Relevant Parameters for the Classification of Reading Books Depending on the Degree of Textual Readability in Primary and Compulsory Secondary Education (CSE) Students

R. GIL ORTEGO, (Senior Member, IEEE), AND I. MARTÍNEZ SÁNCHEZ

[1]Electrical and Computer Engineering Department, Industrial Engineering Technical School, Spanish University for Distance Education, 28040 Madrid, Spain
[2]Department of Methods of Research and Diagnosis in Education I, Faculty of Education, Spanish University for Distance Education, 28040 Madrid, Spain

Corresponding author: R. Gil Ortego (rgil@ieec.uned.es)

ABSTRACT This paper tries to establish the most important parameters of readability when it comes to choosing reading books for students in the second and third stages of primary and compulsory secondary education (CSE). A computational system is postulated, through data-mining techniques, capable of automatically classify reading texts by virtue of syntactic, lexical, semantics, and topological parameters regarding their overall content. The validity of this process is ensured through a careful selection by editorial experts used to accomplish the same validation task among different texts. In this paper, the results in the automatic classification of the degree of readability reached are promising; they allow to continue with a deeper investigation in this field of analysis with the purpose of progressing in the age rank studied, as well as in the text themes defining the corpus.

INDEX TERMS Algorithms, artificial intelligence, communication symbols, data processing, data systems, machine learning, machine learning algorithms, pragmatics, readability metrics, semantics, semiotics, syntactics, text processing, writing.

I. INTRODUCTION

The role of reading in the cognitive, emotional, social and cultural development of the human being constitutes one aspect into which has been significantly delved during the last decades in the academic field, since the importance of the reading process for the achievement of other competences and abilities is a peaceful debate [1], [2], [3], [4]. Likewise, the benefits of an intense and regular reading activity throughout life has been associated with a lesser effect of the advance of age on the cognitive function in general, and on the linguistic abilities in particular [3], [5], [6]; in this way, it is clear that the process of reading is crucial for the individual's executive functions with huge influence throughout the whole process of life [4], [7].

In the context of formal education, reading is an essential tool within the learning process at any curricular subject [3], [8], [9], and its importance increments as the educational level increases. However, the acquisition of knowledge through reading implies the comprehension of what has been read, something which has been deeply demonstrated by scientific evidences, in which the positive correlation existing between reading comprehension and academic performance is conclusively proved [10], [11], [12]. On that subject underline that reading is the skill which has the greatest influence on academic performance, although, in their opinion, it is even more relevant to consider the importance of reading for the access to information in the successive life stages [13]. In this way, comprehensive reading results crucial for both academic and professional success, since it is a vehicle for the social inclusion in the contemporaneous world [10]. In this author's opinion, the access to equality of chances is regulated by the individual's reading competence, in such a way that

The associate editor coordinating the review of this manuscript and approving it for publication was Fan Zhang.

the ability to exercise some of the social rights requires to previously incorporate this ability [10] In this same sense, Reynolds has indicated that the command of the reading skill is basic in the modern world [14].

From a psycholinguistic perspective, reading comprehension is a complex cognitive process of multifactorial nature, which depends as much from the text as from the reader's characteristics, as well as from the objective of reading and from the processing demands that the task implies [2], [6], [15]. Catts and Kamhi have explained that the multidimensionality of the process entails that attention must be paid to a set of intrinsic and extrinsic factors to analyze and conceptualize the student's performance as regards this ability [16]. Among the reader intrinsic factors, linguistic and cognitive abilities, previous knowledge and motivation are some frequently cited. On the other hand, concerning extrinsic factors, physical textual characteristics (such as font type or size) and linguistic textual characteristics (lexical, semantic and syntactic aspects) can be mentioned.

Given the importance of the environment on the development of linguistic abilities, the development of an interest in reading and the construction of a reading identity, another factor to consider is the sociocultural context. In other words, the range of situations in which reading takes place [17]. A relationship has been shown to exist between the degree of exposure to reading materials and reading competence, showing that a richer reading experience is associated with a higher reading competence at different levels throughout development [18]. In this sense, children's literature fulfills an essential role in that it helps to encourage exposure to printed materials from early ages. In fact, it has been observed that the number of books in the home predicts reading performance of readers at primary and secondary school. This is true even after controlling for factors such as family income, parental education level, language used at home and other aspects of the family context, and teaching [19]. Further, it has been found that well equipped school libraries constitute a protective factor against social disadvantage [20].

Children's literature constitutes an important resource in encouraging reading for pleasure outside of the school, especially from third grade of obligatory primary education, when normally developing children have had the opportunity to acquire the ability of automatic and fluent decodification. This makes them more independent readers when engaged in activities of voluntary reading, with the support of their educators and parents. Previous research has established a strong relationship between the time dedicated to reading and the performance of the reader [21]. In fact, reading regularly outside of school is associated with higher reading test scores [22], [23]. However, the benefits of independent reading extend beyond the educational purposes to personal development and quality of life [24].

In general, the mere exposure to printed materials does not necessarily ensure the development of high reading competence [17]. As has been noted, reading is a complex activity for which success is determined by three factors: the reader, the text and the objective/task [17]. Whether or not the voluntary reading activities will work to improve comprehension will largely depend on the degree of compatibility between the reader's competence and the complexity of the text [25]. Complexity of the text is understood as the ease/difficulty with which the text is read/understood and is known as readability [26]. Research proposes that readers do not enjoy reading books that are not appropriately challenging [27]. For contrasting reasons, both books that are too easy and those which are too difficult, can drive the reader to lose interest [28]. Evidence suggests that a positive relationship exists between enjoyment of reading and frequency of reading [29].

Consequently, in reading comprehension, multiple aspects related to the format of the text influence. The text has been translated into other languages, into the text and into the terms of its ease / difficulty to be read and understood [26], which are included within the term readability. Reading is, therefore, a relevant aspect from the pedagogical point of view, insofar as it allows to establish the identity in a text according to its characteristics for an account reader [31], [32], being a dimension that at the same time influences on intrinsic factors, such as interest or motivation towards reading. Reference [13] Have recognized the importance of reading in life in which the child has to be present in the reading of the texts with which he works in the classroom and in the educational axis of the educational action found in the educational context.

The term readability was used as one of the first time in the Spanish context by José Fernández Huerta in his outstanding article of 1958, "Legibilidad y lecturabilidad: dos conceptos básicos en los libros escolares [33], and employed in his also decisive text of 1959, "Medidas sencillas de lecturabilidad" [34]. In the 1959 article, he defined readability as follows:

"Text readability is closely linked to reading comprehension. More specifically, it conveys the possibility that the terms employed in a book and its syntactic structures comprehensively concern readers".

As it can be observed, the term is more connected to the "comprehensibility" determined by the syntactic structure than to the typographical aspects of the text. The term has been successful in several spheres of knowledge, such as graphic design, and it has been more spread in the other side of the Atlantic, in countries such as Colombia, Mexico or Costa Rica. Campos, Contreras, Riffo, Veliz & Reyes have exposed that this term makes reference both to educative and cognitive considerations, since, as these authors underline, although in a first place it was created to allude to the selection of texts that were more suitable to the students' needs in each educational stage, the advance of research revealed the connection existing between this dimension and the different elements composing the text, which indicate the corresponding degree of difficulty [35].

Someone as outstanding in the world of Spanish ortho-typography and lexicography as José Martínez de Sousa

defined the term vehemently, still recognizing its difficulties of acceptance:

"Readability or comprehensibility refers to the ease of comprehension and interpretation of a text related to the style and plot [. . .] and is a function of the structural characteristics and content of the text: interest, difficulty, diversity, density, length of sentences, choice of words, etc. It should not be confused with legibility, which, as we shall see, refers to the ease of reading due to typographical presentation. In fact, a text can be very legible [. . .] and scarcely readable [. . .]" [36].

And José Luis Rodríguez Diéguez (1940–2005), professor in Didactics for the University of Salamanca, and pioneer in our country in research about written text comprehensibility, also expounded the following statement in that regard:

"The sound expression of words requires formal characteristics in the text that facilitate the identification of graphemes. Specially, it refers to the typographical characteristics of the text [. . .]. But the comprehension of the meaning of the text supposes a whole series of cognitive processes that, evidently, can be facilitated from the outside, by means of the use of diverse resources that can denominate, in a lax way, stylistic [. . .]" [37].

Therefore, the term legibility encompasses the set of typographical and linguistic characteristics of the written text [38]. That is to say, legibility has two complementary dimensions: on the one hand, typographical legibility, where size, shape, illustrations, etc. are important aspects to consider; and on the other hand, linguistic legibility, related to the syntactic, lexical and semantic aspects of the text [38]. According to Ripoll, the legibility of a text has influence on reading mistakes, so it emerges as an important question when it comes to plan the materials at reader's disposal [38].

The focal point of the present research lies on linguistic legibility, mentioned under the term readability. This term, in the Anglosaxon context, is sometimes called "legibility" and sometimes "readability". With the purpose of unifying terms between English and Spanish, the term legibility will be employed making reference to linguistic legibility or readability.

The importance of the study about readability in the current context of textual information treatment results relevant in applications such as cataloguing of didactic materials suitable with the student's educational level [37]; calibration of an appropriate language in the medical context to be understandable by laypeople; production of instruction manuals for electric or electronic devices and most of all, in the creation of web texts. Precisely, one of the most relevant factors for the classification of web pages in Google, according to Searchmetrics, is the legibility of a text. It has been verified the existence of a positive correlation between a good positioning in search engines and the web page legibility.

This document is structured in the next sections: in the first place, the problem statement is faced; next, reading comprehension and legibility are connected; later, the methodology employed in this research is addressed; to continue with, the results obtained are shown; consecutively, the discussion and conclusions of the word are presented; and finally, all the references used are delivered.

II. PROBLEM STATEMENT
Among the different applications of readability, this research focuses on a relevant aspect in the educational sphere: the automatic classification of reading texts suitable to the levels of the students' evolutionary development, in the stages of Primary and Compulsory Secondary Education of the Spanish educational system; that is to say, from 8 to 14 years. The selection of this particular rank of ages for the research answers to the need of investigation upon this population, in relation to which few studies dealing with the existing link between the degree of acquisition of the reading skill and readability have been found [4], [39]. The achievement of reading maturity, which is supposed to be reached during the foresaid rank of ages, makes it necessary for selected materials to be adapted to the student's level of development, since the readability rate in this sphere is an indicator upon which the decision concerning the materials must be based [4], [39].

As it has been stated in the preceding section, the formulated problem is relevant, not only in the editorial field to facilitate and validate the personal processes of texts selection, but also in the creation of web pages oriented towards a particular public and, in general, in formative contexts, since this is an emerging issue in which it is necessary to reach a deeper perspective in the investigation field, obtaining solid evidence about readability contributions to these different spheres [39].

III. READING COMPREHENSION AND READABILITY
The approach to the concept of reading comprehension requires firstly to identify which are the essential elements defining this skill and, in line with these elements, to know which the steps are to develop it [2], [40]. As a result, one of the assertions upon which the study is based is the conceptualization of reading, which, according to Jiménez is an action that implies the interaction between the individual and the text, through which the former tries to decode the meaning lying on the integrating words [39]. As reported by Jiménez, reading is an activity that leads the subject to extract and build the meaning of a text, an action which is part of the communicative process [39].

Taking into account the features of the reading process, reading comprehension can be defined as a vehicle through which the reader establishes interactive relationships with the reading content, links the ideas with previous ones, contrasts them, contends them and then, elicits personal conclusions [41], [42]. These conclusions about significant information increase the reader's cognitive baggage when they are assimilated and stored [39]. Akin to Jiménez, reading comprehension is assigned to a broader concept: reading competence, which can be defined as the individual's faculty to employ reading comprehension in an efficient way in the social

sphere to which he/she belongs, something which, in this author's opinion, directly implies the establishment of a bond between the individual and the society [38]. This is a capacity which allows the individual to participate in the different existing spheres.

With no intention to delve deeper into this concept, this article focuses on reading comprehension; in that regard, Morles (states that "without comprehension there is no reading", pointing to the synergies existing between both processes [42]. Consequently, the relevance of reading for an individual involves the fact that, for reading comprehension to develop, certain conditions have to take place and a series of characteristics have to be shown: reading must be active, exploratory and investigative for the reader to have access to the textual meaning and with that, to be involved within the intrinsic cultural scheme of reading.

In this context, three comprehension levels have been defined:

a) Primary comprehension: this is the comprehension of simple affirmations. At this level, lack of vocabulary usually produces difficulties for the reader. This problem develops incomprehension on the part of the reader since he/she does not recognize the meaning of the voices used, aspect that can be solved resorting to dictionaries. Since the concepts are universal and they do not always correlate with objects that can be graphically represented, the limited development of the abstract thinking, which does not take place until maturity is reached, may be the origin of the incomprehension of particular affirmations, since, before the achievement of the foresaid stadium, some difficulties may arise when it comes to assimilate the concepts involved.

b) Secondary comprehension: this is the comprehension of the author's argumentative central concepts, main affirmations, fundaments and how ideas are connected. At this level, failures may have as a cause the non-distinction between the primary and secondary ideas. It is very common that the reader retains the example and forgets the universal affirmation exemplified by it. Lack of agility in the logical thinking also hinders secondary comprehension. The reader must understand the links which join the most important affirmations in the text, and doing so, he/she is inwardly recreating the relationships conceived by the author himself. This results in the reader's development of his/her logical thinking. Consequently, a scarce development of the logical thinking will obstruct or even prevent comprehensive reading at this level.

c) Deep comprehension: this is the comprehension that overtakes the text and that grabs the implications linked both to the context in which it was written and the context in which it is read, and the implications concerning what "it really is" and/or what "it must be". This comprehension involves a previous and wider knowledge on the part of the reader. The broader the baggage of knowledge with which the reader addresses the text, the deeper the comprehension will be.

Being exposed the three existent levels of comprehension, it must be highlighted that for an individual to be able to understand what he/she reads, it is necessary to take into account (besides personal characteristics) readability and legibility, terms which must be analyzed jointly [44], [45]. Both concepts, which are used as synonyms in Spanish, refer to the ease or difficulty with which a written text counts to be read and understood. However, some differences have been appointed: legibility only refers to the typographical qualities of the text [44], [46], while readability informs about the reader's subjective conditions; that is to say, to the possibility for a text to be understood by a particular reader [37], [45]. As it has been previously commented in this article, when dealing with legibility, reference is being made to the text readability, just as it is affirmed by Alliende in his work "La legibilidad de los textos" [47]. In other words, legibility is the set of textual characteristics which favor or complicate a more or less effective communication between them and the readers, depending on the latter's competences and on the conditions in which reading is conducted.

The study and evaluation of the level of legibility is related to the fields of Quantitative Linguistics [37]; Statistic Linguistics [48] and Informetrics, [49] which covers the study of the quantitative effects of information. With both a practical and theoretical scope, these scientific disciplines make emphasis on the development of mathematical models, as well as paying attention to the derivation of measures for the different phenomena studied. In effect, legibility measures have investigated about the prediction of the comprehension attributable to a text as long as it can be read by individuals with particular psychological characteristics or educational levels.

There exist varied systems to measure text legibility [37], [47], [50], which are used as a last resort to provide the situation of the verbal message reception in conditions of textual production. Despite this, the major part of the formulae demands a considerable effort to be applied, as much for word, syllable or sentence re-counting as for other categories that the model exhorts. Time is needed for the application of the foresaid methodologies and, sometimes, also because of the complexity of the employed equation, the resolution becomes accessible only for specialists.

Nevertheless, with the appearance of certain statistical programs, the time spent in analyzing and evaluating texts has been considerably diminished. The employment of textual data-mining techniques (which will be detailed later) has been applied on the evaluation of the text books used in the second year of Primary Education and in the first year of Compulsory Secondary Education to develop the research exposed throughout this article.

IV. MATERIAL AND METODOLOGY

Once the research problem was fixed the decision was to concentrate on the selection of relevant parameters for the

automatic classification of reading texts with a proficiency equivalent to journal experts, and with the aim of helping the student freely and autonomously choose the reading materials that they most like and are compatible with their level of cognitive development. In this section the material used in the present study for data collection, data configuration, the methodological processes used in the selection of the classification models with greatest predictive validity and finally, the results obtained applying the method described. We will begin by reviewing the parameters used to measure the readability of selected texts.

A. LEGIBILITY PARAMETERS

Since more than 70 years ago, a series of works pursuing to provide evidences regarding the calculation of a readability rate and its gradation by reading comprehension levels have been produced [44]. The works Kincaid [51] are prominent, where only simple functions of two or three linguistic variables, such as syllables and words, are dealt with. The important idea about these authors is to develop a regression equation (1) which considers concepts related to the semantic difficulty, such as number of syllables per word; and concepts related to the syntactic difficulty, such as sentence length.

$$RGfk = 0.39(\text{Mean of words per sentence})$$
$$+ 11.8(\text{Mean of syllables per word}) - 15.59 \quad (1)$$

The modification of this formula by different authors [52] has meant a considerable number of rates. This classic approach has been used during a large period of the past century, even when it has certain limitations: a) a text without "noise" must be chosen (that is to say, with well-formed sentences); b) the text must have a minimum extension of words, at least 300 [53] and c) its weakness has been shown in web page texts [54], [55], [56]. Taking these limitations into consideration, other dynamic approaches have been strengthened in the present. Approaches which get automatically adapted to the evolution of both vocabulary and individuals; and mainly, approaches which overlap with current issues such as Internet and social media, in which cohesion, syntactic ambiguity, discourse organization and density of propositions, among other determinants, are taken into account.

This new methodology applied on the evaluation of readability started to be employed in the middle of the 2000 decade and it is substantiated on textual data-mining proceedings, artificial intelligence and machine learning.

In order to address the objective of the present study a set of parametric parameters were selected: lexical, syntactic and semantics. As a novel aspect, topological parameters of a wide set of texts of reading books chosen by editorial experts were also included. A methodology of supervised classification "machine learning" was used, selecting the most relevant parameters to obtain the greatest degree of precision with respect to the classification provided by the editorial experts.

TABLE 1. Sample distribution by age. Source: Own devising.

Ages	Number of texts
8 – 10 years	109
10 – 12 years	78
12 – 14 years	83

B. SAMPLE

To develop this research, a random sample of more than 300 words has been collected from each of the 270 texts of reading books selected by editorial experts. The foresaid experts, with wide professional experience, have graded the reading comprehension level corresponding to each text. The experts labeled each text as 9, 11 or 13, when the text is more suitable for ages (8 – 10), (10 – 12), or (12 – 14) years, respectively. The sample distribution was as shown in Table I.

C. MODEL

The process of selecting relevant parameters for the automatic classification of texts according to their degree of readability (see Fig. 1) in the present research has been devised through the construction of two programs in R language. The first one is headed towards the reading of texts and the calculation of their parameters. In short, this program develops the following steps: a) gathering textual content: consists on the reading of data files and on the pre-processing of data: characters filtration, corpus creation, creation of the matrix Terms-Documents and tokenization; b) calculation of readability parameters: classic legibility rates; syntactic and semantic complexity rates and topological gathering rates; c) devising of parametric vectors with the addition of data to create a file with the parameters of the 270 texts, together with the incorporation of the classification by editorial experts. The second program is designed for the selection of relevant parameters, for automatic learning with training data and for testing parametric vectors, to be able to classify the reading texts depending on the readability level attributed by the experts with the highly valuable information given by themselves.

Next, every one of the steps of the proposed model of analysis will be detailed.

1) GATHERING TEXTUAL CONTENT
Just as it was concisely indicated when dealing with the sample, more than 300 words of each of the 270 texts from reading books chosen by editorial experts have been compiled. Each of the samples obtained from each text has been subjected to a filtration process consisting on erasing blank lines; suppressing strange characters (#, $, etc.); creating the corpus; directing the tagging and lemmatization processes; and finally, creating the matrix Terms-Documents.

2) CALCULATION OF PARAMETERS
In relation to this new paradigm of automatic learning, it has been considered that, to measure the ease or difficulty of a
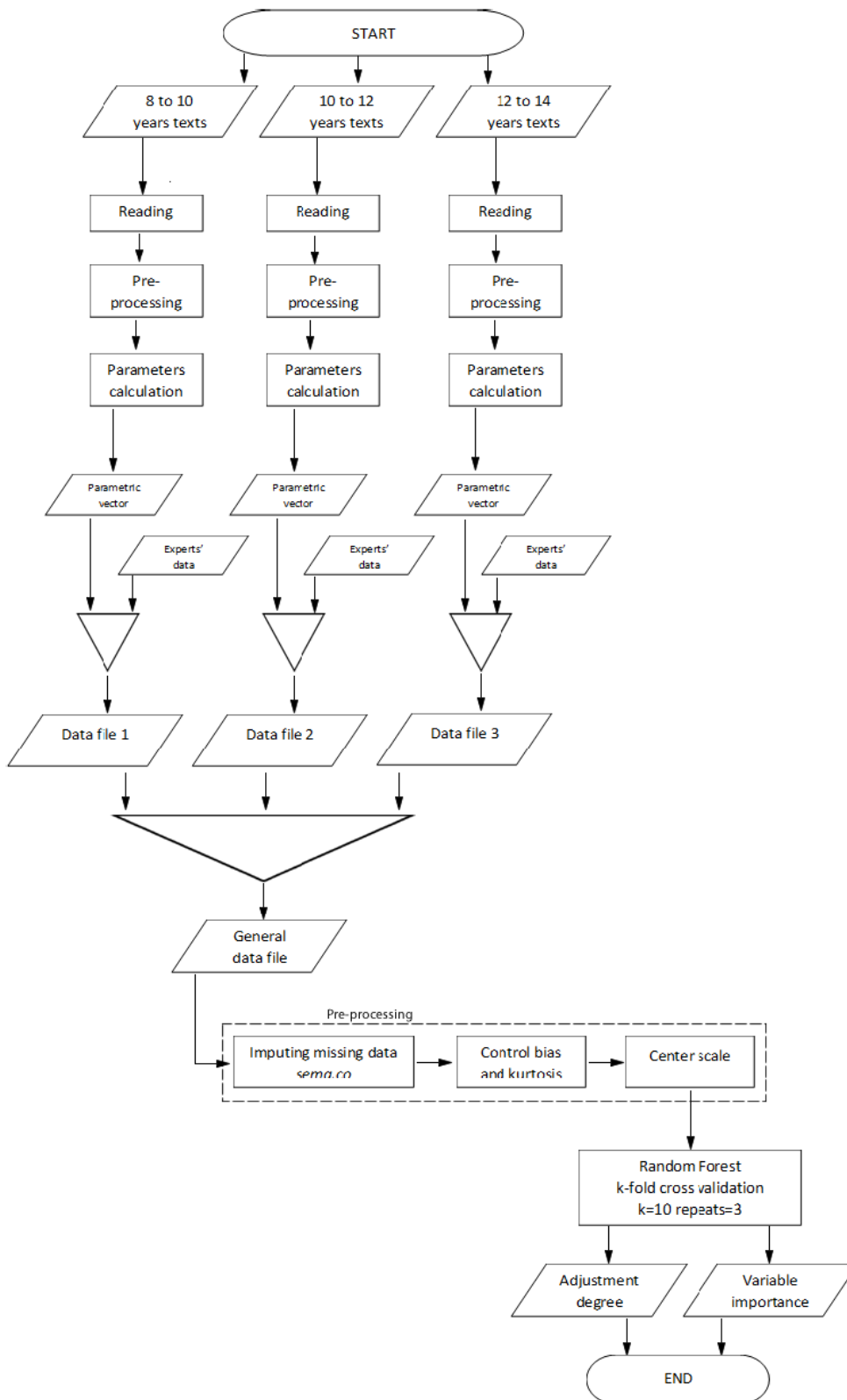
FIGURE 1. Diagram of data processing.

text comprehension, apart from classic readability parameters, others related to syntax, vocabulary and semantics must be also incorporated [57]. Moreover, in this research, a set of rates by Kansky including topological characteristics of

road networks have been used, with the purpose of assessing the connective structures of the paragraphs in the written text, because of their parallelism with these networks. Below, the different types of parameters are explained.

a: CLASSIC PARAMETERS

One of the pioneers in the creation of a readability rate was Rudolf Flesch with the production of the Flesch Reading Ease Score (RES) in 1949. In the Spanish context, one of the first authors in addressing the problem of readability assessment in a Spanish text was the pedagogue José Fernández Huertas based on Flesch's rate, whose adaptation appeared in 1959 [34]. Its expression (2) is as follows:

Fernández Huertas' rate
$$= 206.84 - 60(Sy/W) - 1.02(W/Se) \quad (2)$$

where: Sy = Syllables; W = Words; Se = Sentences

Another classic author in the creation of a readability rate is Peter Kincaid, who, together with his team, developed modified Flesch's rate in 1975 for university texts and texts about technical matters. Again, in Spain, Francisco Szigriszt Pazos introduced in 1993 a modification of Flesch's formula with the Flesch-Szigriszt Rate for Spanish ("formula of perspicuity"). The already stated rate takes the following expression (3):

Flesch-Szigriszt's rate (FSZR) is : 206.835
$$- 62.3(Sy/W) \text{ -(W/Se)} \quad (3)$$

Both Fernández Huertas' and Szigriszt's formulae have influenced the creation of tables attaching the readability rate with the educational level. Thus, for instance, the first ones are located in a rate between 60 and 70 at a regular level, corresponding to the 7th and 8th educational years; and Szigriszt places the normal level between 55 and 65, which corresponds to a kind of publication of basic secondary Education, general press and sports press.

b: SYNTACTIC PARAMETERS

Many times, a document can be illegible due to unusual linguistic constructions or a grammatical language which tends to be expressed in the syntactic properties of the text. Therefore, syntactic characteristics have been employed Bernth [58] to measure the "clarity" of the written text, with the objective of helping writers to improve their writing abilities.

In the child's evolutionary development, grammatical constructions increase progressively in complexity. This sentence formation process of growing intricacy finds its development throughout the whole school period.

Hunt suggests some syntactic maturity rates of quantitative nature [59]. Among these rates, the primary ones are emphasized as they are the most frequently applied: terminal unit length (TUL), clause length (CLL) and subordination rate (SR).

Hunt calls it "terminal" because grammatically, it is usual to start writing with a capital letter and to finish with a full stop or a question mark and minimal because it is the smallest unit in which a fragment of discourse can be divided [59]. This unit, which substitutes the classic concept of sentence, consists of a main clause and any subordinate clause which might be added or attached to it [59], [60]. In this way, simple and complex sentences constitute a terminal unit. Coordinate and juxtaposed sentences, on the contrary, contain two or more terminal units since, according to Hunt, they constitute independent terminal units [60].

As a consequence, sentences have been assimilated as terminal units for the calculation of syntactic parameters; thus: the average length of a terminal unit (sin.lo) will be the quotient between the total number of words and the number of terminal units in the text. The average length of clauses (sin.cla) is calculated by dividing the total number of words by the clauses in the text. Finally, the subordination rate (sin.sub) is obtained by dividing the number of clauses by the number of terminal units in the text.

c: LEXICAL PARAMETERS

The complexity and difficulty of texts depend on their vocabulary and sentence length [14]. To measure the lexical complexity of a text, four parameters have been taken into account: lexical density, lexical diversity, lexical wealth, and average frequency per word, since these are the most employed parameters in diverse researches [57]. Lexical density (lex1) is defined as the quotient between notional words and textual words. Nouns, verbs, adjectives and some adverbs are considered lexical or notional words.

According to Ure [61], lexical density is below 0.4 in spoken texts and higher to 0.4 in written texts. In general texts, lexical density varies between 0.2 and 0.3 and between 0.5 and 0.6 in specialized texts. The lower the value, the easier the text will be.

Lexical diversity (lex2) assesses how many different words are employed in a text. There exist several rates to measure lexical diversity. In general, the determinants indicated by [62] are demanded in those rates: 1) stability to be independent to the chosen text size; 2) sensibility in the rank of values to be able to distinguish between rates of similar texts; 3) coherence with other rates, in such a way that correlation with the same can be established.

The type-token ratio (TTR) is considered a fundamental parameter to measure lexical diversity, where a lower value will be an indicator of the text easiness. For the calculation, the text is divided into 100-word blocks. In each of these blocks, the TTR is the quotient between different words of the text in a 100-word block divided by the total number of words in the block.

Related to lexical diversity, there exist several rates measuring lexical wealth (lex3) depending on the repetition ratio. Out of the five lexical wealth rates which meet the criteria formerly commented (Yule's MTLD, HD-D, K; Honoré and Maas' H) Yule's K has been chosen since it is the widest

used in the incorporation of words into a corpus. The foresaid rate was proposed by G.U. Yule in 1944 and it is based on a probabilistic model, in which it is presupposed that the appearance of the words in a text is merely regulated by chance. Yule's formula (K) measures lexical wealth in a text depending on the lexical repetition ratio. Its expression (4) is the following:

$$K = \frac{10^4 \left(sum \left(fX^* X^2 \right) - N \right)}{N^2} \quad (4)$$

where: N = Number of words; X = Vector of different words; fX = frequency of different words.

Finally, to assess lexical frequency (lex4) (that is to say, the influence that the use of frequent words in linguistic corpora has on reading comprehension) the median of standardize frequencies of words from the CREA base (Corpus de Referencia del Español Actual) has been applied. The higher the value of the rate, the easier the reading of the text will be.

d: SEMANTIC PARAMETERS

To create semantic parameters the cognitive concept of proposition has been used as a base. As it is known, a proposition is formed by a predicate or propositional concept and one or more arguments, each one with a unique semantic role. Predicates may be identified as verbs, adjectives, adverbs and connective sentences, and they refer to properties or relationships. Arguments, on the other hand, can be concepts or other inserted propositions. They accomplish different semantic functions [57].

Propositions possess an important role in the identification and assimilation of the text. Taking this into consideration, the following rates have been created: local coherence (sema.co), regarded as the cosine similarity of the vectors which form two adjacent sentences; and density of information (sema.densi), consisting on the number of nominal words in the text.

e: TOPOLOGICAL PARAMETERS

To assess the connective structure of the paragraphs of the written text, the topological structure of road networks has been employed, since they are considered parallel to the nodes of the network constituting the document (see Fig. 2). For this purpose, parameters associated to centrality, cohesion and network accessibility have been chosen.

Centrality, in this context, seeks to identify the topological position of paragraphs within the network formed by the relationship connections established between their word networks. As a measure for similarity among the aforementioned vectors, the cosine has been used, and also, the existence or non-existence of connection among paragraphs has been fixed in the cases in which similarity exceeds its average. Therefore, each paragraph represents one of the nodes of which the network is formed. Its influence over the surrounding area can be easily noticed in its graphic representation.

To provide a centrality rate of the network, the rate of average centrality (con1) has been used, which we have called
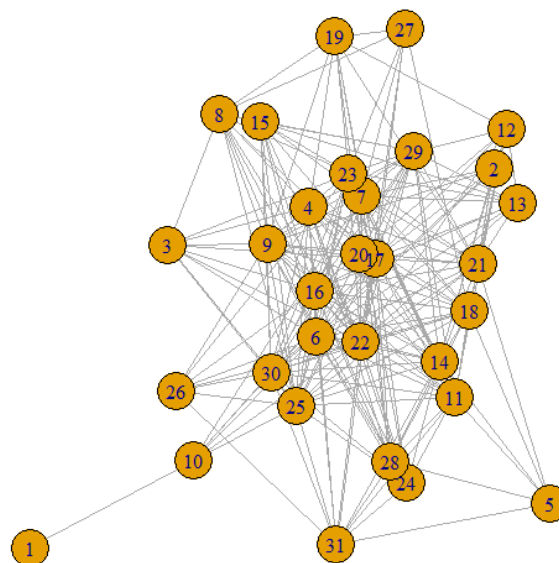

FIGURE 2. Topological relationship between paragraphs.

rate of topological centrality. It is derived from the sum of Shimbel's rates divided by the number of nodes minus 1. In this way, the lower the value of the foresaid centrality rate, the higher the centrality of the network will be.

Cohesion (con2) has been measured by means of Kansky's gamma. This is a rate which connects the number of existent arcs and the highest possible number of arcs. The value of gamma for a network of maximum connectivity is 1. Its expression (5) is the following:

$$\text{con } 2 = e/(v(v-1)/2) \quad (5)$$

where: e = links; v = nodes.

Connectivity (con3), as it is shown in expression (6), allows to analyze the spatial organization of the nodes in a network, providing thus a general idea about the easiness of access of each node in relation with the remaining nodes in the network. In this case, the connectivity rate of the network has been measured by means of the alpha rate, which we have called alpha rate of topological connectivity. Its value varies between 0 and 1 when maximum connectivity occurs.

$$\text{con } 3 = \frac{e - v + p}{(v(v-1))/2 - (v-1)} \quad (6)$$

where: e = links; v = nodes; p = related components.

Therefore, in the example Fig. 2, its topological parameters would be: con1 = 6.67; con2 = 0.48; and con3 = 0.44. It is observed that all the 31 nodes have a similar centrality, an indicator of a good accessibility for the knowledge of the transmitted text; except node 1 which is an introductory paragraph. Nodes 20 and 17 stand out for their centrality, so the value of con1 for them will be lower. Cohesion (con2) and connectivity (con3) are average, which indicates a degree of connectivity between the good sentences, that is, a link between the paragraphs of the text.

"lec.pazos" "lec.fles" "sin.lo" "sin.cla" "sin.sub" "lex1" "lex2" "lex3" "lex4" "sema.co" "sema.densi" "con1" "con2" "con3" "enu" "grupo"

84.78 88.73 7.79 6.97 1.12 0.53 0.4 100.42 952.08 0.91 0.22 0 0.67 0.64 "40 enigmas para fanáticos del fútbol.txt" 9

91.88 95.7 6.4 6.1 1.05 0.54 0.53 79.96 1326.82 0.78 0.18 0 0.71 0.73 "Academia de magia 1.txt" 9

89.33 93.22 6.5 5.35 1.22 0.54 0.46 77.24 1105.96 0.56 0.15 4.28 0.67 0.61 "Academia de magia 2.txt" 9

87.96 91.83 7.45 6.05 1.23 0.55 0.43 83.19 807.88 0.75 0.16 4.96 0.58 0.54 "Academia de Magia 3.txt" 9

84.86 88.6 9.15 6.27 1.46 0.53 0.45 109.15 1218.56 0.52 0.16 4 0.67 0.61 "Academia de magia 4.txt" 9

FIGURE 3. Results from 4 texts for students from 10 to 12 years.

3) CREATION OF PARAMETRICAL VECTORS

A vector representing the textual parameters (following what has been exposed in the previous section) has been provided for each of the texts, together with the title of the book from which the text is extracted; and finally, the gradation provided by experts (Fig.3).

The combination of vectors is attached in a file for the following selection process.

As it can be observed in Fig. 3, in the processing file the names of the variables of readability parameters appear, followed by the name of the book and of experts' gradation.

Reading books are those recommended for students to complete their formation in the curricular subject "lengua castellana" within the Spanish syllabus.

4) SELECTION OF CLASSIFICATION MODEL

In the automatic classification of texts conforming to the criterion grouping variable assigned by the editorial experts, the following two processes were conducted: a) selection of the classification algorithm with the greatest precision when explaining the dependent variable; b) once the classification model was selected, its degree of accuracy was examined, and the independent variables were selected which most contributed to the model.

Once the file is constructed with the parameter vectors of the 270 texts alongside their grouping variable, we proceeded to select the most accurate model from those most commonly used: penalized logistic regression (glmnet), linear discriminate analysis (lda), Naïve-Bayes (nb), partial least squares regression (pls), support vector machine radial basic function kernel (svmR) and random forest (rf). K-fold cross validation was used to validate the methods, with k = 10 and repeats = 3, using the R statistical package. The results of this process are shown in Table II.

As can be seen in the Table, the glmnet and rf procedures give similar results, perhaps due to the use of Bagging in random forests and Boosting in glmnet. Both procedures, in synthesis, are very similar. In addition, reduced sample size

TABLE 2. Results from the selection processes of parameters.

Type	Procedure	Accuracy	Kappa
glmnet	Penalization Logistic regression (alpha=1; lambda = 0,03)	0,781	0,67
lda	Linear discriminant	0,73	0,593
nb	Naive-Bayes (usekernel=TRUE)	0,737	0,605
pls	Partial Least Squares (ncomp=2)	0,717	0,571
svmR	Support vector Machine radial (c=0,50)	0,706	0,555
ranger	Random Forest (mtry=3; nodesize=5; ntree=500)	0,789	0,682

and using k-fold cross validation can be part of the proximity of results.

Before selecting the most adequate classification model, pre-processing of data was conducted: a) the presence of missing cases was observed in the dataset and as only one was found (sema.co) this was imputed; b) bias and kurtosis is controlled in case transformation has occurred and variables are normalized to the same scale prior to selecting the model.

As can be observed, the model with the greatest accuracy is the Random Forest. As a result, it will be chosen for the selection of the most relevant variables for the automatic classification of texts.

V. RESULTS

The Random Forest model was applied to the dataset using the statistical package of the same name. Two elements result from the procedure: a) The most relevant parameters for classification with the selected model, and b) Degree of accuracy achieved using the confusion matrix of the classification.

The ordering of parameters according to their degree of accuracy (see Fig. 4), that is to say, according to the number of elements of the model fitting with the experts' classification criterion, generates the next result.

As it can be observed, the three most important rates are: topological (cohesion –con2- and connectivity –con3-) and vocabulary (lexical frequency –lex4-). The presence of the cohesion and topological connectivity parameters stands out elevantly. The former confirms the importance of cohesion for the comprehension of a text, since the connection between its sentences allows the reader to create a general vision and therefore, to build a mental structure about its content. Connectivity indicates the easiness of access to sentences due to the existence of a related structure of the network nodes in the text.

Lexical frequency confirms the importance of understanding the words in the text for the reader, in this case, reflected by the higher frequency of these in the CREA base.

As regards to the confusion matrix that indicates the degree of adjustment of the model, the results obtained are shown in Tables III and IV.

The degree of adjustment of the model for the validation data reaches 78.9%, which gives an overall idea about the precise adjustment of the model and about its ability for the
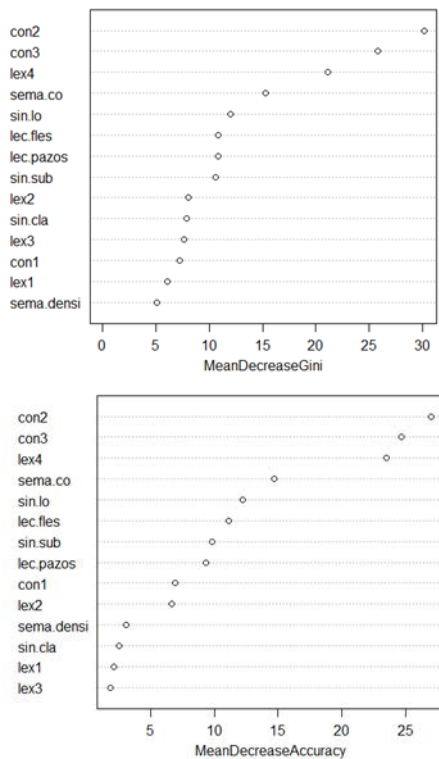
FIGURE 4. Gradation by importance of the different parameters in the random forest model.

TABLE 3. Coding matrix.

Reference Prediction	9	11	13
9	95	8	2
11	6	56	21
13	0	20	62

TABLE 4. Degree of adjustment (acuricity) of the selected model.

Degree of Adjustment	
Accuracy:	0.7889
95% CI	(0.7353, 0.836)
No Information Rate	0.3741
P-Value [Acc>NIR]	<2e-16
Kappa	0.6817
Mcnemar's Test P-Value	0.5106

automatic selection of reading texts for the ages of the sample students.

VI. DISCUSSION AND CONCLUSIONS

The research conducted is framed within a set of works related to data-mining in the educational context, even though evidences offer the possibility of extrapolating the employ-ment of this method into other disciplines. The starting point has been placed with readability and reading comprehension in the age rank of 10 to 14 years, period in which the student completes the full acquisition of reading maturity [63]. In line

with the discoveries by Campos [57] the degree of complexity of a text affects on dimensions such as reading comprehen-sion, something which compels to considerate readability as a way to increase the students' capacity to access to infor-mation, and with that, according to Meltzer [4], to facilitate their unlimited access to social interaction, which, to a greater extent, requires of the command of reading comprehension to be able to exercise the individual's rights in society. Readabil-ity must be regarded, thus, as one of the precepts to ensure democratization in the access to contents [6], which, in the current scenario, is one of the requirements to reach equality among all people.

The results in the automatic classification of the degree of readability reached in this research are promising; they allow to continue with a deeper investigation in this field of analysis with the purpose of progressing in the age rank studied, as well as in the text themes defining the corpus. By means of the Random Forest Procedure, the 78.9% of the texts of Primary and Compulsory Secondary Education have been classified, out of a total of 270 reading texts from Spain schools.

The most relevant readability parameters discovered are: cohesion –con2- and connectivity –con3-. Once literature is revised, it permits to verify that the work makes an origi-nal contribution to the research sphere, providing important elements for the classification of texts.

As a prospective line of research, work will continue on this matter to incorporate literary and press texts to strengthen the model, which will be focused upon a wider age rank. In line with Reynolds' proposal [14], the results find their application in the educational sphere, to contribute to develop the selection of texts made in virtue of their characteristics, to foster a better acquisition of reading competence and to emphasize comprehension.

REFERENCES

[1] R. C. Flores-Macías, J. E. Jiménez, and E. García, "Procesos cognoscitivos básicos asociados a las dificultades en comprensión lectora de alumnos de secundaria," *Revista Mexicana de Investigación Educativa*, vol. 20, pp. 581–605, Jun. 2015.

[2] C. Blachowicz and D. Ogle, *Reading Comprehension: Strategies for Inde-pendent Learners*. New York, NY, USA: Guilford Press, 2017.

[3] C. A. Temple, D. Ogle, A. N. Crawford, and P. Freppon, *All Children Read: Teaching for Literacy in Today's Diverse Classrooms*, 5th ed. London, U.K.: Pearson, 2018.

[4] L. Meltzer, *Executive Function in Education*, 2nd ed. New York, NY, USA: Guilford Press, 2018.

[5] B. R. Payne, X. Gao, S. R. Noh, C. J. Anderson, and E. A. L. Stine-Morrow, "The effects of print exposure on sentence processing and memory in older adults: Evidence for efficiency and reserve," *Aging, Neuropsychol., Cognition*, vol. 19, nos. 1–2, pp. 122–149, Jan. 2012.

[6] S. R. Suleiman, *The Reader in the Text*. Princeton, NJ, USA: Princeton Univ. Press, 2014.

[7] D. J. Follmer, "Executive function and reading comprehension: A meta-analytic review," *Educ. Psychologist*, vol. 53, no. 1, pp. 42–60, Apr. 2017.

[8] E. B. Moje, D. Stockdill, K. Kim, and H.-J. Kim, "The role of text in dis-ciplinary learning," in *Handbook of Reading Research*, vol. 4. Abingdon, U.K.: Routledge, 2011, pp. 453–486.

[9] Y. G. Benítez, K. C. Espinoza, and V. R. Pérez "Niveles de comprensión lectora en alumnos de secundaria. Una comparación por tópico," *Actuali-dades en Psicología*, vol. 29, no. 118, pp. 13–23, May 2015.

[10] A. O. González, "Educación inclusiva y fácil lectura: La comprensión y el texto como vías para llegar a todos nuestros estudiantes," *Revista Electrónica de Investigación y Docencia*, vol. 13, pp. 169–192, Jan. 2015.

[11] N. S. Pinto, A. I. M. Martínez, and L. Jiménez-Taracido, "Learning strategies, reading comprehension and academic achievement in secondary education," *Psicologia Escolar e Educacional*, vol. 20, no. 3, pp. 447–456, Dec. 2016.

[12] M. Hebert, J. J. Bohaty, J. R. Nelson, and J. Brown, "The effects of text structure instruction on expository reading comprehension: A meta-analysis," *J. Educ. Psychol.*, vol. 108, no. 5, pp. 609–629, 2016.

[13] N. Nassiri, A. Lakhouaja, and V. Cavalli-Sforza, "Modern standard arabic readability prediction," in *Proc. Int. Conf. Arabic Lang. Process.* Cham, Switzerland: Springer, 2018, pp. 120–133.

[14] R. Reynolds, "Insights from Russian second language readability classification: Complexity-dependent training requirements, and feature evaluation of multiple categories," in *Proc. 11th Workshop Innov. NLP Building Educ. Appl.*, 2016, pp. 289–300.

[15] D. N. Rapp and P. van den Broek, "Dynamic text comprehension," *Current Directions Psychol. Sci.*, vol. 14, no. 5, pp. 276–279, Oct. 2005.

[16] H. W. Catts and A. G. Kamhi, "Prologue: Reading comprehension is not a single ability," *Lang., Speech, Hearing Services Schools*, vol. 48, no. 2, pp. 73–76, Apr. 2017.

[17] C. E. Snow, *Reading for Understanding: Toward an R&D Program in Reading Comprehension*. Santa Monica, CA, USA: RAND Corporation, 2002.

[18] S. E. Mol and A. G. Bus, "To read or not to read: A meta-analysis of print exposure from infancy to early adulthood," *Psychol. Bull.*, vol. 137, no. 2, pp. 267–296, 2011.

[19] F. Schubert and R. Becker, "Social inequality of reading literacy: A longitudinal analysis with cross-sectional data of PIRLS 2001 and PISA 2000 utilizing the pair wise matching procedure," *Res. Social Stratification Mobility*, vol. 28, no. 1, pp. 109–133, Mar. 2010.

[20] S. Krashen, S. Lee, and J. McQuillan, "An analysis of the PIRLS (2006) data: Can the school library reduce the effect of poverty on reading achievement," *CSLA J. (California School Library Assoc.)*, vol. 34, no. 1, pp. 26–28, 2010.

[21] R. L. Allington, "How reading volume affects both reading fluency and reading achievement," *Int. Electron. J. Elementary Educ.*, vol. 7, no. 1, pp. 13–26, 2014.

[22] M. O. Martin, I. V. Mullis, and A. M. Kennedy, "PIRLS 2006 technical report," TIMSS & PIRLS Int. Study Center, School Educ., Boston College, Chestnut Hill, MA, USA, Tech. Rep., 2007.

[23] OECD, *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*. Paris, France: OECD Publishing, 2010.

[24] C. Clark and K. Rumbold, "Reading for pleasure: A research overview," Nat. Literacy Trust, London, U.K., Tech. Rep., Nov. 2006.

[25] J. T. Guthrie, *Engaging Adolescents in Reading*. Thousand Oaks, CA, USA: Corwin Press, 2007.

[26] E. Dale and J. S. Chall, "The concept of readability," *Elementary English*, vol. 26, no. 1, pp. 19–26, 1949.

[27] B. D. Wright and M. H. Stone, *Making Measures*. Chicago, IL, USA: Phaneron Press, 2004.

[28] R. Routman, *Reading Essentials: The Specifics You Need to Teach Reading Well*. London, U.K.: Pearson, 2003.

[29] C. Clark and J. Douglas, *Young People's Reading and Writing: An In-Depth Study Focusing on Enjoyment, Behaviour, Attitudes and Attainment*. London, U.K.: National Literacy Trust, 2011.

[30] C. Clark, "Writing enjoyment, behaviours and attitudes in 8 to 11-year-olds in 2017/18," Nat. Literacy Trust Res., London, U.K., Tech. Rep., Apr. 2018.

[31] W. Kintsch and D. Vipond, "Reading comprehension and readability in educational practice and psychological theory," in *Perspectives on Learning and Memory*, L.-G. Nilsson, Ed. Hillsdale, NJ, USA: Erlbaum, 2014, pp. 329–365.

[32] E. A. Stevens, M. A. Walker, and S. Vaughn, "The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014," *J. Learn. Disabilities*, vol. 50, no. 5, pp. 576–590, Sep. 2017.

[33] J. F. Huerta, "Legibilidad y lecturabilidad: Dos conceptos básicos en los libros escolares," *Consigna (Revista pedagógica de la sección femenina de Falange ET y de las JONS*, vol. 213, pp. 3–8, 1958.

[34] J. Fernández, "Medidas sencillas de lecturabilidad," *Consigna*, vol. 214, pp. 29–32, 1959.

[35] D. C. Saavedra, P. C. Carmona, B. R. Ocares, M. Véliz, and A. R. Reyes, "Complejidad textual, lecturabilidad y rendimiento lector en una prueba de comprensión en escolares adolescentes," *Universitas Psychologica*, vol. 13, no. 3, pp. 1135–1146, Jan. 2014.

[36] J. M. De Sousa, *La Palabra y su Escritura*. Gijón, Spain: Trea, 2006.

[37] J. R. Diéguez, P. M. Berihuete, and M. V. C. Pérez, "Ecuaciones de predicción de lecturabilidad," *Enseñanza Teaching, Revista Interuniversitaria de Didáctica*, vol. 10, pp. 48–75, 1992.

[38] J. C. Ripoll, "Font legibility in first year primary students/Legibilidad de distintos tipos de letra en alumnos de primero de primaria," *Infancia y Aprendizaje*, vol. 38, no. 3, pp. 600–616, Jun. 2015.

[39] E. J. Pérez, "Comprensión lectora VS Competencia lectora: Qué son y qué relación existe entre ellas," *Investigaciones Sobre Lectura*, vol. 1, pp. 65–74, Jan. 2014.

[40] R. J. Spiro, B. C. Bruce, and W. F. Brewer, *Theoretical Issues in Reading Comprehension*. Abingdon, U.K.: Routledge, 2017.

[41] M. S. Demagistri, "Comprensión y producción lectora en adolescentes. Análisis comparativo de dos franjas etarias," in *Proc. Primer Encuentro de Grupos de Investigacion Sobre Procesamiento del lenguaje*, 2011, pp. 111–234.

[42] M. K. Ness, "Reading comprehension strategies in secondary content area classrooms: Teacher use of and attitudes towards reading comprehension instruction," *Reading Horizons*, vol. 49, no. 2, pp. 58–84, 2016.

[43] A. Morles, "El desarrollo de las habilidades para comprender la lectura y la acción docente," in *Comprensión de la Lectura y Acción Docente*. Madrid, Spain: Germán Sánchez Ruipérez Foundation, 1991, pp. 261–274.

[44] A. Davison and R. N. Kantor, "On the failure of readability formulas to define readable texts: A case study from adaptations," *Reading Res. Quart.*, vol. 17, no. 2, pp. 187–209, 1982.

[45] P. Sigaud-Sellos, "Aproximación a los conceptos de legibilidad y lecturabilidad: Aplicación a la lectura de textos digitales," Univ. Navarra, Pamplona, Spain, Tech. Rep., Jul. 2010.

[46] M. A. Tinker, "Legibility off print," *Optometry Vis. Sci.*, vol. 41, no. 2, p. 113, Feb. 1964.

[47] F. Alliende, *La legibilidad de los textos. Manual para la evaluación, selección y elaboración de textos*, A. Bello, Ed. Santiago, Chile, 2002.

[48] B. Mandelbrot, "Interview with benoît mandelbrot," in *Mathematical People: Profiles and Interviews*, D. Albers and G. L. Alexanderson, Ed. Cambridge, MA, USA: Birkhäuser, 1985.

[49] J. Tague-Sutcliffe, "An introduction to informetrics," *Inf. Process. Manage.*, vol. 28, no. 1, pp. 1–3, Jan. 1992.

[50] A. B. Pérez and U. G. Couto, "Legibilidad de las páginas Web sobre salud dirigidas a pacientes y lectores de la población general," *Revista Española de Salud Pública*, vol. 76, pp. 321–331, Aug. 2002.

[51] J. P. Kincaid, R. P. Fishburne, Jr., R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," Defense Tech. Inf. Center, Fort Belvoir, VA, USA, Tech. Rep. RBR-8-75, Feb. 1975.

[52] W. H. DuBay. (2004). *The Principles of Readability*. [Online]. Available: http://www.impactinformation.com/impactinfo/readability02.pdf

[53] P. Kidwell, G. Lebanon, and K. Collins-Thompson, "Statistical estimation of word acquisition with application to readability prediction," *J. Amer. Stat. Assoc.*, vol. 106, no. 493, pp. 21–30, Mar. 2011.

[54] L. Si and J. Callan, "A statistical model for scientific readability," in *Proc. 10th Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2001, pp. 574–576.

[55] K. Collins-Thompson and J. P. Callan, "A language modeling approach to predicting reading difficulty," in *Proc. HLT-NAACL*, 2004, pp. 193–200.

[56] L. Feng, N. Elhadad, and M. Huenerfauth, "Cognitively motivated features for readability assessment," in *Proc. 12th Conf. Eur. Assoc. Comput. Linguistics (EACL)*, 2009, pp. 229–237.

[57] D. Campos-Saavedra, "Lecturabilidad y Rendimiento lector en una prueba de comprensión en escolares," M.S. thesis, Dept. Basic Gen. Educ., Fac. Educ., Univ. Concepción, Santiago, Chile, 2014.

[58] A. Bernth, "EasyEnglish: A tool for improving document quality," in *Proc. 5th Conf. Appl. Natural Lang. Process.* Washington, DC, USA: Association for Computational Linguistics, Apr. 1997, pp. 159–165.

[59] K. W. Hunt, "Grammatical structures written at three grade levels," Champaign ILL, Nat. Council Teachers English, NCTE Res. Rep. 3, 1965.

[60] K. W. Hunt, "Recent measures in syntactic development," *Elementary English*, vol. 43, no. 7, pp. 732–739, 1966.

[61] J. Ure, "Lexical density and register differentiation," in *Proc. 2nd Int. Congr. Appl. linguistics Sel. Papers*, Cambridge, U.K., 1971, pp. 443–452.

[62] J. Kabatek, Ed. *Lingüística de Corpus y Lingüística Histórica Iberorrománica, Beihefte zur Zeitschrift für Romanische Philologie* (Edition Niemeyer). Berlin, Germany: Walter de Gruyter, 2016, p. 448.

[63] P. S. Nohales, T. M. Giménez, and E. V.-A. Gámez, "Controversia entre madurez lectora y enseñanza precoz de la lecturarevisión histórica y propuestas actuales," *Aula Abierta*, vol. 40, no. 3, pp. 3–14, 2012.

R. GIL ORTEGO (M'07–SM'18) received the M.Sc. degree in telecommunication engineering from Alcalá de Henares University, Madrid, Spain, in 2004, and the Ph.D. degree from the Industrial Engineering Technical School, Spanish University for Distance Education (UNED), Madrid, in 2010.

From 2005 to 2006, she was a Research Assistant with the Industrial Engineering Technical School, UNED. From 2006 to 2010, she received the National Scholarship for the Ph.D. degree. She conducted three research stays during the Ph.D. degree, including the Research and Development of Educational Technology (CRADLE), the Tokyo Institute of Technology, Tokyo, Japan, in 2007 and 2009, and the School of Engineering, Technology and Media, National University, San Diego, CA, USA, in 2010. From 2010 to 2015, she was an Assistant Professor with the Electrical and Computer Engineering Department (DIEEC), UNED, where she has been a Ph.D. Assistant Professor, since 2015.

Dr. Gil was the WIE Spain Chair, from 2011 to 2013. She was a recipient of the Best Paper Award from following conferences including ICCGI 2009, FINTDI 2011, and EDUCON 2012.

I. MARTÍNEZ SÁNCHEZ received the bachelor's degree in pedagogy, the master's degree in pedagogy and educational psychology, and the Ph.D. degree from the University of Granada, in 2013.

She has received the Scholarship for the University Teacher Training granted by the Ministry of Education for the completion of Ph.D. degree from the University of Granada, in 2013. From 2007 to 2015, she was with the University of Granada. Since 2015, she has been with the National University of Distance Education. Her research interests include the field of Higher Education, the evaluation of the quality of education and educational centers, and research methodology. She has presented several works in different international congresses on her research topics.

• • •