A Linguistic Approach To Health Literacy In Brazil: Terminological Aspects[1]

Liana Braga Paraguassu
Universidade Federal
do Rio Grande do Sul
ORCID: 0000-0003-4043-1836
liana@linguatraducoes.com.br

Investigadora i consultora de les
àrees de Comunicació Accessible en Salut i Alfabetització en Salut. Doctoranda i màster en l'àmbit d'Estudis del Llenguatge dins del programa de Postgrau de Lletres de la Universitat Federal de Rio Grande do Sul (UFRGS). També ha format part de l'equip guanyador del Google Latin America Research Awards 2019 pel sistema MedSimples.

Maria José Bocorny Finatto
Universidade Federal
do Rio Grande do Sul
ORCID: 0000-0002-6022-8408
mariafinatto@gmail.com

Abstract

Health Literacy (HL) is a fairly new area of research in Brazil that is usually conducted by healthcare professionals. In this research, we aim to explore the contributions of Language and Terminology Studies to fostering new understandings on the concept of HL considering the socioeconomic and educational challenges experienced in Brazil. In addition to providing an overview of HL in Brazil, this paper also reviews a set of experiments made while developing MedSimples, a system designed to help promote HL using plain Portuguese.

KEYWORDS: terminology; health literacy; plain Portuguese; MedSimples

Professora titular de la Universitat
Federal de Rio Grande do Sul (UFRGS). Membre de la Comissió Brasilera de Terminologies cientificotècniques de l'Institut Internacional de la Llengua Portuguesa, per designació del Ministeri de Relacions Exteriors de Brasil, des de 2019. Col·laboradora del Consell Nacional de Desenvolupament Científic i Tecnològic del Brasil, en matèria de productividad de la recerca des de 2007. Guanyadora del premi Google Latin America Research Awards 2019, per la proposta del sistema MedSimples, una eina d'ajuda per a la simplificació de textos relacionats amb la salut.

Resum

Un enfocament lingüístic de l'alfabetització sanitària al Brasil: aspectes terminològics

L'Alfabetització en Salut (AS) és una àrea d'investigació nova al Brasil que normalment duen a terme professionals de la salut. En aquesta investigació, pretenem explorar les contribucions dels Estudis de Llengua i Terminologia per a fomentar noves visions sobre el concepte d'AS tenint en compte els reptes socioeconòmics i educatius al Brasil. A més de proporcionar una visió general d'AS al Brasil, aquest article també revisa un conjunt d'experiments realitzats durant el desenvolupament de MedSimples, un sistema dissenyat per ajudar a promoure l'AS utilitzant el portuguès planer.

PARAULES CLAU: terminologia; alfabetització en salut; portuguès planer; MedSimples

A Linguistic Approach To Health Literacy In Brazil: Terminological Aspects

Liana Braga Paraguassu i Maria José Bocorny Finatto

1   Introduction

This paper discusses the results of an ongoing doctoral[2] research that relates different contributions from Language and Terminology Studies in order to foster new understandings on the concept of Health Literacy (hereinafter HL) in Brazil. In addition to providing an overview of HL in Brazil, this paper also reviews a set of experiments made while developing MedSimples system (Paraguassu et al. 2020; Zilio et al., 2020). MedSimples, in its design and functionality, is one of the inputs for the ongoing research.

MedSimples is a text processing system available for free on a website dedicated to th e theme of text and terminological accessibility (Finatto and Motta, 2019). The system was built to support communication mediators while writing facilitated texts on Health-related topics. MedSimples helps writers create and edit texts according to the needs of Brazilian adults with limited education and little experience in reading texts in general. MedSimples (hereinafter FMed) puts into practice important theoretical points of our understanding, as linguists and terminologists, of new scopes for the linguistic-terminological mediation of communication.

As a frame to situate FMed, we have the promotion of HL. HL is a multi-, inter and transdisciplinary social and political mission that has taken on a whole new meaning in face of the pandemic. Although, in Brazil, HL is more commonly addressed by health professionals, it is indisputable that language is a fundamental part of it, and for that it is also worthy of attention from linguists and terminologists.

In this trajectory, we argue that our studies on Literacy in general generated in the area of Education and based on robust research on Teaching, Reading and Literacy in Brazil can be combined with Terminology and Translation Studies. This confluence of different perspectives can highlight new paths for the construction of an effective HL in our country, where HL is both an idea and a set of actions that take into account linguistic and cultural diversity, socio-historical heritages and multifaceted socioeconomic aspects of a country-continent like ours.

In order to achieve this goal, we have a peculiar situation to overcome, only 12% of the Brazilian adult population can be considered sufficiently proficient in reading comprehension (INAF, 2018). Therefore, our intention is to collaborate with ideas and proposals to facilitate written information on topics of public interest, especially in the areas of Healthcare, Civil Rights and Science Teaching.

In the next sections of this paper, we will explore the following topics: the Brazilian perspective of HL amidst a historical trajectory of studies; the conception of simplification as a type of intralingual translation connected to the ideal of text and terminological accessibility (hereinafter ATT). After that, we present a summary on how terminology was simplified/translated before being incorporated into the current functionalities of FMed. Finally, we will show preliminary results from a series of tests made with FMed users-writers (content producers) and end readers (readers of the content produced with the help of FMed), as well as the current stages of the research and future perspectives.

2   Health Literacy and other types of literacy in Brazil

According to a fairly recent definition provided by WHO (2016),[3] HL is the ability of individuals to "gain access to, understand and use information in ways which promote and maintain good health for themselves, their families and their communities". As we can see by this definition, HL is much more than an individual's ability to read and understand leaflets and nutrition labels, make doctor appointments, or follow prescriptions. HL means expanding the individual's sphere to reach their community, which involves, for example, being able to mobilize their community to deal with adverse social, economic, and environmental conditions more effectively.

These more recent definitions of HL are in line with a linguistic, psycholinguistic, and educational concept of literacy in general. In Brazil, the concept of literacy [*letramento*] was introduced in the 1980s by linguists to embrace the new Education Studies that were arising (Silva, 2009). The word *letramento* was added to the Brazilian Portuguese educational vocabulary to embrace new concepts by refining and going beyond the ideas of *alfabetização* nurtured by Paulo Freire. Roughly speaking, *alfabetização* is the ability to read and write and is related with some formal school experience.

Alfabetização is a gradual process. For example, it begins when the learner can perceive that a word such as *insecticide* is comprised of two meaning segments: [insect] + [cide] 'that kills'. After a simpler or initial decoding phase, a person, when faced with another word like *fungicide*, will be able to identify and understand the differentiated linguistic value of the *fungi*- segment and give meaning to *-cide*, even though they do not understand that the segment is a derivation of the word *fungo*, which initially refers to something that is not the same as *insect*. Finally, when a person manages to distinguish *fungi* and *insects*, in their different senses and concepts, this means their level of *alfabetização* has advanced significantly, at least in terms of lexicon.

However, our concept of *letramento* is not limited to the ideas of *alfabetização* or formal schooling, nor is strictly conditioned by them. That is, in Brazil, a person can have several years of formal school experience and go to university, but not necessarily be fully literate, as they are not able to apply the knowledge to which they were exposed in real life.

In practice, a vast number of people in Brazil are unable to read and understand informational materials that address the dangers of handling *insecticides* and *fungicides*, even though this type of information is essential to a great portion of our population, as farming is a major source of income in our country. On the other hand, when readers are capable of being aware and responsive to the information received in written and oral format, they become "polyglots" within their own language, and this means they have acquired *letramento*.

According to Estopà (2021), the promotion of Literacy in Health – which the author calls, in the European context, *Alfabetización en Salud* – can be understood as a process of culturalization in Health, as it generates changes in how people deal with their illness and, above all, in the relationship established between doctor and patient.

In Brazil, research in HL tends to be led by health professionals, without a broad participation of linguists. Unfortunately, in our country HL is still an incipient area of research and has not received the same investment as in other countries. Fortunately, in the past few years, the level of interest in HL has grown, and new studies are emerging, especially in our public universities, medical associations, and the Oswaldo Cruz Foundation (Fiocruz). These are important studies that include, for example, the adaptation of international Health Literacy tests (Maragno et al., 2019) to Brazilian Portuguese to be used with patients from our public health system. Nevertheless, we have noticed generic definitions about HL and attempts to replicate experiences validated in North America and Europe without great contextualization for the Brazilian reality.

As we can see in Figure 1, while in countries such as the United States the average citizen, in 2017, spent 13.4 years in school and, in Germany, 14.1 years, in Brazil the mean years of schooling was 7.8 years. Although there is not a definite relationship between years in school and full literacy, as the construction of knowledge does not occur linearly, the chances of a Brazilian developing different literacies – from linguistic to scientific literacy – tend to be hampered by the lack of investments and the low quality of education associated with fewer years of schooling.

Through a series of empirical observations, within Language Studies (Carvalho, 2020) and Health Literacy (Baker, 2006; Sampaio, 2019), evidence support that a substantial portion of the Brazilian population does not have the foundations needed to understand texts on Health topics as the information is currently conveyed. One of the fundamental causes, which has become commonplace, is associated with the historical deficiencies of our formal educational system.

The fragilities of our educational system have long been pointed out by scholars from different fields of knowledge (Soares, 2009; Passamai et al., 2019). Regarding its effects, concerning misinformation in Health, unfortunately it is not difficult to collect data about the missteps of healthcare institutions as well as healthcare professionals. As an emblematic example, among many, we have the various campaigns for vaccination, such as those for HPV, which prevents a type of cancer. For these campaigns, government institutions often distribute "informational" materials that are full of unexplained and hardly understandable terminology. As a result, we have witnessed a frightening lack of support from the population, with low adherence to vaccination, even from people who are considered highly educated and literate in our country.

To better understand the profile of our readers, it is worth examining the different elements that help
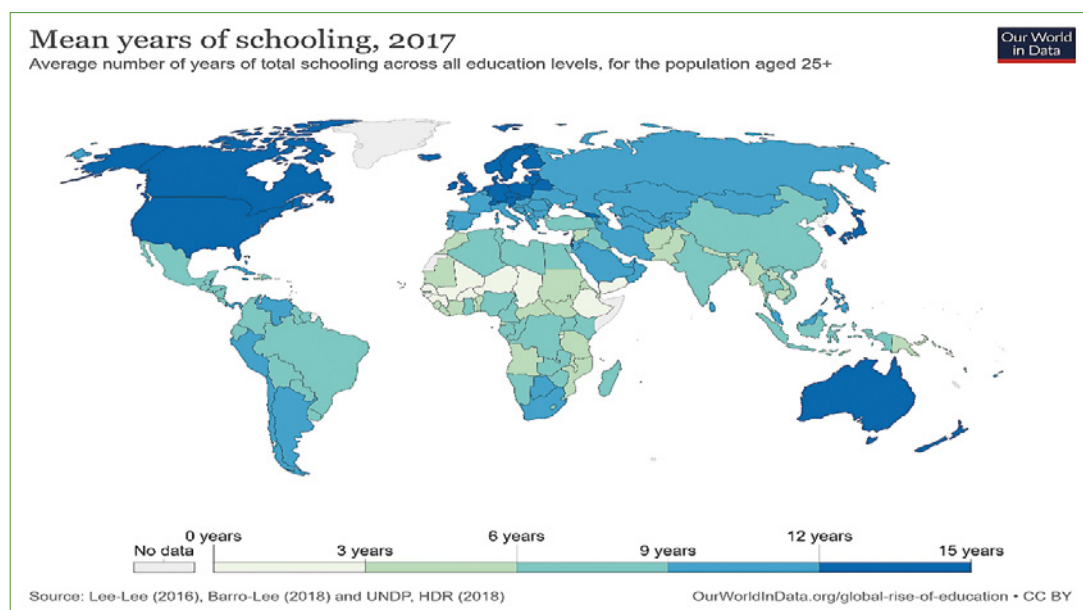

FIGURE 1. Mean years of schooling in the world in 2017. Source: *https://ourworldindata.org/covid-vaccinations*

us to outline a literacy framework of our population. According to a latest study conducted by INAF (Brazilian Institute for Literacy Studies) in 2018, three out of ten (29%) Brazilians (38 million people), aged between 15 and 64 years old, are considered functionally illiterate, which means they have a hard time understanding and expressing themselves through words and numbers in everyday situations, such as giving change at the grocery store or identifying the main information on a vaccination poster (Instituto Paulo Montenegro, c2018, *online* document).

In this context, linguists, whether translators, professional writers or researchers in Reading and Education, who have the necessary knowledge to recognize linguistic barriers in a text and how to overcome them, will have a fundamental role in mediating communication. In Brazil, literacy (*letramento + alfabetização*) is an area of concern in Language Studies, especially Applied Linguistics, while Terminology has been dedicated to different phenomena of specialized communication and recognizes the principle of terminological adequacy (Cabré, 1999). By combining these two research areas, our goal is to try to understand the meaning of HL in Brazil and how it could be implemented in the midst of such a peculiar scenario of adversity.

3   Simplification as intralingual translation

According to Jakobson (1959: 233), there are three basic types of translation: interlingual translation, intersemiotic translation and intralingual translation. From this trinomial, we highlight the type of translation that has to do with the simplification of texts and which, as stated by Zethsen (2016: 693), aims to overcome internal barriers in the same language, the *intralingual translation*.

When we rewrite a text in Brazilian Portuguese, in our case, to overcome certain barriers to reading comprehension, we are, in a way, *translating* this text to those who do not fully comprehend all the aspects involved in such text. When we rewrite a highly specialized text for scientific dissemination, we try to break possible barriers of specialized language. Thus, we see intralingual translation as a form of text simplification that requires rewriting and rewording (Paraguassu, 2019: 80-81). The greatest challenge is transforming a source text – whether it comes from academic scientific writing or even from knowledge that presents itself as scientific dissemination – into something that, although "non-specialized", still preserves the original conceptual adequacy and accuracy.

In different Terminology studies and research, various linguistic strategies aiming to translate or simplify specialized texts in the direction of more or less efficient scientific dissemination have already been described (Ciapuscio, 2008; Saggion 2017) and validated (Silva Cortina, 2018). In our case, as we deal, at the same time, with literacy gaps (*letramento + alfabetização*), we began to explore different strategies for text facilitation. Some of them of a synthetic nature and some of a semantic nature, concerning the vocabulary and the global text and discourse structure.

Here are two examples of lexical simplification by substitution (1) and by explanation (2), which involve terminologies.

> Ex. 1: The patient has <u>*nephrolithiasis*</u>. → The patient has <u>kidney stones</u>.
> Ex. 2: The patient has <u>*nephrolithiasis*</u>. → The patient has <u>*nephrolithiasis*</u>, a disease more popularly known as <u>kidney stones</u>.

Syntactic simplification deals with structural issues in the text that may hinder reading comprehension. For example, a sentence in passive voice (1), which is likely to be more difficult for certain readers, can be changed into active voice, which is usually easier to understand, as the subject of the action becomes more evident. Examples of syntactic changes for text simplification may include changing subordinate clauses (2) into independent and shorter sentences, facilitating the identification of the subject in the sentence.

> Ex. 1: The patient was examined by the attending physician. → passive voice
> The attending physician examined the patient → active voice
> Ex. 2: The patient, who has been experiencing constant pain in the arm, was referred to a specialist. → subordinate clauses
> The patient has been experiencing constant pain in the arm. The patient was referred to a specialist. → independent clauses.

However, the simplification process of a text that is theoretically complex will involve many variables, elements that go beyond simply following programmatic recommendations inherited from the *Plain Language* guidelines (Paraguassu, 2019: 99-103), even if they have already been adapted to Brazilian Portuguese (Cortina Silva, 2018). There is no "magical" formula that guarantees the success of a text simplification. What we can do is to rely on linguistic studies and data that will support the writer's decision-making. For example, we have "structural" linguistic tests valid for Brazilian Portuguese (Cortina Silva, 2018: 178-198). These tests can be applied by the writer to check whether scores that indicate semantic and syntactic complexity were reduced in the simplified version of the text. In this case, these tests use linguistic measurements based on lexical-statistical patterns observed in current *corpora*. These studies on the readability and complexity of texts (Aluísio and Gasperin, 2010) made with computational support and linguistic foundation were conducted by Brazilian researchers

during the PorSimples project for the Simplification of Portuguese Texts.

Below, we exemplify how the analysis of terminologies and texts can help mediate communication (Conceição and Zanola, 2020). To this end, we present FMed, a system that helps users write simplified texts on healthcare topics for people with limited education and literacy.

4 Terminologies in the FMed

The Terminology studies that underlie our research have a textual and communicative perspective (Finatto, 2020), that is, the terminologies are perceived, first of all, in a global discourse setting, which integrates, in its specificities, a given genre and a series of conventional features. Thus, as the Communicative Theory of Terminology (Cabré 1999) emphasizes, the words of a text receive the status and value of a term only in a given discursive context of specialized communication.

With this framework in mind, we have designed and built FMed (Paraguassu et al. 2020; Zilio et al., 2020). The system focuses on two reader profiles: i) Type 1, adult reader with 9 years of formal education; ii) Type 2, adult reader with at least 13 years of formal education. The system *workflow* is summarized as follows:

1. The input is a text provided by the user-writer who selects: *a*) one of the health-related themes available in the system (currently Covid-19, Parkinson's Disease, Newborn Care); *b*) the end reader profile for which they intend to write.
2. The user's text then is transformed into a sequential lemmatized list of words and multiword expressions (MWEs). For the lemmatization, FMed uses a resident parsing system, the PassPort (Zilio et al., 2018). PassPort is a free open system for Brazilian Portuguese lemmatization.
3. Each item in the lemmatized list is syntactically and semantically tagged and connected to its correspondences in three categorized sub-lists – or databases – which contain, respectively: *a*) common words and MWEs that are considered potentially **simple** for the average reader; *b*) common words and MWEs that are considered potentially **difficult** for the average reader *c*) **terminologies** and **specialized MWEs** that are considered potentially complex for the average reader.
4. Each potentially difficult item underlined in *b*) (blue) and *c*) (green) receives a suggestion for simplification. When a common word or MWE is considered difficult, the system suggests a synonym based on an ordered database/listing of facilitated synonyms. If there is no synonym to provide, an alert is displayed on the screen. For the terminologies and/or specialized MWEs, the suggestions come from the system glossaries, that is, a list of terms and their simple definitions/

explanations per domain. If the term is not found in the database, an alert message is displayed.

5. As a result, the output text points out potentially complex words within the common and specialized lexicon that require the user's attention. Suggestions and alerts about the need for facilitation are displayed to the user in the form of pop-ups. The suggestions and alerts can then be assessed by the user-writer, who may or may not include them in the final version of the text.
6. The user-writer exports the text with the suggestions in TXT format so that they can finish editing it outside the system.

Currently, FMed offers simplified definitions or explanations for terminologies within 3 different Health-related subjects: *a*) Parkinson's Disease/Neurology; *b*) Covid-19/Infectiology; *c*) Newborn Care/Pediatrics.

As the FMed system also highlights potentially difficult words that are part of the common lexicon, suggesting more accessible synonyms, it was necessary to build a controlled vocabulary. This vocabulary was based on a word frequency threshold associated with a *corpus* of Popular Brazilian Portuguese, CorPoP (Pasqualini and Finatto, 2018). CorPop is comprised of texts selected based on the average level of literacy of Brazilian readers, serving as reference for Plain Portuguese.

FMed's main users-writers are content producers such as journalists, healthcare professionals, technical writers, among others. In general, highly educated and literate people who are able to recognize that some texts are very difficult to some people and wish to facilitate these texts for an audience that, for different reasons, have comprehension limitations.

The system is not exclusive to users-writers and can also be harnessed by the *end reader* of a health-related text available on the Internet. But, in this case, some skills and knowledge that go beyond basic literacy are required, as the end reader will need to have a certain level of computer and digital literacy.

FMed is part of a project funded by CNPq – Brazilian Council for Scientific and Technological Development – since 2016. The first modules to be developed were Parkinson's Disease (PD) and Pediatrics. In 2019, after receiving a distinction of merit granted by Google (Lara 2019), it was possible to advance the research by: *a*) incorporating new themes, such as Covid-19; *b*) expanding the coverage of PD terms; and *c*) adding specification to the module of Pediatrics with Newborn Care.

Currently, in 2021, the system offers more than 3,000 terms with simplified definitions using CorPop as reference for Plain Portuguese. The definitions/explanations are reviewed and validated by consultant physicians. In FMed internal glossaries, there are 1,793 terminology entries on PD, +450 on Covid-19, and +850 on Pediatrics. Currently, we have been working to expand the themes and sub-themes by adding a module on Oncology. Figure 2 below illustrates FMed processing a text on the Covid-19 Module.
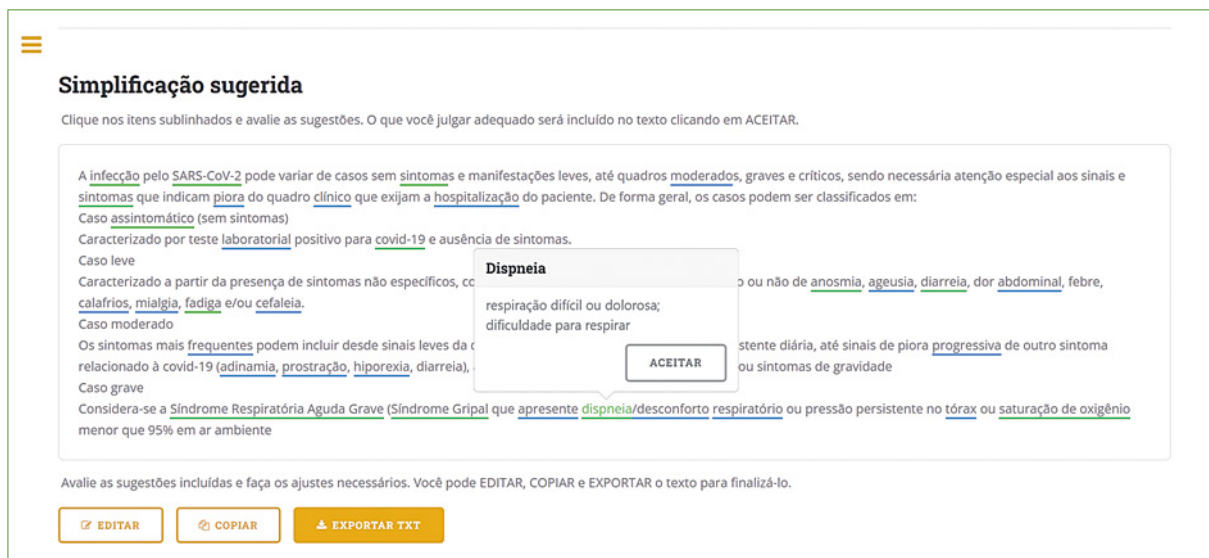
FIGURE 2. A text on Covid-19 processed on FMed system. Source: *http://www.ufrgs.br/textecc/acessibilidade/page/cartilha/*

It is worth mentioning the criteria used to differentiate *terminology* and *difficult words*. The underlined terminologies are those pre-recognized in *corpora* specially collected and studied for each subject area and specialty. The suggested definitions for the terms consider two types of readers, Type 1 reader and Type 2 reader, where Type 1 reader corresponds to the majority of the Brazilian population, that is, people with Elementary School or averagely 8-9 years of formal education. Type 2 reader corresponds to people who have finished High School or more.

FMed combines three types of internal word lists. The content and nature of these lists are summarized in Table 1 on the next page and detailed at *http://www.ufrgs.br/textecc/acessibilidade/files/origem%20sugest%20med-simples%20-%20texto%20final.pdf*.

In view of the proposal of a system to help simplify texts on specific health-related subjects, we have not planned a glossary that covers all the subjects together, but separate computational lexicons for each subject. In the case of PD, for example, we started with a specific glossary containing approximately 100 terms that was later expanded to +1750 items, including nuclear and peripheral terms. The expansion came from the addition of new *corpora* and consideration of two reader profiles, different uses, and related sub-themes in order to improve coverage and the different facilitation needs that had emerged. As a result, new terminologies were added, such as terms from Biochemistry, Nutrition, as well as the frequent comorbidities in the elderly, etc. Figure 3 on the next page illustrates the core terms and the peripheral terms in the PD domain.

As we can see in Figure 4, the greater the number of peripheral terms, the better the system performs. This is because the system is then capable of processing a greater variety of texts on a given theme. Currently, the FMed PD module has shown the best performance

according to a computational test conducted in 2020 (Zilio et al., 2020). The PD Module has a coverage of +1750 terms, of which approximately +1300 are peripheral terms. The Covid-19 module, on the other hand, has a coverage of +450 terms, from which the vast majority is part of the main core, as we have not yet been able to include a new repertoire of peripheral terms.

4.1 Definitions/explanations of FMed internal glossaries by domain

To propose simplified definitions/explanations on the themes covered by FMed, we always rely on trustworthy sources. We select materials available on-line and authored by healthcare professionals or by properly identified health institutions, such as the Brazilian Ministry of Health and Fiocruz. We also use printed materials such as popular science texts written by healthcare professionals, handbooks, glossaries, and dictionaries in the medical field. However, when we use materials that do not target laypersons, we have to make extra effort throughout the simplification process.

As already mentioned, we work with two reader profiles. It is worth emphasizing that each type of reader, Type 1 and Type 2, is associated with a different version of the internal glossary for each domain. The PD module works with these two profiles, while Covid-19 and Newborn Care currently offer only Type 1 Reader (Elementary School).

The construction of the internal glossaries was divided into different stages. At first, we searched for definitions in recognized medical websites, dictionaries, glossaries, and medical handbooks, as well as repertoires of terms for laypersons in websites of institutions such as WHO, CDC, Fiocruz and the Brazilian

List of simple words	List of difficult words	List of terms of the selected domain (internal glossary)
List of most frequent words on CorPop. CorPop texts include tabloids and fictional literature texts formatted for readers with little reading experience. When identified in the text inserted on FMed, the words in this list are not underlined and no suggestions are given, as they are likely to be accessible to a Type1 reader. Examples from this list are SALADA, CADEIRA, LIVRO, DE, E PARA.	List based on the Thesaurus of the Portuguese Language (TeP 2.0) (Maziero, Pardo 2008). This thesaurus was filtered and restructured based on CorPop simple word list. If a given item occurs five (5) times or **less** in CorPop, it is indicated as potentially difficult on FMed. When a difficult word is detected, it is underlined in blue and an ordered series of simpler synonyms is shown, also extracted from the combination of TeP 2.0 and CorPop. When there is not a simpler synonym available, the alert "difficult word, please review" is displayed. Examples from this list are EXARAR, EXIMIR, PROPAGAR, MAGNITUDE, TODAVIA.	Lists created based on popular science texts for the lay audience in general; scientific articles and drug leaflets on the themes covered by FMed. The glossaries (1) Parkinson's Disease, (2) Covid-19 and (3) Pediatrics - Newborn Care have a core of specific terms as well as what we call peripheral terms (See Fig. 3). When any of the glossary items is identified in the text, simplified definitions are shown, which can be dismissed or incorporated into the text by the user-writer. The list of terms extracted from the three Health + Medical *corpora* is more extensive than the list of terms + definitions. Therefore, when a term is identified among the words in the text and there in not a definition in the database, a "Further research required" alert is displayed.

TABLE 1. Content of FMed internal word lists. Source: Created by the authors

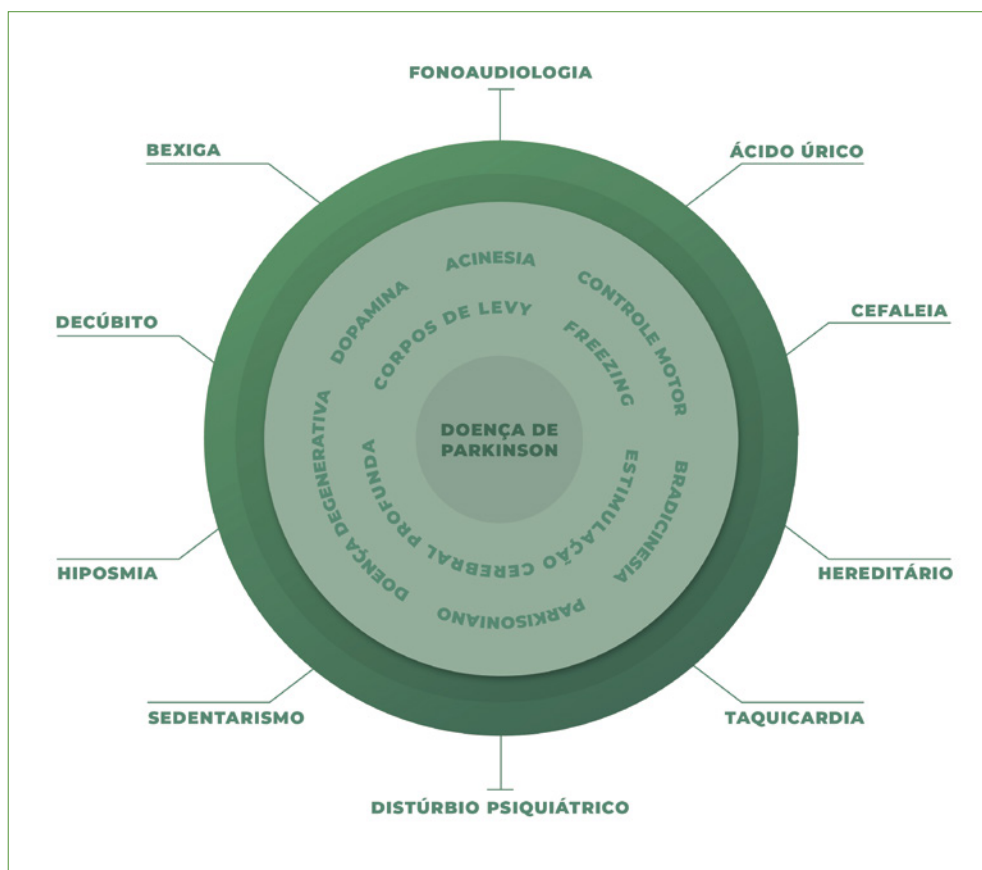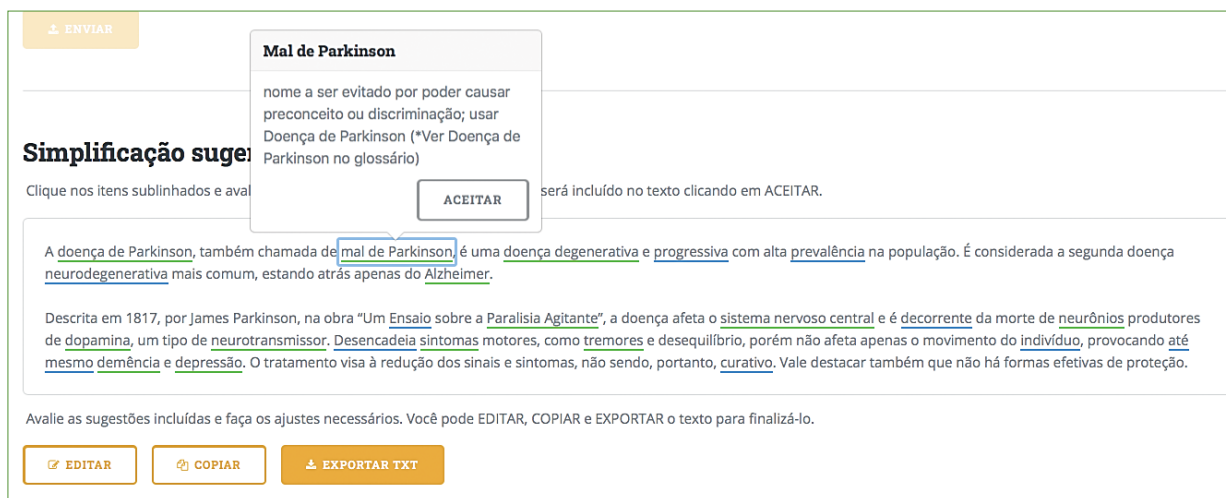FIGURE 3: Terminology core and peripheral terms. Source: Created by the authors

FIGURE 4. A text on Covid-19 processed on FMed system. Source: *http://www.ufrgs.br/textecc/acessibilidade/page/cartilha/*

Ministry of Health. After an initial selection of 3 or 4 definitions for the same term, we have put together a new definition/explanation following the steps below:

1. From the collected definitions, we selected the definition that presented the concept more clearly for a Type 1 or Type 2 reader profile. Often, the definition of the term was pieced together like a puzzle, in which one piece was taken from one source, such as the WHO website, and the other piece came from another source, such as the Fiocruz website.

2. We also mapped and evaluated existing terms and potentially difficult words within the selected defining paraphrase. Frequently, the definitions or explanations had new terms and low-frequency words that had to be dealt with, as seen below:

coluna vertebral 1 → eixo central *do corpo responsável por sustentar a nossa posição* bípede

After further analysis using CorPop, we considered that words such as *eixo central* (*central axis*) and *bípede* (*bipedal*) could be potential barriers to our Type 1 reader's understanding (Elementary School). On CorPop, we found that both *bípede* and *eixo central* had zero occurrences. For this reason, to put together a definition considering a reader profile with limited literacy and education, we replaced the terms with simpler words or expressions. In this case, *eixo central* was suppressed and *bípede* was replaced with an explanation, as follows:

coluna vertebral 2 → *sustenta o nosso corpo e permite que a gente caminhe sobre os dois pés, sem precisar usar as mãos.* **spine** 2 → supports our body and allows us to walk on both feet without using the hands.

In Figure 5, it is possible to see an extract of the work that supported the internal glossary of the Covid-19 module and its definitions/explanations.

Categorizing definitions is not always an easy matter (Felten, Finatto, 2021). Therefore, our intention is not to put FMed definitions/explanations into a strict category but outline some strategies and intentions. FMed internal glossary definitions, when possible, try to use what Marcuschi (2001) has named "*definições instanciativas*". The so-called "*definições instanciativas*" are corpus-based sentence definitions that were first proposed for English learners' dictionaries (Cobuild) and later adopted for school dictionaries in Brazil. In this type of definition, the entry word can be part of the definition/explanation and the language is usually more informal, with the first-person plural and the use of pronouns like you/we, simulating a conversation.

In addition to using Plain Portuguese and simplification/translation strategies, FMed definitions were also constructed with the help of other resources, such as analogy/metaphor and exemplifications preceded by "*como*" (such as). Below are some examples. In item 1, we have an analogy between the *brain* and a *computer*. In item 2, *antibodies* are compared to *fighters*. In item 3, *antigen*, examples are given after "such as".

1. **encéfalo** → centro do sistema nervoso, é o computador do nosso corpo; reúne e processa todas as informações, enviando mensagens para o resto do nosso corpo; é diferente do cérebro, pois inclui outras partes do sistema nervoso; o cérebro faz parte do encéfalo

2. **anticorpos** → defensores do organismo que lutam contra bactérias, vírus e outros corpos estranhos

3. **antígeno** → parte do corpo estranho, como vírus, bactéria etc., que o nosso corpo detecta para gerar defesas

One of the main goals with these types of definitions/explanations is that the user-writer can fully or partially incorporate them into the text, keeping in mind that the end reader should not need to resort to dictionaries or glossaries to understand the meaning of words in the final version of the facilitated text. However, it important to emphasize that sometimes using the whole definition/explanation as it is presented may not be the best solution for a facilitated text.

Autora	Me. Liana Braga Paraguassu - Doutoranda do PPG-Letras-UFGRS
Orientadora	Dra. Maria José Finatto - Docente do PPG-Letras-UFRGS
Consultor médico	Dr. Ricardo Eizerik Machado, CREMERJ - Nº. 510110079-3
*Fontes de consulta:	ao final deste glossário
A	
	ADAPTADO PARA A FERRAMENTA MEDSIMPLES
achatar a curva	tentativa de impedir que o novo coronavírus se espalhe, infectando um grande número de pessoas; essa medida é importante para que os hospitais possam se preparar para receber o maior número de pacientes possível
adenovírus	grupo de vírus que normalmente causa doenças respiratórias, como resfriados, bronquite ou pneumonia; pode causar ainda conjuntivite e doenças nos intestinos, com sintomas como diarreia, vômito e dores de barriga
aerossol	partículas muito pequenas de poluição e poeira ou pequenas gotas que escapam da nossa boca quando falamos, tossimos ou do nosso nariz quando espirramos
afebril	pessoa sem febre
agente etiológico	micro-organismos, como bactérias, vírus, fungos etc. que podem causar doenças
agente infeccioso	micro-organismos, como bactérias, vírus, fungos etc. que podem causar doenças
agente patogênico	micro-organismos, como bactérias, vírus, fungos etc. que podem causar doenças
aglomerações	muitas pessoas juntas num mesmo lugar
ambientes bem ventilados	ambientes em que o ar consegue circular, com janelas abertas
anosmia	perda do olfato, capacidade de sentir cheiros
antebraço	parte do braço que vai do punho até o cotovelo
antibióticos	remédios utilizados para tratar doenças causadas por bactérias
anticoagulante	remédios que impedem a formação de coágulos, que são o sangue na forma sólida
anticorpos	proteínas que trabalham como defensores do organismo contra bactérias, vírus e outros corpos estranhos
antígeno	parte do corpo estranho, como vírus, bactéria etc., que o nosso corpo detecta para gerar defesas
antimicrobiano	remédios e substâncias que têm a capacidade de matar micróbios, como bactérias e fungos
antirretroviral	remédios usados para o tratamento de infecções por um tipo de vírus chamado retrovírus; o HIV é um retrovírus
antiviral	remédios usados para tratar doenças causadas por vários tipos de vírus
Anvisa	órgão do governo que controla a produção, distribuição e venda de produtos e serviços envolvendo alimentos e remédios
arritmia	palpitação; o coração bate num ritmo diferente do normal
asma	doença nos pulmões que faz com que os brônquios (tubos que ligam a traqueia aos pulmões) se fechem por estarem inflamados, fazendo com que a pessoa fique com um chiado e aperto no peito, falta de ar e tosse seca

FIGURE 5. Covid-19 preliminary glossary. Source: Created by the authors

It is important that the user-writer uses such definitions/explanations with discretion and edits the text in order to make the facilitated version more "reader friendly". Hence, users-writers can add or reduce the number of information provided in the definitions as they see fit.

This last point is something that we believe needs to be made clearer to the user-writer, as, according to tests conducted with potential FMed users, such as journalists and physicians, one of the suggestions for improvement was that the definitions/explanations should be leaner. Although we believe that this can be an improvement to be implemented in the construction of these definitions, it is also important that the users-writers understand that they can and should edit their text according to the reader's profile.

5 Tests, study in progress and future perspectives

We are currently involved in a number of trials that evaluate FMed usability and performance. We have already conducted initial tests on FMed usability with *a)* users-writers; and *b)* end readers.

During the first phase of testing, we have invited five potential users-writers, three journalists and two physicians, for the usability tests. The intention was that the primary users would simplify a given text using FMed and that the resulting text could then be tested with end readers. The profile established for the end readers was drawn based on data collected on the average Brazilian reader (INAF, 2018): adult, worker, with primary education.

For this test, we selected a text on Parkinson's Disease (PD), the FMed module that to date has the greatest coverage of terms. Users-writers first simplified the text without using FMed. After, FMed was introduced in order for them to assess its usability and whether the system had contributed in any way to their simplification process as content producers.

As reported by these users, and summarized in Table 2, FMed helped them simplify the PD text to a greater or lesser extent. Question "**Q2.** Now, we would like to know if MedSimples helped you simplify your text" is part of a broader questionnaire, which aims to evaluate the system and identify points for future improvement.

Answer	Percentage based on 5 respondents
Yes, the system helped a lot.	40%
Yes, the system helped a little.	60%
No, the tool didn't help at all.	0%
It didn't help and I liked my version of simplification better.	0%

TABLE 2. Answers from FMed users-writers regarding its effectiveness in simplifying texts Source: Created by the authors based on a survey conducted at *www.surveymonkey.com*

The text selected for this test and processed on FMed is shown in Figure 6 below. The system indicates terms in green and potentially difficult words in blue, offering potentially simpler definitions in comparison to those found in specialized dictionaries or medical handbooks.

After the five users-writers had presented their final simplified versions of the texts using FMed, these texts were analyzed by us. We checked whether the text still had some low-frequent vocabulary and unexplained terminology. We also verified sentence length, use of indirect forms, among other factors. As extra support, we reviewed the texts provided by the users-writers using Coh-metrix 3.0 (McNamara et al., 2014). This tool contains, among other features, a Flesch Index[4][5] (FI) readability calculator. The FI associates a ranking with the Brazilian different levels of education necessary for understanding a given text.

From these analyses, we selected the text with the highest FI (the higher the FI, in theory, the easier to read) for a new initial test, this time with the end readers. Five people were invited, and at first, the original, non-simplified text was presented to them. Three questions were asked, the same questions that would be asked later with the simplified text.

Below, we present the testing in Phase 1, with the original text (Figure 7) before simplification, and the 3 related questions. As shown in Table 3, in addition to the reading comprehension test, the respondents were asked to indicate their level of education. Respondents were instructed to read the text and answer the related questions according to their understanding of the text. If they did not know how to answer a question, they should check "Não sei" (3. I don't know) and if they thought that the information was not provided in the text, they should check "O texto não diz" (4. The text doesn't say).

The test results with the complex and simplified texts are compiled in Tables 3 and 4. As shown in Table 3, with the unfacilitated text, only 2 out of 5 respondents could answer 2 out of 3 questions correctly (respondents with High School), and none of the respondents were able to answer all questions.
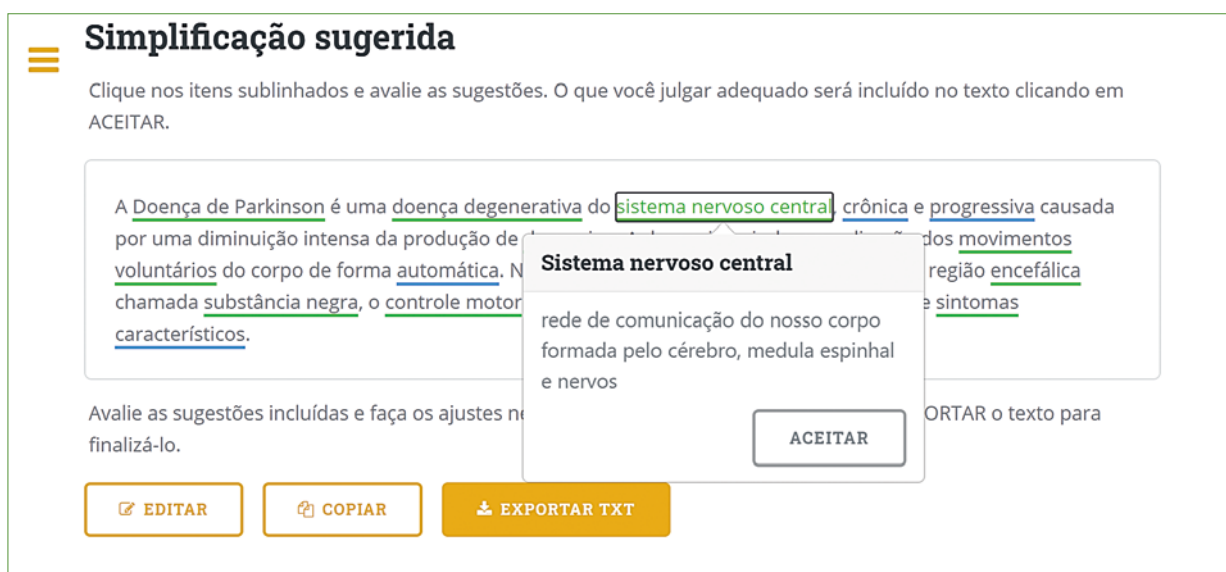

FIGURE 6: Original text on Parkinson's Disease on MedSimples. Source: *http://www.ufrgs.br/textecc/acessibilidade/page/cartilha/*

IMPORTANTE: OS SEUS DADOS PESSOAIS NÃO SERÃO DIVULGADOS
Grau de escolaridade
( ) Ensino fundamental incompleto
( ) Ensino fundamental completo
( ) Ensino médio incompleto
( ) Ensino médico completo
( ) Nível universitário

A Doença de Parkinson é uma doença degenerativa do sistema nervoso central, crônica e progressiva causada por uma diminuição intensa da produção de dopamina. A dopamina ajuda na realização dos movimentos voluntários do corpo. Na falta dela, particularmente numa pequena região encefálica chamada substância negra, o controle motor do indivíduo é perdido, ocasionando sinais e sintomas característicos.

P1 . De acordo com o texto, a Doença de Parkinson:
1. Melhora depois de um certo tempo.
2. Piora com o tempo.
3. Não sei.
4. O texto não diz.

P2 . De acordo com o texto, a dopamina:
1. Ajuda a realizar movimentos que o nosso corpo faz sem a gente controlar, como os batimentos do coração.
2. Ajuda a realizar movimentos que a gente faz porque quer, como pentear o cabelo.
3. Não sei.
4. O texto não diz.

P3 . De acordo com o texto, o que a falta de dopamina causa:
1. A perda de memória.
2. A perda do controle dos movimentos.
3. Não sei.
4. O texto não diz.

FIGURE 7. Text on PD without simplification. Source: Created by the authors

	P1	P2	P3	Tempo de resposta
R1 – E. F. incompleto declarado	Não sei.	Não sei.	Não sei.	8 min.
R2 – E. F. completo declarado	Não sei.	O texto não diz.	A perda do controle dos movimentos.	12 min.
R3 – E. F. completo declarado	Não sei.	Não sei.	Não sei.	5 min.
R4 – E. M. completo declarado	Piora com o tempo.	O texto não diz.	A perda do controle dos movimentos.	8 min.
R5 – E. M. completo declarado	Piora com o tempo.	O texto não diz.	A perda do controle dos movimentos.	10 min.

TABLE 3. Test with end readers - Phase 1 (original text on PD without simplification). Source: Created by the authors

Table 4 shows that in comparison with data in Table 3, there is a significant improvement in the readers' level of reading comprehension between the original text and the simplified version with the help of FMed. Table 4 indicates that after the simplification, 2 out of 5 respondents were able to answer all questions correctly (respondents with secondary education) and that the other respondents (primary education) also improved their reading comprehension by being able to answer 2 out of 3 questions correctly. It is important to point out that the text presented to readers did not have the suggestions proposed by the tool simply incorporated into the text. The final version of the text was edited by the user-writer.

IMPORTANTE: OS SEUS DADOS PESSOAIS NÃO SERÃO DIVULGADOS
Grau de escolaridade
( ) Ensino fundamental incompleto
( ) Ensino fundamental completo
( ) Ensino médio incompleto
( ) Ensino médico completo
( ) Nível universitário

A Doença de Parkinson afeta o cérebro. É uma doença que piora com o tempo. Ela diminui a produção de uma substância do cérebro chamada dopamina. A dopamina ajuda no controle dos movimentos. A falta de dopamina faz com que a pessoa tenha tremedeira e não consiga caminhar direito, amarrar os sapatos etc.

P1 . De acordo com o texto, a Doença de Parkinson:
1. Melhora depois de um certo tempo.
2. Piora com o tempo.
3. Não sei.
4. O texto não diz.

P2 . De acordo com o texto, a dopamina:
1. Ajuda a realizar movimentos que o nosso corpo faz sem a gente controlar, como os batimentos do coração.
2. Ajuda a realizar movimentos que a gente faz porque quer, como pentear o cabelo.
3. Não sei.
4. O texto não diz.

P3 . De acordo com o texto, o que a falta de dopamina causa:
1. A perda de memória.
2. A perda do controle dos movimentos.
3. Não sei.
4. O texto não diz.

FIGURE 8: Text on PD after simplification. Source: Created by the authors

	P1	P2	P3	Tempo de resposta
R1 – E. F. incompleto declarado	Piora com o tempo.	Não sei.	A perda do controle dos movimentos.	7 min.
R2 – E. F. completo declarado	Piora com o tempo.	O texto não diz.	A perda do controle dos movimentos.	12 min.
R3 – E. F. completo declarado	Piora com o tempo.	Não sei.	A perda do controle dos movimentos.	8 min.
R4– E. M. completo declarado	Piora com o tempo.	Ajuda a realizar movimentos que a gente faz porque quer, como pentear o cabelo.	A perda do controle dos movimentos.	8 min.
R5 – E. M. completo declarado	Piora com o tempo.	Ajuda a realizar movimentos que a gente faz porque quer, como pentear o cabelo.	A perda do controle dos movimentos.	9 min.

TABLE 4. Answers with end readers - Phase 2 (simplified text). Source: Created by the authors

It is worth emphasizing that FMed trials are at their preliminary stages. We are working to develop better structured tests, which reflect the needs of both readers and researchers who seek to build alternatives for the facilitation of texts and for the promotion of HL in Brazil. We have not yet included other resources, of a visual nature, such as the promising terminological infographics (Vidal and Estopà 2020), as this would require mobilizing different knowledge that is not part of our expertise.

The first step towards promoting HL is knowing the characteristics of the audience or population we are going to address in the mediation of communication (Osborne, 2013: 115). With this knowledge we will be able to design and develop well-informed and customized strategies. Knowing and mapping the readership that seeks information on Health-related topics in an entire country such as Brazil is not an easy task. Therefore, this reveals the importance of testing and collaborative work on literacy profiles of our population. ✤

Bibliography

ALUÍSIO, Sandra M.; GASPERIN, Caroline (2010) «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*. New York: ACL, vol.1, p. 46-53. <https://aclanthology.org/W10-1607/>

BAKER, David W. (2006). «The Meaning and Measure of Health Literacy». *J Gen Intern Med*, 21(8): 878–883. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1831571/>

CABRÉ, M. Teresa (1999). *La terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra.

CARVALHO, Yuli Souza; REBECHI, Rozane Rodrigues (2020). «Inteligibilidade e convencionalidade em textos de divulgação da área médica: uma análise à luz da Linguística de Corpus». *Revista de Estudos da Linguagem*, 29 (2). <http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/17462>

CONCEIÇÃO, Manuel Celio; ZANOLA, M. Teresa (org.) (2020). *Terminologia e mediação linguística: métodos, práticas e atividades*. Faro: Universidade do Algarve, <http://id.bnportugal.gov.pt/bib/bibnacional/2073879>

CIAPUSCIO, Guiomar (2008). *Textos especializados y terminología*. Barcelona: Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra.

ESTOPÀ, Rosa (2021). «Culturalización en salud: aportaciones desde la lingüística». *Cadernos de Linguística*, 2 (1), p. 1-22 <https://cadernos.abralin.org/index.php/cadernos/article/view/244/271>

FELTEN, Eduardo; FINATTO, Maria José Bocorny (2021). «A Definição Terminológica em Libras: Rumos e Frentes de Pesquisa». *Porto Das Letras*, 6 (6), p. 123-151 <https://sistemas.uft.edu.br/periodicos/index.php/portodasletras/article/view/11441>

FINATTO, Maria José Bocorny (2020). «Acessibilidade textual e terminológica: promovendo a tradução intralinguística». *Revista Estudos Linguísticos*, 49 (1), p. 72-96 <https://revistas.gel.org.br/estudos-linguisticos/article/view/2775>

FINATTO, Maria José Bocorny; MOTTA, Ester (2019). «Terminologia e Acessibilidade». *Revista GTLex*, 2 (2), p. 316-356. <https://doi.org/10.14393/Lex4-v2n2a2017-6>

Instituto Paulo Montenegro (2016). *Habilidades de Leitura, Escrita e Matemática são limitadas em muitos setores da economia brasileira, podendo restringir produtividade e capacidade de inovação* <http://download.uol.com.br/educacao/2016_INAF_%20Mundo_do_Trabalho.pdf>

Instituto Paulo Montenegro. *Indicador de Alfabetismo Funcional* (INAF). C2017 <https://alfabetismofuncional.org.br/>

JAKOBSON, Roman (1959). «On Linguistic Aspects of Translation». In: Achilles Fang *et al. OnTranslation*. Cambridge: Harvard University Press, p. 232–239 <https://web.stanford.edu/ffieckert/PDF/jakobson.pdf>

MARCUSCHI, Luiz Antônio (2001). *Definições*. Recife: UFPE. [Unpublished manuscript].

MARAGNO, Carla Andrea Daros *et al.* (2019). «Teste de letramento em saúde em português para adultos». *Rev Bras Epidemiol* 22: E190025 <https://www.scielo.br/j/rbepid/a/kzZkT67nn6S7rfSxW8nRSLp/?format=pdf&lang=pt>

Maziero, Erick Galiano; Pardo, Thiago A.S.; Di Felippo, Ariani; Dias-Da-Silva, Bento Carlos (2008). «A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Eletrônico para o Português do Brasil». *VI Workshop em Tecnologia da Informação e da Linguagem Humana* (TIL), p. 390-392.

McNamara, Danielle S.; Graesser, Arthur C.; McCarthy, Philip M.; Cai, Zhiqiang. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

Osborne, Helen (2013). *Health Literacy from A to Z*. Burlington: Jones & Bartlett Learning.

Paraguassu, Liana Braga (2018). *Tradução Especializada Acessível (TEA): revisão do tema e proposta de disciplina para cursos de graduação em tradução*. Dissertation <https://lume.ufrgs.br/handle/10183/193093>

Paraguassu, Liana Braga *et al.* (2020). «MedSimples: an automated simplification tool for promoting health literacy in Brazil». *Workshop: Digital Humanities and Natural Language Processing*. Aachen: CEUR- WS, vol. 1. p. 78-80 <http://ceur-ws.org/Vol-2607/abstract1.pdf>

Pasqualini, Bianca (2018). *CorPop: um corpus de referência do português popular escrito do Brasil*. Tese de Doutorado. Porto Alegre: Universidade Federal do Rio Grande do Sul. <https://www.lume.ufrgs.br/handle/10183/177566>

Pasqualini, Bianca; Finatto, Maria José Bocorny (2018). «CorPop: a corpus of popular Brazilian Portuguese». *Latin American and Iberian Languages Open Corpora Forum - OpenCor* <http://www.ufrgs.br/textecc/porlexbras/corpop/files/OpenCor_2018_paper_ref.pdf>

Passamai, Maria da Penha Baião; Sampaio, Helena Alves de Carvalho; Henriques, Eliane Mara Viana (org.) (2019). *Letramento funcional em saúde: as habilidades do usuário eo sistema único de saúde*. Curitiba: CRV.

Saggion, Horacio (2017). «Automatic text simplification». *Synthesis Lectures on Human Language Technologies*, Kaohsiung, 10 (1).

Silva, Asafe Davi Cortina (2018). *Textos de divulgação para leigos sobre o transtorno do estresse pós-traumático em português: alternativas para a acessibilidade textual e terminológica*. Dissertação de Mestrado. Porto Alegre: Universidade Federal do Rio Grande do Sul <http://www.ufrgs.br/textecc/acessibilidade/page/publicacoes/>

Silva, Elisabet M. S. (2009). *Histórico de letramento e práticas letradas em redações de vestibular*. Dissertação de Mestrado. Campina Grande: Universidade Federal de Campina Grande.

Soares, Magda (1998). *Letramento: um tema em três gêneros*. Belo Horizonte: Autêntica.

Zethsen, Karen Korning; Hill-Madsen, Aage (2016). «Intralingual Translation and Its Place within Translation Studies – A Theoretical Discussion». *Meta*, 613, p. 692-708.

Zilio, Leonardo; Wilkens, Rodrigo; Fairon, Cedrick (2018). «PassPort: a dependency parsing model for portuguese». In: PROPOR. *International Conference of Computational Processing of the Portuguese Language*, 13 p. 479- 489 <https://link.springer.com/chapter/10.1007/978-3-319-99722-3_48>

Zilio, Leonardo. et al. (2020). «A Lexical Simplification Tool for Promoting Health Literacy». In: *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI)*. Paris: European Language Resources Association (ELRA), v. 1. p. 70-76 <http://readi-lrec2020.wixsite.com/readi-lrec2020/programme>

Vidal-Sabanés, Laia; Estopà Bagot, Rosa (2020). *La terminologia que utilitzen les pacients de càncer de mama: una anàlisi descriptiva*. *Terminàlia*, 21, p. 7-18 <https://www.raco.cat/index.php/Terminalia/article/view/374426>

Notes:

3. Retrieved from: *https://www.who.int/teams/health-promotion/enhanced-wellbeing/ninth-global-conference/health-literacy*
4. **FLI** Brazilian Portuguese = 248.835 - [1,015 x (Number of words per sentence)] - [84.6 x (Number of text syllables / Number of text words)]
FKLI = ((0.39 x average words per sentence)+(11 .8 x average of syllables per word)) - 15.59
5. Coh-Metrix-Port 3.0 calculates metrics to assess the cohesion, coherence, and difficulty of understanding a text through different levels of linguistic analysis: lexical, syntactic, discursive, and conceptual. For the implementation of all these metrics (+80), several Natural Language Processing (PLN) resources and tools are used. Available at: *http://fw.nilc.icmc.usp.br:23380/cohmetrixport*