictures are useful for understanding a text by helping the user to get an idea of what the text is about, and they also found that participants with low lexical and lower reading skills benefited greatly from the lexical processes performed on the sentences.

Another paper on medical text literacy support is presented in a web-based approach called "SIMPLE" [67]. In this web-based approach, which currently supports English and Italian, different medical terms can be automatically identified. As a next step, substitutions and explanations to these terms are offered, fed by different dictionaries with content related to medical terms such as synonyms or definitions. The evaluations showed moderate to almost perfect agreements in term translation and term detection respectively.

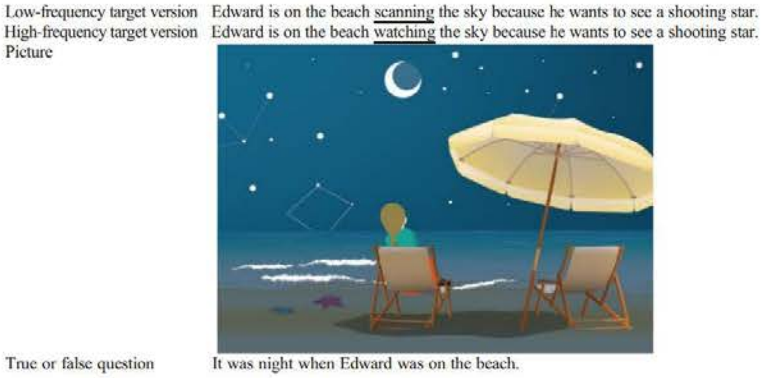

Figure 2.10: Sample of a test for people with dyslexia with the sentences with words with low/high frequency, an image and a comprehension question, as shown in [66]

Research for the deaf and hard of hearing on lexical simplification is also proposed. In a recent paper [68], an evaluation was made of whether there is a benefit of offering automatic lexical simplification procedures to the hearing impaired. In addition, the response of the users was evaluated by providing them with an adaptive approach, giving them greater autonomy in controlling which words should be replaced. The results were positive, with users giving preference to the adaptive approach.

While one of the earliest recorded works on lexical simplification assumed that all words in a text were potentially complex, the method quickly lost popularity as it was realized that a specialized method is needed for each target audience, since some words may be simple depending on the target audience. Lexical simplification methods then quickly evolved to using lexicons to determine whether a word was considered complex [8]. This method demonstrated better results because they were lexicons built specifically for a user, however, they presented the problem of being limited to the size of the system's vocabulary and costly scalability [69]. Machine learning methods try to learn models that predict the complexity of a word. Supervised methods while outputing good results [70], require annotating datasets to achieve their objective [71], which leads to a significant disadvantage when dealing with languages with few annotated corpora for text simplification. Unsupervised approaches emerged because of this disadvantage, requiring a minimum amount of data to create a model but mostly yielding lower results than the best supervised approaches [72]. However, recent approaches have improved by allowing more detailed context information to be obtained [73]. Similarly, hybrid approaches that leverage methods from the latter methods have emerged, such is the case of [74], which uses a corpus-based approach and a combination of a free lexicon, decision trees and context-based rules.

To date, the most comprehensive survey was conducted by [8], which grouped papers focusing on different aspects of lexical simplification into four main steps: complex word identification (CWI), generation of substitutes (SG), selection of substitutes (SS) and substitute ranking (SR). Figure 2.11, illustrates Shardlow´s definition of this task [53]. The following Sections will introduce these steps, which are covered in this Thesis to detect complex words and offer simpler synonyms.

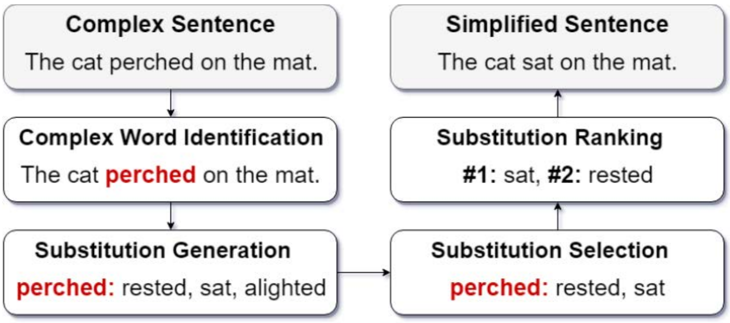

Figure 2.11: Lexical simplification system pipeline as shown in [8]

2.3.1.1 Complex Word Identification (CWI)

CWI aims to select the candidates to be simplified, that is, to detect which words are complex in a given text [65]. Although at the beginning this stage did not exist because the simplifiers considered each word as complex, over the years, several problems were encountered in performing this method. For example, in the work of Devlin and Tait [75], a large number of sentences were found to undergo structural or meaning changes. This also represents a latent problem for the next stages of lexical simplification, since by classifying all words as complex, unnecessary replacements end up being made in all sentences. Therefore, subsequent work quickly moved from this method to those described below:

1. **Threshold-based:** As the name suggests, the threshold-based method consists of establishing a certain simplicity metric (called threshold), which when exceeded by a target word/phrase, it is considered as complex. One of the most popular metrics is based on word frequency, trying to simplify words that are less frequent in a set of texts [76] [70]. For example, a threshold can be applied to instances that are not among the 5000 most frequent words in the Google IT corpus and classify them as complex [8]. Similarly, in Bott's work [64], the decision was made to avoid substitutions for words that appear in more than 1% of a long corpus. Another popular metric is word length, where a maximum word length is defined in order to discern between a simple or complex word [77]. However, developing a single simplicity metric is a difficult task, since using metrics such as word length, it has been shown that a large number of complex words are not detected due to being within the threshold [64] [78].

2. **Lexicon-based:** This method arose due to the limitations presented in the previous one and exploits collections (or lexicon) of words/phrases that are previously detected, where if a word is found in the collection it is represented as complex, otherwise, it is represented as simple. For domain-specific tasks, this method can be

very beneficial, as in the work of Deleger and Zweigenbaum [79] , which presents a method for creating a lexicon of complex words and phrases in the medical domain. Another resource in this domain is the Unified Medical Language System (UMLS), which is a database of medical terms that is very useful in this task [80] [81]. These lexicons can also be used as part of a larger system, as in [82] [65], where a lexicon of words extracted from easy-to-read content is used to be represented as a feature of a machine learning classifier.

While this approach is effective in practice as demonstrated by the FACILITA system [69], at the same time, it is highly costly, given the fact that creating this resource manually takes a lot of time not to mention the challenge of deciding whether a word is complex or not, given the differences of opinion between people.

3. **Implicit Complex Word Identification:** In this method and as the name suggests, the detection is performed implicitly, since CWI is not considered as an initial stage, but as a method included in the other stages of lexical simplification. By performing this method, it is expected to eliminate possible substitutions of a word with a more complex substitute. Some work uses features of a word, such as frequency or length [83] [64], to decide whether a candidate substitution is more or less complex than the original target word. Horn et al. [84] generates a list of candidate replacements in which they add the target word in order to rule out the possibility that the replacement is more complex than the original word.

   Additional research uses parallel corpora to train its algorithms [7]. For example in the work of Zhu et al. [85], a CWI translation method with a Tree-based Simplification model trained with a parallel dataset of Wikipedia and simple Wikipedia is proposed, achieving good readability results. Later, Xu et al. [86] offer an alternative method in their training, dealing with the limitations of manually simplified data by adding large-scale paraphrases learned from bilingual texts.

   This method is considered challenging to evaluate in comparison to traditional CWI approaches [8], although it is clear that good results are shown when it can be ensured that the captured training data corresponds to the needs of the target audience.

4. **Machine learning-assisted:** offers a way to learn a language model based on the complexity of a word. However, a certain number of instances labeled as simple or complex are necessary, where features are extracted from the words to then train a classifier that will have the ability to discern between a simple or complex word.

SemEval 2016		
Team	**Technology/Strategy**	**Features**
AI-KU	Support Vector Machine	Word Embedding features Word context
AKTSKI	Support Vector Machine	Analysis of dataset annotations Semantic features Morphological features
MACSAAR	Random Forest Support Vector Machine	Word probability Word length
PLJUJAGH	SGD Linear SVC Random Forest Gradient Boost Classifier	Document frequency Term frequency Sentence length Word length Position of a word in a sentence Glove Word Embbedding
LTG	SVM Bayesian Classifier	Word Probability Conditional probabilities Joint probability Word length
USAAR	Entropy Perplexity	Entropy of meanings Perplexity of sentences
SV000gg	Support Vector Machine Ada Boosting Passive-Aggressive Learning Stochastic Gradient Descent Decision Trees Gradient Boosting Random Forest	Word length Number of syllables Number of meanings Number of synonyms Number of hyperonyms Number of hyponyms Probabilities of n-grams PosTag with 3-word windows If a word exists in a dictionary

Table 2.3: Description of systems presented at SemEval 2016

Also, additional research is generated thanks to CWI's shared tasks. Such as the SemEval 2016 task [87], in which participants were asked to create systems that perform the CWI task, i.e., detecting complex words in a set of sentences, specifically, identifying words that make comprehension difficult for non-native English speakers. Of the 42 systems presented at SemEval 2016, most used machine learning-based approaches in their processes and their results were evaluated in relation to F-Score and G-Score metrics, with the system with the best score being the system developed by Paetzold and Specia [88], which combines several approaches, such as threshold-based, lexicon-based and machine learning approaches. Another team was AI-KU [89], which used two Support Vector

Machine (SVM) classifiers, one used embbeddings of the target word and the other used embbedding in the context of the word obtaining an F-Score and G-Score of 0.103 and 0.545 respectively. Another system that used SVM classifiers was AKTSKI [87] which used two SVM classifiers, one that analyzes the labels according to the dataset annotations and the other that does not take them into account. In addition, they use semantic and morphological features, where in their "wsys" system they achieved an F-Score and G-Score of 0.100 and 0.534 respectively. In turn, teams using more than one type of classifier were presented, such was the case of MACSSAR [90], which used Random Forest Classifier (RFC) and SVM, based on the Hypothesis that complex words are less likely to be frequent and in general, have more length than simple words. This team achieved in their "RFC" system an F-Score and G-Score of 0.270 and 0.754 respectively. In addition, the team PLJUJAGH [91], used several algorithms for their classification such as a linear SVC, Random Forest and GradientBoostClassifier achieving F-Score and G-Score of 0.252 and 0.767 respectively. Table 2.3 provides an additional summary of the systems involved in the task.

BEA Workshop (NAACL-2018)		
Team	**Technology/Strategy**	**Features**
TMU	Random forest Random forest regressors	Word length Sentence length Frequencies: Wikipedia-WikiNews
NILC	Linear Regression Logistic Regression Decision Trees Gradient Boosting Word Embbedding	Word length Sentence length Number of syllables Number of meanings Probabilities of n-grams
UnibucKernel	Linear Kernel SVM	Number of characters Number of vowels/consonants Percentage of vowels/consonants Number of double consonants N-grams of characters PosTag Word Embedding Features
CAMB	Random Forest AdaBoost	Frequency of speech: Datamuse CEFR levels extracted from CALD Word length Number of syllables Number of meanings Number of hyponyms Number of hyperonyms Word familiarity score (FAM) Number of samples (KFSMP) Written frequency (KFFRQ) Date of acquisition
NLP-CIC	Random Forest Gradient Boosting Tree Ensembles	Word frequency Term frequency Inverse term frequency Document frequency PosTag Familiarity Date of acquisition Characteristics Word Embbedding
SB@GU	Random Forest Extra Trees Convolutional Neural Networks Recurrent Convolutional Neural Networks	Word length Number of syllables Probabilities of n-grams: Wikipedia If number If alphanumeric PosTag Suffix length Number of meanings Number of hyponyms Number of hyperonyms Theme distributions Word Embedding features Word Frequency: British National Corpus CELEX frequency Bigram, trigram frequency

Table 2.4: Description of systems presented at BEA Workshop (NAACL-2018)

The results of this task were reflected in a subsequent task, which was performed at the NAACL 2018 conference, in the BEA workshop [92]. Unlike the 2016 task, this task offered data for several languages such as English, German, Spanish and French. In addition, the participation of machine learning-based approaches was overwhelming [93], introducing novel approaches based on deep learning. This is the case of NILC [94] which presented three approaches for classification, one using traditional machine learning strategies, the second using Word Embbedding by extracting word vectors and the last one modeling the context of the target words supported by a Long short-term memory (LSTM). In the best results in the English language, NILC obtained an F-Score of 0.8636 in the News dataset, in the WikiNews dataset it obtained 0.7961 and 0.7965 in the Wikipedia dataset. One team that used SVM in their process was UnibucKernel [95], which is based on a set of features extracted from the word and the word context, such as word embeddings. Other systems used a set of classifiers to solve the task, such as SB@GU [96], which tested several classifiers such as random forest, extra three, convolutional neural networks and recurrent convolutional neural networks implemented with Keras and PyTorch. In the best results in the English language, SB@GU obtained an F-Score of 0.8325 in the News dataset, in the WikiNews dataset it obtained 0.8031 and 0.7832 in the Wikipedia dataset. In its best result in the Spanish language, SB@GU achieved an F-Score of 0.7281. CAMB [97], combined AdaBoost and Random Forest combined with features extracted from different resources. In the best results in the English language, CAMB obtained an F-Score of 0.8736 in the News dataset, in the WikiNews dataset it obtained 0.84 and 0.8115 in the Wikipedia dataset. Another team that used Random Forest for its binary classification is TMU [98], using conventional features such as word and sentence length. In English language results, TMU obtained an F-Score of 0.7873 on the WikiNews dataset and 0.7619 on the Wikipedia dataset. In Spanish language results, TMU ranked first with an F-Score of 0.7699. An interesting study was presented by the NLP-CIC [99] team, comparing two perspectives, one from feature engineering and the other from deep learning. This team used features such as target word frequency, term frequency, inverse term frequency, document frequency, syntactic and lexical features such as OpenNLP's PosTag, word embedding features, using a pre-trained 300-dimensional model and measuring the distance between the sentence and the target word. In the best results in the English language, NLP-CIC obtained an F-Score of 0.8551 in the News dataset, in the WikiNews dataset it obtained 0.8308 and 0.7722 in the Wikipedia dataset. In its best result in the Spanish language, NLP-CIC achieved an F-Score of 0.7468. Table 2.4 provides an additional summary of the systems involved in the task.

Similarly, tasks specific to the Spanish language arose in subsequent years. This is the case of the ALexS 2020 workshop [100], which had as its main task the detection of difficult terms found within the scope of academic content. To perform this task, the workshop relied on the VYTEDU-CW corpus, generated by the authors of the same competition. An important fact in this task is that no training data was provided, but only dev data to adjust the systems to the file formats. One participant was UDLAP [101], which

ALexS (SEPLN-2020)		
Team	**Technology/Strategy**	**Features**
UDLAP	Lexicon strategy Threshold strategy	Lexicon (CREA) Lexicon (Internet-related terms) Frequent n-grams Word Frequency
Vicomtech	Word level features K-Means	Lemma length Lemma frequency Wordnet synsets Lemma probability Word frequency (Wikipedia) Word probability (Wikipedia)
HULAT	SVM	Word length Boolean features Word2vec vectors BERT vectors

Table 2.5: Description of systems presented at ALexS Workshop (SEPLN-2020)

presented approaches based on lexicons and thresholds, which in its best version obtained the highest macro F1 score of 0.27 and the highest macro precision score of 0.34. The Vicomtech [102] team used a clustering method (K-means) supported by several features for classification such as lemma length, frequency of lemmas extracted from domain-specific resources, number of synsets extracted from Wordnet, lemma probability and word frequency/probability extracted from Wikipedia. Finally, the HULAT [103] team, presented a supervised approach using length features, Boolean features, easy-to-read lexicon-based features and vectors extracted from embeddings models, trained with data from the BEA 2018 workshop. Table 2.5 provides additional information on the systems involved in the task.

Finally, other works approach this customized task to each user independently, performing an adaptive CWI. For example work [104], which presents a system that continuously learns from user feedback and consequently obtains a model that evolves with the user using the system. In a similar way, the work of [105] presents a system that adapts to the user's native language, by differentiating words that are written in a similar way between languages (French, English, German and Spanish) but have divergent semantics, thus obtaining the ability to simplify texts in different languages in a more personalized way.

Thesis approach: To further improve the results of this task and in relation to the described state of the art, an exploration of methods and features for the representation of a word is carried out in this Thesis. During this process, new machine learning methods are proposed. Also, resources from easy-to-read content and word embedding have been proposed and created.

2.3.1.2 Substitute Generation (SG)

The next step is SG, which involves producing substitute candidates for the complex words detected in the previous step [65]. A good generator seeks to have maximum recall by producing possible replacements for a target word in all the contexts it may appear. Paetzold [8] groups substitute generation approaches into two categories: linguistic database querying and automatic generation:

1. **Linguistic database querying strategy:** as its name suggests, uses linguistic databases manually constructed by professionals in which a target word has a number of synonyms or words related to it. The best-known resources in this type of strategy are Wordnet [106] and different Thesaurus [64]. Although this strategy presents reliable replacements, it has the disadvantage of being a time-consuming task to create and of not having a wide coverage, especially in languages other than English, as demonstrated in the work of Shardlow [78], where in analyzing the errors in his results, he found that on several cases it was due to WordNet's failure to provide simpler synonyms. Some studies have explored ways to deal with this disadvantage by combining more than one linguistic resource. Such is the case of [82] [65] which combines resources for the Spanish language such as Babelnet [107] and Thesaurus. Also, the work of Leroy et al. [76] where the advantage of Wordnet information together with additional resources such as Wiktionary and UMLS (Unified Medical Language System) is taken. For the Portuguese language, the PorSimples project [108] aimed to improve its coverage by combining information from different Portuguese databases [109] [1].

2. **Automatic generation:** takes into account the disadvantages presented in the querying strategy. This strategy seeks to gather candidates extracted from less expensive resources. For example, Kajiwara [110] takes advantage of Japanese dictionaries that lack synonyms, but have word descriptions, with the aim of subsequently extracting the POS tag from these descriptions and finally putting as substitution candidates all words containing the same POS as the target word. A common method in a variety of papers is the use of paraphrases as a replacement for synonyms, which can be seen in papers where, supported by aligned resources, document [111] or sentence [112] level extraction methods are used. One of the most widely used resources for obtaining paraphrases has been Simple English Wikipedia (SEW), which contains English Wikipedia texts adapted for easier reading to a wider audience. Paraphrases can be extracted by various filters [113] [83], or more sophisticated methods as in the work of Feblowitz and Kauchak [114], which uses a tree transduction model to extract simplified content. In a similar way Paetzold and Specia [115] investigated the learning of tree transduction rules to extract possible lexical and syntactic simplification rules.

[1] http://www.linguateca.pt/PAPEL

But probably the best-known resource belonging to this strategy is the paraphrase database (PPDB), which has been used in different ways, one of these being the work of Pavlick and Callison-Bruch [116], where they trained a classifier to detect complex-simple paraphrase relations. This work led to the creation of the Simple Paraphrase Database (Simple PPDB), which contains billions of paraphrases in different languages, including Spanish.

Later, with the emergence of word embedding technology, different works to solve this task were presented. Glavas and Stajner [117] used this technology to extract possible substitutions that have the smallest cosine distance with the target complex word, because these models capture different word features, such as synonymy [118]. Other work took this idea further by using contextual models together with POS tag filters to obtain a list of possible replacements of higher quality [58]. Recent approaches use new representations of word embeddings, such as the LSBert [119] system, which uses the novel BERT models to mask the target complex word in the sentence in order to select words with higher probabilistic distribution as candidates for substitution, obtaining results superior to the state of the art, on different datasets.

Thesis approach: Although over the years resources have been produced for the Spanish language, there is a clear difference with other languages such as English, which has a wide variety of linguistic resources. In this Thesis, an exploration of existing linguistic resources for the Spanish language is carried out. At the same time, this Thesis experiments with unconventional methods that do not require a large amount of annotated data to operate, such as the methods offered by word embedding models.

2.3.1.3 Substitute Selection (SS)

In the third step, in which a substitute is selected from the set of synonyms extracted from the previous step, the most suitable synonym is chosen according to its complexity and the context [65]. In this step, the list of possible replacements generated in the previous step is received. This step is the most important in simplification, since a lexical simplification system must be prevented from making substitutions that affect the meaning or fluency of a complex sentence. As in the complex word identification stage, approaches that do not incorporate a selector in their processes by considering all candidates as possible substitutions have been presented [75]. However, in later work was shown that the absence of this stage, negatively influenced the preservation of the meaning of the original sentence [78]. Paetzold [8] describes the following strategies for accomplishing this task:

1. **Explicit sense labelling:** Due to the low precision obtained with the selection of every candidate, many works using strategies from other tasks are found. An example is the explicit sense labelling strategy, which attempts to tackle the selection as a word sense disambiguation (WSD) problem, by searching for the sense of the

word in its respective sentence, and then selecting possible substitutes that follow the same sense. Some papers that follow this strategy use WordNet as their main resource, taking advantage of sense tags [8], synsets [120] or using synonyms [121] extracted from this resource. Unfortunately, as in previous stages that rely on linguistic resources, this method has limitations since these resources are expensive to produce or extend [7].

2. **Implicit sense labelling:** tries to improve on the weaknesses of the previous strategy, by automating the learning of sense labels of complex words, rather than searching from a sense database. In [122], alternative WordNet words and a language model are used. This language model was trained on a large amount of unannotated data, where at training time the context of the words was taken into account. With this combined method, the results showed a precision of 65% and 57.6% versus the baseline using WordNet alone, which obtained 53.2% and 45.9%. Although this strategy obtains better results than the previous one, the algorithms required are more complex to build [7].

3. **Part-of-speech tag filtering:** focuses on providing a simpler selection alternative, especially in languages where there are few or no resources for WSD, by using the POS tags of the target word and possible substitutes. For example, some projects focused on the Portuguese language [69] [108], found a functional alternative with this strategy by selecting candidates that have the same POS tag as the target word. However, this strategy has presented problems when dealing with highly ambiguous words, where a word may represent more than one POS tag [8].

4. **Semantic similarity filtering:** consists of setting a similarity metric between the meaning of a word in a given context and the substitute in the same context in order to discard all the candidates that do not have sufficient similarity with the complex word. Early work shows an interesting strategy [83] by setting a semantic threshold, then calculating the similarity of words in a given context and discarding candidates with similarity below the threshold. A notable advantage of their strategy is that it can be adapted to other languages and does not rely on manually annotated linguistic databases. Later, similar filters were presented, as in [123], where they use a word embedding model to extract semantic similarity values between candidates and the target word context, which is extracted by calculating the cosine distance between the vectors of the words to be analyzed.A similar approach for the Spanish language was presented in the LexSiS system [64] [56] by using a vector space model to extract and select the lowest cosine distance. Another approach for the same language was presented in [82] [65], where the similarity of the candidates with the context of the words being or not being words with lexical content is analyzed. Later, this study was extended in [124], where this strategy was evaluated with different classical embedding models, such as Word2vec or Fasttext, and contextual embeddings such as BERT or Sense2Vec, achieving better precision than

previous works.

Thesis approach: A similar approach as [123] is followed in this Thesis by extracting similarity metrics between the target word and its context. To optimize results, an exploration, creation and comparison of different types of embeddings are performed. These embeddings are, for example, Word2Vec, Fasttext, Sense2Vec and BERT.

Since to this Thesis knowledge there are no Sense2Vec models for the Spanish language, a model is created within the framework of this Thesis.

2.3.1.4 Substitute Ranking (SR)

Finally, the SR step consists in deciding which of the candidate substitutions that fit the context of a complex word is the simplest. This step has the target audience as a priority, because given the needs of the audience, this step quantifies the complexity of the candidates received in the previous stage, with the objective of providing the simplest candidate [125].

1. **Frequency-based:** approaches use the intuition that the more occurrences a word has, the simpler it can be. Despite being a simple method, good and consistent results have been shown by work to date.

   For the English language, the best-known metric is Kucera-Francis [126] which is based on the frequency of words in the Brown [75] corpus. Further work has shown that the resource that supports these metrics can be determinant in the results of the task. As in [127], where frequencies were extracted from a corpus composed of movie subtitles, finding that they captured more common words than in the Brown corpus, and consequently, better results than the Kucera-Francis coefficient were obtained.

   Other works use frequencies extracted from long corpora. Resources such as the Microsoft N-gram Services platform[2], which offered access to several languages, including German, French and Chinese [128]. Another resource is Google 1T Corpus[3], which has versions for European languages[4], including Spanish and contains more than a trillion words [76] [129]. Other works use resources focused on the Spanish language, as in [124] where they use frequencies extracted from the CREA corpus of the RAE, achieving good results when compared with presented systems in the first task of SemEval 2012 [130].

   However, the biggest disadvantage faced by this method is the dependence on the size of the frequency resource. This size may vary depending on the language and

[2]Microsoft Web Language Model API. https://azure.microsoft.com/engb/services/cognitive-services/web-language-model

[3]https://catalog.ldc.upenn.edu/LDC2006T13

[4]https://catalog.ldc.upenn.edu/LDC2009T25

consequently, the performance. Furthermore, while the intuition of this method is good, it is not perfect, because in order to determine the simplicity of a word, there are other features to consider, such as word length [131].

2. **Simplicity measures:** emerge as an alternative to the disadvantages described in the previous method, where different features of a word are combined to determine a specific simplicity metric. Over the years, different combinations were shown, as in [83] where they combine frequency and word length as a simplicity metric. Another work [132] combines word frequencies in different corpora, word length and number of WordNet senses to determine simplicity.

   Later, more complex approaches were built by using additional information outside the word, such as its context. In [110] present a weighting system designed for the Japanese language, where they present five different metrics that take advantage of the relationships of the target word and its sentence. These metrics use resources of frequencies, senses, trigrams, among others. Another similar work [117] tried to deal with this in an automatic way, by averaging several types of features such as semantic or context similarity. Recently, a work that is part of this Thesis [124], similarly proposed a weighting system, which uses frequency features and features extracted from machine learning methods that will be described in the following strategy. The method has the intuition to choose the word with the highest final score as the simplest.

3. As with other methods and given their current popularity, **machine learning-assisted** methods are featured in the rankers. A good example of this strategy is the "UOW-SHEF-SimpLex" [133] system that was presented in SemEval-2012 task 1. This system used an SVM ranker, supported by features extracted from a frequency model, bag-of-words and psycholinguistic features. Another work based on this strategy [123] introduced a different supervised approach called "Boundary Ranking", which combines thresholds and ngram frequencies extracted from subtitles, to finally train a linear model. Years later, more sophisticated approaches have been emerging, as in [8] which calculates the difference in simplicity between a pair of candidates by using a multi-layer perceptron that receives several features from the same pair. A related approach was that of Maddela and Xu [134], where they proposed a novel neural model of readability classification with a Gaussian-based feature vectorization layer that uses human ratings created by the authors (lexicon of 15,000 English words), to measure the complexity of any word or phrase.

As in the CWI stage, shared tasks for the substitute ranking stage have been proposed. One of the most prominent was the task performed at SemEval 2012 [130], where participants were asked to rank a series of alternative substitutes, all of them considered suitable for a target word in context, according to how "simple" these substitutes are. It is important to mention that in the data published for the task, simplicity ties were contemplated in cases where words were considered equally "simple", thus representing an additional challenge to participants in their rankings. Table 2.6 shows a summary of the participants in this task, which have been described in previous points.

SemEval-2012 Task 1		
Team	**Technology/Strategy**	**Features/Resource**
ANNLOR [128]	Language models Association measures SVM ranker	Microsoft Web n-grams BNC Corpus Simple English Wikipedia Bing Search Engine
EMNLPCPH	Binary classifier	WordNet Web corpora Word n-gram probabilities Character n-gram probabilites Distributional differences Candidate length Document syntatic complexity Letter-wise recognizability
SB-mmSystem [135]	Simplicity measures Rule-based approach	WordNet frequency Relevance rules Number of senses
UNT [132]	Simplicity measures	Simple English Wikipedia English dialogues WordNet Google Web1T data
UOW-SHEF-SimpLex [133]	SVM ranker	N-gram frequency model Bag-of-words Psycolinguistic features

Table 2.6: Description of systems presented at SemEval 2012 Task 1

Up to this point, tasks as CWI are treated as a binary problem where the goal is to determine whether a word is simple or complex. However, a recent workshop [136] at SemEval 2021, changes this premise by assigning a continuous scale value to identify the complexity of a word, which they call Lexical Complexity Prediction (LCP), which at the end is very related to both CWI and SR steps. The task offered two sub-tasks, one for uniwords and one for uniwords/multiwords. In addition, the organizers provided an English language resource with 10,800 instances, which was scored on a 5-point Likert scale, with 0: Very Easy, 0.25: Easy, 0.5: Neutral, 0:75: Difficult, 1: Very Difficult. Because the typical evaluation metrics were not sufficient to measure the quality of the systems, the organizers chose new metrics, among which are the "Pearson's Correlation" metric that measures statistical relationships between continuous variables and the "R2" metric that measures the proportion of the variance of the original labels captured by the predicted labels. Table 2.7 shows a brief description of the task's participants. One of the participants with the best scores was the DeepBlueAI [137] team by using a wide variety of pre-trained language models along with different training strategies such as pseudo-labelling and data augmentation to finally apply a stacking method to give the final prediction. With these methods, the team obtained the highest "Pearson's Correlation" in the second task, and the second-best in the first task. Like this team, the use of contextual embedding models stood out as a fundamental part of the presented systems. Such is the case of the JUST BLUE team [138] that leverages context information extracted from BERT and RoBERTa models, achieving the highest "Pearson's Correlation" score in the first task. Similarly, the RG_PA team [139] performs an assembly of RoBERTa models in its classification, obtaining the second highest Pearson's Correlation score in the second task. Other systems followed feature-based approaches [140], such as length, frequency or sentence-level features. Similarly, the "ANDI" system [141] focused on extracting features from a wide variety of embedding models such as Glove, Word2vec, BERT, RoBERTa, ELECTRA, ALBERT, DeBERTa; which go through a Gradient Boosted Regression classifier.

SemEval-2021: Task 1		
Team	**Technology/Strategy**	**Features/Resource**
DeepBlueAI	Emsemble of pre-trained language models	Pseudo Labelling Data Augmentation Stacked Training Models Multi-Sample Dropout
JUST BLUE	Emsemble of pre-trained language models Context and token prediction fine-tuned models	BERT RoBERTa
RG_PA	Emsemble of pre-trained language models	RoBERTa
Mosquera	Feature-based approach Gradient Boosted Regression	Length Frequency Semantic Sentence level readability
Andi	Feature-based approach Emsemble of pre-trained language models Gradiente Boosted Regression	Psycolinguistic features Glove embeddings Word2vec embeddings BERT RoBERTa ELECTRA ALBERT DeBERTa
CS-UM6P	Context and token encoding	BERT RoBERTa
OCHADAI-KYOTO	Context and token prediction fine-tuned models	BERT RoBERTa

Table 2.7: Description of systems presented at SemEval-2021: Task 1

Thesis approach: Considering the related work described above, this Thesis explores different simplicity features of a word. As a next step, a weighting system is proposed. This system combines embedding and frequency dictionary information to determine the simplicity of words in a list of replacement candidates.

2.4. Word Sense Disambiguation (WSD)

As seen in the substitute selection step, the task of WSD is necessary and very helpful, however, this task can help in any task where there is some kind of ambiguity. This task is known as identifying which sense of a word is used in a sentence. As new words continue to be added to the Spanish language, this task becomes increasingly complex.

1. **Knowledge-based** strategies require extensive lexical resources to determine the sense of a target word. Lesk [142] follows a knowledge-based approach by overlapping the word context and the sense definitions from a machine-readable dictionary. Next, the sense that has a greater number of words in common with the context of the target word is chosen. This approach is dependent on finding the exact words in the definition, resulting in poor performance.

2. **Supervised** strategies require sense tagged corpora to train an algorithm to determine which sense is correct. For example, a team in the Senseval-3 competition [143] followed a supervised approach by training SVMs supported by the neighboring word's POS tags, single words around the context and syntactic relations, showing good recall when compared with the other participants in the task.

3. **Unsupervised** are more recent than the previous strategy and most of the research relies on using comparable corpora strategies [144]. Additionally, with the introduction of word embeddings, many investigations combine these concepts. Moradi [145] trained a Word2Vec model to evaluate similarity measures to disambiguate the Persian language by considering semantic relationships between words.

4. **Semi-supervised** strategies appear by attempting to improve the disadvantages of each of the previous approaches. Generally, they consist of training classifiers from a small amount of labeled data and a large amount of unlabeled data. A recent work [146], proposes a semi-supervised method to improve a Long Short Term Memory (LSTM) model using self-learning and constructing the model by using the few labeled data to which they had access.

At the same time, many competitions have been organized with the goal of solving WSD problems. The SemEval-2007 competition [147] was focused on correctly disambiguating and identifying the semantic relationship between words. The SemEval-2013 [148] competition addressed this task by presenting a multilingual sense-annotated corpus, one of which was in Spanish, tagged with WordNet, Babelnet, and Wikipedia. The SemEval-2015 [149] addressed both WSD and Entity Linking (EL) to analyze and find ways to solve these tasks with similar methods by providing resources that integrate encyclopedic knowledge and lexicographic information.

However, others have tackled this problem from another point of view. Such is the case of Google, which uses its BERT language representation model (Bidirectional Encoder Representations from Transformers) [37] to solve different tasks in NLP by fine-tuning their pretrained models. A research project carried out based on this approach [14] consisted in fine-tuning a BERT model for WSD using WordPiece embeddings as part of the entries. Good results were obtained, overperforming the results of current approaches in F1-scores. The core process of this Thesis WSD system follows the aforementioned approach.

This task greatly supports other areas, such as simplification systems [150] oriented to people with cognitive disabilities. For the benefit of people with disabilities with communication and language problems, research has been found on systems that include WSD such as [151], to provide predictive text functionality, and research concerned with the selection of a correct pictogram [152].

Thesis approach: In this Thesis, the accessibility guideline which requires a mechanism that provides definitions to an unusual word is followed. However, this is not a trivial task, as there is a wide ambiguity in the Spanish language, where a word can have several definitions.

Given this scenario, the task of WSD is beneficial, as it provides mechanisms for such problems. In this Thesis, a new process for finding correct definitions for an unusual word is proposed, which is supported by state of the art resources such as BERT models for the Spanish language.

2.5. Resources and Corpora

Lexical resources and corpora play an important role in the development and evaluation of simplification systems. Unfortunately for Spanish, few annotated texts can be used to satisfy these tasks. Manual production of simplified texts is a non-trivial and, at the same time, expensive task [13]. In this Section, various projects related to TS and the improvement of readability are presented. Additionally, annotated text corpora used for Text Simplification and, more specifically, lexical simplification works are described.

As seen in several projects described above, the Internet's amount of information is insufficient to make a stable system capable of simplifying texts for a specific audience. The data must be prepared in a certain way for the computer to easily find patterns and inferences, and some relevant metadata specific to the task has to be added [153].

Table 2.8 shows the described and some additional examples of important resources in TS for English and Spanish.

Resource	Annotated text	Size	Language: English (EN) Spanish (ES)	Annotated method
Simple English Wikipedia	A simplified version of regular Wikipedia	183,000 content pages to date	EN	Pages edited by 1203 active users
SemEval 2012 [130]	English Internet Corpus [154]	2010 instances of simplicity rankings	EN	Native English speakers
LSeval [155]	English Internet Corpus [154]	430 instances of simplicity rankings	EN	46 Amazon Mechanical Turk (turkish), 9 PHD students
LexMTurk [84]	Wikipedia	500 instances with target complex words and simpler synonyms	EN	50 Turkish English speaking
BenchLS [156]	Compilation of LSeval and LexMTurk	929 instances with an average of 7 candidates per complex word	EN	Corrected and filtered by English speakers
NNSeval [58]	Filtered version of BenchLS	239 instances	EN	Non-native english speakers
Wikipedia - Simple Wikipedia	Simple English Wikipedia	167,689 aligned sentences	EN	Language modelling [157]
PWKP (WikiSmall) [85]	Wikipedia and Simple Wikipedia	108,016 aligned sentences	EN	Statistical machine translation
Simplext [56]	News texts	200 aligned news texts	ES	Human editors trained in easy-to-read guidelines
SS Corpus [158]	Wikipedia and Simple Wikipedia	492,993 aligned sentences	EN	Unsupervised method
Newsela [159]	News articles	Parallel simple-complex articles with 11-grade levels	EN, ES	Manually produced by professional editors
RANLP 2017 [92]	Wikipedia	14,280 instances with target complex words	EN, ES	54 turkeys (Native and non-native speakers)
WikiLarge [160]	WikiSmall, Aligned sentences pairs [157][161]	2,000 for dev 359 for test 296,402 for training	EN	Combination of previously created simplification corpora

PPDB-S/M [162]	PPDB	5709 unigrams for S size 15,524 unigrams for M size	ES	Built by filtering and ordering paraphrases pairs from the paraphrases database (PPDB)
CASSA [162]	CASSA dataset	5,640,694 5-grams	ES	Generated by extracting all unique 5-grams pairs from CASSA resource
ASSET [163]	TurkCorpus extension	23,590 human simplifications associated with the 2,359 sentences from TurkCorpus	EN	Amazon Mechanical Turk
VYTEDU-CW [100]	Transcripts of academic videos	9175 words, 723 annotated as complex	ES	430 annotators students

Table 2.8: Text simplification resources for English/Spanish

Thesis approach: At this point, it is easy to determine that many NLP methods, or specifically methods in lexical simplification, need data to operate, either to train algorithms or to validate them. Furthermore, taking into account the limited or in some cases the non-existent amount of resources for lexical simplification in the Spanish language, this Thesis includes the creation of a new resource to support this task.

This corpus differs from existing resources because it possesses complex target words, along with proposed contextualized synonyms. The added value of this corpus regarding state of the art is that it was annotated and validated by linguists who are experts in easy reading and plain language in Spanish to address the scope of cognitive accessibility.

2.6. Conclusions

As part of the main objective and specific objectives of this Thesis, aspects related to the field of accessibility to support audiences with cognitive disabilities and lexical simplification were introduced.

As seen in approaches pertaining to the literature, this Thesis follows accessibility guidelines concerning cognitive disabilities to textual content. Also, the steps of a complete process of lexical simplification in a generic domain for the Spanish language have been described: CWI, SG, SS and SR. For each of the steps, the Thesis approach has been described with the definition and use of new methods and resources to confirm the research hypotheses formulated.

Furthermore, keeping in mind the lack of resources for training, tuning and evaluation of NLP methods, this Thesis describes existing resources and then describes the procedures for the creation of a new resource for the support of lexical simplification methods for the Spanish language to support audiences with cognitive disabilities.

Finally, due to the fact that quite a lot of ambiguity is found in the Spanish language, works related to word sense disambiguation were studied, with the aim of offering a solution to the accessibility requirement of providing a mechanism for generating correct definitions to a target word.

Chapter 3

Lexical simplification process: approaches and resources

In this Chapter, the Thesis proposal is presented, which includes a complete lexical simplification process using new NLP approaches and resources to fulfill the objectives of this Thesis and confirm the formulated research hypotheses. Figure 3.1 illustrates the architecture of the proposal, where one can differentiate the tasks and resources in which this Thesis contributes to state of the art, such as the steps of lexical simplification and word sense disambiguation. In addition, resources created within the framework of the Thesis are shown, such as the easy-to-read dictionary and the EASIER corpus, which is described below.

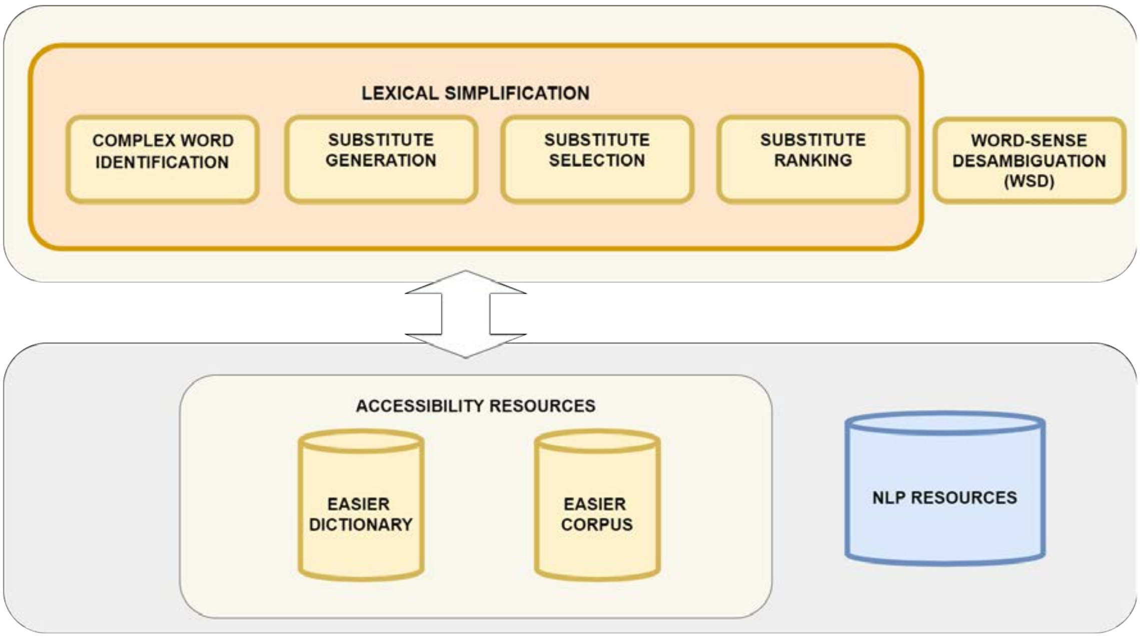

Figure 3.1: Proposal architecture

3.1. Accessibility Resources

This Section presents accessibility resources to support audiences with cognitive disabilities created throughout this Thesis. As shown in Figure 3.1, these resources were created to support lexical simplification steps.

3.1.1. EASIER Corpus

As told in Chapter 2, the need for annotated data has always been present and the creation of new resources to support NLP methods is helpful. That is why, in the framework of this Thesis, the EASIER corpus was created.

This corpus aims to offer evaluation support for CWI tasks and fitting SG/SS tasks contextually. This has been achieved through the assistance of a linguist who is an expert in easy-to-read and plain language guidelines annotating 260 documents. Two additional experts and a target audience have analysed the resulting corpus to assure the quality of the data provided.

3.1.1.1 Corpus building

The corpus building includes three steps (see Figure 3.2). In addition, the evaluation and results concerning these steps are described in Chapter 4.

1. **Initial Step.** Based on the annotator's experience and knowledge in easy-to-read and plain language guidelines, various annotation criteria are established (see Section 3.1.1.2) to detect complex words and the suggestion of simple synonyms.

2. **Annotation Step/ Early evaluation.** The annotator performs the analysis of the texts according to the annotation criteria using the annotation tool. Additionally, to validate the annotation's current state, an initial evaluation with the participation of people with intellectual disabilities of the set of texts annotated to date was performed.

3. **Evaluation.** Once the documents have been fully annotated, the resulting corpus is described in the following Section. A portion of the data set is extracted and annotated by two other annotators for comparison with the original annotations.

Figure 3.2: Corpus building methodology

To complement the information on the different steps in the construction of the corpus, it is necessary to mention the selection process of annotators, together with the materials and instruments necessary for the construction of the corpus, which is described below:

Selection of Annotators: Three human annotators have taken part in this work. One was the annotator of the entire corpus, while the other two annotated part of the corpus to calculate the Inter-Annotator Agreement (IAA). All three annotators are native Spanish speakers, expert linguists and specialists in easy-to-read and plain language guidelines. They have extensive experience of at least 15 years in transforming conventional texts into easy-to-read texts. They belong to Plena Inclusión[5] Madrid and Grupo Amas Fácil[6], two organisations that work to offer resources to people with intellectual and learning disabilities. It should be noted that these annotators in these organizations adapt the texts manually following a methodology. This methodology involves people with cognitive disabilities throughout the process. This means that the experts annotators who have elaborated the EASIER corpus have a great experience in simplifying texts taking into account people with intellectual disabilities.

Materials: News from "60 y más" journal[7] were randomly selected based on its length and the beginning of 2019 to the first months of 2020. A total of 260 documents were set containing long texts concerning a range of different topics in the areas of current affairs, health, guides for seniors and news. Each document had a similar length, and there was an average of 15 sentences per document[8]. This journal belongs to Imserso[9] , the Institute for the Elderly and Social Services in Spain, which is focused on the elderly. This group's main objective is to promote the social integration of the elderly through information in Spanish.

Instruments: Annotators used an annotation tool created as an extension for Google Chrome[10]. It has been developed by the PhD student to perform the following actions:

1. Select and deselect words that are considered complex or unusual in a given text.
2. Propose simple, context-appropriate synonyms for the target word on the following page.

3.1.1.2 Annotation Criteria

As mentioned in the previous Section, the annotators' annotation criteria to guide the creation of this corpus are presented below:

- **Annotation criteria :** Once the annotator has carried out an initial reading, she

[5] plenainclusionmadrid.org/

[6] amasfacil.org/

[7] Inmerso´s Journal for the ederly revista60ymas.es

[8] Example of "60 y más" document revista60ymas.es/60mas_01/actualidad/2019/noviembre /IM_128077?dDocName=IM_128077

[9] Institute for the Elderly and Social Services imserso.es/

[10] Annotation tool git repository github.com/ralarcong/EASIER_AnnotationTool

returns to the beginning of the document, searching for complex words following the criteria given below:

1. Words that are common in verbal communication but which our target audience does not necessarily know. A reference used to help the annotator identify these words is usually found in the Spanish linguistic frequency indexes (Gran Diccionario de Uso del Español Actual, Corpus CREA[11], Corpus CORPES XXI[12]).
2. The length of the word is also considered. In the process of reading long words, people can lose information. The syllable configuration must also be considered. When syllables are long or have more consonants, the effort needed to pronounce said syllables could affect comprehension.
3. Difficulty in reading or pronouncing words, such as “esternocleidomastoideo” (sternocleidomastoid). In addition to being a long word, it isn’t easy-to-read and pronounce.
4. Complexity, as regards technical words. An example would be the terms used in the medical or legal fields.
5. Abbreviations or acronyms are considered when the explanation is not provided in the document. For example, a document that deals with the work of the WHO, but the words “World Health Organization” are not included.
6. Words in other languages are also included as complex words. Since EASIER’s target audience is the elderly and people with disabilities, it should not be assumed that they know other languages.
7. Roman numerals are complex.
8. Idioms and expressions are also included as they can have double meanings that are difficult to understand.
9. Metaphorical expressions are also included among the statements to be considered since their comprehension is very complex.
10. Some uncommon abstract terms such as "justice" or "emotion" whose physical forms cannot be perceived or imagined. Some of the less common terms are difficult to understand.
11. Complex concepts formed by several words, which could give rise to different situations:
    - Complex expression comprised of complex words: key indicators, contractual resources.
    - Complex expression comprised of simple words whose more familiar meaning was modified: social tourism, portfolio of services.

[11]List of frequencies from the RAE corpus.rae.es/lfrecuencias.html
[12]rae.es/banco-de-datos/corpes-xxi CORPES corpus, versión: 0.93

- Complex expressions comprise complex and simple words whose most well-known meaning has been modified: strategic framework, inter-territorial council.

12. Common words that have their most common meaning modified by the context in which they are found (linked to polysemy). For example, the word "active" when meaning the portion of the population either with a job or looking for a job, instead of referring to a person who likes to be active.
13. Percentages and mathematical expressions are complex, as are numbers expressing large quantities.
14. Adverbs ending in "-mente" (-ly) are more complicated because of their prolonged pronunciation.
15. Collective nouns, for example "indumentaria" (clothing). The global concept is more complicated than the enumeration.
16. Words that are antiquated or in disuse.

Table 3.1 provides examples of words selected according to the criteria described in this Section.

Sentence	Criteria	Word
El Ministerio de Sanidad, Consumo y Bienestar Social ha remitido hoy los informes provisionales de cinco de las 66 técnicas sometidas a evaluación dentro del Plan de Protección de la Salud frente a las Pseudoterapias. (Today, the Ministry of Health, Consumer Affairs and Social Well-being has issued the provisional reports on five of the 66 techniques evaluated by the Health Protection Plan against Pseudotherapies.)	Pseudoterapias (Pseudotherapies)	2, 3, 4
El punto de partida de este proceso fue un análisis exploratorio inicial de 138 técnicas o procedimientos, basado en una revisión de las publicaciones científicas (revisiones sistemáticas y ensayos clínicos). (The starting point of this process was an initial exploratory analysis of 138 techniques or procedures, based on a review of scientific publications (systematic reviews and clinical trials))	ensayos clínicos (clinical trials)	1, 2, 11, 12
Las instituciones que lo forman son la OMC, el Consejo General de Enfermería, el Consejo General de Colegios Oficiales de Farmacéuticos, AEMPS, SECA, el Consejo General del Trabajo Social, SEMG, semFYC, AEP, SEMERGEN, EUPATI, CERMI, la SEMI. (The institutions comprising this group are the WTO, the General Council of Nursing, the General Council of Pharmacists Associations, AEMPS, SECA, the General Council of Social Work, SEMG, semFYC, AEP, SEMERGEN, EUPATI, CERMI, and the SEMI.)	semFYC	5
… al tiempo que ha manifestado la necesidad de que el desarrollo del marco estratégico de su abordaje tenga en cuenta “la efectividad, la eficiencia, la accesibilidad, la seguridad y la atención centrada en el paciente”. (…while at the same time stating the need for the development of the approach of the strategic framework to take into consideration “the effectiveness, efficiency, accessibility, security and patient-focused care.”)	marco estratégico (strategic framework)	1, 2, 11, 12
El tercer eje se refiere al derecho, la ética y la dignidad de la persona y pretende mejorar los servicios, apoyos y prestaciones para avanzar en la atención a las personas en distintos ámbitos. (The third pillar refers to the right, ethics and dignity of the person and aims to improve services, support and benefits in order to improve care for people in different areas.)	Ética (ethics)	10
… así como recambios de componentes externos de implantes quirúrgicos (componentes externos del estimulador diafragmático o electroestimulador del nervio frénico). (… as well as the replacement of external surgical implant components (external components of the diaphragmatic stimulator or phrenic nerve electro-stimulator).)	Diafragmático (diaphragmatic)	2, 3, 4
“España está trabajando en este objetivo en el marco de la iniciativa Connecting Europe Facility”, ha señalado la ministra. (“Spain has been working on this goal within the framework of the Connecting Europe Facility initiative”, the minister pointed out.)	Connecting	6
Desde los recogidos por encuestas de salud de la población, que obtienen periódicamente información en cerca de 30.000 hogares; las altas hospitalarias o el seguimiento en la Atención Primaria de una cohorte de 4,7 millones de personas. (From data collected by population health surveys, which periodically obtain information from around 30,000 households; hospital discharges or the follow-up of a cohort of 4.7 million people in primary care.)	Periódicamente (periodically)	2, 14
Una mirada a la indumentaria tradicional. (A look at traditional clothing.)	Indumentaria (clothing)	15
El spot de la campaña se emite en televisiones autonómicas y estatales desde el 14 al 24 de noviembre. (The campaign spot will be broadcast on both state and regional television stations from November 14th to the 24th.)	spot	1, 6

Table 3.1: Annotation Criteria examples

3.1.1.3 Corpus Statistics

A total of 260 documents were annotated, from which an average of 15 sentences per document was obtained. As a result, approximately 8100 complex words were gathered. At the same time, it should be mentioned that more than 5100 words for which at least one synonym was proposed were also obtained (see Table 3.2).

	EASIER
Documents	260
Sentences	3,778
Tokens	134,528
Average number of sentences per document	15
Average number of tokens per document	517
Total instances for CWI	44,975
Complex Words	8,155
Total instances for SG/SS	5130
Proposed synonyms	7,892
Average of complex Words per document	30
Average of proposed synonyms per document	29
Complex Words with at least one substitute	5,130

Table 3.2: EASIER corpus statistics

3.1.1.4 Corpus Structure

Two distinct datasets could be distinguished, one to support CWI tasks and another for SG/SS tasks. Each instance of the CWI dataset (Table 3.3) has six columns and are represented as follows:

- The first column shows the ID of the document.
- The second column shows the ID of the sentence for a particular word.
- The third column shows the sentence.
- The fourth and fifth columns show the offset of the target word.
- The sixth column shows the target word.
- The seventh column shows the gold-standard label.

Doc ID	Sentence ID	Sentence	Start offset	End offset	Word	Label
1	136	La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	3	14	Importancia (importance)	0
1	136	La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	18	22	Leer (reading)	0
1	136	La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	31	41	Etiquetado (labelling)	1
1	136	La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	51	58	Comprar (purchasing)	0
1	136	La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	62	70	Alimento (foodstuffs)	0

Table 3.3: CWI dataset instance examples

For the second dataset (Table 3.4), each instance has five columns and are represented as follows:

- The first column shows the ID of the document.
- The second column shows the ID of the target word.
- The third column shows the target word.
- The fourth column shows the sentence.
- The fifth column shows the suggested synonyms for the target word separated commas.

Doc ID	Word	Word ID	Sentence	Proposed Substitutions
1	167	Etiquetado (labelling)	La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	Letrero (sign), inscripción (inscription), rótulo (banner)
1	168	Etiqueta (label)	La campaña ‘Esta Navidad… Que tu mesa se vista de etiqueta’ con el objetivo de recordar a la población la información que debe tener en cuenta a la hora de comprar alimentos envasados o a granel así como los datos que deben figurar en las páginas web que venden online (The “This Christmas… dress up your table*” campaign aims to remind people of what information they need to keep in mind when purchasing packaged foodstuffs or in bulk, as well as the information that must be included on webpages that sell online. *in the original Spanish, the name of this campaign uses a play on words with the different meanings of “etiqueta”, which can mean either formal dress or a label.)	Ceremonia (ceremony), protocolo (protocol)
1	169	Envasados (packaged)	La campaña ‘Esta Navidad… Que tu mesa se vista de etiqueta’ con el objetivo de recordar a la población la información que debe tener en cuenta a la hora de comprar alimentos envasados o a granel así como los datos que deben figurar en las páginas web que venden online (The “This Christmas… dress up your table*” campaign aims to remind people of what information they need to keep in mind when purchasing packaged foodstuffs or in bulk, as well as the information that must be included on webpages that sell online. *in the original Spanish, the name of this campaign uses a play on words with the different meanings of “etiqueta”, which can mean either formal dress or a label.)	Empaquetados (packaging)
1	170	a granel (in bulk)	La campaña ‘Esta Navidad… Que tu mesa se vista de etiqueta’ con el objetivo de recordar a la población la información que debe tener en cuenta a la hora de comprar alimentos envasados o a granel así como los datos que deben figurar en las páginas web que venden online (The “This Christmas… dress up your table*” campaign aims to remind people of what information they need to keep in mind when purchasing packaged foodstuffs or in bulk, as well as the information that must be included on webpages that sell online. *in the original Spanish, the name of this campaign uses a play on words with the different meanings of “etiqueta”, which can mean either formal dress or a label.)	Suelto (loose), sin envase (without packaging)
1	171	online	La campaña ‘Esta Navidad… Que tu mesa se vista de etiqueta’ con el objetivo de recordar a la población la información que debe tener en cuenta a la hora de comprar alimentos envasados o a granel así como los datos que deben figurar en las páginas web que venden online (The “This Christmas… dress up your table*” campaign aims to remind people of what information they need to keep in mind when purchasing packaged foodstuffs or in bulk, as well as the information that must be included on webpages that sell online. *in the original Spanish, the name of this campaign uses a play on words with the different meanings of “etiqueta”, which can mean either formal dress or a label.)	en línea (online), conectado a Internet (connected to the Internet)

Table 3.4: SG/SS dataset instance examples

It is worth mentioning that this corpus is currently being used to give evaluation support for the lexical simplification steps described in the following Sections.

3.1.2. EASIER Dictionary

There is an important amount of texts manually simplified following easy reading guides and plain language. These texts have the potential to offer assistance to NLP methods by providing simple lexical content to support audiences with cognitive disabilities. Therefore, this Thesis takes advantage of these resources by extracting and filtering them to create a resource containing simple vocabulary, which was tested in one step of the lexical simplification described in Section 3.2.1.

3.2. Lexical Simplifier

This Section presents different approaches that have been taken throughout this Thesis to accomplish the task of lexical simplification. Figure 3.3 shows the approaches performed in this Thesis in a more specific way, where for each step, resources and outputs are displayed. Additionally, to make the content more understandable, as the accessibility guidelines suggest that the definition of an unusual word should be provided and since many words in Spanish are polysemic, a word sense disambiguation module has been created.

3.2.1. Complex Word Identification (CWI)

As indicated in Chapter 2 to improve results of state of the art, an exploration of methods and features for the representation of a word is carried out, and new machine learning methods are proposed. Also, resources from easy-to-read content to support audiences with cognitive disabilities and word embedding have been proposed and created.

In this stage, an algorithm needs to distinguish which words are complex and which are not for a certain audience. As this Thesis presents machine learning CWI approaches, datasets with words labelled as either complex or simple are required, to train and validate the algorithms. These datasets are described as follows:

As a first resource, this Thesis uses the EASIER corpus CWI dataset, which is described in Section 3.1.

As a second dataset, this Thesis uses the shared CWI task dataset from the BEA Workshop [92] [13]. This dataset is composed of 17603 instances that were annotated by 54 Spanish speakers, most of whom were native. Each instance contains a uniword/multiword target which is selected by annotators. This dataset allows us to evaluate the Thesis approaches and at the same time to compare with other works that have used the same dataset. The structure of this dataset is as follows:

[13] https://sites.google.com/view/cwisharedtask2018

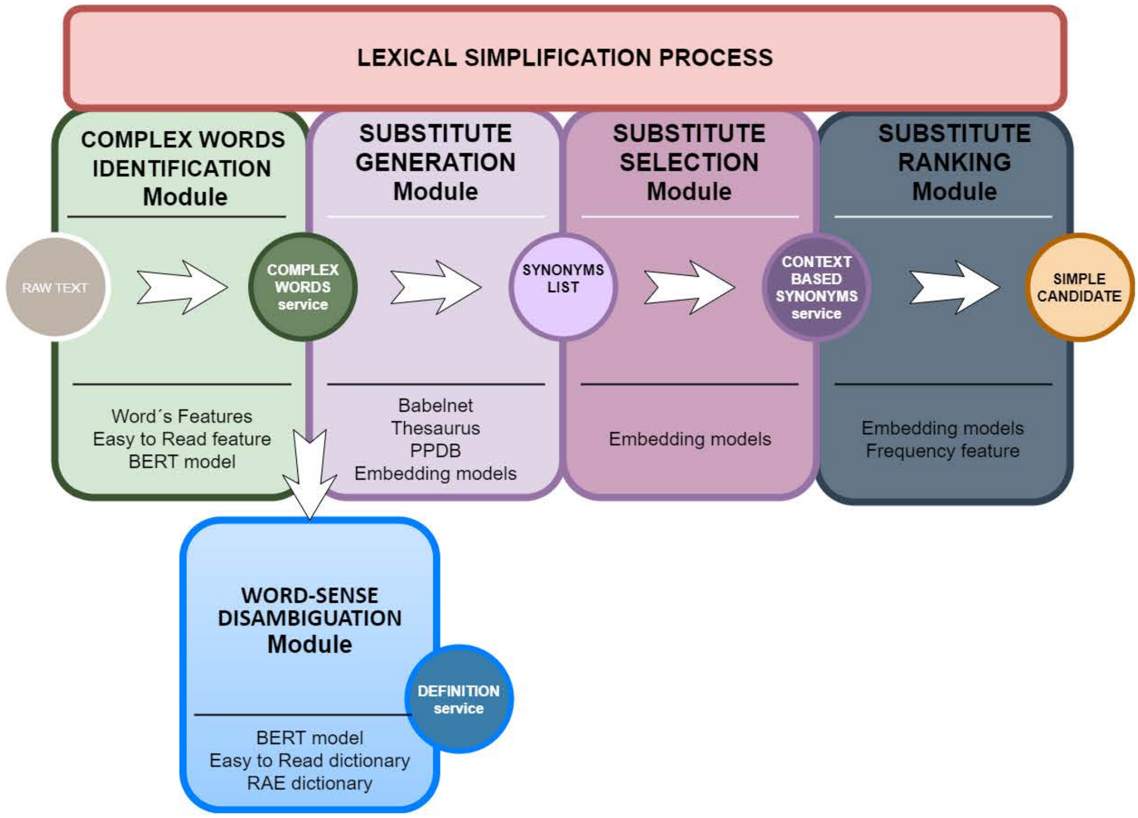

Figure 3.3: Proposal modular view of the simplification process

- The first column shows the ID of the sentence.
- The second column shows the actual sentence in which a complex phrase annotation exists.
- The third column shows the initial char offset of the target word in the sentence.
- The fourth column shows the end Char offset of the target word in the sentence.
- The fifth column shows the target word.
- The sixth and seventh columns show the number of native annotators and the number of nonnative annotators who examined the sentence.
- The eighth and ninth columns show the number of native annotators and the number of nonnative annotators who marked the target word as difficult.
- The Tenth Column shows the gold-standard label for the binary task (0: simple and 1: complex).
- The Eleventh Column shows the gold-standard label for the probabilistic task (<the number of annotators who marked the word as difficult>/<the total number of annotators>).

This experimentation does not consider the information in some columns (the sixth, seventh, eighth, ninth and eleventh columns) because the information in these columns is intended for use in the probabilistic task.

3.2.1.1 Feature Exploration

With these datasets and for the purposes of training the algorithm, each word (instance) must be represented as a set of features that help to distinguish between complex and simple words. Below, the features that were used in different approaches throughout this Thesis are described. This list of features is the result of several experiments [82] [103] [65] with combinations of them:

1. **Length Features:** Word Length, Sentence Length, Number of syllables of the word.
2. **Probability Features** using an Ngram Corpus: the Probability of the word(unigram), the Probability of bigram (word and left/right word), the Probability of trigram (word and left/right two next words).
3. **Boolean Features:** if the word is lowercase, if the word is Uppercase, if the word is a digit, if the word has uppercase characters, if the word is composed of punctuation symbols, if the word contains punctuation symbols.
4. **E2R Feature:** As an important contribution to this Thesis, resources from the domain of easy-to-read were taken as an advantage in the classification. As a result, this Thesis proposes a new feature with the creation of an E2R dictionary. For each word, if a target word exists in the E2R dictionary, it is classified as 0; otherwise, it is designated as 1. The dictionary is fed from a range of sources that provide easy-to-read literature developed by experts. Subsequently, this text is cleaned to preserve only the content words (noun, verbs, adjectives, adverbs). Presently, this dictionary contains 13,400 simple words.
5. **Word Embedding (Word2vec) feature:** for each word, vectors from a Word2vec model trained on The Spanish Billion Words Corpus are extracted [164].
6. **Word Embedding (Fasttext) feature:** or each word, vectors from a Fasttext model trained on Common Crawl and Wikipedia with the FastText tool with character n-grams of length 5 are extracted [165].
7. **Word Embedding (Sense2Vec) feature:** for each word, vectors from a Sense2Vec model are extracted. Since there are no Spanish Sense2Vec models, in this Thesis a model trained on The Spanish Billion Words Corpus was created [164]. The main diference between Sense2Vec and Word2Vec vectors is that the latter fails to encode the context by assigning a single key regardless of the context in which it appears. This does not happen in a Sense2vec model, because it generates vectors of words with contextual keys.
8. **Word Embedding (BERT) feature:** for each word, vectors from a Spanish BERT pre-trained model are extracted [166]. The main difference between this type of embedding and others, such as Word2Vec or FastText, is that BERT produces word

representations that are dynamically informed by the words around them, whereas Word2Vec the words are represented as unique indexed values. Therefore, with BERT embeddings, each word could have several vectors, one for each of its possible meanings. Therefore, these models allow us to deal with the task of word disambiguation when complex words are identified.

As said before, a machine learning approach is followed. In particular, Support Vector Machine (SVM) is used because its successful performance for text classification tasks. In principle, this Thesis used the radial basis function (RBF) kernel [82]. However, later on, an experimentation with a linear kernel was also performed, which is much faster and has the additional advantage that SVM has shown good performance in classifying sparse instances [65].

3.2.1.2 BERT for Complex Word Identification

In addition to previous traditional machine learning approaches. This Thesis proposes a BERT [37] for the NER approach. BERT is a powerful NLP tool but using it for NER without fine-tuning it won't give good results.

In this Thesis, a process of fine-tuning of two models is performed, with the help of the datasets described in the CWI Section. These BERT models are: Google´s multilingual BERT pre-trained model (mBERT) [37] and a Spanish BERT pre-trained model (BETO) presented in 3.2.1.

The process[14] was modified so the model could predict the entities "COMPLEX" and "SIMPLE" in a given text. The default parameters for the fine-tuning are the following:

- train_batch_size: 32
- max_seq_length: 128
- learning_rate: 2e-5
- num_train_epochs: 4.0
- do_lower_case: False
- CRF: True

As shown in Figure 3.4, the modified process takes an input where contextual embeddings are extracted from the embedding model and the Conditional Random Forest (CRF) layer gives the final decision on the target word´s entity.

3.2.2. Substitute Generation (SG)

The substitute generation step generates substitution candidates for complex words, considering all the contexts in which they may appear. To obtain better results than those of

[14] github.com/kyzhouhzau/BERT-NER

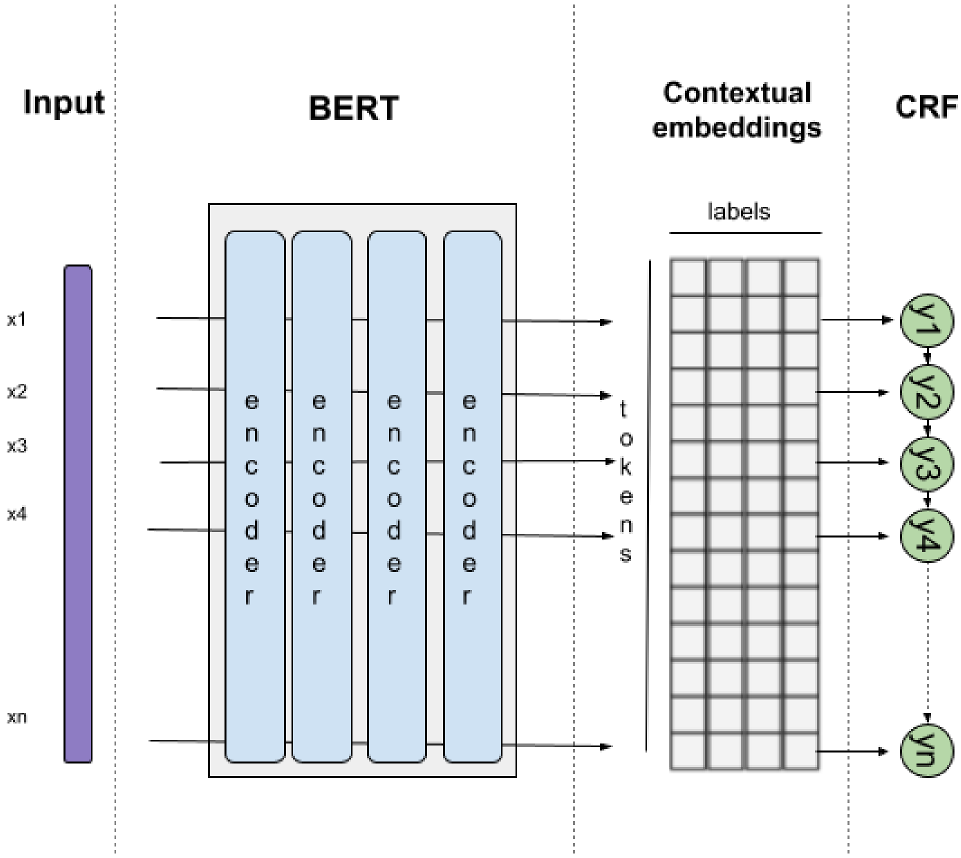

Figure 3.4: Fine-tuning process to a BERT model for the Complex Word Identification step

state of the art, this Thesis explores Spanish linguistic resources for this task. In addition, automatic solutions that are not dependent on manually annotated content, such as embedding resources are explored.

3.2.2.1 Resource Exploration

In this step, candidates for substitution for a target word from a variety of linguistic resources are extracted. Also, the extraction of candidates for unconventional resources such as word embeddings is investigated. These resources are described in the following:

1. **Thesaurus:** Synonym database [15].
2. **Babelnet:** Synonym database [16].
3. **PPDB:** Paraphrase database [17].
4. **Embedding models:** described at CWI Section (3.2.1)

[15] thesaurus.altervista.org/
[16] live.babelnet.org/
[17] paraphrase.org/

3.2.2.2 Generators

This Thesis tests the performances of different substitute generation strategies by using the resources mentioned above and applying rules to search for a better result. The combinations are described as follows:

(1) **Thesaurus database:** synonym search for the target word.

(2) **Thesaurus database:** search for synonyms for the target word and its lemma.

(3) **Babelnet database:** search for synonyms for the target word.

(4) **Babelnet database:** search for synonyms for target word and its lemma.

(5) **PPDB:** search for replacements for the target word.

(6) **PPDB:** search for replacements for target word and its lemma.

(7) **Babelnet + Thesaurus:** concatenate the extracted values from (2) and (4).

(8) **Babelnet + Thesaurus + PPDB:** concatenate the extracted values from (2), (4) and (6).

(9) **Babelnet + Thesaurus:** in addition to the procedure described in (7), the target word's lemma and stem are extracted. Subsequently, the candidate words that contain the stem or match the extracted lemma are deleted.

(10) **Babelnet + Thesaurus + PPDB:** in addition to the procedure described in (8), the target word's lemma and stem are extracted. Subsequently, candidate words that contain the stem or match the extracted lemma are deleted.

(11) **Embedding approach:** the performance of different embedding models were tested by extracting the nearest neighbors of each target word (50 neighbors).

3.2.3. Substitute Selection (SS)

As described in Chapter 2, state of the art methods are refined in this Thesis, by evaluating the context of the target word with similarity metrics supported by word embeddings. This procedure is evaluated with different types of embeddings to determine the best outcome.

This step takes the list of synonyms extracted from the previous step and selects the most suitable synonym according to its simplicity and context.

3.2.3.1 Selectors

This Thesis selectors use different types of word embedding models, from static to contextualized. The same embedding models as in Section 3.2.1 are used. These models allow this Thesis to calculate the cosine distance between word vectors to perform the following procedures:

1. **No selections :** selects all candidates.

2. **Any Window :** As shown in Figure 3.5, the procedure obtains three similarity values (candidate and target word, candidate and target word's context words in the sentence (previous and subsequent words)). Next, these values are added and stored. Finally, this process is repeated for every candidate, and the selector picks the three candidates with the highest values.

3. **Lexical window :** Similar to (2), but instead of selecting the first context word, the first word with lexical content (previous and subsequent words) is selected.

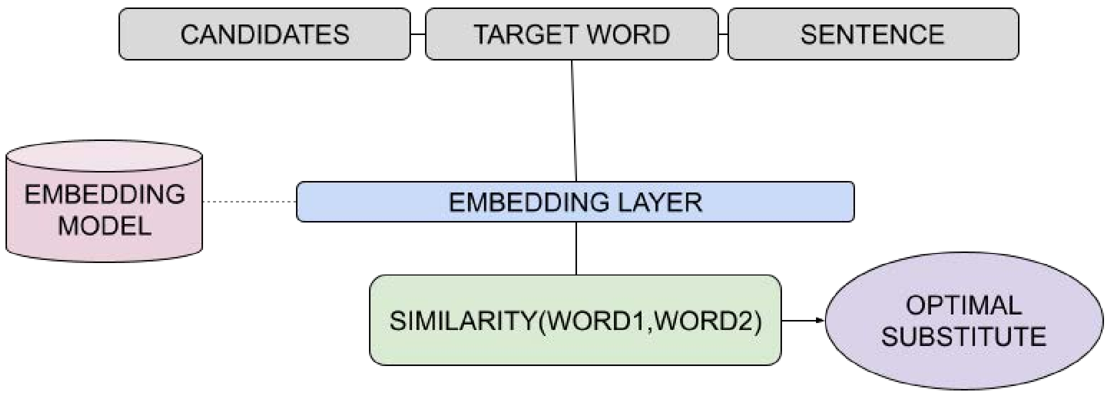

Figure 3.5: Substitute selection procedure

Additionally, this Thesis exploits the functionality of CWI models to detect complex words:

4. **CWI Model filter :** Before performing the selection, the candidate list is filtered, excluding the complex words predicted by the CWI model observed in the CWI Section (3.2.1). Then, the same process described in (3) is performed.

3.2.4. Substitute Ranking (SR)

Taking into account word simplicity metrics, this Thesis proposes a weighting system that uses frequency features and information provided by embedding models.

The SR step takes the list of synonyms extracted from the previous step and chooses which candidate that fits the context is the simplest, taking into account the target user.

3.2.4.1 Rankers

At this step, a combination of frequency-based and machine learning-assisted strategies has been implemented by developing a weighting module that uses the following features to rank a word:

- **BERT prediction:** Probability distribution of the candidate. This can be obtained from the vocabulary corresponding to the masked word. The higher the probability, the more relevant the candidate for the original sentence.
- **Semantic similarity:** Cosine distance between the original word vectors and the candidate vectors in the list. The shorter the distance, the more similar the two words. To extract these vectors, different embedding models are used.
- **Frequency Feature:** Because frequency-based approaches have shown good results at this stage, the decision was made to incorporate it as a feature in the ranker. A dictionary of the Real Academia de la Lengua Española (RAE) is used to extract the frequency of each candidate, which is made up of 10000 terms ordered by their frequency. The more frequent a word is, the simpler it is supposed to be.

The ranker makes the decision to choose the simplest candidate based on the candidates that obtained the best results in each of the features.

3.3. Word Sense Disambiguation method (WSD)

As described in Chapter 2, ambiguity in the Spanish language is present in many terms. Therefore, a method for determining word definitions had to be developed. In this Thesis, a method that uses an embedding model to disambiguate a word by taking information from the context is proposed.

The objective of this step is to select the correct definition for a specific word. Taking advantage of the versatility offered by the BERT models, this Thesis presents a procedure that takes a Spanish BERT model as its core.

The definitions are extracted from the following two dictionaries: the "Real Academia de la Lengua Española" Dictionary (RAE) [18] and the "Diccionario Facil" [19], with the latter being a dictionary of Easy Reading definitions created by the "Plena Inclusión Madrid" association's experts and users with cognitive disabilities.

The procedure creates a list of definitions for the target word extracted from the RAE and the "Diccionario Facil". With the help of the model in the system, the word is masked in the sentence to which it belongs, and then the model predicts which words can be substituted for the masked word. This results in a list of words that share a common meaning, thus disambiguating the target word. With the help of Spacy [20], these words are lemmatized to enrich the list. The words in the sentence with lexical content are then extracted and added to the list.

[18] www.rae.es
[19] www.diccionariofacil.org
[20] www.spacy.io/

Since the first list created by this Thesis system contains words with similar semantics, these two lists are compared, and the coincidences are counted. The Hypothesis followed is that the definition provided by the second list, which has more coincidences of words than the first list, is the correct definition associated with the target word and, consequently, is chosen by the WSD system. If no coincidences are found, the system selects the first definition on the list.

3.4. Conclusions

This Chapter has described the architecture, methods, resources and tools pertaining to the Thesis proposal. Regarding the accessibility resources, this Chapter described the construction process of the EASIER corpus and the E2R dictionary, which aim to provide support to LS stages. Likewise, in this Thesis, different methods that offer solutions to these LS stages are described, combining approaches from the NLP discipline and resources focused on accessibility. Finally, regarding the accessibility requirement concerning the need for a mechanism to provide correct definitions to unusual words, a method that relies on the WSD task to provide definitions is proposed in this Thesis.

In the following Chapter, different experiments are presented in order to evaluate the methods and resources proposed above.

Chapter 4

Evaluation. Experiments and results

In this Chapter, in order to evaluate the proposal, different scenarios for each of the steps to apply lexical simplification are studied. This Chapter is divided into subsections as in Chapter 3 and this subsection contains a description of the experiment, the dataset used, the methodology employed and results obtained. Also, a study with target users is presented. Finally, error analysis and discussion are presented to complement the results.

4.1. Accessibility Resources

This accessibility resources Section describes the experiments performed on the accessibility resources resulting from this Thesis, with the objective of ensuring quality and usefulness for the LS stages.

4.1.1. EASIER Corpus

In this Section, a quality evaluation of the EASIER corpus is described. This evaluation follows the recommendations of Pustejovsky & Stubbs book [153] on resource annotation for the NLP discipline.

As told before, this corpus is composed of two datasets: one for the task of CWI and one for substitute related tasks, such as SG and SS. To evaluate these datasets, two additional experts were added to assist the corpus evaluation.

4.1.1.1 Methodology

In this Section, the methodology and validation metrics necessary to ensure the quality of the EASIER corpus data are described.

- **Datasets Evaluation :** The evaluation methodology for both resulting data sets is different due to the different structure between them, which are described below:

1. **CWI Dataset Inter-Annotator Agreement:** It is important to evaluate the annotation task, so it is very common to perform AAI scores. IAA scores provide a way to evaluate the accuracy of the annotation task that can be performed by two or more annotators. Commonly used metrics for these types of evaluations are as follows:
   - **Cohen's Kappa:** This metric evaluates the agreement between two annotators, taking into account the possibility of chance agreements.
   - **Fleiss Kappa:** This metric is similar to the previous one, with the difference that this metric is designed for scenarios where there are more than two annotators in the evaluation.

   To obtain the portion to be evaluated, 10% of the corpus was randomly extracted. As a result, 26 documents were obtained, from which 390 sentences to evaluate were obtained.
2. **SG/SS Dataset Evaluation:** In contrast, for the evaluation of the substitute dataset, the decision was made to use a different methodology, because the idea of an annotator having to propose new substitutes for a target word was very costly. Therefore, to evaluate this dataset and in order to verify the quality of the proposed synonyms, the two additional annotators were asked to assign two types of labels for each synonym: "0: poorly defined synonym" and "1: well-defined synonym".

   To obtain the portion to be evaluated, 10% of the total number of instances in the dataset were randomly extracted, on the condition that they had to have at least three proposed synonyms. As a result, the evaluation portion consisted of 513 target words with their respective proposed substitutes.

4.1.1.2 Results

In this Section, the results of the methods described above are presented.

- **Initial evaluation :** An initial analysis of the corpus (approximately a quarter of the total) was carried out with people with cognitive disabilities belonging to the target group to evaluate and refine the expert linguist's annotation guidelines.

  1. **Validators :** Following the methodology in the process of adapting texts in an easy reading, validation sessions should be carried out in which people with disabilities are the validators who ensure that the adaptation is being made correctly. Eight people with mild intellectual disabilities (group 1) and elderly people (group 2), with five women and three men were chosen to participate in the initial evaluation. Of the five women, three were people with intellectual disabilities and two were elderly. In the group of men, two were people with intellectual disabilities, and one was an older adult. The validators' age ranged from 25 to 86, seven with primary education and one with secondary schooling.

Target Word	Synonyms	Conclusion
Etiquetado (Labelling)	Letrero (sign) , inscripción (inscription), rótulo (banner)	Explanation required for both groups
Etiqueta (formal/label)	Ceremonia (ceremony), protocolo (protocol)	Explanation required for group 1 - Known for group 2
Envasados (packaged)	Empaquetados (packaging)	Known by both groups
A granel (in bulk)	Suelto (loose), sin envase (without packaging)	Known for both groups
On-line (Online)	en línea (online), conectado a Internet (connected to the Internet)	Known by group 1 - Explanation required for group 2
Comensales (diners)	Invitados (guests)	Unknown by group 1 - Known by group 1
Saludables (salubrious)	Sanos (healthy), beneficiosos (beneficial)	Explanation required for both groups
Copiosa (copious)	Abundante (abundant)	Unknown by both groups
Crudos (raw)	sin cocinar (not cooked)	Known by both groups
Denominación (denomination)	Nombre (name)	Explanation required for both groups
Reclamar (claim)	Demandar (sue), quejarse (complain), exigir (demand)	Explanation required for both groups
Irregularidades (irregularities)	Anomalía (anomaly), alteración (alteration), variación (variation)	Unknown by both groups
Óptimas (optimum)	Buenas (good), excelentes (excellent)	Explanation required by both groups
Embalaje (packaging)	Envase (container), envoltorio (wrapping)	Known for both groups
Íntegro (exhaustive)	Entero (whole), completo (complete), intacto (intact)	Known for both groups
Consumidor (consumer)	Comprador (buyer/purchaser), cliente (client), usuario (user)	Explanation required for group 1 - Known group 2
Provisional (provisional)	Temporales (temporary)	Unknown for both groups
Consejo (Council)	Asambleas (assembly), juntas (board), comisiones (commission/committee)	Known for both groups
Proporcionar (provide)	Dar (give), proporcionar (provide)	Known for both groups
Ciudadanía (citizens)	Sociedad (society), población (populace), nacionalidad (nationality)	Known for both groups
Veraz (veracious)	Real (real), cierta (certain), verdadera (true)	Unknown by both groups
Eficacia (efficiency)	Utilidad (usefulness), efectividad (effectiveness)	Unknown by both groups
Contrastar (contrast)	Comprobada (proven), comparada (compared), verificada (verified)	Unknown for both groups
Soporte (base)	Base (basis), fundamento (foundation), apoyo (support)	Unknown for both groups
Evidencias (evidence)	Certeza (certainty), demostración (demonstration), seguridad (security), prueba (proof)	Known for both groups

Table 4.1: An extract of the target/synonym dataset for human evaluation with group 1 (people with mild intellectual disabilities) and group 2 (older people).

2. **Procedure:** The validation session lasted three hours, including a twenty-minute break, and was moderated by a psychologist and the expert in easy reading who was annotating the EASIER corpus. The validators were provided with documents containing twenty-five adverse words. These documents belong to the current affairs Section (see Table 4.1), all framed within sentences and the corresponding synonyms. The moderator projected the document on a screen, then read each sentence aloud and asked the group whether they knew the adverse word or not and its meaning. This was an important step that allowed for assessing the participants' comprehension capacity and clarifying the concepts if there were doubts. Each validator gave his or her opinion and was free to make comments as they saw fit. The moderator then read the synonyms and reread each sentence aloud, substituting each synonym's adverse word. Finally, the validators commented on the meaning of each synonym, determined the most appropriate option and, if there were several synonyms, ordered them according to their comprehension criteria, which is as follows:
   - Known: The validator understands the meaning of the word.
   - Explanation required: The validator has an idea of the meaning of the word due to its context but needs an explanation.
   - Unknown: The validator does not know/understand the word.
3. **Initial evaluation results :** Table 4.1 shows a portion of the dataset used for evaluation. The human evaluation showed that most of the words represented a challenge for the participants to comprehend (84%), either because they were unfamiliar with said words or needed additional explanation by the moderators. This demonstrates moderate results regarding the quality of the corpus in the decision making of word complexity criteria. For the synonyms proposal, the validators responded well, showing a better understanding of the text with the proposed synonyms. However, users gave a different priority to the suggested synonyms. For example, they understood the word "alteraciones" (alterations) better than the word "irregularidades" (irregularities). Also, users experienced increased difficulty in understanding when more than three synonyms were proposed. Thanks to the validation session, the need for several resources or elements to assist in understanding the meaning of a complex word was confirmed. In some cases, it was found that merely showing possible substitutions for a word was not enough for participants to fully understand it, as the user required additional information about the word, such as a definition or an example. This requirement reaffirms the objectives of the EASIER project in which the Thesis proposal has been applied (see Section 1.3). This project, which in addition to satisfying the processes of LS (CWI, SG/SS), also offers additional comprehension aids such as providing disambiguated definitions and pictograms.

- **CWI Dataset Inter-Annotator Agreement:** Table 4.2 presents the evaluations for the CWI dataset, sorted by POS tags. It can be said that a moderate result was obtained with a Fleiss Kappa coefficient of 0.641. The highest agreement was reached when analysing the multiwords since words or phrases of great length imply difficulty comprehension.

POSTAG	**Cohen's Kappa (Rater 1-2)**	**Cohen's Kappa (Rater 1-3)**	**Cohen's Kappa (Rater 2-3)**	**Fleiss Kappa**
N	0.4750	0.4114	0.5711	0.484
V	0.4082	0.5218	0.4385	0.454
A	0.2011	0.1942	0.4640	0.31
I	0.5002	0.1545	0.2658	0.3
PN	0.2263	0.2441	0.5338	0.347
N-V	0.4667	0.4365	0.5586	0.487
N-V-A	0.4628	0.4374	0.5602	0.487
N-V-I	0.4689	0.4342	0.5559	0.486
N-V-I-PN	0.4330	0.4228	0.5530	0.471
N-V-M	0.6455	0.6079	0.6728	0.641
N-V-A-M	0.6422	0.6094	0.6739	0.641
N-V-I-M	0.6465	0.6060	0.6707	0.64
N-V- I- PN-M	0.6067	0.5926	0.6597	0.619

Table 4.2: EASIER corpus - CWI dataset results where N: nouns, V: verbs, A: adverbs, I: Interjections, PN: proper nouns, M: multiwords

- **SG/SS Dataset results:** In addition, Figure 4.1 shows the results for the substitute dataset. As can be seen, the results can be defined as positive, due to the difference present between well-defined and incorrectly defined substitutes. The annotators reviewed 1026 different substitutes, of which annotator 2 rated 987 as well-defined and 37 as incorrectly defined. Annotator 3 rated 913 synonyms as well-defined and 113 as incorrectly defined.

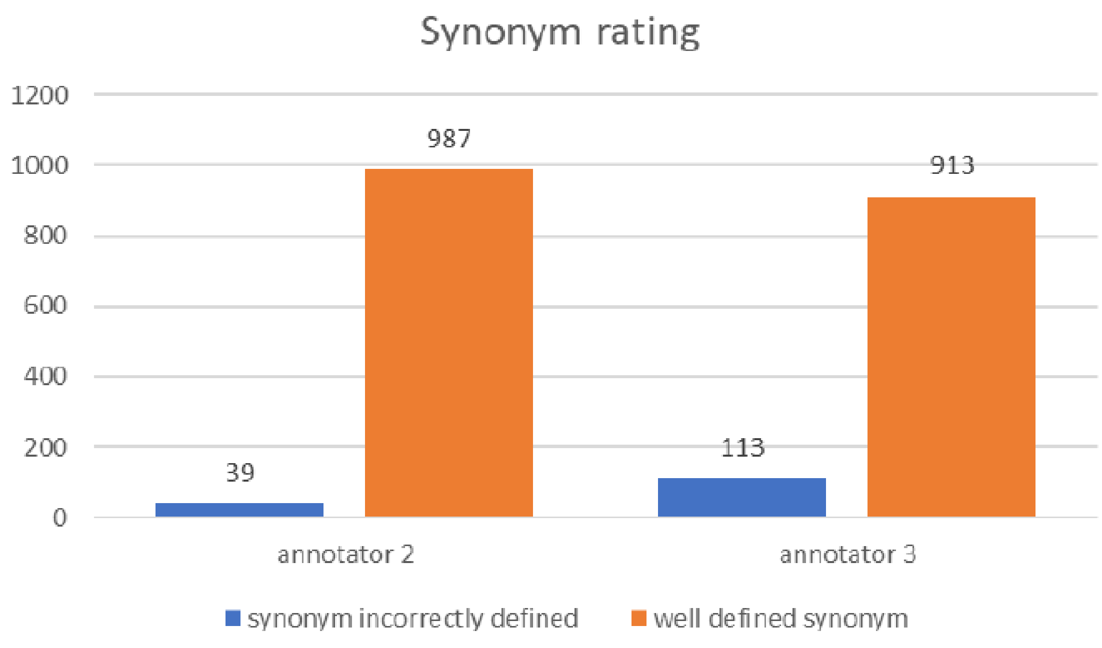

Figure 4.1: EASIER corpus - Substitute dataset results

In addition, as a contribution of this Thesis, these resources and the evaluations described above can be found publicly in web repositories [21] [22].

4.1.1.3 Discussion

Given the rapid growth of research in different areas of NLP, the creation of resources to support different processes is always helpful. In this Section, the processes of annotation and evaluation of the EASIER corpus are described, which has content that supports the task of lexical simplification in the CWI and SG/SS subtasks.

In the CWI dataset, the IAA showed a Fleiss Kappa score of 0.641, considering it a moderate score. When this score was obtained, the decision was also made to calculate the Cohens Kappa score, in order to see if there was a greater agreement between the two annotators. When analyzing the agreements, it was found that between scorer 1 and scorer 2, a maximum score of 0.6465 was obtained, between scorer 1 and 3, a maximum score of 0.6094 and between scorer 2 and 3, a maximum score of 0.6739. When looking at these scores it can be seen that the results are similar between annotators and it can be concluded that the task of CWI is not a trivial task, so a high inter-annotator agreement is a complicated task.

On the other hand, in the substitute dataset, a large percentage of substitutes were evaluated as well-defined, specifically scorers 2 and 3 scored over 96% and 88% of the

[21] http://dx.doi.org/10.17632/ywhmbnzvmx.2
[22] https://github.com/LURMORENO/EASIER_CORPUS

total instances as well-defined respectively. Subsequently, an analysis was carried out of the instances in which the substitutes were rated as incorrectly defined. It was found that in several cases, these words were qualified in this way due to the fact that, although they could fit in the context, they presented some ambiguity with regards to their meaning. An example of this is the word "salubrity" in the sentence “Tiempos en los que la salubridad era escasa.” (Times when salubrity was scarce). The well-defined replacements were “limpieza” (cleanliness) and “hygiene” (hygiene). However, the incorrectly defined replacement was “salud” (health), which may work within the context of the sentence, but which modifies its semantics.

Finally, to conclude and synthesize what has been described in this Section, the Thesis describes the EASIER corpus that possesses complex target words, together with proposed contextualized synonyms, answering research Hypothesis 1 and implicitly supporting all the hypotheses related to lexical simplification, since thanks to this resource it was possible to validate the methods proposed in this Thesis.

4.1.2. EASIER Dictionary

As mentioned above, this resource takes advantage of easy-to-read content to create a dictionary with simple vocabulary. The following Sections will show the practical effects of this novel resource. However, in this Section, a comparison with other resources that aim to support lexical simplification is presented. Table 4.3 shows these comparisons, where it can be seen that the content of the EASIER dictionary has a 38% of matches in the totality of the uniwords of the BEA resource and a 44% of matches with the uniwords annotated as simple from the same resource. While, compared to the EASIER corpus, it presents a 64% match in the totality of uniwords and a 70% match with the uniwords annotated as simple from the same resource.

Resource	% Simple	% Total
BEA	44	38
EASIER	70	64

Table 4.3: Percentage of matches between datasets and the EASIER dictionary.

4.2. Lexical Simplifier

Continuing the structure of the previous Chapter, this Section will present the results obtained throughout this doctoral Thesis, divided by each stage of lexical simplification.

4.2.1. Complex Word Identification (CWI)

The first step in lexical simplification is the complex words identification. In this Section different classifiers and features presented in the previous Chapter are evaluated.

4.2.1.1 Methodology

As a first step, a method to represent a word must be defined, in order to distinguish between a simple and a complex word. As described in the previous Chapter, as first approaches, this Thesis implements supervised machine learning strategies [82] [65], which were trained and evaluated with the datasets from the BEA workshop.

In subsequent evaluations, transfer learning strategies were evaluated by tuning a BERT model to perform the current task [124]. To perform this tuning task, a sample corpus is necessary. Therefore, this Thesis uses the datasets from the BEA workshop and those from the EASIER corpus.

About the typical metrics to be used when working with these methods, in this Thesis the following are taken into account:

- **Precision:** the proportion of correct positive predictions.
- **Recall:** the proportion of actual positives correctly identified.
- **F1-Score:** the harmonic average between precision and recall.

4.2.1.2 Results

As a first classifier, a support vector machine with an RBF kernel was chosen. Table 4.4 shows the results of different combinations of features using this classifier. As can be seen, the best results were obtained by combining length, Boolean, frequency, easy-to-read content and embedding features, reaching a maximum score of 0.7497. At the same time, it can be seen that the embedding features alone obtained a high score, with a score of 0.7283.

Feature	**Precision**	**Recall**	**F1**
L+F	0.6819	0.6834	0.6614
B+L+F	0.7887	0.7071	0.7122
E+B+L+F	0.8015	0.7143	0.7314
E+B+L+F+W	0.8544	0.7141	0.7341
E+B+L+F+W+F*	**0.8636***	**0.7257***	**0.7497***
W+F	0.7920	0.7214	0.7283
E+W+F	0.8250	0.6911	0.6982
E+L+F	0.8095	0.7018	0.7205
B+E	0.7205	0.7057	0.7097
B+E+W+F	0.7299	0.7599	0.7299

Table 4.4: System CWI scores for feature combinations where L:Length, F:Frequency, B:Boolean, E:E2R, W:Word2Vec, F:FastText

As a further evaluation, a change in the RBF kernel to linear of the classifier was done, because this kernel has been shown to be faster and perform better with binary tasks of this type. Along with this change, a wider variety of features were evaluated for the representation of a word. Table 4.5 shows evaluations of this classifier with different features, in which it can be seen that an even higher score of 0.794 was achieved with the combination of specific length and boolean features, along with embedding and easy-to-read content.

Feature	**Precision**	**Recall**	**F1 Score**
WL	0.74	0.70	0.702
SN	0.719	0.695	0.700
SL	0.296	0.5	0.372
P	0.709	0.689	0.693
E	0.296	0.5	0.372
F	0.659	0.589	0.569
W2V	0.71	0.70	0.700
S2V	0.797	0.777	0.783
BT (400 dimensions)	0.730	0.717	0.720
BT (450 dimensions)	0.735	0.720	0.725
BT (480 dimensions)	0.74	0.72	0.727
WL+SN	0.749	0.693	0.698
SN+SL	0.744	0.695	0.700
WL+SN+SL	0.748	0.697	0.702
WL+SN+SL+P	0.768	0.710	0.716
WL+SN+SL+B+E+W2V+F	0.789	0.771	0.776
W2V+BT	0.76	0.75	0.752
WL+BT	0.78	0.77	0.778
WL+B+BT	0.79	0.78	0.783
WL+B+E+BT	0.80	0.78	0.787
WL+B+E+W2V+BT*	**0.80***	**0.79***	**0.794***
WL+B+E+W2V+S2V+BT	0.797	0.788	0.791
WL+B+E+S2V+BT	0.794	0.779	0.784

Table 4.5: CWI results of feature combinations, where WL: Word Length, SN: Syllable number SL: Sentence Length, P: Probability, B: Boolean, E: E2R, F: Fasttext, W2V: Word2vec, S2V: Sense2Vec, and BT: BERT

Table 4.6 shows the results for the CWI task and also shows the results of the models trained with the combination of the data of the Spanish datasets. Also, the results are compared with a traditional machine learning approach, described above.

	Precision	Recall	F1
mBERT_EASIER_EASIER-test	0.695	0.694	0.694
BETO_EASIER_EASIER-test	0.696	0.691	0.693
mBERT_BEA_BEA-test	0.669	0.628	0.643
BETO_BEA_BEA-test	0.653	0.640	0.640
mBERT_EASIER-BEA_BEA-test	0.676	0.675	0.674
BETO_EASIER-BEA_BEA-test	0.639	0.603	0.598
mBERT_EASIER-BEA_EASIER-test	0.685	0.687	0.685
BETO_EASIER-BEA_EASIER-test	0.695	0.677	0.685
SVM approach_BEA_BEA-test	**0.80**	**0.79**	**0.794**

Table 4.6: Results for BERT models on CWI task where the structure is Model_TrainDataset_TestDataset

In addition, as a contribution of this Thesis, the implementation of the methods and their evaluations described above can be found publicly in web repositories [23] [24] [25].

4.2.1.3 Discussion

In this Section, the evaluations of the first stage of LS (CWI), are described.

In the first scopes, different features were used to represent a word, supported by a support vector machine with RBF kernel. The highest F1-score of 0.7497 was achieved by combining different types of features. By performing ablation studies shown in Table 4.4 and 4.7, hints as to which feature is more relevant in the classification process are obtained. As expected, length and frequency features are a fundamental choice in classification, as they alone obtained an F1-score of 0.6614. Boolean features helped the classifier to obtain morphological information of the target word increasing the previous score by 0.0508 points reaching an F1-score of 0.7122. As the next feature and being the feature created in the framework of this Thesis, the E2R feature was introduced, by increasing the F1-score to 0.7314, demonstrating that this resource has the potential to be useful for different tasks due to the fact that it is a resource containing simple words annotated by experts. Similarly, the vectors given by the embedding models proved beneficial in the classification, achieving a 0.7341 with Word2vec and a final F1-score of 0.7497 with Fasttext. These results gave us an indication that the use of embedding of different types could give a benefit when classifying words.

[23] https://github.com/ralarcong/EASIER_lexical_simplification
[24] https://github.com/LURMORENO/easier
[25] https://github.com/ralarcong/EASIER_EVALUATIONS

Independent Feature	F1
Lenght-Frequency	0.6614
Boolean	0.4485
E2R	0.4011
Word2vec	0.7012
FastText	0.4901

Table 4.7: F-1 scores for every feature alone

Upon further investigation at this stage, a kernel change in the classifier from RBF to linear because the linear one is much faster [167] and has the additional advantage that SVM has shown good performance in classifying sparse instances [168].

In addition, new types of features were included in the evaluations, while others were discarded. As shown in Table 4.5, a maximum F-1 score of 0.794 was achieved, which greatly outperforms the previous classifier. Furthermore, when performing the ablation studies, the importance of embedding models in the classification was noted, since with the exception of the Fasttext model, F1-scores greater than 0.7 points were obtained. Using Word2Vec and the BERT models, an F1-score of 0.752 is obtained. Furthermore, when evaluating the F1-scores independently for each feature, the Word2Vec feature yields a score of 0.70, proving to be a valuable resource for this task. Also by evaluating individually the length and Boolean features, it allowed us to recognize which of the features belonging to these categories are better at classification, such is the case of the word length feature that independently offers a higher score than combined with other features of the same category. Also, it is worth mentioning that the E2R feature is still relevant in this new configuration as it offers a benefit in the final score.

The reasons why a feature was discarded are varied, for example, reasons related to the vocabulary belonging to the feature, such is the case of the probability features, which in many cases presented null values because some words were not found in its dictionary, resulting in an F1-score of 0.69 and having no synergy with other features. Another case of discarding was given by the embeddings models, since at the beginning it was believed that the combination of several types of embeddings models would result in a higher score, however, this was not the case. This Thesis believes that these negative results occur because the models were created with different resources in the case of Sense2vec[26], Fasttext[27] or BERT[28] and consequently, each model presented different vocabularies and vectors, confusing the classifier.

Also, in recent evaluations, transfer learning methods were tested. Fair results were obtained by tuning a BERT model with the BEA and EASIER corpora, obtaining a maximum F-1 score of 0.694 when tuning and evaluating the BERT model with the EASIER corpus datasets. While when evaluated with the BEA dataset, the highest value was reached by the model trained with EASIER and BEA data with an F1-score of 0.674. It

[26]https://github.com/explosion/sense2vec
[27]https://github.com/facebookresearch/fastText
[28]https://github.com/shehzaadzd/pytorch-pretrained-BERT

is worth mentioning that as an initial scope, in the training stage the multiwords were excluded, taking into account only the uniwords of each dataset, consequently, it is believed that this is one of the causes of a low F1-score.

As complementary information and for comparative purposes, Table 4.8 shows the results of the systems that were submitted to the BEA workshop task. As can be seen, the approach using a linear support vector machine, ranks above all the systems with an F1-score of 0.794. One of the closest systems in the score is the TMU system [98], which presented an approach that uses the frequency of the word in different corpora to then train it with a random forest classifier, on the other hand the NLP-CIC system [169], used deep learning methods combined with word/character embeddings, word length and frequencies.

SPANISH	F-1
Thesis approach (Linear)	**0.794**
TMU	0.7699
NLP-CIC	0.7672
ITEC	0.7637
Thesis approach (RBF)	**0.7497**
NLP-CIC	0.7468
-	-
-	-
-	-
CoastalCPH	0.6918
Gillin Inc	0.6804
Gillin Inc.	0.6784
Thesis approach (BERT)	0.674
Gillin Inc.	0.6722

Table 4.8: F-1 scores for the CWI task on BEA Workshop 2018

Finally, it is important to mention that the advantage of combining accessibility resources with different types of word embedding helps to answer hypotheses 1 and 2. Additionally, applying a transfer learning method to refine an embedding model to perform the CWI task supports Hypothesis 3.

4.2.2. Substitute Generation (SG)

As the next stage in lexical simplification, in this Section, the methods and evaluations for the SG stage are described.

4.2.2.1 Methodology

As described in the previous Chapter, in this Thesis, as a first scope, strategies based on linguistic databases and automatic generation were evaluated [65], which are briefly mention below:

(1) **Thesaurus database:** synonym search for the target word.

(2) **Thesaurus database:** search for synonyms for the target word and its lemma.

(3) **Babelnet database:** search for synonyms for the target word.

(4) **Babelnet database:** search for synonyms for target word and its lemma.

(5) **PPDB:** search for replacements for the target word.

(6) **PPDB:** search for replacements for target word and its lemma.

(7) **Babelnet + Thesaurus:** concatenate the extracted values from (2) and (4).

(8) **Babelnet + Thesaurus + PPDB:** concatenate the extracted values from (2), (4) and (6).

(9) **Babelnet + Thesaurus:** in addition to the procedure described in (7), the target word's lemma and stem are extracted. Subsequently, the candidate words that contain the stem or match the extracted lemma are deleted.

(10) **Babelnet + Thesaurus + PPDB:** in addition to the procedure described in (8), the target word's lemma and stem are extracted. Subsequently, the candidate words that contain the stem or match the extracted lemma are deleted.

Recent evaluations show an exploration of embedding models to perform this task [124]. The performance of different embedding models was tested by extracting the nearest neighbors of each target word (50 neighbors). The tested models are described in the CWI Section (3.2.1).

To evaluate these methods, a portion of the substitute dataset from the EASIER corpus is extracted, represented by 575 instances [29], of which the first 500 are used. Each instance contains a sentence, a target word and three substitutes proposed by a linguistic expert. On the other hand, the evaluation metrics used for this stage are those proposed by Paetzold [8]:

- **Potential:** the proportion of instances for which at least one of the candidates generated is contained within the gold standard.
- **Precision:** the proportion of generated substitutions that are contained within the gold standard.

[29] http://dx.doi.org/10.17632/ywhmbnzvmx.2

- **Recall:** the proportion of gold-standard substitutions that are among the generated substitutions.

- **F-Score:** the harmonic average between precision and recall.

4.2.2.2 Results

Table 4.9 shows results for the initial methods described. As can be seen, the method of concatenating the outputs of different resources for the Spanish language obtained the best potential with 0.898 and recall with 0.597, but lower results in precision and consequently F1-score. However, it is important to emphasize that recall is important for this stage, since the greatest number of substitutes is required in all the contexts that may appear.

	Potential	**Precision**	**Recall**	**F1**
(1)	0.288	0.047	0.170	0.074
(2)	0.5	**0.070**	0.248	**0.109**
(3)	0.312	0.042	0.156	0.066
(4)	0.760	0.051	0.426	0.091
(5)	0.796	0.048	0.480	0.087
(6)	0.808	0.050	0.485	0.090
(7)	0.644	0.059	0.335	0.099
(8)	**0.898**	0.043	**0.597**	0.080
(9)	0.890	0.060	0.564	**0.109**
(10)	0.896	0.054	0.589	0.098

Table 4.9: Substitute generation results

When the embedding models were evaluated, the results described in Table 4.10 were obtained. As can be seen, the results were not higher than previous approaches, with the Sense2vec model obtaining a maximum recall of 0.298.

	Potential	Precision	Recall	F-1
Word2vec	0.358	**0.191**	0.188	0.034
FastText	0.464	0.0294	0.289	0.053
Sense2Vec	**0.506**	0.056	**0.298**	0.095
BERT	0.348	0.030	0.282	0.054

Table 4.10: Results for SG - Embedding models

In addition, as a contribution of this Thesis, the implementation of the methods and their evaluations described above can be found publicly in web repositories [30] [31] [32].

[30]https://github.com/ralarcong/EASIER_lexical_simplification
[31]https://github.com/LURMORENO/easier
[32]https://github.com/ralarcong/EASIER_EVALUATIONS

4.2.2.3 Discussion

In this Section, the evaluations of the second stage of lexical simplification (SG) are described.

In initial approaches, the use of resources generated by linguistic database generation strategies and automatic generation was evaluated, where resources such as Thesaurus, Babelnet and PPDB (Paraphrase Database) were used. The results in Table 4.9 showed that by combining these resources a high potential of 0.898 and a recall of 0.597 were obtained, however, due to the high number of false positives, a low precision (0.043) was obtained. To deal with this, an analysis of the generators outputs was performed, where cases where the generators proposed as a substitute the same original target word but in a different grammatical form, were found. Although the recall is more important than precision at this stage, the decision was made to incorporate cleaning techniques in the generators (described in (9) and (10)), which showed an increase in precision and a small decrease in potential and recall.

In subsequent work, the versatility of different embeddings was evaluated for this stage. The models described in section 3.2.1 were used to extract the 50 nearest neighbors of each target word. Table 4.10 shows the results of this evaluation, where the Sense2vec model obtained the best potential and recall of 0.506 and 0.298, respectively. Unfortunately, these results did not represent an improvement to previous results. When analyzing negative results, several cases where the models proposed the target word in different grammatical forms as substitutes, were found. In addition, because these models provide semantic similarity between words, cases where the lists of substitutes contained antonyms of the target word were found.

Finally, it is important to mention that the exploration of different embedding resources for candidate extraction supports the answer to Hypothesis 2.

4.2.3. Substitute Selection (SS)

Moving on to the next stage, SS takes the substitutes generated in the previous stage and prioritizes them taking into account the context of the target word.

4.2.3.1 Methodology

As described in the previous Chapter, embedding models are used as a fundamental resource in the proposed methods, which are describe briefly below:

1. **No selections :** selects all candidates.
2. **Any Window :** obtains three similarity values (candidate and target word, candidate and target word's context words in the sentence (previous and subsequent words)). Next, these values are added and stored. Finally, this process is repeated

for every candidate, and the selector picks the three candidates with the highest values.

3. **Lexical window :** Similar to (2), but instead of selecting the first context word, the first word with lexical content (previous and subsequent words) are selected.

4. **CWI Model filter :** Before performing the selection, the candidate list is filtered, excluding the complex words predicted by the CWI model observed in the CWI Section. Then, the same process described in (3) is performed.

In early evaluations [65], only one type of embedding model was used in the evaluations, which is Word2vec. However, in recent evaluations [124], different types of embedding are incorporated in the evaluations, such as Fasttext, Sense2vec and BERT. Regarding evaluation metrics and resources, the dataset and metrics from the substitute generation stage are used again.

4.2.3.2 Results

Unlike the previous stage, in this stage a higher precision is searched, since the list of substitutes is shortened in order to have a new list with words adjusted to the context. Table 4.11 shows the first evaluations of this stage, by taking the best generator from the previous stage (named 8) and evaluating the selectors from its output. These initial evaluations showed an increase in precision values, however, there was still a lot of room for improvement.

	Potential	**Precision**	**Recall**	**F1**
(1)	0.896	0.054	**0.589**	0.098
(2)	0.022	0.008	0.006	0.007
(3)	0.406	**0.172**	0.121	**0.142**
(4)	0.006	0.003	0.001	0.002

Table 4.11: Substitute selection results – generator (8)

In order to improve the results of the selectors, the best selector from Table 4.11 was taken and evaluated with the different generators, along with filtering strategies. Table 4.12 shows these results, where an increase in precision can be seen with generator 8 complemented with the lexical window selector.

	Potential	**Precision**	**Recall**	**F1**
(1)	0.234	0.111	0.090	0.10
(2)	0.368	0.174	0.122	0.144
(3)	0.226	0.167	0.086	0.095
(4)	0.376	0.164	0.113	0.134
(5)	0.382	0.165	0.114	0.135
(6)	0.392	0.168	0.116	0.137
(7)	0.36	0.157	0.114	0.132
(8)	0.406	0.172	0.121	0.142
(9)	0.504	**0.226**	0.154	**0.183**
(10)	0.502	0.222	0.153	0.181

Table 4.12: Substitute selection results – different generators

Once these evaluations were seen, it was concluded that the "Lexical window" strategy was the best among the proposed strategies. Therefore, in recent evaluations, an experimentation with this strategy is performed along with different embedding models. Table 4.13 shows the results of the above, where it was found that the Fasttext model performed better than the previously used Word2vec model with a final precision of 0.338.

	Potential	Precision	Recall	F-1
Word2vec-Easier	0.692	0.304	0.304	0.304
FastText	0.736	**0.338**	0.338	**0.338**
Sense2Vec	0.69	0.308	0.308	0.308
BERT	0.266	0.125	0.125	0.125

Table 4.13: Results for SS

In addition, as a contribution of this Thesis, the implementation of the methods and their evaluations described above can be found publicly in web repositories [33] [34] [35].

4.2.3.3 Discussion

In this Section, the evaluations of the third stage of lexical simplification (SS) are described.

At this stage, a semantic similarity strategy was evaluated, due to the great versatility of embedding models. In initial evaluations, a Word2Vec model was taken as a base model, and different methods were evaluated based on it. These methods took the context of the target word and were then semantically evaluated with the possible substitutes. In addition, the usefulness of a model intended for the CWI task in performing this task was

[33] https://github.com/ralarcong/EASIER_lexical_simplification
[34] https://github.com/LURMORENO/easier
[35] https://github.com/ralarcong/EASIER_EVALUATIONS

tested. When evaluating these methods (Table 4.11) with the information from the best generator of the previous stage, it was concluded that the best method is "lexical window", as it obtained the highest potential and precision among all the methods with 0.406 and 0.172 points respectively. Later, in order to obtain higher scores, the decision to evaluate this last method with the other generators was made. When analyzing the results, it could be seen that by performing cleaning strategies prior to the selector, a better precision was obtained. As shown in Table 4.12, when evaluating the "lexical window" method with the results of the (9) generator, better results were obtained, with a potential and precision of 0.506 and 0.226 respectively.

In recent works, the embedding side was explored for this stage by evaluating different types and determining which was the best. Table 4.13 shows the results of these evaluations, where the FastText model proved to be better than others by obtaining a potential and recall of 0.736 and 0.338 respectively. These results represented an improvement over previous work presented with the Word2vec model, which obtained a potential and recall of 0.692 and 0.304 respectively. This Thesis assumes that this higher score was obtained because the FastText model provides char and ngrams embeddings to face the problem of OOV (Out-of-vocabulary) words.

Finally, it is important to mention that the exploration of different embedding resources for the measurement of candidate similarity in the context of the target word supports the answer to Hypothesis 2.

4.2.4. Substitute Ranking (SR)

Finally, when evaluating the last stage of SR, the strategy that can select the easiest candidate for a specific audience is aimed.

4.2.4.1 Methodology

1. **English language evaluation:** To this Thesis knowledge, there were no public datasets in Spanish to evaluate this procedure, therefore, the decision to adapt these procedures to evaluate it with English language datasets was made [124], specifically, datasets from the English Lexical Simplification task of SemEval 2012 [170]. The trial set is composed of 300 instances, and the test set, 1, 710 instances. Each instance contains a sentence, a target complex word, and candidates ranked by their simplicity. This resource is divided into several files:

    - File where sentences (contexts) are displayed, along with a complex word highlighted by instance.
    - File showing possible replacements to the complex word randomly ordered by instance.
    - File showing the same possible replacements sorted by simplicity.

Table 4.14 shows an example of the content found in this resource, showing the sentence, complex word and some replacements ordered by simplicity. Note that in this resource there are complexity ties where replacements are considered to have a similar level of simplicity (e.g., first instance - fourth annotation).

The evaluation metric is the TRank measure, proposed in the shared task. This metric calculates the proportion of instances for which the highest ranked candidate produced by a ranker is the same as the one in the gold-standard.

Sentence	**Complex Word**	**Annotation 1**	**Annotation 2**	**Annotation 3**	**Annotation 4**	...
The roses have grown out of control, wild and carefree, their bright blooming faces turned to bathe in the early autumn sun.	bright	colourful	bright	brilliant	gleam, luminous	...
He was bright and independent and proud .	bright	intelligent	clever	bright	-	...
Snow covered areas appear bright blue in the image which was taken in early spring and shows deep snow cover .	bright	vivid	deep, bright, shining	vibrant	luminous	...
The packed screening of about 100 high-level press people loved the film as well .	film	movie	film	picture	-	...
That's not to say the process of actual negotiating isn't taking place .	taking	take	happen	occur	-	...

Table 4.14: Sample of the English Lexical Simplification task of SemEval 2012

2. **Spanish language evaluation:** Although at the time of the Thesis, no datasets were found to evaluate this stage. Recently, access to an unpublished dataset for the Spanish language was acquired.

   This dataset is called ALEXSIS[36] [37] which is a spanish dataset for lexical simplification, that contains 381 instances. The sentences and complex words of this dataset were extracted from the CWI Shared Task 2018 datasets[38] for Spanish. Each instance is composed by a sentence, a target complex word (one-word), and 25 candidate substitutions proposed by human annotators. For this evaluation, only uniwords substitutes were taken into account, since multiwords did not perform well in the English language. The sample dataset has the following structure (see Table 4.15), separated by tabulations:

   - The first column shows the sentence of the target word.
   - The second column shows the target complex word.
   - From the third column, the proposed replacements are shown separated by tabs.

   On the other hand, the gold dataset contains the substitutions sorted by simplicity for each instance.

   Initially, it was intended to evaluate the procedures with the same metric as for the English language (TRank). However, in this case it is not possible, since the TRank metric takes into account the number of complexity ties in its measurements and the ALEXSIS dataset for the Spanish language does not present instances in which complexity ties are present. Therefore, by eliminating ties from the TRank measure, it becomes a traditional accuracy metric.

[36] github.com/LaSTUS-TALN-UPF/ALEXSIS
[37] 10.5281/zenodo.5837149
[38] https://sites.google.com/view/cwisharedtask2018/datasets

Sentence	Complex Word	Annotation 1	Annotation 2	Annotation 3	...
Anteriormente el equipo jugaba las ligas de la extinta República Democrática Alemana y actualmente está jugando en la 2.	extinta	muerta	desaparecida	finada	...
Antes de aquello, el estadio albergaba una capacidad para más de 130.000 espectadores.	albergaba	alojaba	acogía	tiene	...
Balbo era guardaespaldas de Cesare Battisti durante las manifestaciones realizadas a favor de la guerra.	guardaespaldas	guardia	segurata	escolta	...
El representativo chileno obtuvo una muy buena participación al conquistar los tres primeros lugares del citado certamen.	representativo	representante	característico	famoso	...
El representativo chileno obtuvo una muy buena participación al conquistar los tres primeros lugares del citado certamen.	conquistar	ganar	colonizar	vencer	...

Table 4.15: Sample of the ALEXSIS dataset for spanish lexical simplification

4.2.4.2 Results

As shown in Table 4.16, the results of the English dataset describe that the frequency-based approach alone obtained good results with a TRank of 0.513, outperforming a strong baseline with TRank of 0.454 and being close in TRank to the best team (UOW-SHEF-SimpLex) presented in the task which developed a supervised approach with contextual and psycholinguistic features.

English approach	**TRank**
Baseline-L-Sub Gold	0.454
Frequency	**0.513**
Word2vec	0.168
FastText	0.1882
Sense2Vec	0.142
BERT	0.177
Frequency-BERT-FastText	0.37
UOW-SHEF-SimpLex	**0.602**

Table 4.16: Results for SR on English test dataset

In contrast, the results of the Spanish dataset show an improvement compared to the previous ones (Table 4.17), obtaining a maximum accuracy of 0.514 with the Fasttext model. Similarly, the next best score was achieved by another embedding model (BERT) with a score of 0.509. A more extensive analysis of the overall results is provided in the discussion Section.

Spanish approach	**Accuracy**
Frequency	0.264
Word2vec	0.134
FastText	**0.514**
Sense2Vec	0.074
BERT	0.509
Frequency-BERT-FastText	0.442

Table 4.17: Results for SR on Spanish test dataset

4.2.4.3 Discussion

In this Section, the evaluations performed for the proposed methods to perform the substitute ranking stage (SG) are described.

As a first approach to address this stage, a weighting system was proposed that extracts word frequencies from a frequency dictionary and also extracts information from embeddings models such as semantic similarity and probability distribution. Table 4.16

shows the results for the English language, where the proposed method obtained a TRank score of 0.37. By performing an ablation study, it was found that the frequency feature obtained better results independently with a score of 0.513, while, on the embeddings side, the FastText model obtained a score of 0.1882, being the best score among the embeddings. Additionally, Table 4.16 includes Trank scores of participating systems in the task belonging to the dataset, where a strong baseline was presented with a TRank score of 0.454, which the proposal managed to outperform and also with the frequency feature, a score close to the best team in the task that obtained a TRank score of 0.602 (UOW-SHEF-SimpLex) was obtained.

When analyzing errors, problems were detected with the classification of multiwords, because the classical embedding models receive uniwords as inputs, they did not assign a weight to the multiwords, consequently, classifying it as the most complex term in the list and therefore, obtaining wrong results in many cases. In the case of the results for the BERT model, this Thesis believes that by performing a fine-tuning process as was done in the CWI stage, it could improve the results in this task.

Similarly, Table 4.17 shows the results of the proposed rankers for the Spanish language. Using the weighting system, an accuracy score of 0.442 was obtained. Unlike the English language, the frequency-based ranker did not perform well, obtaining 0.264 accuracy points. This is partly due to the fact that the RAE frequency dictionary is smaller in size than the one used for the English language.

Similar to previous results, the Sense2vec and Word2vec models did not perform well. When analyzing the results, it was noted that there were many cases of OOV (Out of Vocabulary Word) and the results were similar because both models were trained under the same corpus. However, the contextual BERT model and the classical FastText model obtained better results than in the English language, achieving an accuracy score of 0.509 and 0.514 respectively. These results reflect that the vocabulary had a higher number of occurrences than in the other models and also took advantage of the morphological/contextual information offered by these models compared to the classical Word2Vec model.

Finally, the proposed weighting system combining linguistic resources and embedding supports the Thesis in answering hypotheses 2 and 3.

4.3. Word Sense Disambiguation (WSD) method

Following the accessibility guidelines, a mechanism to provide definitions had to be implemented, however, due to the ambiguity present in the language, a method of word sense disambiguation had to be incorporated. In this Section, the evaluations of the method proposed in the previous Chapter are shown.

4.3.0.1 Methodology

The evaluation of this stage was conducted by a linguist expert, specialized in easy-to-read content and plain language [171]. About the evaluation dataset, a set of sentences was constructed, associated with a target word and a definition selected by the proposed method. As a result, a dataset with a length of 525 instances was obtained.

The evaluator had the task of verifying whether the definition proposed by the method was correct, taking the context of the word in the sentence into account. The unit of measurement to be retrieved was the percentage of correct definitions.

4.3.0.2 Results

Table 4.18 shows this evaluation, where of the 525 instances evaluated, 70.48% of definitions were qualified as correct. The method using the BERT model processed 117 instances of which the expert rated 64.95% of the total as correct, while the "First in" method processed 408 instances, of which 72.06% were rated as correct.

	# Instances	**% Correct**
BERT Model	117	64.95
First in	408	72.06
Total	525	**70.48**

Table 4.18: Results in WSD System

In addition, as a contribution of this Thesis, the implementation of the methods and their evaluations described above can be found publicly in web repositories [39] [40].

4.3.0.3 Discussion

In this Section, the evaluations made to the proposed approach for the detection of a correct definition of target words are described.

To perform this task, a WSD method supported by a Spanish BERT model was proposed. Table 4.18 shows the results of the evaluation performed by the expert linguist. As mentioned above, stable results were obtained, with a total of 70.48% correct answers out of 525 examples.

Table 4.19 shows an example of a positive result, finding matches in the words "declarar" (state) and "exponer" (present). This result is given because the system usually selects longer definitions, since there is a higher probability that in a longer definition there is a higher number of matches. Another similar case occurs in the example of Table 4.19,

[39] https://github.com/LURMORENO/easier
[40] https://github.com/ralarcong/EASIER_EVALUATIONS

where it was noticed that the system tends to choose the definitions provided by the "Diccionario Fácil" resource, since this resource offers examples along with the definitions themselves. Table 4.21 shows that instead of choosing a short definition that does not offer any explanation (named 1), the system chooses a complete definition with an extensive explanation (named 3).

When analyzing negative results, it was detected that the system presented problems when dealing with sentences with a generic sense. Table 4.20 shows an example related to this problem, where due to the context where the word "redunda" is found, the model provides generic words and consequently selecting the wrong definition. Moreover, in some cases, it was found that the system does not find coincidences because the definition is in another grammatical form. This issue can be corrected by lemmatizing the words in the definition and searching for coincidences.

The analysis showed another type of negative result, when the system was dealing with short sentences and generic sense. Table 4.20 shows an example where the model did not have much context to analyze and consequently produced generic words such as "results", "consists" and "helps".

Finally, given the good results of the contextual embedding model, this stage provides support for Hypothesis 2.

Sentence	"… confianza para todos los ciudadanos", ha explicado. ("… trust for all citizens," he explained)
Target Word	"explicado" (explained)
Definition options	1. Declarar o exponer cualquier materia, doctrina o texto difícil, con palabras muy claras para hacerlos más perceptibles. (state or present any material, doctrine or difficult text with very clear words so as to make it more understandable) 2. Enseñar en la cátedra. (teach from a podium) 3. Justificar, exculpar palabras o acciones, declarando que no hubo en ellas intención de agravio. (justify, exculpate words or actions, stating that there was no insult or injury intended) 4. Dar a conocer la causa o motivo de algo. (present the cause of or motive for something)
Definition selected by the system	Declarar o exponer cualquier materia, doctrina o texto difícil, con palabras muy claras para hacerlos más perceptibles. (state or present any material, doctrine or difficult text with very clear words so as to make it more understandable)

Table 4.19: WSD system positive results example

Sentence	"Y todo ello **redunda** en una mejor salud" ("And all of this **results** in improved health.")
Target Word	"redunda" (results)
Definition options	Dicho especialmente de un líquido: Rebosar, salirse de sus límites o bordes por demasiada abundancia. (Used especially when referring to a liquid: spill over the edges or borders because the quantity exceeds the capacity.)
Definition selected by the system	Dicho de una cosa: Venir a parar en beneficio o daño de alguien o algo. (Used when referring to a thing: to create a benefit for or a damage to someone or something.)

Table 4.20: WSD system negative result

Sentence	“Deben ser reconocidos por su contribución a la mejora en el intercambio y procesamiento de datos e información entre todos los agentes del sistema.” (They should be recognized for their contribution to the improvement in the exchange and processing of data and information between all of the system’s agents.)
Target Word	“procesamiento” (processing)
Definition options	1. Acto de procesar. (the act of processing) 2. Acto por el cual se declara a alguien como presunto autor de unos hechos delictivos a efectos de abrir contra él un proceso penal. (an action by which an individual who has allegedly committed a crime is formally charged and criminal proceedings are initiated) 3. Aplicación sistemática de una serie de operaciones sobre un conjunto de datos, generalmente por medio de máquinas, para explotar la información que estos datos representan. (The systematic application of a series of operations on a set of data, normally by machines, to exploit the information said data represents.)
Definition selected by the system	Aplicación sistemática de una serie de operaciones sobre un conjunto de datos, generalmente por medio de máquinas, para explotar la información que estos datos representan. (The systematic application of a series of operations on a set of data, normally by machines, to exploit the information said data represents.)

Table 4.21: WSD system on selecting a more explanatory definition

4.4. System User Evaluation

This Section presents an evaluation of the complete proposal through real users. The impact of the proposal, specifically on lexical simplification in the target user, is evaluated in order to identify possible future improvements in this Thesis. In the following Sections, more details about this experimental study will be given, indicating a description of the participants, materials, measures, procedure, results, and discussion.

4.4.1. Participants

The participants were recruited by the HULAT group[41] in collaboration with the AMAS Group[42], an organization that works for people with intellectual disabilities. 50 participants have been involved in this experimental study.

Table 4.22 shows an overview of the demographic information of the participants. The participants were divided into two groups: Group 1 represented 25 elderly people (50%), and group 2 represented 25 people with intellectual disabilities (50%). Across the entire population (both groups), the largest number of participants was reached by the age range between younger than 44 years and 67 to 70 years (28% for both ranges), and the smallest number of participants was reached by the age range between 45 to 66 years (12%) and participants older than 71 years (20%). There was a minimal difference between the number of men (52%) and women (48%) with 26 and 24 participants respectively.Regarding the educational level of the participants, the majority had a high school level of education with 22 participants (44%), followed by primary level with 16 participants (32%), and the least number of participants were registered for people with no registered studies and people with a university degree with 8 (16%) and 4 (8%) participants respectively. In addition, due to the objectives of the study, the participants' reading level was evaluated through the number of books read per year, where the largest number of participants was reached by participants who did not read any book and participants who read 1 to 3 books, with 16 (32%) participants for both cases; followed by 8 (16%) participants who read 3 to 6 books, 7 (14%) participants who read 6 to 12 books and finally 3 (6%) participants who read more than 12 books per year.

When analyzing the groups separately, group 1 was represented by participants between the ages of 65 and 75 years. On the gender side, there was a minimal difference with 13 (52%) and 12 (48%) male and female participants, respectively. Regarding the educational level of the participants in this group, the majority had a high school level of education with 13 participants (52%), followed by 5 (20%) participants with no registered studies, 4 (16%) participants with a university degree and 3 (12%) participants with a primary degree. Finally, when collecting reading data from the participants, 10 (40%) participants who read 1 to 3 books per year were found, followed by 6 (24%) participants who read 6 to 12 books, 5 (20%) participants who read 3 to 6 books, 3 (12%) participants

[41] hulat.inf.uc3m.es/
[42] grupoamas.org/

Features	All Participants		Group 1 (Elder)		Group 2 (Int)	
	N=50	%	N=25	%	N=25	%
Age						
44 or younger	14	(28)	-	-	14	(56)
45-66	12	(24)	4	(16)	8	(32)
67-70	14	(28)	11	(44)	3	(12)
71+	10	(20)	10	(40)	-	-
Gender						
Female	24	(48)	12	(48)	12	(48)
Male	26	(52)	13	(52)	13	(52)
Education (Highest completed)						
None	8	(16)	5	(20)	3	(12)
Primary school	16	(32)	3	(12)	13	(52)
High school	22	(44)	13	(52)	9	(36)
University education	4	(8)	4	(16)	-	-
Reading experience (books per year)						
None	16	(32)	3	(12)	13	(52)
1-3	16	(32)	10	(40)	6	(24)
3-6	8	(16)	5	(20)	3	(12)
6-12	7	(14)	6	(24)	1	(4)
12+	3	(6)	1	(4)	2	(8)

Table 4.22: Participant demographic information (Group 1: Elder people, Group2: People with intellectual disabilities).

who did not have the habit of reading and finally only one (4%) participant who read more than 12 books per year.

On the other hand, group 2 was represented by 14 (56%) participants under 44 years of age, 8 (32%) participants between 45 and 66 years of age, and 3 (12%) participants between 67 and 70 years of age. Similar to the previous group, there was a minimal difference with 13 (52%) and 12 (48%) male and female participants, respectively. Regarding the educational level of the participants in this group, the majority had a primary school level of education with 13 participants (52%), followed by 9 (36%) participants with a high school level and 3 (12%) participants with no recorded educational level. Finally, concerning the reading level, the majority had no reading habit with 13 (52%) participants, followed by 6 (24%) participants who read 1 to 3 books per year, 3 (12%) participants who read 3 to 6 books, 2 (8%) participants who read more than 12 books and one (4%) participant who read 6 to 12 books per year.

4.4.2. Materials

In order to bring the experimental study closer to a real case, texts were selected from the public website of the Community of Madrid[43] aimed at providing useful information to all citizens, in which there are texts about culture and leisure, health, among others. A total of 30 sentences with similar lengths were randomly selected from different articles belonging to the year 2021.

4.4.3. Procedure

First, participants were informed about the study and required to sign a consent form. After, the participants filled a simple demographic questionnaire. Finally, each participant was asked to complete tasks. The entire experimental study was supervised by an ethics committee of the Carlos III University of Madrid.

The sessions were held at the facilities of the AMAS group, where the researcher was together with the professional from the AMAS group. Both for the people with disabilities and for the elderly people, sessions were held in small groups of 2 or 3 people where, before starting to perform each task, the statement was read aloud of each task and resolved if there was any doubt. The same mechanics were followed for all sessions: reading and explanation prior to carrying out the task.

The main steps were:

1. Demographic questions about age, gender, education level and reading habits (see Appendix 6.3.5.
2. Explanation and performance of task 1, referring to the CWI task (see Section 4.4.3.1).
3. Explanation and performance of task 2, referring to remaining tasks in the lexical simplification process where a substitute needs to be provided (see Section 4.4.3.1).

4.4.3.1 Tasks

In order to evaluate the different steps of the proposal, the following tasks were defined.

1. **Task 1:** : This task aims to measure the CWI task, that is to say, predictions that the proposed system produces in texts when discerning between simple or complex words.

   Each participant had to analyze 15 randomly selected sentences with a length between 15 and 30 words. In each sentence, the participant had to select uniwords or multiwords that he/she considered complex or difficult to understand. Appendix 6.3.5 shows the entire set of instances for task 1.

[43] comunidad.madrid/

2. **Task 2:** The objective of this task is to measure the quality of the substitutes to the detected complex words. In order to verify if, in fact, the substitutes proposed by the system are more adequate than the original complex word.

   Each participant had to analyze 15 sentences with a length between 15 and 30 words, randomly selected. In each sentence, a detected complex word is highlighted and 2 substitution candidates extracted from the system are proposed. Thus, each participant was tasked to analyze the sentences with each substitute and as a next step, answer yes/no questions about whether the substitute helps to better understand the sentence. Appendix 6.3.5 shows the entire set of instances for task 2.

4.4.4. Measures

The measures utilized were metrics used in the domain of machine learning methods (accuracy, precision, recall and F-1) in order to be able to compare the proposal to other related work [172] [8].

4.4.5. Results and discussion

Table 4.23 shows the results for task 1. The results were moderate when compared to other systems in the English language [172], obtaining an overall F1-score of 0.536 points, with a better recall than precision with 0.665 and 0.558 respectively. When evaluating the proposal by groups, there was a slight difference in precision between groups 1 and 2, with 0.54 and 0.576 points, respectively. On the other hand, with respect to recall, there was a greater difference between the groups, with 0.72 points for group 1 and 0.65 points for group 2.

GROUP 1					GROUP 2				
ID	**AC**	**PR**	**RC**	**F-1**	**ID**	**AC**	**PR**	**RC**	**F-1**
11	0.691	0.513	0.844	0.434	1	0.669	0.537	0.574	0.519
12	0.691	0.513	0.844	0.434	2	0.686	0.546	0.613	0.524
21	0.720	0.585	0.709	0.575	3	0.729	0.620	0.696	0.625
22	0.691	0.531	0.640	0.486	4	0.703	0.555	0.677	0.527
23	0.708	0.554	0.711	0.522	5	0.695	0.548	0.642	0.522
24	0.725	0.588	0.726	0.578	6	0.725	0.592	0.717	0.585
25	0.695	0.520	0.845	0.447	7	0.716	0.582	0.694	0.572
27	0.686	0.517	0.631	0.454	8	0.720	0.631	0.673	0.639
28	0.843	0.803	0.826	0.813	9	0.644	0.561	0.571	0.562
30	0.708	0.554	0.711	0.522	10	0.716	0.578	0.701	0.565
31	0.695	0.524	0.746	0.458	13	0.674	0.529	0.575	0.500
32	0.686	0.510	0.677	0.432	14	0.725	0.631	0.681	0.639
33	0.708	0.540	0.850	0.486	15	0.703	0.665	0.661	0.663
34	0.695	0.527	0.704	0.469	16	0.720	0.610	0.680	0.614
35	0.674	0.501	0.508	0.426	17	0.712	0.586	0.673	0.581
36	0.691	0.513	0.844	0.434	18	0.712	0.625	0.660	0.632
37	0.691	0.520	0.679	0.456	19	0.691	0.588	0.626	0.590
38	0.703	0.533	0.848	0.473	20	0.712	0.575	0.685	0.562
39	0.699	0.537	0.700	0.491	26	0.703	0.533	0.848	0.473
40	0.699	0.534	0.724	0.481	29	0.695	0.520	0.845	0.447
42	0.703	0.533	0.848	0.473	45	0.682	0.525	0.597	0.481
43	0.691	0.513	0.844	0.434	48	0.703	0.569	0.658	0.556
44	0.691	0.513	0.844	0.434	49	0.720	0.574	0.741	0.554
46	0.682	0.518	0.595	0.462	50	0.686	0.539	0.613	0.508
47	0.682	0.500	0.341	0.406	51	0.720	0.588	0.703	0.581
GROUP 1 SCORES					**GROUP 2 SCORES**				
ID	**AC**	**PR**	**RC**	**F-1**	**ID**	**AC**	**PR**	**RC**	**F-1**
ALL	0.702	0.54	0.72	0.494	ALL	0.703	0.576	0.65	0.57

OVERALL SCORE			
ACURACCY	**PRECISION**	**RECALL**	**F-1**
0.702	0.558	0.665	0.536

Table 4.23: Result metrics for both groups in Task 1 where ID = User Id, AC = Acuraccy, PR = Precision

Figure 4.2 shows a graphical comparison of the precision scores between the study groups, where a clear difference can be seen between the groups where group 2 obtained better results than group 1. This indicates that the proposed CWI model obtained a higher number of quality predictions for people with intellectual disabilities than for elderly people, by obtaining a higher number of true positives. Although the difference in scores between the groups is minimal (about 0.036 points), this suggests that the proposal makes higher quality predictions for people with intellectual disabilities.

	PRECISION		RECALL	
	GROUP 1	GROUP 2	GROUP 1	GROUP 2
25%	0.513	0.545	0.679	0.626
50%	0.523	0.575	0.724	0.673
75%	0.537	0.591	0.844	0.696
Max	0.803	0.665	0.85	0.848
Mean	0.54	0.576	0.73	0.672
Std	0.059	0.038	0.125	0.07

Table 4.24: Precision and recall statistics ordered by groups.

Table 4.24 complements this graph by showing the data related to the precision scores divided by quartiles, which also shows an average of 0.54 points for group 1, compared to 0.576 points for group 2. Additionally, the standard deviation of each group is specified with a lower deviation of group 2 with 0.038 points, compared to group 1 with a higher deviation of 0.0.59 points.

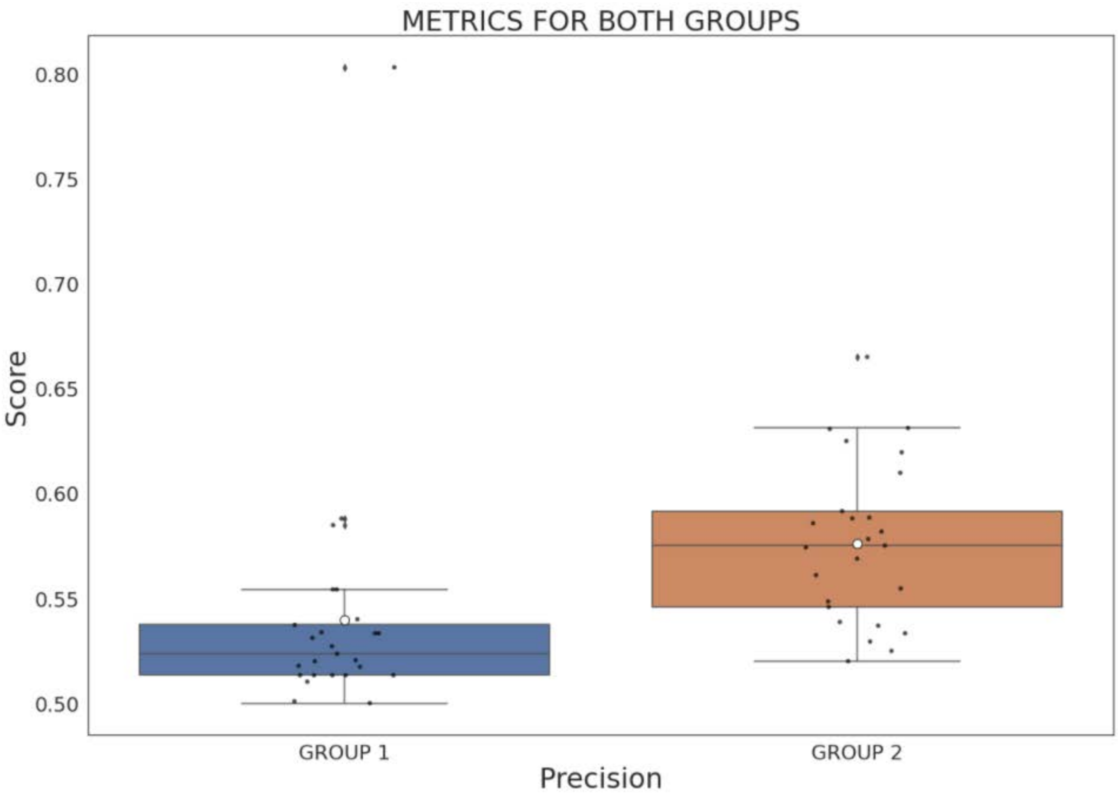

Figure 4.2: Precision scores among every participant divided by groups.

On the other hand, when analyzing recall scores, an increase compared to the precision was noted. Figure 4.3 shows another graph comparing the recall scores of the study groups, where clearly a larger data dispersion can be noticed in the first group than in the second. This is confirmed by the data provided in Table 4.24, where the same data as in the previous metrics can be seen. However, unlike the previous metric, group 1 obtained a higher score than group 2, with a recall mean of 0.73 and 0.672 points respectively; additionally, group 2 presents a standard deviation of 0.07 points and group 1 presents an even higher standard deviation with 0.125 points, which can be seen graphically in Figure 4.3. These results are encouraging, as this Thesis aims to find a model that generalizes correctly to both target audience groups and the general public. This is confirmed by comparing the precision and recall metrics, the latter being higher.

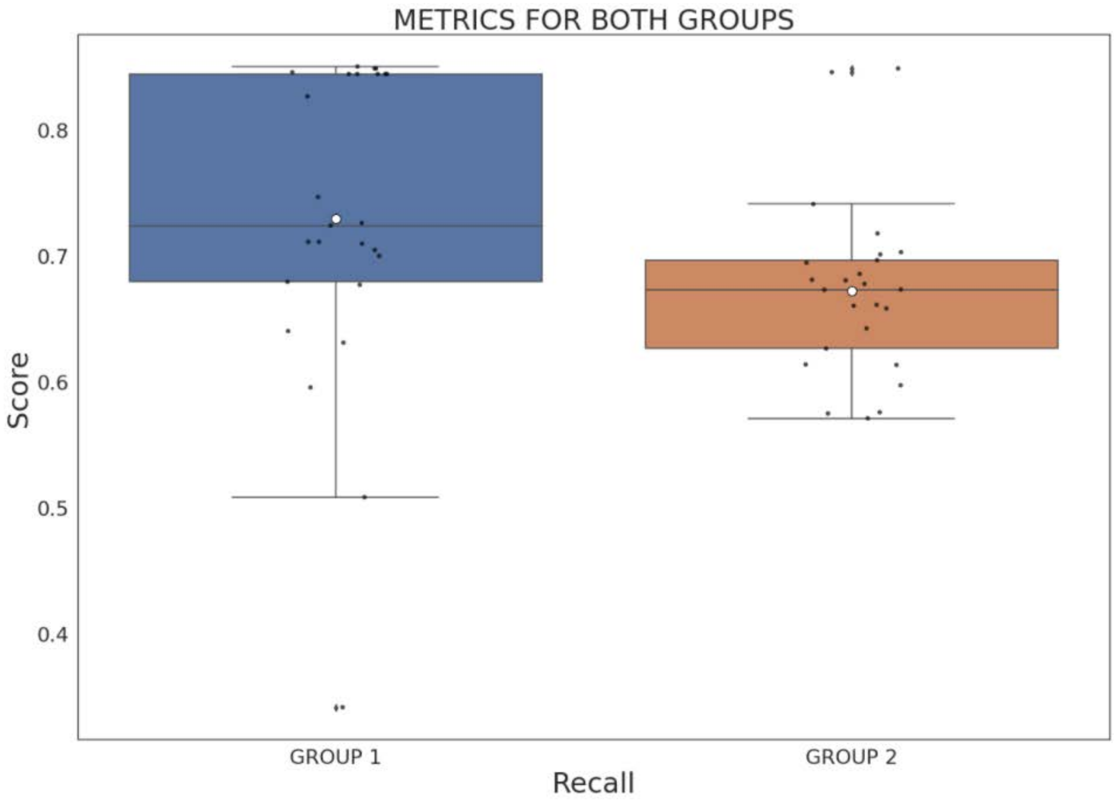

Figure 4.3: Recall scores among every participant divided by groups.

An overall F1-score of 0.536 points was obtained, and when divided by groups, a better F1-score was noted for group 2 with 0.57 points, compared to the F1-score of group 1 with 0.494 points.

In addition, Figure 4.4 shows the number of words that each participant considered as complex, which are divided by groups. In this Figure, it can be clearly distinguished that people in group 2 considered the texts to contain a higher number of complex words, while for group 1, only one person had a higher density of detected words (73 words), which helps this Thesis to make more sense of the precision and recall scores described above. This indicates that the selected texts had a more complex content for people with intellectual disabilities than for older people and consequently, it is suggested that the proposal had greater benefit for the group with intellectual disabilities.

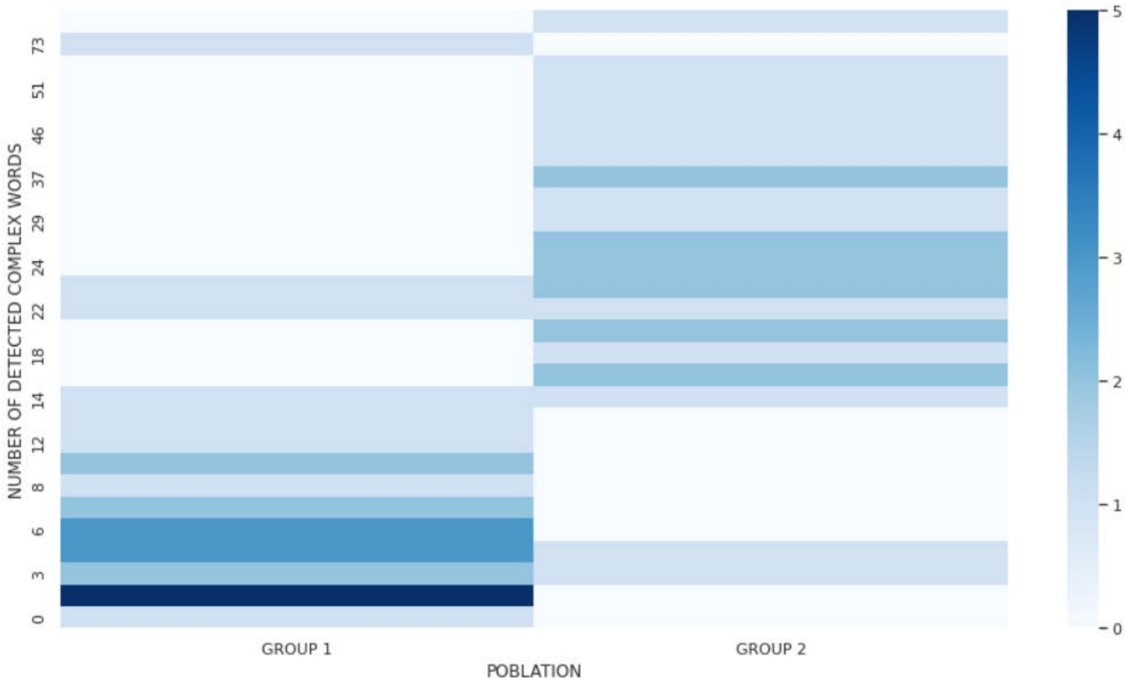

Figure 4.4: Number of detected complex words, divided by groups.

About task 2, results were moderate too compared to similar tasks in the English language. The quality of the substitutes generated by the proposal was evaluated and as described above, each participant was tasked to evaluate 2 candidate substitutes for each sentence in the dataset. Table 4.25 shows two types of results divided by groups, the first one where the number of users who accepted at least one of the candidates presented for each sentence is recorded and the second one where the number of users who accepted both candidates presented for each sentence is recorded.

For the first result, a percentage greater than half of the participants in both groups was obtained, where for group 1 an acceptance of 59.7% was obtained, while for group 2 a greater benefit was obtained with 70.1% acceptance. This suggests that the proposal helps to reduce the level of complexity of the sentences, at least with a suggested candidate, and although good acceptance was achieved in both groups, the group with intellectual disabilities benefited the most.

It is important to mention that the data shown in Table 4.25, shows the number of cases divided by sentence and also shows the mean number of cases per group and type of result, together with their percentage of acceptance.

	At least one candidate ranked as correct		**Both candidates ranked as correct**	
Sentence-ID	**Group 1 (N:25)**	**Group 2 (N:25)**	**Group 1 (N:25)**	**Group 2 (N:25)**
S1	22	20	15	13
S2	14	22	5	4
S3	12	17	8	9
S4	18	17	10	5
S5	21	19	7	5
S6	17	16	3	9
S7	15	15	12	11
S8	8	17	6	11
S9	14	16	7	9
S10	12	19	3	4
S11	23	20	14	7
S12	15	15	12	6
S13	7	16	5	12
S14	8	16	7	11
S15	18	18	7	9
MEAN	14,9	17,5	8,1	8,3
ACCEPTANCE (%)	59,7	70,1	32,3	33,3

Table 4.25: Task2: Number of cases where at least one candidate and both candidates were ranked as correct, sorted by groups and sentences.

These results in turn, helped this Thesis to detect which replacement candidates were the least accepted, for example in sentences like S8, S13 and S14 the acceptance percentages for people in group 1 were very low, while for people in group 2, they were better accepted. Other sentences, such as S6, had only three cases in which both candidates were accepted. Table 4.26 shows this specific example, where in the sentence it is shown that the candidate generated by the proposal is equal to one of the words present in the sentence.This happens because when evaluating the semantic similarity between these words, the score is higher than with any other candidate and consequently, it is chosen by the selector, so in the future the application of a filter for these cases is planned.

Sentence	La contaminación de ambientes interiores de los inmuebles es un factor determinante en la salud y **bienestar** de sus usuarios. (The contamination of indoor environments of buildings is a determining factor in the health and **well-being** of its users.)
Target Word	bienestar (well-being)
Candidate 1	salud (health)
Candidate 2	sanity (sanidad)

Table 4.26: Proposal negative result example on the substitute management

Figure 4.5 shows these results graphically, dividing the instances where at least one replacement was accepted and instances where no replacement was accepted, divided by group and educational level. As can be seen in the graph, for group 1 there is a high number of accepted replacements in participants with high school level education for group 1 and primary school level for group 2. It is important to note that there is a higher concentration of participants with these levels of education for each group. It is for this same reason that there are cases in which the number of acceptances is low, as in the university level, which only had participants in group 1. This confirms the results described above (Table 4.25), where the proposal suggests that it is more beneficial for people with intellectual disabilities, obtaining a high acceptance in this group by only having 33 cases where candidates were rated as incorrect (out of a total of 375 cases), obtaining better results for people with a primary education level, with only 3 cases with candidates indicated as incorrect (out of a total of 180).

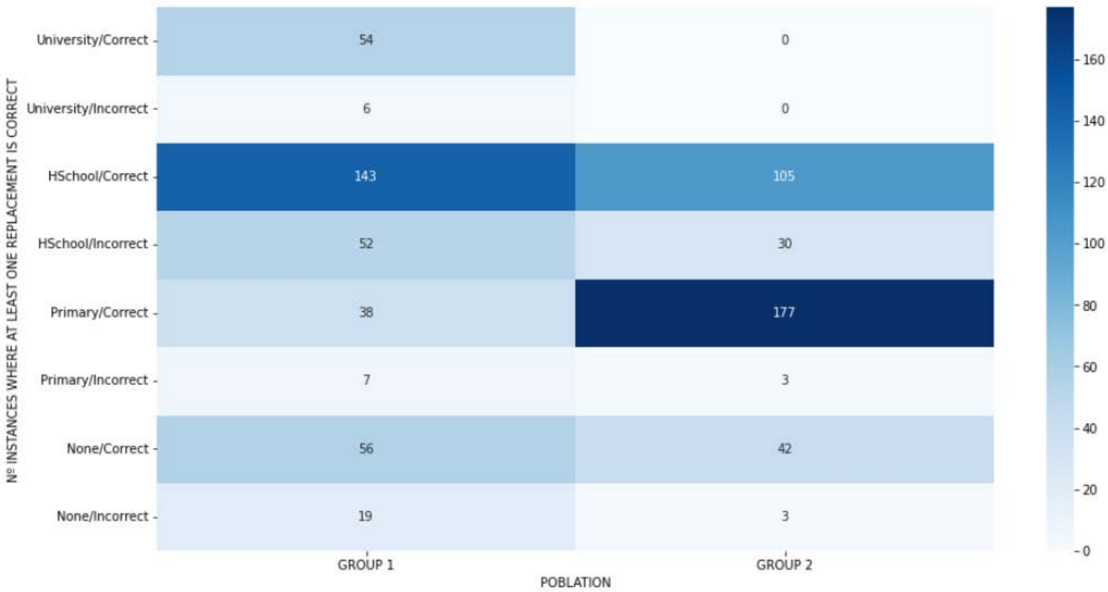

Figure 4.5: Number of instances where at least one substitute was taken as correct of incorrect, divided by group and education level

4.5. Conclusions

This Chapter has described the experiments carried out in order to evaluate the Thesis proposal. To support Hypothesis 1, the data quality of the EASIER corpus was validated, which obtained moderate results in the CWI and substitute datasets. In addition, this resource supports experimentation and provides answers to other hypotheses in this Thesis.

In the CWI stage, a supervised machine learning method combined with different features to represent a word were explored. Among these features, E2R and embedding resources were experimented with, obtaining better results than systems using the same data, thus supporting hypotheses 1 and 2. Later, transfer learning methods were experimented with, obtaining promising results and thus supporting Hypothesis 3.

For the SG stage, methods based on linguistic resources were experimented with, and then compared with methods based on automatic generation supported by embedding models. The results showed that linguistic resources at this stage obtained better results, thus supporting Hypothesis 2.

For the SS stage, methods based on the context of the target word were experimented with to reduce the list of the previous stage. To achieve this, an exploration of different embeddings was performed in order to discover which of these best captured the context of the target word. The results showed an improvement to previous research, thus answering Hypothesis 2 of this Thesis.

In the last stage of LS (SR), an experimentation with a weighting system to rank words according to their simplicity is described, which is based on frequency and embedding features. Initially the methods were validated with data for the English language, obtaining fair results by obtaining scores close to other systems in the same task. Later, an experimentation with data for the Spanish language is performed, obtaining better results than the previous ones and, consequently, providing an answer to hypotheses 2 and 3 of this Thesis.

In addition, this Thesis proposes a disambiguation method for the selection of correct definitions to detect complex words, supported by contextual embeddings. The validation was performed by an expert linguist in easy-to-read content, showing fair results in the disambiguation task, thus supporting Hypothesis 2.

Finally, it was essential to conduct an experimentation with the end user, so to validate the proposal, a study was conducted with elderly people and people with intellectual disabilities. The results were moderate compared to other works in the English language, which will allow the research of this Thesis to improve the user experience. These results, in turn, support each Hypothesis of this Thesis, since this study covers a large part of the contributions of this Thesis.

Chapter 5

Contribution to innovation. EASIER system

This Chapter presents the EASIER platform as a contribution of the Thesis to innovation. EASIER is a platform that helps people to understand texts better and works based on artificial intelligence methods. This platform basically provides a compilation of the contributions presented in this Thesis by performing lexical simplification of Spanish texts offering different comprehension aids. The following Sections describe the motivation for this innovation, along with a description of the modules that compose the platform.

5.1. Motivation

The Internet is a vital means of access to information, as well as an important participatory tool for society. However, people with cognitive and learning disabilities are not adequately supported by current Internet accessibility efforts [173].

Therefore, it is essential to address the issue of cognitive accessibility from two different points of view: the lack of comprehension and readability of texts and the design of user interfaces. Therefore, as a contribution of this Thesis to innovation, a web platform is presented that aims to develop innovative technological solutions that favor access to clear, simple and easy to understand content for people with intellectual disabilities specifically, and for all people in general. To achieve this, the platform contains an underlying system with the contributions of this Thesis, offering support for simplification tasks, within a cognitively accessible user interface [174].

In the following Section, this web platform with an accessible interface that provides automatic lexical simplification of Spanish texts is presented.

5.2. System Overview

This Thesis research has been integrated into a web platform to test and show the effectiveness of the proposals[44]. A typical interaction of the system would be as follows:

1. A Spanish text is entered by the user (Figure 5.1).

2. Complex words are displayed in the following interface (Figure 5.2).

3. For each complex word detected, simpler replacements, a definition and a pictogram are offered, supported by the language and accessibility resources such as an easy-to-read dictionary [175] (Figure 5.2).

Figure 5.1: Sample of the EASIER system user interface

The system´s pictogram service is provided because this Thesis aims to improve the accessibility of texts, and to achieve this, the easy reading guide that says "To illustrate your text, you can use: pictures, drawings or symbols" is followed, along with the plain language guide that says "Use pictorial representation and other media: as illustration, as support while reading". This additional information is provided through an API of the ARASAAC web resource[45], which offers graphic support for people with communication disabilities[174].

[44] github.com/LURMORENO/easier
[45] arasaac.org/developers/api

Figure 5.2: Sample of the results for the EASIER system user interface.

Additionally, the EASIER platform has been designed to comply with WCAG 2.1 (Level AA). Cognitive and Learning Disabilities Accessibility (COGA) guidelines have also been followed, such as using clear and understandable content and making each step of the simplification process as clear as possible, including instructions. Moreover, a consistent visual design using symbols that assist the user has been used. Finally, to validate the user interface of the platform, an experimental study with elderly subjects and subjects with intellectual disabilities was conducted, obtaining satisfactory results [174].

The webpage's user interface has been designed responsively, and a user interface for mobile devices is also provided (Figure 5.3). Moreover, browser extensions have been developed for both Chrome and Mozilla browsers that offer the function of identifying complex words and providing synonyms for text users to select on any webpage using the EASIER system.

Figure 5.3: Sample of the EASIER system mobile user interface.

5.3. Technical Information

This Section briefly explains the procedures required to enter structured data into the system, with the ultimate intention of complying with the textual content accessibility guidelines.

5.3.1. System Architecture

As described above, this system seeks to achieve lexical simplification. To achieve this, the system receives the raw text to be pre-processed into a structured format. This new text is then passed through the CWI module in order to detect which words are complex. These words are then passed to the Substitute Generation and Selection module in order to find synonyms appropriate to the context. Finally, the system provides a simplified content fulfilling the stated objectives. This Section describes the tasks to be carried out to achieve text simplification:

1. **Data pre-processing** for the conversion of the text that is entered into the system (raw Text)

2. **Complex Word Identification (CWI)**, supported by a linear SVM model and methods described in Chapter 3.

3. **Candidate generation/selection**, supported by linguistic and automatic resources and methods based on semantic similarity described in Chapter 3.

4. **Definition search**, supported by the WSD methods described in Chapter 3.

5. **Pictogram search**, supported by the ARASAAC web resource[46].

5.3.2. Pre-Processing

The system must receive a structured input for its correct operation, since most of the information on the Internet comes from heterogeneous sources and unstructured data, it is necessary to create a data pre-processing module for the system. Therefore, this module is supported by a Spanish Spacy model[47] and consequently divided into three main processes:

1. **Segmentation:** As a first step, when any text is received, it is divided into sentences, in order to have a first structuring filter for the following steps.

2. **Tokenization:** The objective is to obtain a list of words from a given text.

3. **POS (Part-of-Speech) Tagging:** In order to distinguish words, the Spacy model has POS tagging functionality. The result is a list of words labeled in pairs, for example ('run','VERB'). In addition, to finish the structuring stage, the system is left with specific Pos tags to avoid words that lack semantic content.

Once these tasks are completed, the words are processed by the strategies proposed in previous Chapters, in order to obtain the final product (Figure 5.2).

5.4. Conclusions

In this Chapter, as a contribution of the Thesis to innovation, the EASIER platform has been presented, which compiles the best performing contributions of this Thesis with the aim of supporting the understanding of textual content to the target audience and the general public, with a cognitively accessible interface.

This platform has an underlying system that is mainly supported by the methods and resources described in this Thesis (Chapter 3). For the CWI task, a linear SVM model is used. For the generation of candidate substitutions, the system is supported by linguistic and automatic generation resources. Finally, for the selection of the most suitable candidate for the context, the system is based on semantic similarity methods supported by

[46]arasaac.org/developers/api
[47]/spacy.io/models/es

embeddings models. On the other hand, the system offers additional services, such as the search for definitions to the detected complex words, which is supported by contributions described in this Thesis. Another service is represented by a pictogram search method, supported by a pictogram database

Chapter 6

Conclusions

There are accessibility barriers to textual content due to understanding problems faced by people with cognitive disabilities. Faced with these issues, there are accessibility requirements that must be met. With this motivation arises the objectives of this Thesis.

6.1. Main Contribution

Driving by the objectives of this Thesis, this research has carried out a series of studies, analyses, and development of NLP techniques and resources designed to address the improvement of the Spanish texts simplification of a generic domain to support audiences with cognitive disabilities. In this way, the set objectives have been achieved.

- In this Thesis, a study and analysis of the state of the art in NLP, text simplification and accessibility issues was carried out, which is described in Chapter 2.
- In addition, as a contribution generated by this Thesis, an architecture supported by methods and resources was designed to support lexical simplification in the framework of cognitive accessibility, which is described in Chapter 3.
- In order to validate the Thesis proposal, different experiments were performed on the resulting methods and resources. Subsequently, for each experimentation, an analysis of these was carried out, in order to achieve a comparison with state-of-the-art research (see Chapter 4).
- Finally, this analysis of results allowed the Thesis to discover open research questions through discussions for each contribution. In addition, it allowed the Thesis to offer conclusions to the proposals, along with future lines of research, which are described in this Chapter.

This has resulted in several contributions to the research that are based on the research hypotheses formulated in Section 1.4.

- This Thesis explores resources in Spanish for lexical simplification. Due to the limited resources for this task compared to other languages, the decision was made to create a new resource for the evaluation of lexical simplification methods. Consequently, this Thesis presents a corpus (EASIER)[48] developed in the scope of cognitive accessibility to support audiences with cognitive disabilities. This corpus possesses complex target words, along with proposed contextualized synonyms, giving response to the research Hypothesis 1 and implicitly supporting all the hypotheses related to lexical simplification, because thanks to this resource, it was possible to validate the methods proposed in this Thesis [49] [50].

- Also, this Thesis explored the possible benefits of easy-to-read resources to the NLP area. As a result, this Thesis presents a dictionary of simple words extracted from these resources aimed at improving cognitive accessibility. This new resource supports the CWI task and in consequence, research Hypothesis 1.

- In addition, the usefulness of different types of embeddings for the lexical simplification task is explored. Additionally, new embedding models [50] were created with the help of the previously described corpus. These models supports research Hypothesis 2 and 3.

- As for the selection of substitutes, a method presented in related works was modified to obtain a better detail of the context of a word [50]. This is achieved thanks to the information of existing embeddings models and models created within the framework of this Thesis. Similarly, this support research hypotheses 2 and 3.

- In the case of the substitute ranking stage, a weighting system was proposed to evaluate complexity among a list of candidates[50]. At first, the proposed methods were evaluated only for the English language due to the scarcity of resources to evaluate this stage, showing fair results. Later, access was obtained to a Spanish dataset for this stage, obtaining good results with the embedding features. These results support research hypotheses 2 and 3.

- Also, following accessibility guidelines, a method for finding correct definitions to target words was proposed. This method showed a good percentage of success, thus supporting Hypothesis 3[51].

- Additionally, to evaluate the reliability of the proposal of this Thesis with the target user, a study with elderly users and users with intellectual disabilities was conducted. The results were moderate compared to other works in the English language, which allowed this Thesis to find points of improvement with the user. At the same time, this evaluation supported in answering each of the hypotheses of this Thesis.

[48] https://github.com/LURMORENO/EASIER_CORPUS
[49] https://github.com/ralarcong/EASIER_lexical_simplification
[50] https://github.com/LURMORENO/easier
[51] https://github.com/ralarcong/EASIER_EVALUATIONS

- Finally, the Thesis proposal has been integrated into a web platform within the framework of the EASIER project [52] [53]. This platform is currently in operation.

As a result of the confirmation of the research hypotheses, the general Hypothesis that it is possible to improve the accessibility of Spanish texts in a generic domain using NLP techniques such as word embedding combinations using accessibility resources to support audiences with cognitive disabilities can be confirmed.

6.2. Future Work

For future work, the performance of other types of classification approaches will be studied, such as recent deep learning approaches (e.g., graph-based neural networks).

In addition, other text simplification approaches will be explored, such as syntactic simplification, with the aim of complementing the lexical simplifier. This is expected to create a hybrid approach that can provide better textual comprehension.

Also, since this Thesis has been conducted for a specific language, in future work, the extension of these methods to different languages such as English is planned, in order to see the usefulness of the proposed methods for languages other than Spanish.

Finally, as has been demonstrated throughout this Thesis, the production of accessibility resources is beneficial to the NLP discipline. For this reason, as future work, the extension of existing resources or creation of new resources is planned.

6.3. Publications

Throughout this Thesis, work has been carried out in different workshops and tasks. As a result of this, publications in high impact journals and conferences have been produced (in chronological order):

6.3.1. Journals

1. **Rodrigo Alarcón**, Lourdes Moreno, Isabel Segura Bedmar, Paloma Martínez Fernández. (2019). Lexical simplification approach using easy-to-read resources. Procesamiento del Lenguaje Natural. 63, 95-102. Sociedad Española para el Procesamiento del Lenguaje Natural, 1989-7553. 2019, Septiembre. [SCImago Journal Rankings (SJR), Computer Science Applications, 0,270, 3].

2. **Rodrigo Alarcón**, Lourdes Moreno, Paloma Martínez. (2021). Lexical Simplification System to Improve Web Accessibility. IEEE Access. 9, 58755-58767.

[52] http://163.117.129.208:8080/

[53] https://hulat.inf.uc3m.es/noticia/PlataformaEASIERunaayudaenlacomprensióndelostextos

2169-3536. 2021, Abril. 10.1109/ACCESS.2021.3072697. [Computer Science, Information Systems, 3.745, Q1].

6.3.2. Conferences

1. Lourdes Moreno, **Rodrigo Alarcón**, Isabel Segura Bedmar, Paloma Martínez Fernández. (2019). Lexical simplification approach to support the accessibility guidelines. Proceedings of the XX International Conference on Human Computer Interaction, Interacción 2019. Donostia, Gipuzkoa, Spain. 2019, Junio. ACM, 978-1-4503-7176-6. 14:1-14:4. 10.1145/3335595.3335651.

2. **Rodrigo Alarcón**, Lourdes Moreno, Paloma Martínez Fernández. (2020). Hulat - ALexS CWI task - CWI for Language and Learning Disabilities applied to University Educational Texts. ALexS 2020: Lexicon Analysis Task. IberLEF (Iberian Languages Evaluation Forum) co-located with SEPLN 2020. Málaga, Spain. 2020, Septiembre. CEUR-WS.org, 1613-0073. 2664, 24-30.

3. Lourdes Moreno, **Rodrigo Alarcón**, Paloma Martínez Fernández. (2020). EASIER system. Language resources for cognitive accessibility. 22nd International ACM SIGACCESS Conference on Computers and Accessibility. ASSETS 2020. (Virtual). 2020, Octubre. ACM Digital Library, [GII-GRIN-SCIE, GGS Class 3].

4. **Rodrigo Alarcón**, Lourdes Moreno, Paloma Martínez Fernández. (2020). Word-Sense disambiguation system for text readability. DSAI 2020 (9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion). 2020, Diciembre. ACM Digital Library. 147–152. https://doi.org/10.1145/3439231.3439257.

5. **Rodrigo Alarcón**. (2020). Simplificación Léxica para la Mejora de la Accesibilidad Cognitiva (Lexical Simplification to Improve Cognitive Accessibility). PLNnet-DS-2020 (Proceedings of the Doctoral Symposium on Natural Language Processing). Jaén, Spain. 2020, Diciembre. CEUR-WS.org, 1613-0073. 2802, 1-7.

6. Lourdes Moreno, **Rodrigo Alarcon**,Paloma Martnez. (2021). Designing and Evaluating a User Interface for People with Cognitive Disabilities. Interacción '21.XXI International Conference on Human Computer Interaction. September 22–24, 2021. Málaga, Spain

7. **Rodrigo Alarcón**, Lourdes Moreno, Paloma Martínez. (2021). Exploration of Spanish Word Embeddings for Lexical Simplification. First Workshop on Current Trends in Text Simplification (CTTS-2021) (SEPLN 2021). Málaga, Spain. 2021, Septiembre. CEUR. Vol-2944.

6.3.3. Publications in progress

1. **Rodrigo Alarcón**, Lourdes Moreno, Paloma Martínez. EASIER Corpus: Annotated news with complex words and synonyms to enhance lexical simplification.

2. **Rodrigo Alarcón**, Lourdes Moreno. Experimental study of an application based on NLP to aid to reading for people with cognitive disabilities. Autores:

6.3.4. Participation in Research Projects

1. **DeepEMR Project** (Extracción de información clínica usando deep learning y técnicas de Big Data) with reference TIN2017-87548-C2-1-R, supported by the Research Program of the Ministry of Economy and Competitiveness - Government of Spain.

2. **EASIER Project** funded by INDRA Technologies and the Fundación Universia [54]

3. **ACCESS2MEET Project** (Accesibilidad cognitiva y sensorial a los sistemas de videoconferencia) with reference PID2020-116527RB-I00, supported by the Knowledge Generation and Research Challenges Projects (2020) of the Ministry of Science and Innovation.

6.3.5. Program Committee

- Member of Committee in "First Workshop on Current Trends in Text Simplification (CTTS-2021)"

- Member of Committee in "Workshop on Text Simplification, Accessibility, and Readability (TSAR) - EMNLP 2022"

[54]https://www.tecnologiasaccesibles.com/es/content/proyecto-easier

Bibliography

[1] L. Moreno, R. Alarcon, I. Segura-Bedmar, and P. Martínez, "Lexical simplification approach to support the accessibility guidelines," in *Proceedings of the XX International Conference on Human Computer Interaction*, 2019, pp. 1–4.

[2] OCDE, *Resultados del informe piaac de la ocde.* 2013. [Online]. Available: http://www.educacionyfp.gob.es/prensa/actualidad/2013/10/20131008-piaac.html.

[3] R. Alarcon, "Simplificación léxica para la mejora de la accesibilidad cognitiva," 2020.

[4] UNE, *Guidelines and recommendations for the elaboration of documents, standard, une 153101:2018 ex*, 2018. [Online]. Available: https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0060036.

[5] D.-G. for Translation (European Commission), *How to write clearly.* Mar. 2011. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/c2dab20c-0414-408d-87b5-dd3c6e5dd9a5.

[6] W3C, *Web Content Accesibility Guidelines (WCAG)*, 2019. [Online]. Available: https://www.w3.org/WAI/standards-guidelines/wcag/.

[7] P. Sikka and V. Mago, *A survey on text simplification*, 2020. arXiv: 2008.08612 [cs.CL].

[8] G. H. Paetzold and L. Specia, "A survey on lexical simplification," *Journal of Artificial Intelligence Research*, vol. 60, pp. 549–593, 2017.

[9] E. Barbu, M. T. Martín-Valdivia, E. Martínez-Cámara, and L. A. Urena-López, "Language technologies applied to document simplification for helping autistic people," *Expert Systems with Applications*, vol. 42, no. 12, pp. 5076–5086, 2015.

[10] S. Devlin and G. Unthank, "Helping aphasic people process online information," in *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, 2006, pp. 225–226.

[11] L. Sauvan *et al.*, "Text simplification to help individuals with low vision read more fluently," in *LREC 2020-Language Resources and Evaluation Conference*, 2020, pp. 11–16.

[12] R. Wilkens, B. Oberle, and A. Todirascu, "Coreference-based text simplification," in *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI)*, 2020, pp. 93–100.

[13] H. Saggion, "Automatic text simplification," *Synthesis Lectures on Human Language Technologies*, vol. 10, no. 1, pp. 1–137, 2017.

[14] J. Du, F. Qi, and M. Sun, "Using bert for word sense disambiguation," *arXiv preprint arXiv:1909.08358*, 2019.

[15] S. S. Al-Thanyyan and A. M. Azmi, "Automated text simplification: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–36, 2021.

[16] P. Gupta and M. Jaggi, "Obtaining better static word embeddings using contextual embedding models," *CoRR*, vol. abs/2106.04302, 2021. arXiv: 2106.04302. [Online]. Available: https://arxiv.org/abs/2106.04302.

[17] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, IGI global, 2010, pp. 242–264.

[18] J. McCarthy, "What is artificial intelligence?," 2007.

[19] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," *arXiv preprint arXiv:1708.05148*, 2017.

[20] T. Mitchell, "Introduction to machine learning," *Machine learning*, vol. 7, pp. 2–5, 1997.

[21] J. M. Hilbe, *Logistic regression models*. Chapman and hall/CRC, 2009.

[22] S. C. Bagley, H. White, and B. A. Golomb, "Logistic regression in the medical literature:: Standards for use and reporting, with particular attention to one medical domain," *Journal of clinical epidemiology*, vol. 54, no. 10, pp. 979–985, 2001.

[23] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *The annals of mathematical statistics*, vol. 41, no. 1, pp. 164–171, 1970.

[24] L. Rabiner and B. Juang, "An introduction to hidden markov models," *ieee assp magazine*, vol. 3, no. 1, pp. 4–16, 1986.

[25] A. McCallum, D. Freitag, and F. C. Pereira, "Maximum entropy markov models for information extraction and segmentation.," in *Icml*, vol. 17, 2000, pp. 591–598.

[26] H. M. Wallach, "Conditional random fields: An introduction," *Technical Reports (CIS)*, p. 22, 2004.

[27] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*, Springer, 1998, pp. 137–142.

[28] M. A. Aizerman, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and remote control*, vol. 25, pp. 821–837, 1964.

[29] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[30] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[31] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1746–1751. DOI: 10.3115/v1/D14-1181. [Online]. Available: https://aclanthology.org/D14-1181.

[32] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[33] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. DOI: 10.3115/v1/D14-1162. [Online]. Available: https://aclanthology.org/D14-1162.

[34] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017. DOI: 10.1162/tacl_a_00051. [Online]. Available: https://aclanthology.org/Q17-1010.

[35] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[36] M. E. Peters *et al.*, "Deep contextualized word representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 2227–2237. DOI: 10.18653/v1/N18-1202. [Online]. Available: https://aclanthology.org/N18-1202.

[37] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[38] F. Paternò and A. G. Schiavone, "The role of tool support in public policies and accessibility," *interactions*, vol. 22, no. 3, pp. 60–63, 2015.

[39] WAI-W3C, *Diverse abilities and barriers*. [Online]. Available: https://www.w3.org/WAI/people-use-web/abilities-barriers/.

[40] ONU, *Convention on the rights of persons with disabilities*, 2006. [Online]. Available: https://www.un.org/esa/socdev/enable/documents/tccconvs.pdf.

[41] BOE, *Directiva (ue) 2016/2102 del parlamento europeo y del consejo de 26 de octubre de 2016 sobre la accesibilidad de los sitios web y aplicaciones para dispositivos móviles de los organismos del sector público.* 2016. [Online]. Available: https://www.boe.es/doue/2016/327/L00001-00015.pdf.

[42] ——, *Real decreto 1112/2018, de 7 de septiembre, sobre accesibilidad de los sitios web y aplicaciones para dispositivos móviles del sector público.* 2018. [Online]. Available: https://www.boe.es/eli/es/rd/2018/09/07/1112.

[43] ETSI-CEN-CENELET, *En 301 549 v3.2.1 (2021–03). accessibility requirements for ict products and services.* 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_en/301500_301599/301549/03.02.01_60/en_301549v030201p.pdf.

[44] P.A.E., *Disponible la norma une-en 301549:2019 en el pae. administracionelectronica.gob.es.* 2019. [Online]. Available: https://administracionelectronica.gob.es/general/error.htm;jsessionid=F41965C8BA5C164E8093AADD30C5C65D.node2_paeaplic.

[45] W3C, *Web content accesibility guidelines (wcag)*, 2019. [Online]. Available: https://www.w3.org/WAI/standards-guidelines/wcag/.

[46] WAI-W3C, *Web content accessibility guidelines (wcag) 2.1 w3c recommendation 05 june 2018. www.w3.org.* 2018. [Online]. Available: https://www.w3.org/TR/WCAG21/.

[47] ——, *Web content accessibility guidelines (wcag) 2.2.* 2021. [Online]. Available: https://www.w3.org/TR/WCAG22/.

[48] ——, *Cognitive and Learning Disabilities Accessibility Task Force (Coga TF) Work Statement. www.w3.org/WAI.* 2021. [Online]. Available: https://www.w3.org/WAI/GL/task-forces/coga/work-statement.

[49] Plainlanguage-gov, *Plain language action and information network (plain).* [Online]. Available: https://www.plainlanguage.gov/about/.

[50] D.-G. for Translation (European Commission), *How to write clearly. op.europa.eu.* 2016. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/725b7eb0-d92e-11e5-8fea-01aa75ed71a1/language-en.

[51] K. Smith, G. Hallam, and S. Ghosh, “Guidelines for professional library/information educational programs-2012,” *IFLA Education and Training Section, IFLA, The Hague*, 2012. [Online]. Available: www.ifla.org/publications/guidelinesfor-professionallibraryinformationeducational-programs-2012.

[52] Plainlanguage-gov, *Plain english- free guides (co.uk)*, 2017. [Online]. Available: http://www.plainenglish.co.uk/free-guides.html.

[53] M. Shardlow, "A survey of automated text simplification," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 1, pp. 58–70, 2014.

[54] P. Sikka, M. Singh, A. Pink, and V. Mago, "A Survey on Text Simplification," vol. 37, no. 4, 2020. arXiv: arXiv:2008.08612v1.

[55] A. Siddharthan, "Syntactic simplification and text cohesion," *Research on Language and Computation*, vol. 4, no. 1, pp. 77–109, 2006.

[56] H. Saggion, S. Štajner, S. Bott, S. Mille, L. Rello, and B. Drndarevic, "Making it simplext: Implementation and evaluation of a text simplification system for spanish," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 6, no. 4, pp. 1–36, 2015.

[57] Y. Sen and Y. Fuping, "Chinese automatic text simplification based on unsupervised learning," in *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, IEEE, vol. 5, 2021, pp. 1666–1672.

[58] G. Paetzold and L. Specia, "Unsupervised lexical simplification for non-native speakers," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, 2016.

[59] E. Barbu, M. T. Martín-Valdivia, and L. A. U. Lopez, "Open book: A tool for helping asd users' semantic comprehension," in *Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility*, 2013, pp. 11–19.

[60] R. Evans, C. Orasan, and I. Dornescu, "An evaluation of syntactic simplification rules for people with autism," Association for Computational Linguistics, 2014.

[61] J. Carroll, G. Minnen, Y. Canning, S. Devlin, and J. Tait, "Practical simplification of english newspaper text to assist aphasic readers," in *Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, 1998, pp. 7–10.

[62] L. Rello, R. Baeza-Yates, S. Bott, and H. Saggion, "Simplify or help? text simplification strategies for people with dyslexia," in *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, 2013, pp. 1–10.

[63] H. Saggion, E. Gómez-Martínez, E. Etayo, A. Anula, and L. Bourg, "Text simplification in simplext: Making texts more accessible," *Procesamiento del lenguaje natural*, no. 47, pp. 341–342, 2011.

[64] S. Bott, L. Rello, B. Drndarević, and H. Saggion, "Can spanish be simpler? lexsis: Lexical simplification for spanish," in *Proceedings of COLING 2012*, 2012, pp. 357–374.

[65] R. Alarcon, L. Moreno, and P. Martínez, "Lexical simplification system to improve web accessibility," *IEEE Access*, vol. 9, pp. 58 755–58 767, 2021. DOI: 10.1109/ACCESS.2021.3072697.

[66] M. Rivero-Contreras, P. E. Engelhardt, and D. Saldaña, "An experimental eye-tracking study of text adaptation for readers with dyslexia: Effects of visual support and word frequency," *Annals of Dyslexia*, vol. 71, no. 1, pp. 170–187, 2021.

[67] M. Alfano, B. Lenzitti, G. Lo Bosco, C. Muriana, T. Piazza, and G. Vizzini, "Design, development and validation of a system for automatic help to medical text understanding," *International Journal of Medical Informatics*, vol. 138, p. 104 109, 2020. DOI: https://doi.org/10.1016/j.ijmedinf.2020.104109. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1386505618309419.

[68] O. Alonzo, M. Seita, A. Glasser, and M. Huenerfauth, "Automatic text simplification tools for deaf and hard of hearing adults: Benefits of lexical simplification and providing users with autonomy," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.

[69] W. M. Watanabe, A. C. Junior, V. R. Uzêda, R. P. d. M. Fortes, T. A. S. Pardo, and S. M. Aluísio, "Facilita: Reading assistance for low-literacy readers," in *Proceedings of the 27th ACM international conference on Design of communication*, 2009, pp. 29–36.

[70] M. Shardlow, "A comparison of techniques to automatically identify complex words.," in *51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop*, 2013, pp. 103–109.

[71] S. Štajner, I. Calixto, and H. Saggion, "Automatic text simplification for spanish: Comparative evaluation of various simplification strategies," in *Proceedings of the international conference recent advances in natural language processing*, 2015, pp. 618–626.

[72] S. Surya, A. Mishra, A. Laha, P. Jain, and K. Sankaranarayanan, "Unsupervised neural text simplification," *arXiv preprint arXiv:1810.07931*, 2018.

[73] J. Qiang, Y. Li, Y. Zhu, Y. Yuan, and X. Wu, "Lexical simplification with pretrained encoders," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 8649–8656.

[74] D. Ferrés, H. Saggion, and X. G. Guinovart, "An adaptable lexical simplification architecture for major ibero-romance languages," in *Proceedings of the first workshop on building linguistically generalizable NLP systems*, 2017, pp. 40–47.

[75] S. Devlin, "The use of a psycholinguistic database in the simplification of text for aphasic readers," *Linguistic databases*, 1998.

[76] G. Leroy, J. E. Endicott, D. Kauchak, O. Mouradi, and M. Just, "User evaluation of the effects of a text simplification algorithm using term familiarity on perception, understanding, learning, and information retention," *Journal of medical Internet research*, vol. 15, no. 7, e144, 2013.

[77] R. Keskisärkkä, *Automatic text simplification via synonym replacement*, 2012.

[78] M. Shardlow, "Out in the open: Finding and categorising errors in the lexical simplification pipeline.," in *LREC*, 2014, pp. 1583–1590.

[79] L. Deléger and P. Zweigenbaum, "Extracting lay paraphrases of specialized expressions from monolingual comparable medical corpora," in *Proceedings of the 2nd Workshop on Building and Using Comparable Corpora: from Parallel to Non-parallel Corpora (BUCC)*, 2009, pp. 2–10.

[80] N. Elhadad, "Comprehending technical texts: Predicting and defining unfamiliar terms," in *AMIA annual symposium proceedings*, American Medical Informatics Association, vol. 2006, 2006, p. 239.

[81] N. Elhadad and K. Sutaria, "Mining a lexicon of technical terms and lay equivalents," in *Biological, translational, and clinical language processing*, 2007, pp. 49–56.

[82] R. Alarcon, L. Moreno Lopez, I. Segura-Bedmar, and P. Martínez Fernández, "Lexical simplification approach using easy-to-read resources," 2019.

[83] O. Biran, S. Brody, and N. Elhadad, "Putting it simply: A context-aware approach to lexical simplification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, pp. 496–501.

[84] C. Horn, C. Manduca, and D. Kauchak, "Learning a lexical simplifier using wikipedia," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2014, pp. 458–463.

[85] Z. Zhu, D. Bernhard, and I. Gurevych, "A monolingual tree-based translation model for sentence simplification," in *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, 2010, pp. 1353–1361.

[86] W. Xu, C. Napoles, E. Pavlick, Q. Chen, and C. Callison-Burch, "Optimizing statistical machine translation for text simplification," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 401–415, 2016.

[87] G. Paetzold and L. Specia, "Semeval 2016 task 11: Complex word identification," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 560–569.

[88] ——, "Sv000gg at semeval-2016 task 11: Heavy gauge complex word identification with system voting," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 969–974.

[89] O. Kuru, "AI-KU at SemEval-2016 task 11: Word embeddings and substring features for complex word identification," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, San Diego, California: Association for Computational Linguistics, Jun. 2016, pp. 1042–1046. DOI: 10.18653/v1/S16-1163. [Online]. Available: https://aclanthology.org/S16-1163.

[90] M. Zampieri, L. Tan, and J. Genabith, "Macsaar at semeval-2016 task 11: Zipfian and character features for complex word identification," Jan. 2016. DOI: 10.18653/v1/S16-1155.

[91] K. Wróbel, "PLUJAGH at SemEval-2016 task 11: Simple system for complex word identification," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, San Diego, California: Association for Computational Linguistics, Jun. 2016, pp. 953–957. DOI: 10.18653/v1/S16-1146. [Online]. Available: https://aclanthology.org/S16-1146.

[92] S. M. Yimam, S. Stajner, M. Riedl, and C. Biemann, "Multilingual and cross-lingual complex word identification.," in *RANLP*, 2017, pp. 813–822.

[93] S. M. Yimam *et al.*, "A report on the complex word identification shared task 2018," *arXiv preprint arXiv:1804.09132*, 2018.

[94] N. Hartmann and L. B. dos Santos, "NILC at CWI 2018: Exploring feature engineering and feature learning," in *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 335–340. DOI: 10.18653/v1/W18-0540. [Online]. Available: https://aclanthology.org/W18-0540.

[95] A. Butnaru and R. T. Ionescu, "Unibuckernel: A kernel-based learning method for complex word identification," Mar. 2018.

[96] D. Alfter and I. Pilán, "SB@GU at the complex word identification 2018 shared task," in *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 315–321. DOI: 10.18653/v1/W18-0537. [Online]. Available: https://aclanthology.org/W18-0537.

[97] S. Gooding and E. Kochmar, "CAMB at CWI shared task 2018: Complex word identification with ensemble-based voting," in *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 184–194. DOI: 10.18653/v1/W18-0520. [Online]. Available: https://aclanthology.org/W18-0520.

[98] T. Kajiwara and M. Komachi, “Complex word identification based on frequency in a learner corpus,” in *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 195–199. DOI: 10.18653/v1/W18-0521. [Online]. Available: https://aclanthology.org/W18-0521.

[99] S. T. Aroyehun, J. Angel, D. A. Pérez Alvarez, and A. Gelbukh, “Complex word identification: Convolutional neural network vs. feature engineering,” in *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 322–327. DOI: 10.18653/v1/W18-0538. [Online]. Available: https://aclanthology.org/W18-0538.

[100] J. A. Ortiz-Zambranoa and A. Montejo-Ráezb, “Overview of alexs 2020: First workshop on lexical analysis at sepln,” 2020.

[101] A. R. Sulayes, “General lexicon-based complex word identification extended with stem n-grams and morphological engines,” in *IberLEF@SEPLN*, 2020.

[102] “Vicomtech at alexs 2020: Unsupervised complex word identification based on domain frequency,” pp. 7–14, 2020.

[103] R. Alarcon, L. Moreno, and P. Martínez, “Hulat - alexs cwi task - cwi for language and learning disabilities applied to university educational texts,” 2020.

[104] J. Bingel, G. Paetzold, and A. Søgaard, “Lexi: A tool for adaptive, personalized text simplification,” in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 245–258.

[105] A. Palmero Aprosio, S. Menini, and S. Tonelli, “Adaptive complex word identification through false friend detection,” in *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 2020, pp. 192–200.

[106] J. Burstein, J. Shore, J. Sabatini, Y.-W. Lee, and M. Ventura, “The automated text adaptation tool,” in *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, 2007, pp. 3–4.

[107] R. Navigli and S. P. Ponzetto, “Babelnet: Building a very large multilingual semantic network,” in *Proceedings of the 48th annual meeting of the association for computational linguistics*, 2010, pp. 216–225.

[108] S. Aluísio and C. Gasperin, “Fostering digital inclusion and accessibility: The porsimples project for simplification of portuguese texts,” in *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, 2010, pp. 46–53.

[109] E. G. Maziero, T. A. Pardo, A. Di Felippo, and B. C. Dias-da-Silva, “A base de dados lexical ea interface web do tep 2.0: Thesaurus eletrónico para o portugués do brasil,” in *Companion Proceedings of the XIV Brazilian Symposium on Multimedia and the Web*, 2008, pp. 390–392.

[110] T. Kajiwara, H. Matsumoto, and K. Yamamoto, “Selecting proper lexical paraphrase for children,” in *Proceedings of the 25th Conference on Computational Linguistics and Speech Processing (ROCLING 2013)*, Kaohsiung, Taiwan: The Association for Computational Linguistics and Chinese Language Processing (ACLCLP), Oct. 2013, pp. 59–73. [Online]. Available: https://aclanthology.org/O13-1007.

[111] N. Elhadad and K. Sutaria, “Mining a lexicon of technical terms and lay equivalents,” in *Biological, translational, and clinical language processing*, Prague, Czech Republic: Association for Computational Linguistics, Jun. 2007, pp. 49–56. [Online]. Available: https://aclanthology.org/W07-1007.

[112] L. Deléger and P. Zweigenbaum, “Extracting lay paraphrases of specialized expressions from monolingual comparable medical corpora,” in *Proceedings of the 2nd Workshop on Building and Using Comparable Corpora: from Parallel to Non-parallel Corpora (BUCC)*, Singapore: Association for Computational Linguistics, Aug. 2009, pp. 2–10. [Online]. Available: https://aclanthology.org/W09-3102.

[113] M. Yatskar, B. Pang, C. Danescu-Niculescu-Mizil, and L. Lee, “For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia,” in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles, California: Association for Computational Linguistics, Jun. 2010, pp. 365–368. [Online]. Available: https://aclanthology.org/N10-1056.

[114] D. Feblowitz and D. Kauchak, “Sentence simplification as tree transduction,” in *Proceedings of the second workshop on predicting and improving text readability for target reader populations*, 2013, pp. 1–10.

[115] G. Paetzold and L. Specia, “Text simplification as tree transduction,” in *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, 2013.

[116] E. Pavlick and C. Callison-Burch, “Simple ppdb: A paraphrase database for simplification,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 143–148.

[117] G. Glavaš and S. Štajner, “Simplifying lexical simplification: Do we need simplified corpora?” In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Beijing, China: Association for

Computational Linguistics, Jul. 2015, pp. 63–68. DOI: 10.3115/v1/P15-2011. [Online]. Available: https://aclanthology.org/P15-2011.

[118] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," in *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*, 2013, pp. 746–751.

[119] J. Qiang, Y. Li, Y. Zhu, Y. Yuan, Y. Shi, and X. Wu, "Lsbert: Lexical simplification based on bert," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3064–3076, 2021. DOI: 10.1109/TASLP.2021.3111589.

[120] S. R. Thomas and S. Anderson, "Wordnet-based lexical simplification of a document," in *Proceedings of KONVENS 2012*, J. Jancsary, Ed., Main track: oral presentations, ÖGAI, Sep. 2012, pp. 80–88. [Online]. Available: http://www.oegai.at/konvens2012/proceedings/13_thomas12o/.

[121] B. P. Nunes, R. Kawase, P. Siehndel, M. A. Casanova, and S. Dietze, "As simple as it gets - a sentence simplifier for different learning levels and contexts," in *2013 IEEE 13th International Conference on Advanced Learning Technologies*, 2013, pp. 128–132. DOI: 10.1109/ICALT.2013.42.

[122] J. De Belder and M.-F. Moens, "Text simplification for children," Jan. 2010.

[123] G. Paetzold and L. Specia, "Lexenstein: A framework for lexical simplification," in *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, 2015, pp. 85–90.

[124] R. Alarcón, L. Moreno, and P. Martínez, "Exploration of spanish word embeddings for lexical simplification," 2021.

[125] G. Paetzold and L. Specia, "Lexical simplification with neural ranking," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 2017, pp. 34–40.

[126] A. P. Rudell, "Frequency of word usage and perceived word difficulty: Ratings of kucera and francis words," *Behavior Research Methods, Instruments, & Computers*, vol. 25, pp. 455–463, 1993.

[127] M. Brysbaert and B. New, "Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english," *Behavior research methods*, vol. 41, no. 4, pp. 977–990, 2009.

[128] A.-L. Ligozat, C. Grouin, A. Garcia-Fernandez, and D. Bernhard, "Annlor: A naive notation-system for lexical outputs ranking," in *First Joint Conference on Lexical and Computational Semantics (* SEM)*, 2012, pp. 487–492.

[129] R. Baeza-Yates, L. Rello, and J. Dembowski, "Cassa: A context-aware synonym simplification algorithm," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015, pp. 1380–1385.

[130] L. Specia, S. K. Jauhar, and R. Mihalcea, “Semeval-2012 task 1: English lexical simplification,” in ** SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, 2012, pp. 347–355.

[131] L. Rello, R. Baeza-Yates, L. Dempere-Marco, and H. Saggion, “Frequent words improve readability and short words improve understandability for people with dyslexia,” in *IFIP Conference on Human-Computer Interaction*, Springer, 2013, pp. 203–219.

[132] R. Sinha, “UNT-SimpRank: Systems for lexical simplification ranking,” in **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, Montréal, Canada: Association for Computational Linguistics, Jul. 2012, pp. 493–496. [Online]. Available: https://aclanthology.org/S12-1069.

[133] S. K. Jauhar and L. Specia, “Uow-shef: Simplex–lexical simplicity ranking based on contextual and psycholinguistic features,” in ** SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, 2012, pp. 477–481.

[134] M. Maddela and W. Xu, “A word-complexity lexicon and a neural readability ranking model for lexical simplification,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium: Association for Computational Linguistics, Oct. 2018, pp. 3749–3760. DOI: 10.18653/v1/D18-1410. [Online]. Available: https://aclanthology.org/D18-1410.

[135] M. Amoia and M. Romanelli, “Sb: Mmsystem-using decompositional semantics for lexical simplification,” in ** SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, 2012, pp. 482–486.

[136] M. Shardlow, R. Evans, G. H. Paetzold, and M. Zampieri, “SemEval-2021 task 1: Lexical complexity prediction,” in *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, Online: Association for Computational Linguistics, Aug. 2021, pp. 1–16. DOI: 10.18653/v1/2021.semeval-1.1. [Online]. Available: https://aclanthology.org/2021.semeval-1.1.

[137] C. Pan, B. Song, S. Wang, and Z. Luo, “DeepBlueAI at SemEval-2021 task 1: Lexical complexity prediction with a deep ensemble approach,” in *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, On-

line: Association for Computational Linguistics, Aug. 2021, pp. 578–584. DOI: 10.18653/v1/2021.semeval-1.72. [Online]. Available: https://aclanthology.org/2021.semeval-1.72.

[138] T. Bani Yaseen, Q. Ismail, S. Al-Omari, E. Al-Sobh, and M. Abdullah, "JUST-BLUE at SemEval-2021 task 1: Predicting lexical complexity using BERT and RoBERTa pre-trained language models," in *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, Online: Association for Computational Linguistics, Aug. 2021, pp. 661–666. DOI: 10.18653/v1/2021.semeval-1.85. [Online]. Available: https://aclanthology.org/2021.semeval-1.85.

[139] G. Rao, M. Li, X. Hou, L. Jiang, Y. Mo, and J. Shen, "RG PA at SemEval-2021 task 1: A contextual attention-based model with RoBERTa for lexical complexity prediction," in *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, Online: Association for Computational Linguistics, Aug. 2021, pp. 623–626. DOI: 10.18653/v1/2021.semeval-1.79. [Online]. Available: https://aclanthology.org/2021.semeval-1.79.

[140] A. Mosquera, "Alejandro mosquera at SemEval-2021 task 1: Exploring sentence and word features for lexical complexity prediction," in *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, Online: Association for Computational Linguistics, Aug. 2021, pp. 554–559. DOI: 10.18653/v1/2021.semeval-1.68. [Online]. Available: https://aclanthology.org/2021.semeval-1.68.

[141] A. Rotaru, "ANDI at SemEval-2021 task 1: Predicting complexity in context using distributional models, behavioural norms, and lexical resources," in *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, Online: Association for Computational Linguistics, Aug. 2021, pp. 655–660. DOI: 10.18653/v1/2021.semeval-1.84. [Online]. Available: https://aclanthology.org/2021.semeval-1.84.

[142] M. Lesk, "Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone," in *Proceedings of the 5th annual international conference on Systems documentation*, 1986, pp. 24–26.

[143] Y. K. Lee, H. T. Ng, and T. K. Chia, "Supervised word sense disambiguation with support vector machines and multiple knowledge sources," in *Proceedings of SENSEVAL-3, the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, 2004, pp. 137–140.

[144] H. Kaji and Y. Morimoto, "Unsupervised word-sense disambiguation using bilingual comparable corpora," *IEICE transactions on information and systems*, vol. 88, no. 2, pp. 289–301, 2005.

[145] B. Moradi, E. Ansari, and Z. Žabokrtský, "Unsupervised word sense disambiguation using word embeddings," in *2019 25th Conference of Open Innovations Association (FRUCT)*, 2019, pp. 228–233. DOI: 10.23919/FRUCT48121.2019.8981526.

[146] R. Cao, J. Bai, W. Ma, and H. Shinnou, "Semi-supervised learning for all-words wsd using self-learning and fine-tuning," in *Proceedings of the 33rd Pacific Asia Conference on Language, Information and Computation*, Waseda Institute for the Study of Language and Information, 2019, pp. 356–361.

[147] S. Pradhan, E. Loper, D. Dligach, and M. Palmer, "SemEval-2007 task-17: English lexical sample, SRL and all words," in *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic: Association for Computational Linguistics, Jun. 2007, pp. 87–92. [Online]. Available: https://aclanthology.org/S07-1016.

[148] R. Navigli, D. Jurgens, and D. Vannella, "Semeval-2013 task 12: Multilingual word sense disambiguation," in *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, 2013, pp. 222–231.

[149] A. Moro and R. Navigli, "Semeval-2015 task 13: Multilingual all-words sense disambiguation and entity linking," in *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, 2015, pp. 288–297.

[150] L. SEVENS, J. J. O'FLAHERTY, and A. VLIEGHER, "Towards integrating people with intellectual disabilities in the digital world," in *Intelligent Environments 2016: Workshop Proceedings of the 12th International Conference on Intelligent Environments (2016)*, vol. 348, 2016.

[151] H. Al-Mubaid and P. Chen, "Application of word prediction and disambiguation to improve text entry for people with physical disabilities (assistive technology)," *International Journal of Social and Humanistic Computing*, vol. 1, no. 1, pp. 10–27, 2008.

[152] L. Sevens, G. Jacobs, V. Vandeghinste, I. Schuurman, and F. Van Eynde, "Improving text-to-pictograph translation through word sense disambiguation," in *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, 2016, pp. 131–135.

[153] J. Pustejovsky and A. Stubbs, *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. " O'Reilly Media, Inc.", 2012.

[154] S. Sharoff, "Open-source corpora: Using the net to fish for linguistic data," *International journal of corpus linguistics*, vol. 11, no. 4, pp. 435–462, 2006.

[155] J. De Belder and M.-F. Moens, "A dataset for the evaluation of lexical simplification," in *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, 2012, pp. 426–437.

[156] G. Paetzold and L. Specia, "Benchmarking lexical simplification systems," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 3074–3080.

[157] D. Kauchak, "Improving text simplification language modeling using unsimplified text data," in *Proceedings of the 51st annual meeting of the association for computational linguistics (volume 1: Long papers)*, 2013, pp. 1537–1546.

[158] T. Kajiwara and M. Komachi, "Building a monolingual parallel corpus for text simplification using sentence similarity based on alignment between word embeddings," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 1147–1158.

[159] W. Xu, C. Callison-Burch, and C. Napoles, "Problems in current text simplification research: New data can help," *Transactions of the Association for Computational Linguistics*, vol. 3, pp. 283–297, 2015.

[160] X. Zhang and M. Lapata, "Sentence simplification with deep reinforcement learning," *arXiv preprint arXiv:1703.10931*, 2017.

[161] K. Woodsend and M. Lapata, "Learning to simplify sentences with quasi-synchronous grammar and integer programming," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, UK.: Association for Computational Linguistics, Jul. 2011, pp. 409–420. [Online]. Available: https://aclanthology.org/D11-1038.

[162] S. Štajner, H. Saggion, and S. P. Ponzetto, "Improving lexical coverage of text simplification systems for spanish," *Expert Systems with Applications*, vol. 118, pp. 80–91, 2019.

[163] F. Alva-Manchego, L. Martin, A. Bordes, C. Scarton, B. Sagot, and L. Specia, "Asset: A dataset for tuning and evaluation of sentence simplification models with multiple rewriting transformations," *arXiv preprint arXiv:2005.00481*, 2020.

[164] C. Cardellino, *Spanish Billion Words Corpus and Embeddings*, Aug. 2019. [Online]. Available: https://crscardellino.github.io/SBWCE/.

[165] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, *Learning word vectors for 157 languages*, 2018. arXiv: 1802.06893 [cs.CL].

[166] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez, "Spanish pre-trained bert model and evaluation data," in *PML4DC at ICLR 2020*, 2020.

[167] I. Segura-Bedmar, C. Colón-Ruíz, M. Á. Tejedor-Alonso, and M. Moro-Moro, "Predicting of anaphylaxis in big data emr by exploring machine learning approaches," *Journal of Biomedical Informatics*, vol. 87, pp. 50–59, 2018. DOI: https://doi.org/10.1016/j.jbi.2018.09.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046418301874.

[168] J. Kivinen, M. K. Warmuth, and P. Auer, “The perceptron algorithm versus winnow: Linear versus logarithmic mistake bounds when few input variables are relevant,” *Artificial Intelligence*, vol. 97, no. 1-2, pp. 325–343, 1997.

[169] D. De Hertog and A. Tack, “Deep learning architecture for complexword identification,” in *Thirteenth Workshop of Innovative Use of NLP for Building Educational Applications*, Association for Computational Linguistics (ACL); New Orleans, Louisiana, 2018, pp. 328–334.

[170] L. Specia, S. K. Jauhar, and R. Mihalcea, “SemEval-2012 task 1: English lexical simplification,” in **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, Montréal, Canada: Association for Computational Linguistics, Jul. 2012, pp. 347–355. [Online]. Available: https://aclanthology.org/S12-1046.

[171] R. Alarcon, L. Moreno, and P. Martínez, “Word-sense disambiguation system for text readability,” in *9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, 2020, pp. 147–152.

[172] G. Paetzold and L. Specia, “Benchmarking lexical simplification systems,” in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, Portorož, Slovenia: European Language Resources Association (ELRA), May 2016, pp. 3074–3080. [Online]. Available: https://aclanthology.org/L16-1491.

[173] L. Seeman and C. Lewis, “Cognitive and learning disabilities,” in *Web Accessibility*, Springer, 2019, pp. 49–58.

[174] L. Moreno, R. Alarcon, and P. Martínez, “Designing and evaluating a user interface for people with cognitive disabilities,” in *Proceedings of the XXI International Conference on Human Computer Interaction*, 2021, pp. 1–8.

[175] ——, “Easier system. language resources for cognitive accessibility.,” in *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, 2020, pp. 1–3.

Appendix A

Thesis Proposal - User Evaluation: First Task (CWI)

Usuario Número:
Fecha

TAREA 1

Subraya o rodea con un círculo las palabras que consideres complejas o difíciles de los siguientes textos.

[Texto 1]
Con motivo de las medidas establecidas por la crisis del Covid-19, el comercio madrileño caracterizado por su amplitud y diversificación, está sometido a limitaciones de aforo.

[Texto 2]
El objetivo es facilitar la accesibilidad, mejorar la atención desde los Centros de Salud, y mantener al ciudadano en su domicilio hasta la primera valoración, reduciendo los desplazamientos innecesarios.

[Texto 3]
Se ofertarán en total más de 7300 plazas gratuitas para las visitas guiadas y para los ocho conciertos del ciclo flamenco en palacio: una mirada diversa al patrimonio jondo, que reivindica el flamenco como legado patrimonial y cultural.

[Texto 4]
Conoce todas las actividades que se realizan en las bibliotecas: talleres de escritura, clubes de lectura, cuentacuentos, exposiciones, presentaciones de libros, itinerarios culturales y proyecciones.

[Texto 5]
Para que una familia goce de los beneficios que tiene ser familia numerosa es necesario que esta condición sea reconocida mediante un título oficial. Este título se concede, previa solicitud.

[Texto 6]
La caza y la pesca en la Comunidad de Madrid están sujetas a regulación especial y requieren una serie de trámites, entre ellos la obtención de licencia. Además, los nuevos cazadores deben superar un examen o prueba de aptitud.

[Texto 7]
La violencia contra las mujeres es una vulneración de los derechos humanos que no respeta fronteras geográficas, culturales ni económicas.

[Texto 8]
Cuando una mujer está inmersa en un proceso de maltrato, le es difícil discernir entre si realmente lo es o no.

[Texto 9]
Si tiene una alergia o una intolerancia alimentaria, es importante que planifique su viaje de forma anticipada y tome algunas precauciones para disfrutar de su estancia en el extranjero.

[Texto 10]
La contaminación de ambientes interiores de los inmuebles es un factor determinante en la salud y bienestar de sus usuarios. Como novedad, la presente guía aporta un modelo de gestión integral de la Sanidad Ambiental.

[Texto 11]
El edadismo tiene consecuencias graves y amplias para la salud y el bienestar de las personas. El edadismo se produce cuando la edad se utiliza para categorizar y dividir a las personas provocando injusticias como prejuicios, discriminación y estereotipos.

[Texto 12]
Pautas de desinfección de superficies y espacios habitados por casos en investigación, cuarentena, probables o confirmados de COVID-19, viviendas, residencias, espacios de pública concurrencia y transportes de viajeros.

[Texto 13]
Está dirigido a todas las entidades que intervienen en el proceso de abastecimiento y control de la calidad de las aguas de consumo: gestores del agua, ayuntamientos, técnicos sanitarios, personal de mantenimiento, manipuladores de aguas, etc.

[Texto 14]
La finalidad de esta guía es proporcionar a los gestores de los abastecimientos de agua de consumo humano una herramienta de apoyo para la elaboración del protocolo de autocontrol requerido por la normativa vigente.

[Texto 15]
Este documento indica los requisitos esenciales que deben tenerse en cuenta en la elaboración de los pliegos de prescripciones técnicas para la contratación pública del Servicio de mantenimiento dirigido a la prevención y control de la legionela.

Appendix B

Thesis Proposal - User Evaluation: Second Task (Substitutes)

Usuario Número:
Fecha

TAREA 2

Lee despacio los siguientes textos. Como puedes ver en cada texto hay una palabra subrayada y resaltada. Después de cada texto, responde a la pregunta relacionada con esa palabra.

[Texto 1] Cuando una mujer está inmersa en un proceso de maltrato, le es difícil **<u>discernir</u>** entre si realmente lo es o no.

Si sustituimos la palabra "**<u>discernir</u>**" por las siguientes palabras, ¿lees y comprendes mejor el texto?

- distinguir ☐ SI ☐ NO
- determinar ☐ SI ☐ NO

[Texto 2] El objetivo es facilitar la accesibilidad, mejorar la atención desde los Centros de Salud, y mantener al ciudadano en su domicilio hasta la primera **<u>valoración</u>**, reduciendo los desplazamientos innecesarios.

Si sustituimos la palabra “**<u>valoración</u>**” por las siguientes palabras, ¿lees y comprendes mejor el texto?

- evaluación ☐ SI ☐ NO
- apreciación ☐ SI ☐ NO

[Texto 3]
El objetivo es facilitar la accesibilidad, mejorar la atención desde los Centros de Salud, y mantener al ciudadano en su domicilio hasta la primera valoración, reduciendo los **desplazamientos** innecesarios.

Si sustituimos la palabra “**desplazamientos**” por las siguientes palabras, ¿lees y comprendes mejor el texto?

- viajes ☐ SI ☐ NO
- traslados ☐ SI ☐ NO

[Texto 4]
La contaminación de **ambientes** interiores de los inmuebles es un factor determinante en la salud y bienestar de sus usuarios.

Si sustituimos la palabra “**ambientes**” por las siguientes palabras, ¿lees y comprendes mejor el texto?

- lugares ☐ SI ☐ NO
- entornos ☐ SI ☐ NO

[Texto 5]
La contaminación de ambientes interiores de los **inmuebles** es un factor determinante en la salud y bienestar de sus usuarios.

Si sustituimos la palabra “**inmuebles**” por las siguientes palabras, ¿lees y comprendes mejor el texto?

- edificios ☐ SI ☐ NO
- bienes ☐ SI ☐ NO

[Texto 6]
La contaminación de ambientes interiores de los inmuebles es un factor determinante en la salud y **bienestar** de sus usuarios.

Si sustituimos la palabra "**bienestar**" por las siguientes palabras, ¿lees y comprendes mejor el texto?

- salud ☐ SI ☐ NO
- sanidad ☐ SI ☐ NO

[Texto 7]
Como novedad, la presente guía **aporta** un modelo de gestión integral de la Sanidad Ambiental.

Si sustituimos la palabra "**aporta**" por las siguientes palabras, ¿lees y comprendes mejor el texto?

- proporciona ☐ SI ☐ NO
- ofrece ☐ SI ☐ NO

[Texto 8]
Pautas de desinfección de **superficies** y espacios habitados por casos en investigación, cuarentena, probables o confirmados de COVID-19, viviendas, residencias, espacios de pública concurrencia y transportes de viajeros.

Si sustituimos la palabra "**superficies**" por las siguientes palabras, ¿lees y comprendes mejor el texto?

- terrenos ☐ SI ☐ NO
- zonas ☐ SI ☐ NO

[Texto 9]
Pautas de desinfección de superficies y espacios habitados por casos en investigación, cuarentena, probables o confirmados de COVID-19, viviendas, residencias, espacios de pública **concurrencia** y transportes de viajeros.

Si sustituimos la palabra “**concurrencia**” por las siguientes palabras, ¿lees y comprendes mejor el texto?

- reunión ☐ SI ☐ NO
- participación ☐ SI ☐ NO

[Texto 10]
Pautas de desinfección de superficies y espacios habitados por casos en investigación, cuarentena, probables o confirmados de COVID-19, viviendas, residencias, espacios de pública concurrencia y **transportes** de viajeros.

Si sustituimos la palabra “**transportes**” por las siguientes palabras, ¿lees y comprendes mejor el texto?

- traslados ☐ SI ☐ NO
- envíos ☐ SI ☐ NO

[Texto 11]
El edadismo se produce cuando la edad se utiliza para **categorizar** y dividir a las personas provocando injusticias como prejuicios, discriminación y estereotipos.

Si sustituimos la palabra “**categorizar**” por las siguientes palabras, ¿lees y comprendes mejor el texto?

- clasificar ☐ SI ☐ NO
- dividir ☐ SI ☐ NO

[Texto 12]
Se ofertarán en total más de 7300 plazas gratuitas para las visitas guiadas y para los ocho conciertos del ciclo Flamenco en palacio: una mirada diversa al patrimonio jondo, que reivindica el flamenco como **legado** patrimonial y cultural.

Si sustituimos la palabra "**legado**" por las siguientes palabras, ¿lees y comprendes mejor el texto?

- patrimonio ☐ SI ☐ NO
- herencia ☐ SI ☐ NO

[Texto 13]
Para que una familia goce de los beneficios que tiene ser familia numerosa es necesario que esta condición sea **reconocida** mediante un título oficial. Este título se concede, previa solicitud.

Si sustituimos la palabra "**reconocida**" por las siguientes palabras, ¿lees y comprendes mejor el texto?

- conocida ☐ SI ☐ NO
- aceptada ☐ SI ☐ NO

[Texto 14]
La caza y la pesca en la Comunidad de Madrid están sujetas a regulación especial y requieren una serie de **trámites**, entre ellos la obtención de licencia.

Si sustituimos la palabra "**trámites**" por las siguientes palabras, ¿lees y comprendes mejor el texto?

- procesos ☐ SI ☐ NO
- papeleos ☐ SI ☐ NO

[Texto 15]
La caza y la pesca en la Comunidad de Madrid están sujetas a regulación especial y requieren una serie de trámites, entre ellos la obtención de licencia. Además, los nuevos cazadores deben superar un examen o prueba de **aptitud**.

Si sustituimos la palabra "**aptitud**" por las siguientes palabras, ¿lees y comprendes mejor el texto?

- capacidad ☐ SI ☐ NO
- habilidad ☐ SI ☐ NO

Appendix C

Thesis Proposal - User Evaluation: Demographic Survey

CUESTIONARIO DEMOGRÁFICO

Usuario Número:
Fecha

1.- Indica tu sexo

- ☐ Mujer
- ☐ Hombre

2.- ¿Qué edad tienes?

3.- ¿Cuál es tu nivel de estudios?

- ☐ Sin estudios
- ☐ Primario (Certificado escolar)
- ☐ Secundario (Graduado en E.S.O., Bachiller - Formación Profesional)
- ☐ Universitario (Diplomado, Licenciado, Graduado, Doctorado)

4.- ¿Cuántas novelas lees al año?

- ☐ Ninguna
- ☐ de 1 a 3 novelas
- ☐ de 3 a 6 novelas
- ☐ de 6 a 12 novelas
- ☐ más de 12 novelas

5.- ¿Tiene usted alguna discapacidad?

- ☐ No
- ☐ Si
  - ☐ ¿Qué tipo de discapacidad?