TOWARDS MULTILINGUAL READABILITY ASSESSMENT

by

Ion Madrazo Azpiazu

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Computer Science

Boise State University

December 2014

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Ion Madrazo Azpiazu

Thesis Title: Towards Multilingual Readability Assessment

Date of Final Oral Examination: 1st December 2014

The following individuals read and discussed the thesis submitted by student Ion Madrazo Azpiazu, and they evaluated the presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Sole Pera, Ph.D.	Chair, Supervisory Committee
Tim Andersen, Ph.D.	Member, Supervisory Committee
Nere Lete, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Sole Pera, Ph.D., Chair of the Supervisory Committee. The thesis was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

ACKNOWLEDGMENTS

The author wishes to express gratitude to Foo. This work would have been partially supported by some particular grant, if there was one.

ABSTRACT

An *abstract* is a brief summary of the document. A typical abstract provides a brief introduction, enough to provide context for the document, explains the purpose of the thesis or project, and summarizes the major results and conclusions. Keep in mind that a casual observer is likely to judge the content of the document by the abstract and title alone. (There is an old addage: "in a joke, the punchline comes at the end; in a paper [or thesis], it comes in the abstract.") A single concise paragraph usually suffices for the abstract. If it spills onto a second page, it is probably too long.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

LOL – Laughing Out Loud

OMG – Oh My God! now is the time for all good men to come to the aid of their country.

LIST OF SYMBOLS

$\sqrt{2}$	square root of 2
λ	lambda symbol, normally used in lambda calculus but it sometimes gets used for wavelength as well

CHAPTER 1

INTRODUCTION

Reading is an important skill in the academic environment, a competence that can be critical for students' educational opportunities and their careers [41]. As reported by Lennon and Burdick [32] reading for learning takes place when the reader comprehends 75% of a text. This represents an appropriate balance that allows the reader to positively understand the text, while also finding challenges in the reading process that will motivate him to improve his skills [32]. Outside the educational environment, reading generally takes place for comprehension rather than for learning. In this context, it is critical to provide people with texts they can fully understand. For example, patients that properly understand documents disclosed to them before surgery, are known to be less anxious before the operation and obtain more satisfactory results during posterior treatment [40]. However, recent studies[31, 40, 38] show that even medical documents that are supposed to be suited for average readers, tend to be too specialized and even well-educated adults have trouble understanding them. Whether for learning or understanding, the complexity of texts to be read needs to be determined.

Every reader has different reading skills and the levels of difficulty of the texts they need depends also upon their personal objective. Therefore, providing institutions and readers with tools that can measure the complexity of a text so that they can assess whether it is adequate for a user is imperative. *Readability Assessment (RA)* tools [1] are certainly aimed for handling such a task by providing a mean to determine the degree of ease with which a reader can understand a given text, i.e. the *Readability Score (RS)* of the text.

Historically, teachers have been the main stakeholders of RA formulas, using them to select new materials for their courses and curriculum design. However, lately, more stakeholders have found benefits in using RA tools outside the academic environment. Automatic text simplification[44, 42], summarization for people with reading difficulties[25], book recommendation [39], literacy assessment[47], or legal and medical document complexity assessment[31, 35, 40, 38] are only a few examples of applications that take advantage of the complexity levels generated by existing RA tools. Even in commercial environments, book publishers require professional linguistic services in order to tag their publications with a readability level required for their intended audience, a task that could similarly be completed by an automatic tool.

In estimating the complexity of texts, traditional formulas, such as Flesh [27], became very popular in the late 1940's among educators for manually determining text difficulty. Most of these formulas relied on *shallow features*, which could easily be adapted to multiple languages and provide a simple way of determining text complexity. The multilingualism achieved by traditional formulas offered numerous benefits in contexts where the readability of more than one language was needed, i.e., book translation or learning a second language. However, traditional formulas were known to lack precision. For example, they could classify nonsense text as *simple to read*, just because it contained short and frequently-used words [22]. The insufficient

[1] RA tool and RA formula are used interchangeably in this document.

precision encouraged researchers to study and develop better and more sophisticated methods for RA that depended upon more in-depth text analysis [28, 13]. These new formulas continued taking advantage of shallow features, but incorporated more complex features based on the syntax and semantics of text. With the addition of new text complexity indicators, the tools became more precise, but at the same time more constrained regarding their language adaptability [15, 26]. In fact, they used increasingly more language-dependent techniques, which made the systems unadaptable to estimate RS for text in languages other than the one they were designed for. As a result, the multilingualism that was possible in the early stages disappeared.

With multilingualism and precision in mind, we propose to develop **MRAS**, a **M**ultilingual **R**eadability **A**ssessment **S**ystem. This tool should both show results comparable to monolingual state-of-the-art systems and maintain the multilingualism the early tools in the RA field had. For doing so, we will (1) explore features and methods used in literature, (2) design novel features that positively influence the readability level estimating process and (3) analyze how all those features can be adapted to be used in multilingual RA. MRAS will be *open source* and *easily connected* to different applications that require RA as a service, potentially permitting the analysis of all sorts of texts, including text snippets, books, websites and even short and unstructured texts, such as the ones found in social media. In doing so, we will create a system that will adapt itself to the input text language and use an adequate subset of features for the corresponding language for readability prediction, creating, to the best of our knowledge, the first multilingual readability assessment system.

As a byproduct of our research work, we will create a leveled dataset with readability-labeled documents for different languages, which currently is unavailable. In addition, we will create an in-depth report surveying existing strategies for readability prediction.

It is important to note that, for practical purposes, the proposed application will only be tested in three different languages: *English*, for state of the art comparison purposes and as reference of germanic languages. *Spanish*, as a reference for romance languages, and *Basque* as an example of a pre-indoeuropean and minority language.

CHAPTER 2

RELATED WORK

From the past six decades, different RA systems have been developed with high diversity in terms of both languages and features [26, 15]. Initial readability formulas, such as Flesh [27], Dale-Chall [17], and Gunning FOG [12] made use of **shallow features**, mostly based on ratios of characters, terms, and sentences. These formulas, were basic enough even to be computed manually, providing a simple way of estimating a text's complexity, even if the formulas lacked precision in some cases [22]. This simplicity, however, made them easy to be adapted to estimate readability scores in different languages [43].

In recent years, readability formulas have evolved to supervised learning based systems that use a combination of traditional shallow features and new natural language processing based ones, which consider language aspects, such as syntax or semantics of texts. However, incorporating new features has brought a drawback to the area, evidenced by the fact that current systems are too focused in certain languages, making them only functional in the languages they were created for. Current state-of-the-art is composed by methods focused on specific languages, as discussed below:

For **English**, the RA system presented in [13] predicted only two levels of difficulty, simple or complex, using elaborated features, such as ambiguity among the terms in the texts. Other authors [25], oriented their system for assessing the difficulty level of a text for people with intellectual disabilities by developing features that were intended to detect how well a text was structured. A readability prediction system for finalcial documents was presented in [16], which was based on features such as the presence of active voice or number of hidden verbs. It is also important to mention two commercial RA tools, Lexile[1] and AR [2], which are widely used among English speaker academic professionals. Even if their algorithms are not public, they are known to use shallow features showing how common terms of a text are and how long sentences are in average [32]. The literature pertaining to RA for text in English is abundant. For more in-depth discussion on RA formulas refer to [26, 15].

In contrast to English, **Spanish** RA has not seen any significant improvement regarding features in recent years, as most of the existing works are still based on shallow features. Among the well-known RA tools for Spanish, SSR [43] was based on the analysis of sentence length and number of rare words per sentences, whereas LC and SCI [14] were based on density of low frequency words in text. Other systems [45, 24] presented strategies to combine the aforementioned methods to improve RA estimation.

Compared to other languages, **Basque** RA is reduced to only one system. Due to the fact that Basque is considered a minority language and shares little similarity with most spoken languages, limited research has been done in the area. So far, ErreXail [30] is the only system created for Basque RA. ErreXail was developed to predict two different readability values, simple or complex, using features mostly based on ratios of common natural language processing labels, such as Part-of-Speech tags or

[1] https://www.lexile.com/

[2] http://www.renaissance.com/products/accelerated-reader/atos-analyzer

morphology annotations.

Similar to Basque, the literature for **Arabic** RA is limited as well. Al-Ajlan and Al-Khalifa [11] developed a RA tool based on only two features: average letters per term and average terms per sentence. These, features were analyzed using a Support Vector Machine classifier in order to classify text as simple or complex.

Opposed to RA tools for previous languages, structural features do not seem to have such a success for **Chinese** RA. Therefore, most of the research works related to Chinese RA have been focused only on lexical features, such as Tf-Idf of terms [18, 19].

In contrast to the aforementioned techniques, the authors of [23] presented a RA system for **Italian** aimed at assessing readability at sentence level, which combined traditional, lexical, and syntactical methods.

Rather than focusing on the general reader, François and Fairon [28] developed a RA system for **French** with foreign language learners in mind. The objective was to determine which features were more important for a foreign language learner to understand a text. They tested lexical, syntactical and semantic features and showed that semantic ones performed poorly in their case.

Even if the number of RA systems that tackle individual languages is high, they are usually focused on a specific set of features and materials they can analyze. In addition, to best of our knowledge, none the RA systems presented are **multilingual**. MRAS will not only be multilingual, but will also be based on a comprehensive set of existing and novel features which will be general enough to potentially be able to handle all sorts of reading materials. All those characteristics will make MRAS a unique system in the area.

CHAPTER 3

PROPOSED METHOD

3.1 Overview

MRAS is based on a supervised learning approach that relies on knowledge acquired from a leveled corpora. In designing MRAS we followed the steps illustrated in Figure 3.1 and discussed below.

MRAS receives two different inputs: a collection C of documents for each of which a readability label is already assigned, and a document d which its readability is unknown for MRAS and, thus, will be predicted. Both inputs are taken through a preprocessing step described in section 3.3, which cleans, filters and normalizes their content, to prepare them for the feature extraction step described in section 3.4. These features, serve as a numeric representation for each document. MRAS is capable of learning patterns over the representations extracted from each document in C and use these patters to predict readability scores for new unlabelled documents such as d.

3.2 Tools

Whether for preprocessing or for feature extraction, MRAS takes advantage of several existing tools and techniques. Each of those tools is described below.

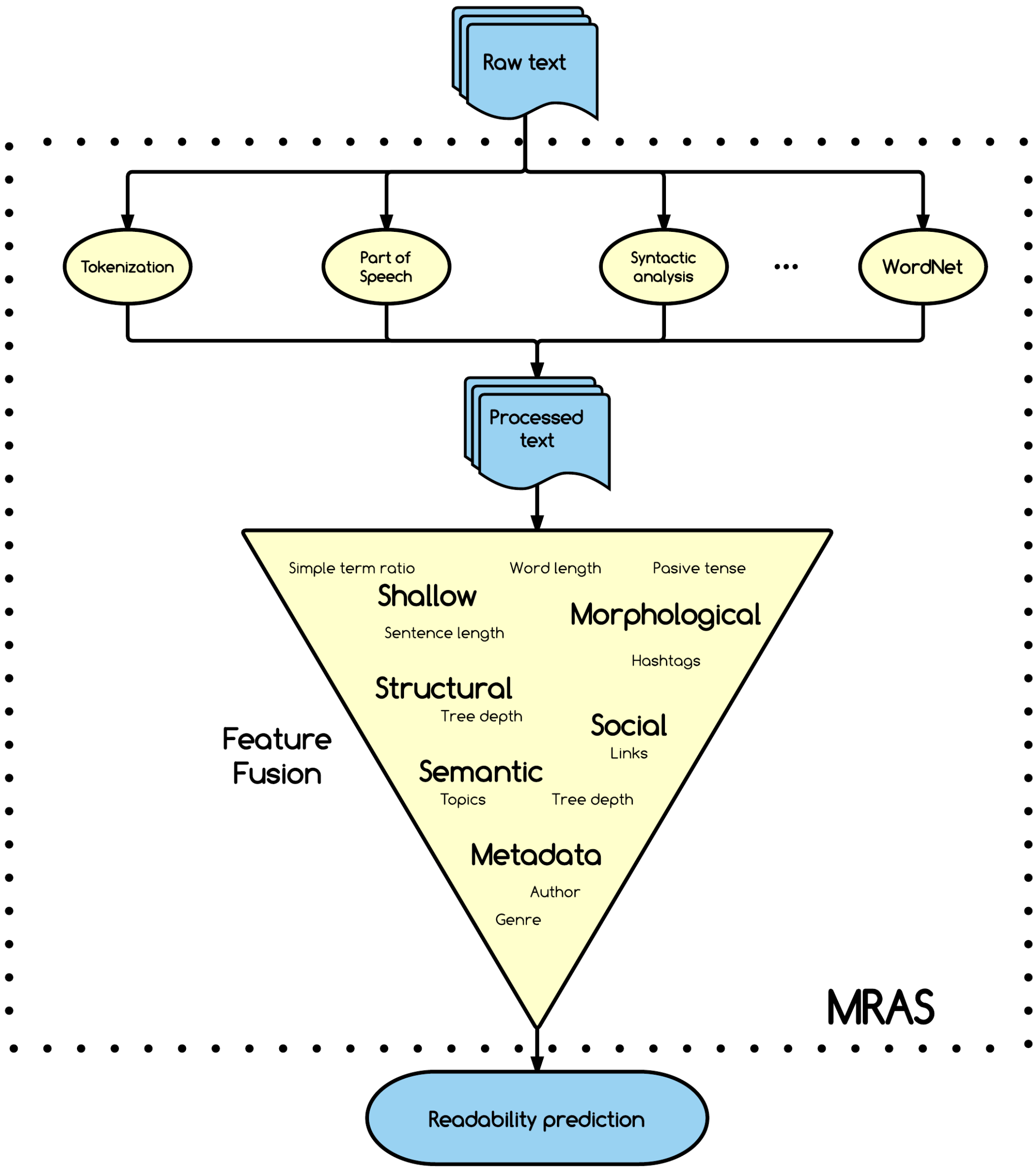

Figure 3.1: MRAS

Freeling NLP

WordNet

Latent Semantic Analysis

3.3 Preprocessing

3.3.1 Document type detection

variety of documents, each of them cannot be treated the same way, different strategies need to be applied for different texts. Therefore, each document used by MRAS is classified using 3 criteria: format, length and language.

should we go deeper? explain how each criteria is determined?

3.3.2 Tokenization

Tokenization is the process of splitting a text into smaller parts, i.e. tokens. A token represents each sensical part of a text and usually corresponds to a term, a number, or a punctuation mark. However, sometimes tokens can be formed by a combination of the previous, i.e., *aren't* or *people's*

Did they win the olimpics?					
did	they	win	the	olimpics	?

Table 3.1: Tokenization example

3.3.3 Part of Speech Tagging

sda

3.3.4 Shallow parsing

das

3.3.5 Depencecy parsing

asd

3.3.6 Named entity detection

sdfsadf

3.4 Feature extraction

3.4.1 Text processing

As the main focus of MRAS is to analyze text, we have identified different text processing methods and tools that will be used in its development. **Freeling NLP** [37, 36] is a multilingual natural language processing (NLP) toolkit that supports 11 different languages. This tool solves common NLP tasks such as, tokenization, sentence detection, part of speech tagging or dependency parsing. **WordNet** is a lexical database that takes advantage of semantic relations between terms to build a graph that is very convenient for semantic analysis tasks. **Latent semantic analysis** is also a commonly used strategy for semantic analysis, which takes advantage of concurrences among terms for determining similarities between them. All those tools, along with others that we will incorporate during the research process, will be used in the text processing step of MRAS.

3.4.2 Feature extraction

Exploring features will be one of the main tasks of this thesis. MRAS should be able to extract a wide range of features that satisfy the needs of each language it will tackle. A general description of the categories of features that we expect to incorporate in MRAS is presented as follows:

Shallow features [27, 17, 12] have historically shown to be of good use when predicting readability. Therefore, they will be incorporated into MRAS and used as a baseline for improvement. Sentence length, word length, or ratio of simple terms are examples of the features that will be included among this category.

Morphological features capture how terms are formed from their root. Even if this aspect is not relevant in some languages, such as English, it has been shown to be a strong predictor for readability scores in morphology rich languages such as Basque [30]. Different morphological phenomena will be analyzed in order to create features in this category.

Structural features are the ones that describe how a text is organized. They can both describe structure within the sentence (syntactical structure) or structure between sentences (pragmatical structure). Depth of the syntactic tree or ratios of different types of connectors between sentences are some examples of the features that are going to be explored under this category.

Semantic features go beyond the tokens and structure of the text in order to analyse the concepts laying on it. Features such as concept density or concept follow-ability are some examples of the features that will be analyzed under this category.

RA can be used in more than just plain text. Internet is evolving into a new social era and so are text resources. Increasingly more resources contain **social information**, such as hashtags, mentions, or links, a type of information that is usually ignored by readability formulas. We propose to investigate how the aforementioned information can be used for readability prediction.

Metadata based features can be useful in environments where text access is limited (i.e. copyrighted material). An exploration of this type of features will also be done in order expand the types of texts MRAS can handle.

3.4.3 Prediction

Individually, each of the aforementioned features can only provide a rough estimate of the readability of a text. However, considering these features in-tandem can lead to a more accurate and robust readability assessment. Consequently, we will analyze different fusion strategies for MRAS, which will make possible to identify the most suitable strategies for readability prediction. The problem of assessing readability can be seen as a classification problem where a discrete categorical class needs to be predicted. Therefore, we would like to explore different **classification** algorithms, such as bayesian networks [34] or support vector machines [20] for readability level prediction. The RA task can also be seen as a **regression** problem, given that the class contains an inherent order on it. Therefore, we would also like to test different regression algorithms, including, but not limited to, linear regression [33] and logistic regression [46]. Finally, we would also like to explore a **hybrid** approach by using classification algorithms that take order in the class into account, such as the ordinal classification approach presented in [29].

CHAPTER 4

EVALUATION

Even if MRAS is designed to be language independent, for practical purposes the evaluation will only be conducted in three languages that we think can faithfully represent the diversity of existing languages. For this purpose, we have chosen a germanic, a romance, and a pre-indioeuropean language, i.e. English, Spanish, and Basque respectively.

4.0.1 Datasets

The ideal dataset for developing MRAS would be a multilingual leveled dataset that would contain the exact same documents written in different languages, as well as human judgments, in terms of readability scores for each document. However, to the best of our knowledge, such a dataset does not currently exist. Consequently, we have identified various sets of leveled documents for each individual language that can suit MRAS' needs and can be used for evaluation purposes. Details on the datasets considered for evaluation purposes can be seen in Table 4.1.

4.0.2 Metrics

The performance of MRAS will be evaluated by means of (1) common classification evaluation methods, such as absolute error [21], (2) regression evaluation methods

	Dataset	Description
English	Lexile [1]	Contains book titles associated with its readability level
	Stantarized tests [2, 3]	Tests for English level, they contain various texts per test
	Other [4, 5, 6]	News for kids, exercises for learning English
Spanish	Lexile [1]	Contains book titles associated with its readability level
	Learning resources [7, 8, 9]	Various exercises for learning Spanish
Basque	Learning resources [1]	Various exercises for learning Basque
Multilingual	Parallel corpus [10]	Contains same texts translated into Spanish and English

Table 4.1: Data resources identified for MRAS development and validation

such as MSE (Mean Square Error) [21] and (3) methods common in the readability assessment domain, such as adjacent accuracy [28].

4.0.3 Overall Assessment

The study and performance analysis of this thesis will aim at answering the following questions:

- Which learning model performs better for MRAS? Which feature subset?
- Which features add more value in terms of predicting readability? Do they add same value for each language?
- How does MRAS perform compared to baseline shallow feature based formulas? and compared to state of the art systems?
- Would MRAS give the same prediction for a text that is translated manually into another language? and for a text that is automatically translated?
- How efficiently can MRAS predict the readability levels of written text in a language for which it has not been trained? If we train MRAS for two languages can we use it to predict the readability of a text in a third one?

- If we have a really small dataset for one language, would adding more data from another language improve the prediction results of the first one?

CHAPTER 5

MRAS IN ACTION

hashtag rec in twitter

CHAPTER 6

CONCLUSIONS

REFERENCES

[1] http://www.ikasbil.eus.

[2] http://www.flo-joe.co.uk/exams.htm.

[3] http://learnenglishteens.britishcouncil.org/content?field_language_level_tid=50&field_section_tid=1129&field_topics_tid=&language=en.

[4] https://www.readinga-z.com/books/leveled-books/.

[5] http://www.breakingnewsenglish.com/news-for-kids.html.

[6] http://www.newsinlevels.com/.

[7] http://cvc.cervantes.es/aula/lecturas/.

[8] http://aprenderespanol.org/lecturas/lecturas-ejercicios.html.

[9] http://www-k6.thinkcentral.com/content/hsp/reading/Senderos/na/common/online_senderos_libros_graduables_para_lectores/senderos_SE/launch.html.

[10] http://albalearning.com/audiolibros/textosparalelos.html.

[11] Amani A Al-Ajlan, Hend S Al-Khalifa, and A Al-Salman. Towards the development of an automatic readability measurements for arabic language. In *Digital Information Management, 2008. ICDIM 2008. Third International Conference on*, pages 506–511. IEEE, 2008.

[12] Judith Albright, Carol de Guzman, Patrick Acebo, Dorothy Paiva, Mary Faulkner, and Janice Swanson. Readability of patient education materials: implications for clinical practice. *Applied Nursing Research*, 9(3):139–143, 1996.

[13] Sandra Aluisio, Lucia Specia, Caroline Gasperin, and Carolina Scarton. Readability assessment for text simplification. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 1–9. Association for Computational Linguistics, 2010.

[14] Alberto Anula. Tipos de textos, complejidad lingüística y facilicitación lectora. In *Actas del Sexto Congreso de Hispanistas de Asia*, pages 45–61, 2007.

[15] Rebekah George Benjamin. Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, 24(1):63–88, 2012.

[16] Samuel B Bonsall, Andrew J Leone, and Brian P Miller. A plain english measure of financial reporting readability. *Available at SSRN 2560644*, 2015.

[17] Jeanne Sternlicht Chall and Edgar Dale. *Readability revisited: The new Dale-Chall Readability Formula*. Brookline Books, 1995.

[18] Yaw-Huei Chen, Yi-Han Tsai, and Yu-Ta Chen. Chinese readability assessment using tf-idf and svm. In *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on*, volume 2, pages 705–710. IEEE, 2011.

[19] Kevyn Collins-Thompson and James P Callan. A language modeling approach to predicting reading difficulty. In *North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 193–200, 2004.

[20] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[21] W Bruce Croft, Donald Metzler, and Trevor Strohman. *Search engines: Information Retrieval in Practice*, volume 283. Addison-Wesley Reading, 2010.

[22] Alice Davison and Robert N Kantor. On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly*, pages 187–209, 1982.

[23] Felice Dell'Orletta, Simonetta Montemagni, and Giulia Venturi. Read-it: Assessing readability of italian texts with a view to text simplification. In *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies*, pages 73–83. Association for Computational Linguistics, 2011.

[24] Biljana Drndarević, Sanja Štajner, Stefan Bott, Susana Bautista, and Horacio Saggion. Automatic text simplification in spanish: a comparative evaluation of complementing modules. In *Computational Linguistics and Intelligent Text Processing*, pages 488–500. Springer, 2013.

[25] Lijun Feng. Automatic readability assessment for people with intellectual disabilities. *ACM Special Interest Group on Accessible Computing*, (93):84–91, 2009.

[26] Lijun Feng, Martin Jansche, Matt Huenerfauth, and Noémie Elhadad. A comparison of features for automatic readability assessment. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 276–284. Association for Computational Linguistics, 2010.

[27] Rudolph Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221, 1948.

[28] Thomas François and Cédrick Fairon. An ai readability formula for french as a foreign language. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 466–477. Association for Computational Linguistics, 2012.

[29] Eibe Frank and Mark Hall. *A Simple Approach to Ordinal Classification*. Springer, 2001.

[30] Itziar Gonzalez-Dios, Marıa Jesús Aranzabe, Arantza Dıaz de Ilarraza, and Haritz Salaberri. Simple or complex? assessing the readability of basque texts. In *Proceedings of International Conference on Computational Linguistics*, volume 2014, 2014.

[31] Ahmed MS Ibrahim, Christina R Vargas, Pieter GL Koolen, Danielle J Chuang, Samuel J Lin, and Bernard T Lee. Readability of online patient resources for melanoma. *Melanoma Research*, 26(1):58–65, 2016.

[32] Colleen Lennon and Hal Burdick. The lexile framework as an approach for reading measurement and success. *Electronic publication on https://cdn.lexile.com/m/resources/materials/Lennon__Burdick_2004.pdf*, 2004.

[33] John Neter, Michael H Kutner, Christopher J Nachtsheim, and William Wasserman. *Applied Linear Statistical Models*, volume 4. Irwin Chicago, 1996.

[34] Thomas Dyhre Nielsen and Finn Verner Jensen. *Bayesian Networks and Decision Graphs*. Springer Science & Business Media, 2009.

[35] James RP Ogloff and Randy K Otto. Are research participants truly informed? readability of informed consent forms used in research. *Ethics & Behavior*, 1(4):239–252, 1991.

[36] Llus Padr, Miquel Collado, Samuel Reese, Marina Lloberes, and Irene Castelln. Freeling 2.1: Five years of open-source language processing tools. In *Proceedings of 7th Language Resources and Evaluation Conference (LREC'10)*, La Valletta, Malta, May 2010.

[37] Llus Padr and Evgeny Stanilovsky. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey, May 2012. ELRA.

[38] Chirag R Patel, Saurin Sanghvi, Deepa V Cherla, Soly Baredes, and Jean Anderson Eloy. Readability assessment of internet-based patient education materials related to parathyroid surgery. *Annals of Otology, Rhinology & Laryngology*, pages 523–527, 2015.

[39] Maria Soledad Pera and Yiu-Kai Ng. Automating readers' advisory to make book recommendations for k-12 readers. In *Proceedings of the 8th ACM Conference on Recommender Systems*, pages 9–16. ACM, 2014.

[40] Jennifer Petkovic, Jonathan Epstein, Rachelle Buchbinder, Vivian Welch, Tamara Rader, Anne Lyddiatt, Rosemary Clerehan, Robin Christensen, Annelies Boonen, Niti Goel, et al. Toward ensuring health equity: Readability and cultural equivalence of omeract patient-reported outcome measures. *The Journal of Rheumatology*, 42(12):2448–2459, 2015.

[41] Richard D Robinson, Michael C McKenna, and Judy M Wedman. Issues and trends in literacy education. 2000.

[42] Horacio Saggion, Sanja Štajner, Stefan Bott, Simon Mille, Luz Rello, and Biljana Drndarevic. Making it simplext: Implementation and evaluation of a text simplification system for spanish. *ACM Transactions on Accessible Computing (TACCESS)*, 6(4):14, 2015.

[43] Seth Spaulding. A spanish readability formula. *The Modern Language Journal*, 40(8):433–441, 1956.

[44] Sanja Štajner, Ruslan Mitkov, and Gloria Corpas Pastor. Simple or not simple? a readability question. In *Language Production, Cognition, and the Lexicon*, pages 379–398. Springer, 2015.

[45] Sanja Štajner and Horacio Saggion. Readability indices for automatic evaluation of text simplification systems: A feasibility study for spanish. In *Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013), Nagoya, Japan*, pages 374–382, 2013.

[46] Strother H Walker and David B Duncan. Estimation of the probability of an event as a function of several independent variables. *Biometrika*, 54(1-2):167–179, 1967.

[47] Barry D Weiss, Mary Z Mays, William Martz, Kelley Merriam Castro, Darren A DeWalt, Michael P Pignone, Joy Mockbee, and Frank A Hale. Quick assessment of literacy in primary care: the newest vital sign. *The Annals of Family Medicine*, 3(6):514–522, 2005.