Umanistica Digitale – ISSN: 2532-8816 – n. 16, 2023
O. Carlino – *Enhancing readability and comprehension of medical reports*
DOI: http://doi.org/10.6092/issn.2532-8816/15987

Enhancing readability and comprehension of medical reports: a comparison between two online text simplification tools

Ottavia Carlino

Università del Salento
ottavia.carlino@unisalento.it

Abstract

Medical reports are a fundamental text type in the life of patients, but also one of the most difficult ones to understand by non-specialists. For this reason, software tools specifically devoted to the simplification of medical texts, including reports, have been developed.

In this paper, we analysed the performance of two different text simplification tools in a sample of fifty medical reports in English. Both tools support lexical simplification, while one of the tools also provides suggestions for syntactic simplification. We compared the two tools by calculating the readability index of each medical report before and after being processed by the tools. To this aim, we used two different readability formulas.

Both tools succeeded in producing some degree of simplification; however, their performance strongly varied depending on the text and the readability formula considered.

Key words: Text Simplification, Medical Reports, Doctor-Patient Communication, Readability, Accessibility.

I referti medici rappresentano una tipologia di testo fondamentale nella vita dei pazienti, ma anche una delle più difficili da comprendere per i non addetti ai lavori. Per questo motivo, sono stati sviluppati software specificamente dedicati alla semplificazione dei testi medici, compresi i referti.

In questo lavoro abbiamo analizzato le prestazioni di due diversi strumenti di semplificazione testuale su un campione di cinquanta referti medici in lingua inglese. Entrambi gli strumenti supportano la semplificazione lessicale, mentre solo uno di essi fornisce anche suggerimenti per la semplificazione sintattica. Abbiamo confrontato i due strumenti calcolando l'indice di leggibilità di ciascun referto medico prima e dopo l'elaborazione da parte degli strumenti. A tal fine, abbiamo utilizzato due diverse formule di leggibilità.

Entrambi gli strumenti sono riusciti a produrre un certo grado di semplificazione; tuttavia, le loro prestazioni variavano fortemente a seconda del testo e della formula di leggibilità considerata.

Parole chiave: semplificazione testuale, referti medici, comunicazione medico-paziente, leggibilità, accessibilità.

1. Introduction

Communication is fundamental to the doctor-patient relationship, and its quality can greatly influence this relationship at the level of the individual, the community and even the entire health care system ([43]). Effective communication has a positive effect on patient satisfaction, compliance, and medical outcomes, thus contributing to cut health care costs ([43]).

Written materials are still one of the most inexpensive and informative means of providing information to patients. Such written texts should ideally meet the needs of each person as regards both content and accessibility, e.g. readability. However, two major challenges exist in creating personalised health information texts ([1]: 1-8). Firstly, text generation is expensive and requires the time of a medical professional. Secondly, the text has to be intelligible to the consumers of the text, who often have different backgrounds and education than the producers of the text.

Although medical professionals are provided with top-level guidelines on how to make text simpler, their intuition is not always sufficient to produce texts that are comprehensible for a specific target group or single patient ([1]: 1-8). Hence, software tools specifically devoted to the simplification of medical texts have been developed, to support medical professionals in writing texts and patients in understanding them. These software tools primarily operate at the lexical level, though some also suggest changes at syntactic level.

In this paper, we compare two text simplification tools specifically developed for processing a medical report (hereafter MR) by contrasting the results obtained by each tool on fifty reports in English. While the first tool considered in this study is intended for readers (patients or lay readers in general) who wish to simplify an MR, the second tool is aimed at those who have to write the report, in order to help them produce a more comprehensible text for a non-medical audience. Despite this, comparing these two tools, which were conceived with different objectives and target audiences in mind, could clarify some aspects of textual simplification. More specifically, this study set out to answer the following questions:

1. Which of the two selected tools succeeds in making MRs easier to read?
2. Is purely lexical simplification sufficient to guarantee an improvement in the readability of a text, or does syntactic simplification play a significant role in this respect?

To these aims, the remainder of this paper is organized as follows: Section 2 illustrates the function and linguistic features of MRs. Section 3 discusses accessibility and text simplification, with specific reference to medical written discourse. Section 4 illustrates the materials and methods, including a detailed description of the text simplification tools selected for this study, readability formulas, and the corpus of MRs used to test the tools. Finally, Section 5 presents the results of this study and Section 6 attempts to draw some conclusion and suggests possible extensions of the study.

2. MRs and their linguistic features

An MR is a document written by a health professional explaining the results of a clinical or instrumental examination that was carried out on a patient. According to Cappelletti ([10]: 197-208), the report is the product of a complex process of analysis in several steps, aimed at transforming raw data from diagnostic tests and/or objective examinations into clinical information useful for patient care. The report bypasses information overload by providing only the pieces of information that concretely respond to the clinical question under investigation. The main addressee of an MR is another physician or healthcare provider. Therefore, problems of accessibility by lay readers are mainly due to the patient not being considered the preferred target reader.

In general, MRs include the following parts ([37]: 151-156):

1. The 'Patient Information' part, which contains the patient's demographic data and clinical information.
2. The 'Diagnostic Tests' part, which includes laboratory tests and comments by the doctor.
3. The 'Medical Diagnosis' part, which is the medical interpretation text entered by the specialist.
4. The Header and Footer of the report, which usually contains information about the medical institution.

MRs show all the typical features of medical specialised discourse ([7]: 17-37). We can group these features into five categories, as illustrated in Table 1.

Category	Features
Lexical	Technical terms (including Abbreviations and Acronyms) Standard phrases
Semantic	Terminologization Metaphors
Syntactic	Syntactic conciseness: - Nominalisation - Pre-modification instead of relative clauses - Stacked nominal phrases - Omission of phrasal elements
Stylistic	Modality Passive and Impersonal style
Textual	Anaphoric references Conjunctions Thematic progression

Table 1: Main linguistic features typical of medical language ([7]: 17-37)

The following paragraphs briefly present each of the listed features.

2.1 Lexical features

All specialised languages are characterised by high frequency of technical terms, i.e. words with a well-defined, monoreferencial meaning in the given specialised context ([20]: 2003). English medical technical terms often derive from Latin or Greek and appear as loan words or compound words composed of Latin or Greek roots, prefixes, suffixes and connecting vowels ([49]). For instance, *mammography* can be divided into the following parts: the root *mamm-* (from Latin *mamma* meaning 'breast'), the root *-graphy* (from Greek γϱάφειν, *gráfein*, meaning 'write, express by written characters'), and the connecting vowel -o.[1] Loan words can be typically observed with reference to organisms such as bacteria (*Bacillus anthracis*, *Bordetella pertussis*, *Escherichia coli*, etc.) or viruses (*Phlebovirus*, *Lolium latent*, *Herpes simplex virus*, etc.) (Maglie 2009), but also to disorders and diseases (e.g. *xerosis*, *photalgia*, *procidentia*, *bruxism*, *aphthous stomatitis*).

Alongside technical terms, MRs are rich in fixed standard phrases. Examples of standard phrases used in MRs include *on examination*, *within normal limits*, *differential diagnosis*, etc. ([46]: 453-469).

Finally, as Kuzmina, Fominykh and Abrosimova ([33]: 548-554) observe, abbreviations are among the most frequently used features of both written and oral medical communication. This is particularly true for MRs, where a number of terms, such as the names of diseases, chemical compounds or therapies almost never occur in their full form. The extensive diffusion of abbreviations in the language of medicine is undoubtedly attributable to its historical tradition, but also to the economy of space and time they offer, which is necessary in many medical scenarios and text types. 'Abbreviation' is often used as an umbrella term subsuming several different forms. Kuzmina, Fominykh and Abrosimova identified the following categories of abbreviations on the basis of their components:

1) syllables: *magtape* (magnetic tape), *MEDEVAC* (medical evacuation), *URAC* (Uric Acid);
2) abbreviated words: *Dx* (diagnosis), *Fx* (fracture), *inj.* (injection), *post-op* (postoperative);
3) letter (or syllable) + word(s): *MUGA scan* (MUltiGAted radionuclide scan);
4) letters + numbers: *T4* (thyroxine), *25-OH* D_3 (25-hydroxyvitamin D3).
5) letters (i.e. acronyms): *ADHD* (Attention Deficit Hyperactivity Disorder), *HPS* (Hantavirus Pulmonary Syndrome), *IDDM* (Insulin-Dependent Diabetes Mellitus).

2.2 Semantic features

The use of words or phrases belonging to general language which have acquired a particular meaning in the medical field is a frequent semantic phenomenon in medical language ([6]: 63-73). Such a phenomenon is referred to as 'terminologization' ([41]; [4]). For example, in an MR, the term *consolidation* does not refer to the state of being consolidated or united but to a *pathological alteration of lung tissue from an aerated condition to one of solid consistency.*[2] As a consequence, the likelihood of misinterpretation is high.

1 McMaster University Online Etymology Dictionary.

2 Merriam Webster Dictionary, 2021.

Furthermore, technical terms in scientific discourse in general and medical language in particular are often the result of metaphorical processes whose stylistic colouring has then vanished with time and use ([44]). In the medical domain, this kind of semantic event is frequent in the naming of body parts, and is hence largely visible in MRs. For example, the almond-shaped mass of grey matter in the front part of the temporal lobe of the brain is called 'amygdala' from Ancient Greek ἀμυγδάλη (*amugdálē*, meaning 'almond'); 'cortex' indicates the outer layer of the brain and the term comes from the Latin word *cortex* referring to the outer layer of a tree; the 'coronary arteries' are the vessels surrounding the heart like a crown and the term comes from the Latin word *corōna*, meaning 'crown'.

2.3 Syntactic features

MRs are characterised by extreme conciseness ([38]), achieved through a range of syntactic choices including nominalisation, pre-modification instead of relative clauses, high frequency of stacked nominal phrases, and omission of phrasal elements.

Nominalisation is a syntactic process typical of specialised texts and consists in using a noun in place of a verb to convey concepts concerning actions and practices. According to Maglie ([38]), in medical language nominalisation is connected to the following reasons:

a) the derivation of a noun from the corresponding verb appears to echo the inference of results from trials;

b) nominalisation places the concept in thematic position, thus making the transfer of information from a condition of novelty to one that is already known more natural. Thematic positioning is also functional to emphasising the action expressed by the verb.

Example 1 is an instance of nominalisation from our MRs:

Dehydration of the intervertebral disc between L5 and S1.

Example 1: Nominalisation

The same sentence without nominalisation would have read as follows: *The intervertebral disc between L5 and S1 is dehydrated.*

Pre-modification is another linguistic device that makes the sentence syntactically condensed. In English MRs it appears to be often connected to a desire to avoid relative clauses ([11]: 31-48). For example, the wording 'a previously sprained ankle' is used in place of 'an ankle that has been previously sprained'. This event is common to other types of specialised medical texts, where relative clauses are not used for reasons of economy and simplicity of the syntactic structure ([38]). Indeed, besides giving the sentence syntactic compression, pre-modification allows for what Gotti ([20]: 79) calls "an easier flow of information".

Hand in hand with pre-modification goes the use of long stacked or compound nominal phrases which serve as ad hoc designations for notions that will be referred to again in the text, hence obviating long descriptions. This event has been largely observed in academic writing ([25]; [45]), where this special kind of noun phrases are employed for purposes of discourse cohesion and coherence ([25]). In the medical domain, stacked noun phrases are commonly composed of no more than six nouns. In addition to adjectives, the pre-modifying elements of the noun phrase may comprise nouns, adverbs, participles, hyphenated phrases, or a combination of these ([38]). The type of specification can relate to the material of which an object is made (e.g. *squamous tissue*, i.e. a tissue made up of flat cells), its function (e.g. *connective tissue*, i.e. a tissue that connects, separates, and supports all other types of tissues), or its mode of functioning (e.g. *dartoic tissue*, i.e. a tissue similar to the tunica dartos in its slow involuntary contractions). Stacked noun phrases made up of two short nouns quite frequently become a single word; at first they are used in the hyphenated form and then written as a single compound word (e.g. *sternocleidomastoid muscle*, i.e. the most superficial muscle of the neck). To decipher them the reader should know the semantic values of each lemma making up the stacked noun phrase as well as the overall topic of the text. For this reason, stacked noun phrases are not necessarily an issue for the specialist reader ([38]), but could represent a major obstacle for the lay reader.

Finally, MRs show a significant number of sentences where generally fundamental phrasal elements, such as verbs, are missing. Example 2, taken from our corpus, is a typical case:

Moderately good health condition; dyspnoeic patient; normally coloured conjunctivas; conscious and cooperative.

Example 2: Missing verbs

A complete sentence would read as follows: *The patient is in moderately good health condition; s/he is dyspnoeic and has normally coloured conjunctivas. S/he is conscious and cooperative.* In other cases, as in Example 3, the missing elements are several and more varied in type:

Cardiac auscultation: rhythmic and regular heart sounds, no murmurs.

Example 3: Missing phrasal elements

A complete sentence would read as follows: *During the cardiac auscultation, we have found that heart sounds are rhythmic and regular and that there are no murmurs.* The omission of phrasal elements results in compressed sentences and is a means for communicating more quickly. This trait is characteristic of the medical domain in general and is particularly prominent in MRs, where economy of time and space is crucial. In specialised communication, omission of phrasal elements does not affect the comprehension of a text, as the value of the absent elements can be restored by the specialist. It may, however, pose problems to the lay reader.

2.4 Stylistic features

An important stylistic feature to consider is modality. As Vihla ([47]) has shown, in contemporary medical writing modality differs depending on the function and pragmatic purpose of the genre. Thus epistemic modality – which expresses the speaker's degree of commitment to the proposition – is prominent in research articles, while deontic modality – expressing the speaker's authority in terms of permission given, obligation imposed or suggestion provided – is prevalent in manuals and MRs. Within MRs, modals are mainly found with a deontic value in recommendations, prescriptions, and instructions (e.g. *The patient is advised to continue rehabilitation at home*), while they have epistemic value in diagnostic questions, hypotheses, disease presentation and differential diagnosis (e.g. *Patients with hypothyroidism may exhibit a number of physiologic alterations*).

Finally, a typical stylistic feature characterising MRs is passive and impersonal style. The regular use of passive voices meets the need for depersonalising the utterances by the specialist, who is more interested in emphasising the consequences of an action than in pointing out who the perpetrator of the action is ([6]: 63-73). This is why in MRs passive forms are rarely followed by mention of the agent. Naturally, passive and impersonal constructions cannot be completely avoided in MRs and active utterances are used in cases where the patient has to act (e.g. *take the tablet, exercise daily, or drink plenty of water)* ([7]: 17-37).

2.5 Textual features

Anaphora is a primary means of textual cohesion. In general language, it is used both for reasons of clarity and to eliminate ambiguity. Conversely, anaphora occurs less commonly in specialised texts, where it is generally replaced by lexical repetition ([38]). Nevertheless, anaphoric references can be occasionally observed in MRs ([7]: 17-37), as in Example 4, where the phrase 'this procedure' refers to an already introduced referent, i.e. 'endovascular recanalization':

An endovascular recanalization was indicated. The patient underwent this procedure without complications.

Example 4: Anaphoric reference

Metatextual reference to portions of the text itself can also be extensively observed. In Example 5, there is an explicit reference to an earlier section of the text (i.e. the Physical Examination section):

As mentioned in the PE section, the patient is still experiencing pain in her left shoulder.

Example 5: Metatextual reference

Other important textual devices are conjunctions, used not only for cohesive, but also for pragmatic purposes, in so far as they indicate logical relations between sentences and clauses. In Example 6, the conjunctive adverb 'therefore' is used to express a result or consequence, while 'nonetheless' is used to introduce oppositional or contrastive information:

He started experiencing generalized tonic seizures today. Therefore, diazepam was administered, improving a little his clinical picture. Nonetheless, he continued experiencing focal seizures.

Example 6: Conjunctions

Finally, textual organization is supported by thematic organization. In specialised texts this involves sequencing thematic parts (where the topic is introduced) and rhematic parts (which show what is said about the topic) by placing the known/given datum in thematic position ([38]). MRs take advantage of all possible types of thematic progression, for the sake of clarity, conciseness, and straightforwardness. Danes ([15]: 106-128) identified 3 main types of thematic progression: constant thematic progression, linear progression, and derived thematic progression.

In constant thematic progression, the same theme or part of the theme appears in consecutive propositions (even if the wording is not always identical). Example 7 is an instance of constant progression from our data, con where the theme of the first sentence re-occurs as the theme of the next one:

The patient consulted the AandE Department today. He was then admitted to the Internal Medicine Unit.

Example 7: Constant thematic progression

In linear progression, the rheme or part of the rheme of one sentence becomes the theme of the next sentence. Example 8 is an instance of linear progression from our data, where 'fall' is the rheme of the first sentence and the theme of the second:

The patient reported that her symptoms started abruptly today after an accidental fall. This fall caused pain and functional impairment.

Example 8: Linear progression

In cases of derived thematic progression themes are derived from a hyper-theme or a hyper-rheme. Example 9 is an instance of derived progression from our data, where the paragraph title 'Vital Signs' serves the role of hyper-theme connecting the different themes of the propositions that follow the title:

Vital Signs:

- Blood Pressure: 140/90 mmHg
- Heart Rate: 75 bpm
- Body Temperature: 37.6 °C

Example 9: Derived thematic progression

3. Accessibility of MRs to patients

The features described above and observed in MRs confirm that this type of text targets a specialist reader ([10]: 197-208). However, MRs are first given to the patient, who should have the right to understand what the text says without requiring the, generally oral, explanation of a doctor (the specialist giving the MR, another specialist, or the GP).

Asymmetrical communication between specialist and lay person requires a less intricate use of language and a lower degree of abstraction. However, as Koch-Weser, DeJong, and Rudd ([30]: 371-382) observe, the majority of medical specialists struggle to adapt their way of communicating to the needs of patients. They have become so accustomed to specialised terminology and utterances that they have a tendency to use the same words and expressions in any context, including medical records. Hence, rather than generating user-friendly information for readers, they create texts that are only accessible to people from the medical field.

In addition, there are two factors that contribute to making MRs difficult to access for patients ([7]: 17-37):

1) Often the reports are written quickly by the doctor who cannot or does not want to spend too much time writing this document. Moreover, in order to optimise time and facilitate the work of the doctor, pre-printed templates are often used and filled in with information about the individual patient, which contributes to the concise and condensed style of MRs.

2) Most hospital MRs, while containing information about the patient or recommendations to be followed by the patient, are addressed to the general practitioner who has to deal with the follow-up of the patient on the basis of the content of the report; in other words, despite an obvious desire of the patient to be able to understand the contents of a what is written about them, patients are not seen as the addressee of MRs.

To partly address the issue of the patient's lack of understanding of MRs, software aimed at simplifying medical texts in general and reports in particular has been developed.

3.1 Text simplification of MRs

Text simplification is the act of editing natural language to lower its complexity and enhance both readability and comprehensibility. For some technical fields, guidelines for text simplification exist. One of the most developed ones is the ASD Simplified Technical English (STE) guide for the aerospace industry. This is a set of rules and guidelines developed by the

AeroSpace and Defence Industries Association of Europe (ASD) aimed at standardising and simplifying the language used in technical documentation ([5]). ASD STE guidance's key objective is to ensure that technical information is easily understood by a heterogeneous audience, regardless of language skills or technical background. This is achieved by providing a controlled vocabulary of approved words and phrases, strict rules for sentence structure and grammar, and guidelines for consistent use of terminology. The ASD STE guide also applies rules of sentence construction, emphasising simplicity and clarity. It limits the use of complex sentence structures, encourages the use of the active voice and discourages the use of idiomatic expressions, jargon and cultural references. In addition, the guide offers guidelines for punctuation, capitalisation, abbreviations and numerical expressions to ensure consistent formatting throughout the documents.

Unfortunately in the medical field no single set of suggestions exists. Leroy, Endicott, Mouradi, Kauchak, and Just ([35]: 522-531) suggest that modifying the words and grammar of a text can result in a simplification of the text, and appropriate changes will improve readability and understanding. The use of computer tools has also been recommended in the medical field, to aid the simplification process and improve comprehension of clinical data ([34]: 169-172).

According to previous studies ([42]: 1-27; [31]: 1-8), the substitution of difficult words with simpler synonyms may lower the difficulty level of a medical text. Additionally, when it comes to medical texts, comprehending technical terms is much more difficult than catching the general meaning of the sentences ([48]: 503-516). This is one of the reasons why software tools for medical text simplification primarily operate at lexical level by offering a lay term or, in their absence, a definition. Indeed, according to Wang, Miller, Schmitt and Wen ([48]: 503-516), it is important for a lay reader to get both the corresponding everyday synonym of a technical term (if any) and additional information on the term itself in order to really understand the meaning of the term. Furthermore, some simplification tools also suggest syntactic reformulations, as is the case with one of the tools tested in this study.

4. Materials and methods

We collected fifty MRs in English (the medical reports we used are described in section 4.3) and used two online, freely usable simplification tools – respectively known as SIMPLE® and 'Text simplification editor for medical and health-related information' (hereafter referred to as TSE) – to produce simplified versions of each text. These two tools were selected over others for the following reasons: they are freely usable online; they share a common simplification method based on lexical substitution and provision of dictionary definitions; one of them also includes syntactic suggestions.

With each software, we replaced the technical terms with the consumer terms or definitions suggested by the tool. In cases where the consumer term was not available, we entered one or more definitions proposed by the tool, depending on the kind of definitions offered by the software and their sequential order. In fact, while TSE never offers more than one definition, SIMPLE® often proposes several definitions (see Section 4.1). In order to replicate the behaviour of typical readers who are obliged to read through several definitions before finding one that suits them, we entered all the relevant listed definitions up to the first useful one. Furthermore, with TSE, we also implemented the sentence-level simplifications it suggested. More specifically, whenever suggested, we:

- replaced the passive form with the active form;

- eliminated double negations (e.g. negation adverbs such as 'not' or 'without' associated with a morphologically negative word, such as 'abnormal');
- removed nominalisations by rephrasing the sentence.

Syntactic substitutions were implemented on their own, in order to assess the impact of syntactic reformulation alone, as well as alongside lexical substitutions, in order to mirror the way users would take advantage of the tools.

Next, we calculated the average readability scores of the newly obtained texts, and compared them with those of the original texts. To this aim, we used two different readability formulas – the SMOG formula and the Fog index – and compared their results.

The following paragraphs provide a detailed description of the two simplification tools and the two readability formulas selected for the study and of and our corpus of MRs.

4.1 The text simplification tools

The first tool selected for this study was developed by the Department of Mathematics and Computer Science of the University of Palermo and it is called SIMPLE® (SImplifier of Medical texts for Patients and Less Expert users). SIMPLE® is a web app aimed at automatically simplifying any kind of medical text, through a straightforward and intuitive user interface. It identifies health-related terms and supplies lexical simplification and extra information. It is primarily meant to 'empower' patients or lay people in general, but may also be useful for people with different levels of expertise (e.g. medical students).

SIMPLE's architecture features the following three main modules, as shown in Figure 1 ([3]: 1-8):

- The HIGHLIGHT module takes as input an arbitrary text, uses a medical vocabulary to find technical terms and highlights them when there is a consumer term and/or a consumer explanation.
- The MAP module links technical terms found by the HIGHLIGHT module to equivalent consumer terms using a medical thesaurus.
- The DEFINE module provides a simple explanation of the term using a consumer medical dictionary.

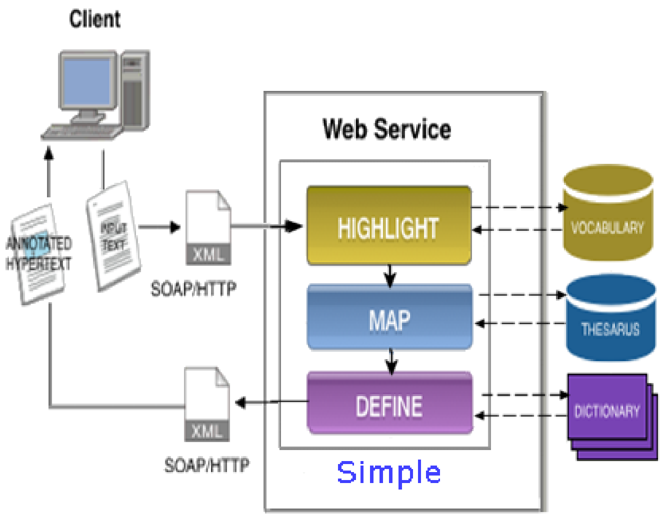

Figure 1: SIMPLE architecture (retrieved from [3]: 1-8)

SIMPLE was implemented using a SOAP/HTTP approach. Specifically, a client makes a request to a web service - written in php and javascript and using a MySql database - which gets the input text to be processed and runs the job via the HIGHLIGHT, MAP and DEFINE modules.

For the HIGHLIGHT module, the 'Unified Medical Language System' (UMLS)1 set up and managed by the US National Library of Medicine, was mostly used. This is a comprehensive multilingual vocabulary collection that includes information on biomedical and health topics. It uses a 'Concept Unique Identifier – CUI' to create a unique identifier for each concept and a mapping between vocabularies, thus enabling translation between different terminology systems.

Consumer Health Vocabularies (CHVs) were used for the MAP module with the aim of translating medical terms into corresponding consumer terms. Among the most well-known examples of CHVs is the Open Access Collaboratory Consumer Health Vocabulary (OAC-CHV), created and maintained by the Consumer Health Vocabulary Initiative. The MAP module employs the OAC-CHV as its thesaurus. The mapping from technical to consumer terms is done via the CUI, when available. Such is the case for technical terms found in the UMLS. For terms with no CUI, a customised concept identifier was created.

For the DEFINE module, an analysis was made of many health consumer websites which often contain health and medical dictionaries created specifically for health consumers and which therefore use language easily understood by the latter. Therefore, the Italian dictionaries for health consumers were used and about 15,000 entries were obtained. In addition, some entries were added from a few somewhat more technical dictionaries in order to provide as many definitions as possible, thus creating a meta-dictionary of about 100,00 entries. For the English definitions, the 28,000 entries of the WebMD online dictionary were used.

As shown in Figure 2, in the processed text, technical terms are marked in yellow; in addition, there is an Info button next to each technical term (see red box in the figure). Clicking on it, the consumer term is displayed together with one or more definitions (see green boxes in the figure). As Figure 1 shows, some definitions are inappropriate (e.g. LAP) while others are of little use (e.g. the first definition), and the reader is often obliged to read through several definitions before they can find a useful one.

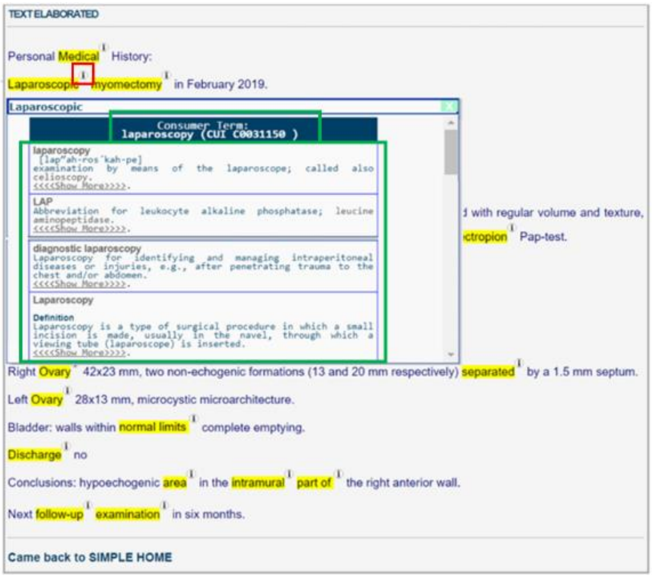

Figure 2: Consumer term and definitions for the technical term 'laparoscopic'

SIMPLE® application is capable of:

1) identifying medical terms in a medical text through the use of medical vocabularies;
2) converting medical terms into consumer terms through medical consumer thesauri;
3) providing explanations of terms using health-consumer dictionaries.

SIMPLE® may be used as a stand-alone web application or may be incorporated into common healthcare platforms for real-time identification and clarification of medical terms.

The second tool, known as 'Text simplification editor for medical and health-related information' and referred to as TSE, was developed by the Department of Computer Science of Pomona College and the University of Arizona. The user inputs an existing document and the tool identifies problematic text parts and suggests possible corrections. The user receives one or more suggestions in a drop-down menu and chooses how to replace a given term, thus modifying the source text. When finished, an overview of the changes made and their impact on text attributes is provided. The user can then distribute the final version of the text to patients and other content consumers. The main constituents of the tool are the text simplification features that are used to identify challenging texts and provide suggestions for improvement ([26]: 3749-3757). The tool features can be divided into two general categories: word-level features and sentence-level features. As shown in Figure 3, at the word level, the tool highlights potentially difficult words (see red boxes in the figure); by clicking on the underlined item, the user is offered a potential replacement in the form of synonym or explanation (see green box in the figure):

Figure 3: Definition of the technical term 'prostatectomy' provided by TSE

Difficult words are identified according to their frequency in the Web. The idea of viewing the Web not only as a source of material for the composition of a corpus, but also as a corpus in itself, has become a discussion topic since 2001 ([28]; [19]: 35-58). By virtue of its unlimited content, the worldwide web has become a powerful resource for analysing the frequency of use of specific language patterns that signify the natural, 'living' language. Google can be implemented as a concordance programme to study the frequency and context of use of particular language patterns and to apply them more naturally ([32]: 173-179). It is assumed that low-frequency words are less familiar and therefore more difficult because the reader does not encounter them as often and is less likely to know their meaning. In the same way, a text that uses more low-frequency words may be more difficult ([34]: 169-172).

As shown in Figure 4 a, beside word-level features, the tool also identifies sentence-level properties that could be modified to improve the comprehensibility/readability of the text (see green box in the figure):

Figure 4 a: Sentence-level suggestions for the phrase 'was provided'

For the purposes of our study, we first needed to select a readability formula that was appropriate to MRs.

4.2 Selecting a readability formula

Readability is an attribute of written text, generally determined by various elements that supposedly make the text more or less difficult to read (for instance, lexicon, sentence complexity). Mathematical formulas using these factors have been developed over the last century to quantify readability ([23]). Based on these formulas, readability is calculated and usually conveyed as an estimated reading level. The aim of the estimated reading level is to establish the difficulty of the text, meaning that an average reader of that level would be able to read or 'cope' with the text or passage without excessive frustration ([8]: 285-301; [12]: 176-184; [23]: 1980).

Critics have often claimed that readability formulas are not accurate or useful measures of text difficulty since they correspond to the surface structure of a text and do not address the interaction between readers and texts, ignoring reader-specific factors such as reading interest and purpose (43-53). Nonetheless, this is not to say that they do not measure anything. Some educators and researchers use them to check the difficulty of text passages. Klare ([27]) regarded readability formulas as much more accurate than human judgement. Furthermore, as Bailin and Grafstein ([8]: 299) state, readability formulas "may well allow us to assess the difficulty of at least certain aspects of the texts present for certain readers."

Two specific properties that stand out in readability formulas are their (1) emphasis on the ease of understanding a text and (2) emphasis on quantification. These emphases made readability a highly attractive concept for educators. The appeal of using formulas to gauge readability is based on the assumption that, as a matter of principle, they assess the difficulty of written material in an objective and quantifiable manner without taking into account the reader's characteristics.

Furthermore, a readability formula can provide a numerical score, thus granting the user the feeling of knowing the exact level of difficulty of a text ([8]: 285-301).

The numerous formulas employed to calculate readability consider a combination of textual components, including one or more of the following: percentage of easy high-frequency words (i.e. words on a preset list described as familiar to most students in a given grade), percentage of difficult words (i.e. words not on the list of familiar words), average number of words per sentence, average number of syllables per word, number of monosyllabic words or number of words with more than one syllable. Nowadays, more than 40 formulas are used to measure readability, yet some of them are more widely known and popular than others.

To the best of our knowledge, the readability formulas most commonly applied to assess health information materials are those reported in Table 2. In the table, column one reports the name of the formula and the paper in which each formula was introduced, column two specifies the equation that is used to calculate a given readability index, column three indicates the reference criteria for establishing each reading level, and column four provides the expected level of comprehension for each formula.

Table 2: Readability formulas commonly used for assessing healthcare texts
(adapted from [48])

Formula	Equation	Criteria for establishing reading grade level	Expected comprehension
Dale-Chall (Dale and Chall, 1948)	$0.1579\left(\frac{difficult\ words}{words}\times 100\right)+0.0496\left(\frac{words}{sentences}\right)$	McCall-Crabbs Standard Test Lessons in Reading (1926)	50% on a multiple-choice test
Flesch Reading Ease (Flesch, 1948)	$206.835-1.015\left(\frac{total\ words}{total\ sentences}\right)-84.6\left(\frac{total\ syllables}{total\ words}\right)$	McCall-Crabbs Standard Test Lessons in Reading (1926)	75% on a multiple-choice test
Flesch-Kincaid (Kincaid, Fishburne, Rodgers, and Chissom, 1975)	$0.39\left(\frac{total\ words}{total\ sentences}\right)+11.8\left(\frac{total\ syllables}{total\ words}\right)-15.59$	Lowest reading grade level in which 50% of tested navy personnel scored at least 35% on a cloze test	35% on a cloze test
Fog - Gunning Fog Index (Gunning, 1952)	$0.4\left[\left(\frac{words}{sentences}\right)+100\left(\frac{complex\ words}{words}\right)\right]$	McCall-Crabbs Standard Test Lessons in Reading (1926)	90% on a multiple-choice test
SMOG - Simple Measure of Gobbledygook (McLaughlin, 1969)	$1.0430\sqrt{number\ of\ polysyllables\times\frac{30}{number\ of\ sentences}}+3.1291$	McCall-Crabbs Standard Test Lessons in Reading (1961)	100% on a multiple-choice test

As Table 2 shows, most of these readability formulas use the McCall-Crabbs Standard Test Lessons in Reading ([39]) as the gold standard for grading understanding. The McCall-Crabbs Standard Test Lessons in Reading is a book designed to test reading comprehension skills of students in grades 1-12. It comprises five levels, with each level consisting of seventy eight readings on a wide range of topics, followed by multiple-choice questions based on facts, implications, or general reasoning. Each text is normed for grade level ability. The reading level is estimated based on the percentage of questions answered correctly. Consequently, this book has been used to establish a range of skill levels, from beginner to post-graduate. Test scores from students of different known levels are used to determine cut-off scores and assess reading levels. Cut-off scores show the predicted understanding for the majority of readers of a given reading level for the material provided.

While in overall literacy grasping the essence of a story may often be considered enough, in healthcare contexts superficial understanding may result in misunderstanding of recommended therapies, thus posing safety risks and leading to substandard care. For this reason, according to Wang, Miller, Schmitt and Wen ([48]: 503-516), averaging reading level estimates from formulas with predicted comprehension levels below 100% to determine the reading level of health-related written materials is not optimal.

Among the formulas that are commonly used in healthcare (see Table 2), the SMOG formula – which performed most consistently in a comparative experiment ([48]: 503-516) – is also the only one considering 100% predicted comprehension. Next in line comes the Fog index, with 90% predicted comprehension. Interestingly, they employ slightly different parameters in calculating readability. If on the one hand they both consider word length (i.e. called 'number of polysyllables' in SMOG and 'complex words' in Fog), on the other SMOG counts the number of sentences, while Fog considers average sentence length. For this reason, we decided to test both indexes in this study.

Several tools are available online to calculate the SMOG index of a text, but application of those tools to one sample text from our corpus showed that they return highly different results (see Table 3). For this reason, we decided to use all of these tools, calculate the arithmetic mean of the six results obtained and consider that value as the SMOG score of our texts.

Table 3: Selected online readability calculators and their returned SMOG scores.

Readability calculator	***Sample text SMOG score***	***Link to the calculator***
Automatic Readability Checker	10.8	readabilityformulas.com/free-readability-formula-tests.php
Online Utility Readability calculator	13.72	online-utility.org/english/readability_test_and_improve.jsp
Readability Analyzer	9.81	datayze.com/readability-analyzer
Text Compare Readability Calculator	14.61	textcompare.org/readability/smog-index
WebFX Readability test tool	10.8	webfx.com/tools/read-able/
Wordcalc	12	wordcalc.com/readability/

Table 4 shows the SMOG scores and the corresponding educational levels. In the table, the first column indicates the SMOG score range, while the second indicates which educational level each range corresponds to.

SMOG Score	***Education Level***
1-4	Elementary School
5-8	Middle School
9-12	High School
13-16	Undergraduate
17+	Graduate

Table 4: Conversion table of SMOG Scores to Education Levels
(adapted from Derguech, Sana E Zainab 2018, 247-256)

The Fog index, presented by Gunning in *The Technique of Clear Writing* ([21]) is a readability formula designed for adults, which gained popularity due to its ease of use. It employs two variables, the average sentence length and the number of words with more than two syllables per 100 words:

$$\text{Grade Level} = 0.4 \times \left[\left(\frac{words}{sentences}\right) + 100\left(\frac{complex\ words}{words}\right)\right]$$

where the ratio of the number of words to the number of sentences indicates the average length of a sentence, and a complex word is defined as a word with more than two syllables.

Table 5 shows what level of education each Fog index corresponds to. In the table, the first column indicates the Fog indices, the danger line (beyond which a text becomes particularly difficult to read) and the easy level (beyond which a text is particularly easy to read), while the second column indicates the estimated reading grades as educational levels.

Table 5: Conversion table of Fog indices to Education Levels
(Adapted from [24]: 423-435)

Fog Index		**Estimated Reading Grades**
	17	College graduate
	16	College senior
	15	College junior
	14	College sophomore
Danger line	13	College freshman
	12	High school senior
	11	High school junior
Easy	9	High school sophomore

8	Eighth grade
7	Seventh grade
6	Sixth grade

Using the same online readability calculators that we used to calculate the SMOG score (see Table 3), we calculated the Fog index before and after processing the reports with the two tools. As in the case of the SMOG score calculation, the different calculators available online returned different results for the Fog index, so we considered the arithmetic mean of the six results obtained.

4.3 The MRs used in this study

To assess the quality of the software output, we collected 50 MRs in English[3]. Each report is related to one of the following medical specialities: cardiovascular diseases; dermatology; diagnostic imaging; emergency medicine; gastroenterology; gynaecology; infectious diseases; internal medicine; laboratory medicine; neurology; oncology; ophthalmology; orthopaedics and traumatology; paediatrics; pneumology; psychiatry; surgery; and urology.

The medical records used in this study – named 1-50 – are described in Table 6. For each file, the table provides the following information: medical speciality, topic, and word count.

Table 6: List of the medical records used in this study

#	***Medical Speciality***	***Topic***	***Word count***
1	Cardiovascular diseases	Acute ischaemic stroke	737
2	Cardiovascular diseases	Atrial flutter and fibrillation	217
3	Cardiovascular diseases	Coronary Artery Disease	701
4	Cardiovascular diseases	Mitral valve disease	1654
5	Cardiovascular diseases	Thoracic Outlet Syndrome	423
6	Diagnostic Imaging	Abdominal ultrasound	371
7	Diagnostic Imaging	Breast biopsy	173
8	Diagnostic Imaging	Breast thermography	140
9	Diagnostic Imaging	Brain CT-scan	216
10	Diagnostic Imaging	Cervicothoracic CT-scan	383
11	Diagnostic Imaging	Colonoscopy and Esophagoscopy	119

3 The reports used in this study were obtained by the author of this paper who works as a freelance translator of medical reports for different translation agencies. To safeguard the privacy of all parties involved, any reference to personal data was removed from the reports prior to their use.

12	Diagnostic Imaging	GI Endoscopy	129
13	Diagnostic Imaging	Echocardiogram	141
14	Diagnostic Imaging	GI Motility test	113
15	Diagnostic Imaging	Right knee MRI	343
16	Diagnostic Imaging	Abdominal CT-scan	394
17	Diagnostic Imaging	Spine MRI	279
18	Diagnostic Imaging	Thyroid ultrasound	191
19	Diagnostic Imaging	Thoracic and abdominal CT-scan	408
20	Diagnostic Imaging	Colon endoscopic biopsy	579
21	Diagnostic Imaging	Retrograde cystography	137
22	Emergency Medicine	Fever and chills	652
23	Emergency Medicine	Seizures	340
24	Gastroenterology	Stomach, duodenum, and sigmoid colon biopsy	108
25	Gynaecology	Amenorrhea	183
26	Gynaecology	Vaginal ultrasound scan	444
27	Infectious diseases	Bilateral COVID-19 Pneumonia	294
28	Infectious diseases	Human Papilloma Virus	195
29	Infectious diseases	SARS-CoV-2 PCR test	441
30	Internal Medicine	AKI and hypokalaemia	288
31	Internal Medicine	Systemic Erythematosus Lupus	358
32	Internal Medicine	Tendinitis	137
33	Neurology	Low-grade glioma	1641
34	Neurology	Right hemiparesis	1228
35	Ophthalmology	Retinal detachment	743
36	Orthopaedics and Traumatology	Cervical root disorder	432
37	Orthopaedics and Traumatology	Intramuscular hematoma	719
38	Orthopaedics and Traumatology	Mechanical lumbago	67
39	Orthopaedics and Traumatology	Polytrauma	455
40	Orthopaedics and Traumatology	Sacrococcygeal contusion	258
41	Orthopaedics and Traumatology	De Quervain's disease	133
42	Orthopaedics and Traumatology	Hip fracture	419
43	Orthopaedics and Traumatology	Tibial eminence injury	271

44	Paediatrics	Cranioencephalic trauma	311
45	Paediatrics	Catatonia and loss of consciousness	697
46	Pneumology	Acute Respiratory Failure	245
47	Pneumology	Acute Respiratory Distress Syndrome	348
48	Surgery	Thoracotomy	389
49	Urology	Genitofemoral neuralgia	206
50	Urology	Varicocele	122

5. Results

Tables 7 and 8 provide a comparative view of the source and target (i.e. simplified) texts, by applying lexical and (when suggested) also syntactic substitutions. In each table, column 1 lists the texts by number, column 2 specifies the title of the medical report, columns 3 and 4 show pre- and post-processing word counts, columns 5 and 6 report pre- and post-processing average SMOG scores, while columns 7 and 8 report pre- and post-processing average Fog scores. In columns 6 and 8, red and green indicate whether the readability index of the simplified text is, respectively, higher or lower than that of its original counterpart.

Table 7: Comparison of wordcount and average SMOG and Fog scores in the reports before and after being processed by SIMPLE® by applying both lexical and syntactic substitutions (whenever possible)

#	*Medical Record*	*Pre-processing Word count*	*Post-processing Word count*	*Pre-processing SMOG score*	*Post-processing SMOG score*	*Pre-processing Fog index*	*Post-processing Fog index*
1	Acute ischaemic stroke	737	888	10.95	10.80	12.32	12.28
2	Atrial flutter and fibrillation	217	296	12.35	11.51	14.60	14.34
3	Coronary Artery Disease	701	940	12.65	12.15	15.95	15.30
4	Mitral valve disease	1654	2221	11.85	11.24	14.6	13.72
5	Thoracic Outlet Syndrome	423	604	10.38	9.62	12.29	11.29
6	Abdominal ultrasound	371	447	11.49	10.67	11.96	11.62
7	Breast biopsy	173	224	11.83	11.89	12.71	13.64

8	Breast thermography	140	156	12.07	11.51	13.98	13.5
9	Brain CT-scan	216	244	12.43	11.17	14.88	13.34
10	Cervicothoracic CT-scan	383	462	13.64	11.91	17.48	15.7
11	Colonoscopy and Esophagoscopy	119	191	11.67	12.08	13.43	14.44
12	GI Endoscopy	129	197	11.45	11.04	14.13	13.58
13	Echocardiogram	141	166	10.76	9.02	12.67	10
14	GI Motility test	113	128	11.57	10.85	16.71	13.66
15	Right knee MRI	343	421	11.19	10.65	19.95	12.57
16	Abdominal CT-scan	394	458	12.44	11.60	15.02	14.53
17	Spine MRI	279	331	12.60	11.87	15.20	14.82
18	Thyroid ultrasound	191	242	9.23	9.51	11.11	10.49
19	Thoracic and abdominal CT-scan	408	521	11.81	11.01	14.31	13.28
20	Colon endoscopic biopsy	579	924	11.58	11.98	13.64	15.36
21	Retrograde cystography	137	187	13.15	12.89	14.95	15.02
22	Fever and chills	652	879	10.59	10.56	13.89	13.42
23	Seizures	340	444	12.39	12.57	14.79	16.47
24	Stomach, duodenum, and sigmoid colon biopsy	108	212	9.50	9.57	11.47	10.51
25	Amenorrhea	183	275	11.33	11.40	14.36	13.66
26	Vaginal ultrasound scan	444	562	11.96	11.52	14.37	14.47
27	Bilateral COVID-19 Pneumonia	294	425	14.27	14.58	18.56	20.49

28	Human Papilloma Virus	195	283	8.76	8.93	10.82	10.53
29	SARS-CoV-2 PCR test	441	541	11.70	11.13	13.30	12.47
30	AKI and hypokalaemia	288	351	12.10	11.67	13.82	14.70
31	Systemic Erythematosus Lupus	358	553	13.12	14.00	16.13	18.34
32	Tendinitis	137	180	9.05	8.03	8.61	7.02
33	Low-grade glioma	1641	1869	10.33	12.58	15.55	16.22
34	Right hemiparesis	1228	1362	11.60	11.54	14.57	14.26
35	Retinal detachment	743	858	11.71	11.47	14.61	14.50
36	Cervical root disorder	432	574	12.97	12.54	16.39	16.55
37	Intramuscular hematoma	719	865	11.78	11.86	14.37	15.21
38	Mechanical lumbago	67	101	10.36	10.66	13.30	12.76
39	Polytrauma	455	533	11.23	10.29	12.79	11.47
40	Sacrococcygeal contusion	258	330	10.87	10.02	11.68	11.18
41	De Quervain's disease	133	156	12.21	12.03	14.70	16.24
42	Hip fracture	419	509	11.95	10.99	13.58	13.15
43	Tibial eminence injury	271	308	11.69	12.26	14.56	15.96
44	Cranioencephalic trauma	311	404	10.35	10.00	11.78	11.05
45	Catatonia and loss of consciousness	697	925	10.98	10.86	12.99	13.25
46	Acute Respiratory Failure	245	340	9.92	9.40	9.01	9.07

47	Acute Respiratory Distress Syndrome	348	509	13.92	13.99	16.73	18.98
48	Thoracotomy	389	486	14.73	13.71	18.55	18.46
49	Genitofemoral neuralgia	206	260	10.87	10.10	12.37	11.29
50	Varicocele	122	202	11.41	11.98	15.15	14.59

Table 8: Comparison of wordcount and average SMOG and Fog scores in the reports before and after being processed by TSE, by applying both lexical and syntactic substitutions (whenever possible)

#	*Medical Record*	*Pre-processing Word count*	*Post-processing Word count*	*Pre-processing SMOG score*	*Post-processing SMOG score*	*Pre-processing Fog index*	*Post-processing Fog index*
1	Acute ischaemic stroke	737	827	10.95	11.03	12.32	12.11
2	Atrial flutter and fibrillation	217	223	12.35	12.21	14.60	13.78
3	Coronary Artery Disease	701	726	12.65	14.09	15.95	15.28
4	Mitral valve disease	1654	1799	11.85	11.76	14.60	13.72
5	Thoracic Outlet Syndrome	423	481	10.38	9.65	12.29	10.84
6	Abdominal ultrasound	371	389	11.49	11.42	11.96	11.41
7	Breast biopsy	173	183	11.83	12.01	12.71	12.67
8	Breast thermography	140	142	12.07	11.86	13.98	13.97
9	Brain CT-scan	216	217	12.43	11.72	14.88	13.57
10	Cervicothoracic CT-scan	383	406	13.64	13.25	17.48	16.72
11	Colonoscopy and Esophagoscopy	119	127	11.67	12.45	13.43	13.18
12	GI Endoscopy	129	132	11.45	11.19	14.13	12.67
13	Echocardiogram	141	171	10.76	10.44	12.67	11.52

14	GI Motility test	113	116	11.57	11.65	16.71	15.08
15	Right knee MRI	343	521	11.19	13.97	19.95	17.74
16	Abdominal CT-scan	394	411	12.44	12.29	15.02	14.64
17	Spine MRI	279	309	12.60	11.24	15.20	13.35
18	Thyroid ultrasound	191	191	9.23	9.23	11.11	9.92
19	Thoracic and abdominal CT-scan	408	444	11.81	11.22	14.31	12.95
20	Colon endoscopic biopsy	579	621	11.58	11.67	13.64	13.27
21	Retrograde cystography	137	148	13.15	13.42	14.95	14.59
22	Fever and chills	652	701	10.59	11.80	13.89	13.77
23	Seizures	340	351	12.39	12.06	14.79	13.91
24	Stomach, duodenum, and sigmoid colon biopsy	108	119	9.50	10.46	11.47	10.94
25	Amenorrhea	183	223	11.33	11.68	14.36	13.78
26	Vaginal ultrasound scan	444	481	11.96	12.24	14.37	14.63
27	Bilateral COVID-19 Pneumonia	294	318	14.27	15.51	18.56	19.57
28	Human Papilloma Virus	195	247	8.76	10.02	10.82	10.90
29	SARS-CoV-2 PCR test	441	460	11.70	11.88	13.30	13.07
30	AKI and hypokalaemia	288	308	12.10	12.07	13.82	13.4
31	Systemic Erythematosus Lupus	358	388	13.12	13.86	16.13	15.90
32	Tendinitis	137	147	9.05	8.83	8.61	7.85
33	Low-grade glioma	1641	1711	10.33	12.33	15.55	14.42
34	Right hemiparesis	1228	1280	11.60	11.78	14.57	13.87

35	Retinal detachment	743	752	11.71	12.81	14.61	14.72
36	Cervical root disorder	432	451	12.97	14.03	16.39	16.64
37	Intramuscular hematoma	719	750	11.78	11.70	14.37	13.68
38	Mechanical lumbago	67	77	10.36	10.63	13.30	12.44
39	Polytrauma	455	479	11.23	10.53	12.79	11.27
40	Sacrococcygeal contusion	258	279	10.87	10.54	11.68	10.78
41	De Quervain's disease	133	134	12.21	12.15	14.70	14.54
42	Hip fracture	419	453	11.95	11.96	13.58	13.20
43	Tibial eminence injury	271	281	11.69	11.62	14.56	13.75
44	Cranioencephalic trauma	311	323	10.35	10.25	11.78	11.03
45	Catatonia and loss of consciousness	697	729	10.98	10.73	12.99	12.01
46	Acute Respiratory Failure	245	269	9.92	10.05	9.01	8.86
47	Acute Respiratory Distress Syndrome	348	381	13.92	13.86	16.73	16.32
48	Thoracotomy	389	414	14.73	14.41	18.55	18.10
49	Genitofemoral neuralgia	206	217	10.87	10.10	12.37	10.89
50	Varicocele	122	144	11.41	11.63	15.15	13.95

Table 9 provides a comparative view of the source and target (i.e. simplified) texts, by applying syntactic substitutions only in TSE. In the table, column 1 lists the texts by number, column 2 specifies the title of the medical report, columns 3 and 4 show pre- and post-processing word counts, columns 5 and 6 report pre- and post-processing average SMOG scores, while columns 7 and 8 report pre- and post-processing average Fog scores. In columns 6 and 8, red and green indicate whether the readability index of the simplified text is, respectively, higher or lower than that of its original counterpart. The resulting scores are shown in black when exactly identical to the original ones.

Table 9: Comparison of wordcount and average SMOG and Fog scores in the reports before and after applying syntactic substitutions suggested by TSE

#	*Medical Record*	*Pre-processing Word count*	*Post-processing Word count*	*Pre-processing SMOG score*	*Post-processing SMOG score*	*Pre-processing Fog index*	*Post-processing Fog index*
1	Acute ischaemic stroke	737	734	10.95	11.04	12.32	12.36
2	Atrial flutter and fibrillation	217	217	12.35	12.35	14.60	14.60
3	Coronary Artery Disease	701	701	12.65	12.65	15.95	15.95
4	Mitral valve disease	1654	1654	11.85	11.85	14.60	14.60
5	Thoracic Outlet Syndrome	423	421	10.38	9.81	12.29	12.28
6	Abdominal ultrasound	371	371	11.49	11.49	11.96	11.96
7	Breast biopsy	173	173	11.83	11.83	12.71	12.71
8	Breast thermography	140	140	12.07	12.07	13.98	13.98
9	Brain CT-scan	216	214	12.43	12.45	14.88	14.85
10	Cervicothoracic CT-scan	383	383	13.64	13.64	17.48	17.48
11	Colonoscopy and Esophagoscopy	119	120	11.67	12.14	13.43	13.47
12	GI Endoscopy	129	129	11.45	11.45	14.13	14.13
13	Echocardiogram	141	140	10.76	10.72	12.67	12.71
14	GI Motility test	113	113	11.57	11.57	16.71	16.71
15	Right knee MRI	343	485	11.19	12.00	19.95	16.70
16	Abdominal CT-scan	394	392	12.44	12.47	15.02	15.05
17	Spine MRI	279	278	12.60	12.67	15.20	15.32
18	Thyroid ultrasound	191	191	9.23	9.23	11.11	11.11

19	Thoracic and abdominal CT-scan	408	408	11.81	11.81	14.31	14.31
20	Colon endoscopic biopsy	579	575	11.58	11.55	13.64	13.94
21	Retrograde cystography	137	137	13.15	13.15	14.95	14.95
22	Fever and chills	652	652	10.59	10.59	13.89	13.89
23	Seizures	340	342	12.39	12.44	14.79	14.82
24	Stomach, duodenum, and sigmoid colon biopsy	108	108	9.50	9.50	11.47	11.47
25	Amenorrhea	183	204	11.33	11.41	14.36	14.36
26	Vaginal ultrasound scan	444	443	11.96	12.00	14.37	14.39
27	Bilateral COVID-19 Pneumonia	294	250	14.27	12.67	18.56	14.77
28	Human Papilloma Virus	195	192	8.76	9.63	10.82	10.91
29	SARS-CoV-2 PCR test	441	442	11.70	11.38	13.30	12.81
30	AKI and hypokalaemia	288	288	12.10	12.10	13.82	13.82
31	Systemic Erythematosus Lupus	358	358	13.12	13.12	16.13	16.13
32	Tendinitis	137	137	9.05	9.05	8.61	8.61
33	Low-grade glioma	1641	1641	10.33	10.33	15.55	15.55
34	Right hemiparesis	1228	1226	11.60	12.00	14.57	14.48
35	Retinal detachment	743	743	11.71	11.71	14.61	14.61
36	Cervical root disorder	432	432	12.97	12.97	16.39	16.39
37	Intramuscular hematoma	719	717	11.78	12.22	14.37	5.06
38	Mechanical lumbago	67	68	10.36	10.73	13.30	13.42

39	Polytrauma	455	463	11.23	11.19	12.79	12.76
40	Sacrococcygeal contusion	258	259	10.87	10.82	11.68	11.57
41	De Quervain's disease	133	133	12.21	12.21	14.70	14.70
42	Hip fracture	419	421	11.95	12.15	13.58	13.63
43	Tibial eminence injury	271	274	11.69	12.1	14.56	14.43
44	Cranioencephalic trauma	311	311	10.35	10.35	11.78	11.78
45	Catatonia and loss of consciousness	697	693	10.98	10.95	12.99	12.90
46	Acute Respiratory Failure	245	245	9.92	9.92	9.01	9.01
47	Acute Respiratory Distress Syndrome	348	347	13.92	14.69	16.73	16.72
48	Thoracotomy	389	389	14.73	14.73	18.55	18.55
49	Genitofemoral neuralgia	206	206	10.87	10.87	12.37	12.37
50	Varicocele	122	121	11.41	11.32	15.15	14.61

5.1 Discussion of results considering the SMOG index

Considering the SMOG index SIMPLE® improved the readability of thirty four out of fifty texts (68% of cases) and worsened the readability of the remaining sixteen texts (32% of cases), as shown in Figure 4 b. On the other hand, TSE improved the readability of twenty five texts (50% of cases), worsened the readability of twenty four (48% of cases) and left the readability of 1 text unchanged (2% of cases), when applying both lexical and syntactic adjustments (Figure 5), while the SMOG score remained unchanged in half of the cases, worsened in sixteen (i.e. in 32% of cases) and improved in only nine cases (i.e. in 18% of cases) when only the syntactic substitutions suggested by the TSE were applied (Figure 6).

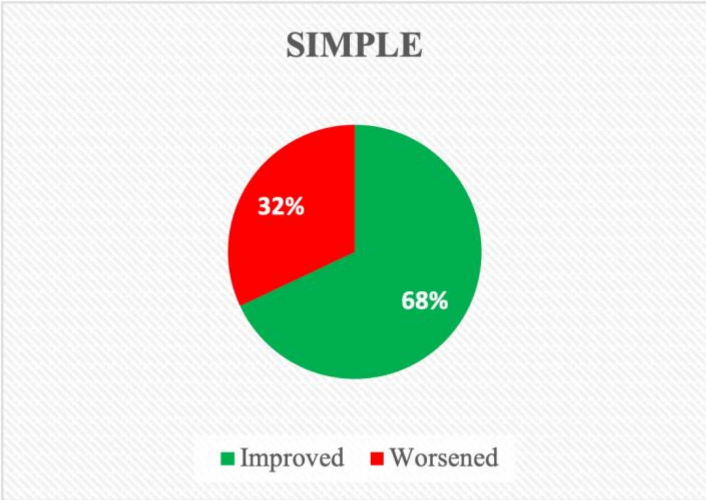

Figure 4 b: Resulting SMOG score after using SIMPLE®

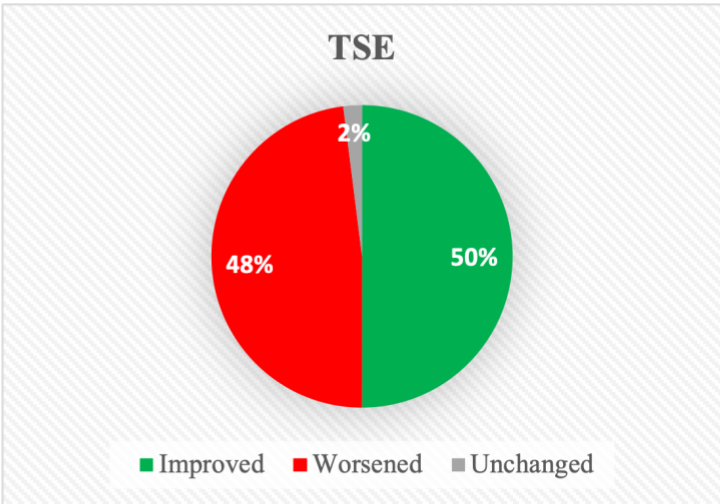

Figure 5: Resulting SMOG score after using TSE

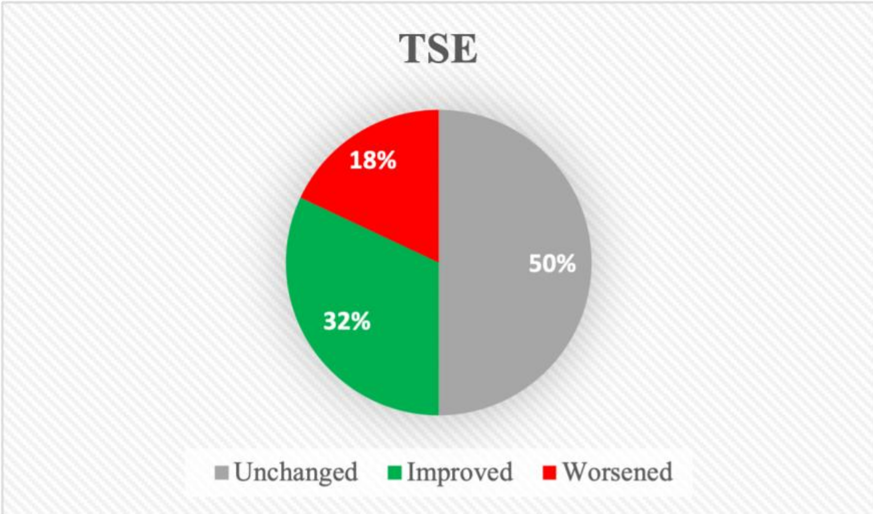

Figure 6: Resulting SMOG score after applying syntactic substitutions suggested by TSE

Thus, considering the SMOG index, the first tool, based solely on lexical simplification, appears slightly more efficient than the second tool. In any case, whichever tool we use, and whichever type of simplification we apply, differences in reading scores are never sufficient to guarantee passing from the initial education level to a lower one. Furthermore, the role of syntactic simplification seems marginal, as some form of improvement in readability is shown in only 18% of cases.

5.2 Discussion of results considering the Fog index

Considering the Fog index, SIMPLE® (Figure 7) showed improved the readability of thirty three out of 50 reports (i.e. in 66% of the cases), and worsened readability in seventeen reports (i.e. 34% of the cases). TSE performed well in forty five reports (i.e. in 90% of the cases), while in the remaining five cases (i.e. 5% of the cases) produced texts with higher Fog scores when applying both lexical and syntactic substitutions (Figure 8), while the Fog score remained unchanged in twenty six cases (i.e. in 52% of cases), worsened in eleven cases (i.e. in 22% of cases) and improved in thirteen cases (i.e. in 26% of cases) when applying syntactic substitution only (Figure 9).

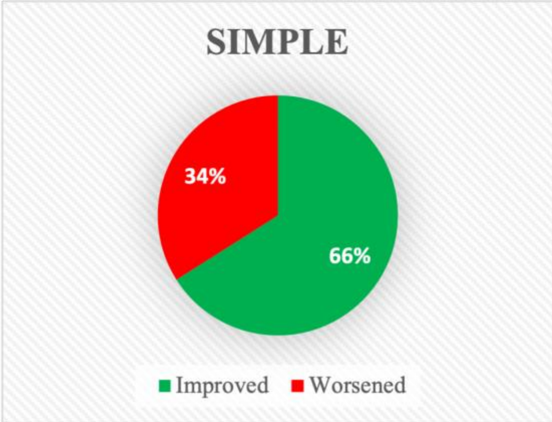

Figure 7: Resulting Fog index after using SIMPLE®

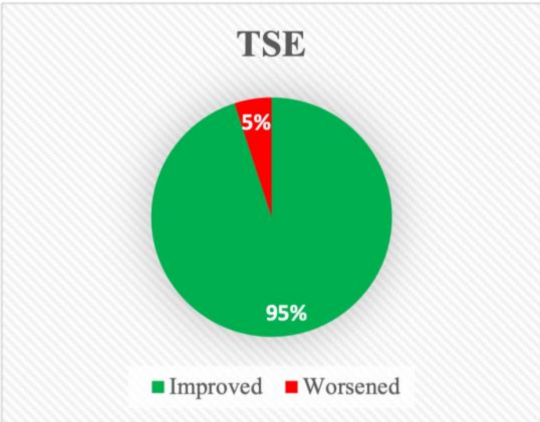

Figure 8: Resulting Fog index after using TSE

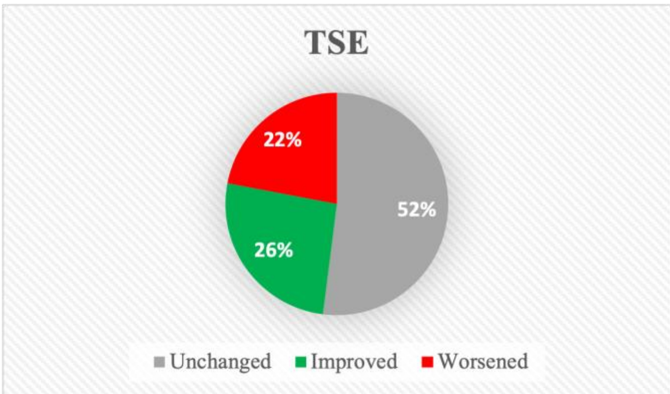

Figure 9: Resulting Fog score after applying syntactic substitutions suggested by TSE

In other words, while the Fog and SMOG formulas returned substantially similar scores for the texts simplified using SIMPLE® (with a positive result in about two thirds of the cases), the Fog formula returned scores corresponding to improved readability for most of the reports simplified by applying both the lexical and syntactic changes suggested by TSE. Finally, as with the SMOG score, the role of syntactic simplification remains marginal, as some form of improvement in readability is shown in only 22% of cases.

5.3 Further assessment of the results achieved with the two tools

Qualitative analysis of the simplified texts shows that the different results achieved by the two tools largely depend on the different dictionaries used for lexical substitution. In fact, an analysis of the suggestions provided by the two tools for a set of technical terms shows that the type and amount of information they provide to the user is quite different. More specifically, in many cases, TSE suggested only other highly technical terms and no explanation. We will here illustrate this with two sample sentences, i.e. Example 10 and Example 11 below, selected from an MR in which SIMPLE® had a good result (reducing the SMOG score of the original text) while TSE had a poor result (increasing the SMOG score of the original text):

In Example 10, taken from the 'Acute Ischaemic Stroke' report, both tools suggest an easier alternative for the technical term 'oedema' (Figure 10 and Figure 11).

No oedema in the left limbs.

Example 10: Oedema

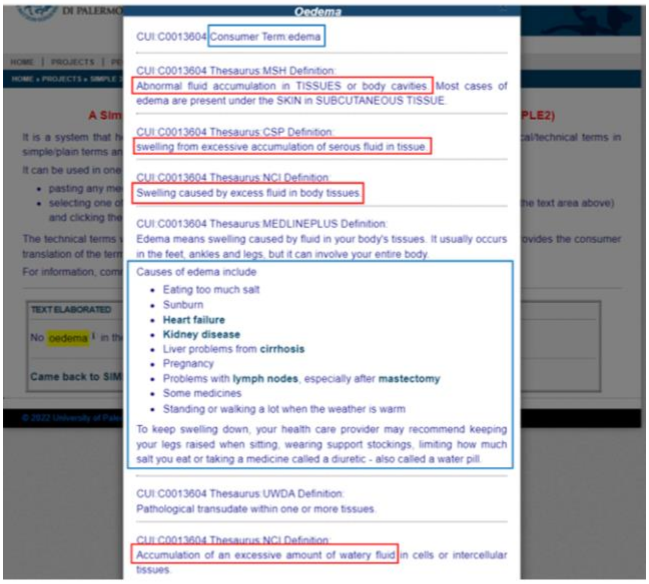

Figure 10: Definitions for 'oedema' provided by SIMPLE®

As shown in Figure 10, SIMPLE® suggests 'edema' as an equivalent consumer term, which does not particularly help the lay reader. The tool, however, also proposes several short definitions that make the technical term easy to understand (see red boxes in the figure). In addition, it provides a comprehensive explanation of the possible causes of this clinical condition and common medical solutions (see blue box in the figure).

The second tool provides several equivalent consumer terms, as shown in Figure 11.

Figure 11: Consumer equivalents of the term 'oedema' provided by TSE

Although this tool does not provide a detailed explanation of ‘oedema’, the two consumer terms marked in Figure 11 with a red box enable the lay reader to understand the meaning of this technical term.

Example 11, taken from the same MR as Example 10, i.e. Acute Ischaemic Stroke, shows how different the suggestions proposed by the two tools are for acronym ‘DVT’.

No oedema in the left limbs, no signs of DVT.

Example 11: DVT

Figure 12and 13 show the suggestions provided by the two tools for acronym ‘DVT’. SIMPLE® makes the acronym explicit and offers several explanations (Figure 12; see red boxes in the figure).

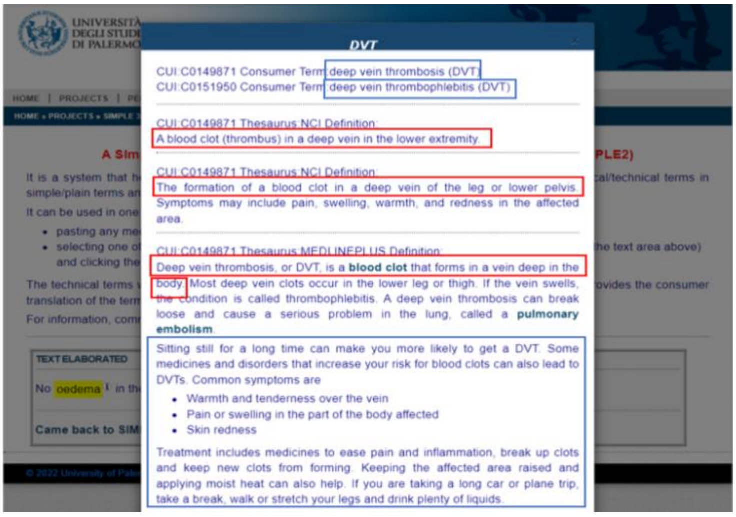

Figure 12: Consumer terms and definitions for ‘DVT’ provided by SIMPLE®

Furthermore, it also describes the possible symptoms and treatment options for this clinical condition (see blue box in the figure).

On the other hand, as shown in Figure 13, TSE makes the acronym explicit (see red box), but does not provide any kind of explanation; furthermore, the list of meanings offered includes suggestions which are unsuitable in the current context (see blue box).

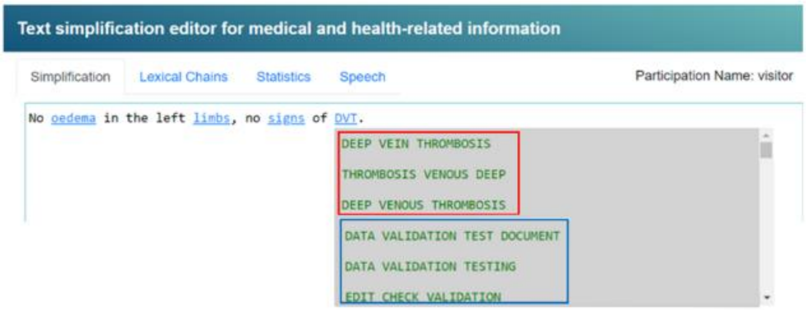

Figure 13: Suggestions to replace 'DVT' provided by TSE

In this case, the lay reader is not helped in any way and the technical term remains unclear.

A different case, but equally useless in terms of simplification, is illustrated by means of Example 12, taken from the same MR as in Example 10 and 11:

There is no sign of hemianopia.

Example 12: Hemianopia

As Figure 14 shows, the second tool only breaks down the technical term 'hemianopia' into its components.

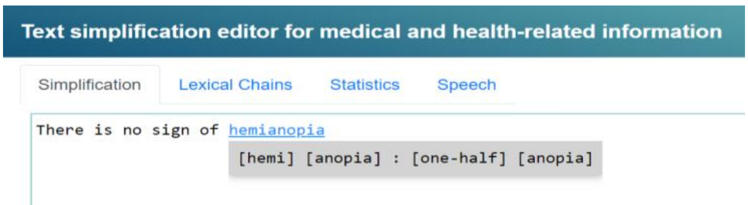

Figure 14: Suggestion to replace 'hemianopia' provided by TSE

In contrast, by submitting the same portion of text to SIMPLE®, we obtained an equally technical synonym ('hemianopsia'), but also explanations of the term, as shown in Figure 15 (see blue and red box respectively in the figure):

Figure 15: Consumer term and definitions for 'hemianopia' provided by SIMPLE®

Thus, with regard to facilitating comprehension of technical terms, SIMPLE® can be said to perform in a more fruitful way, as it provides the reader with a range of types of suggestions and information, thus catering for the different needs of a varied population of users.

The types of syntactic simplification suggested by TSE, on the other hand, are exemplified in Figure 16 and Figure 17 (see blue boxes in the figures). The following example is an excerpt from the 'Acute ischaemic stroke' report:

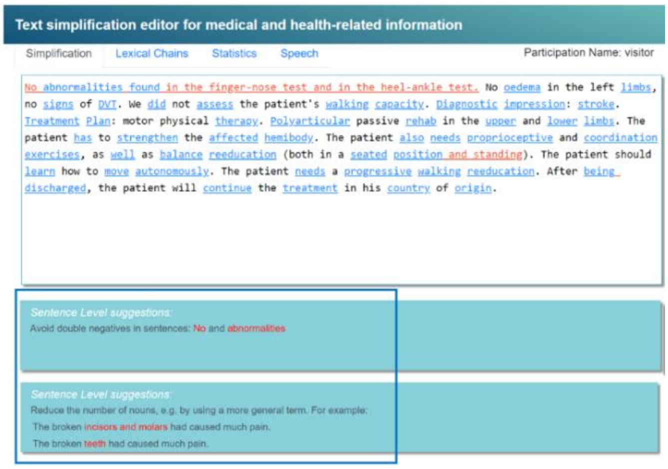

Figure 16: Sentence-level suggestions provided by TSE: double negations and number of nouns

The tool suggests eliminating the double negation consisting of a negation adverb ('no') and a morphologically negative word ('abnormalities'); furthermore, later in the text it suggests reducing the number of nouns, e.g. by using hypernyms.

The following example is an excerpt from the 'Atrial flutter and fibrillation' report:

Figure 17: Sentence-level suggestions provided by TSE: passive to active

Here the tool suggests turning the passive form into the active one. In neither case, does the tool actually clarify how to make these substitutions, but, considering that the addressee of this tool is the editor of the MR and not a lay reader, we can assume that whoever is writing the text can easily make these substitutions.

It can also be observed that in Tables 7 and 8 the post-processing word count is higher in all texts processed, regardless of the tool used. Since the impact of syntactic reformulation is practically zero (see Table 9), we can only infer that such increased word count is due to the fact that when no consumer term is available for the highlighted technical term, both tools display a definition, thus contributing to making the text longer and complicating its syntactic structure.

This result largely depends on the fact that while both formulas take into account word length, only the Fog index considers sentence length (see Section 4.2). In fact, although the processing of reports with both tools consistently produced texts with a higher word count, in the overwhelming majority of the reports analysed, the percentage change in word count produced by SIMPLE® is significantly higher than that produced by TSE, as shown in Table 10.[4]

Table 10: Percentage changes produced by SIMPLE® and TSE in the analysed MRs

#	MR	Δ_{WC} produced by SIMPLE®	Δ_{WC} produced by TSE

4 To calculate the percentage change in word count the following formula is used: ΔWC = [(WCpost – WCpre) / WCpre] × 100, where WCpre indicates the word count of the report before processing by either tool, and WCpost indicates the resulting word count. For example, the percentage change in the word count of the ‘Brest Biopsy’ report after being processed by SIMPLE®, is as follows: ΔWC = [(224 – 173) / 173] × 100 = 29.48%

1	Acute ischaemic stroke	20.49%	12.21%
2	Atrial flutter and fibrillation	36.40%	2.76%
3	Coronary Artery Disease	34.09%	3.57%
4	Mitral valve disease	34.28%	8.77%
5	Thoracic Outlet Syndrome	42.78%	13.71%
6	Abdominal ultrasound	20.48%	4.85%
7	Breast biopsy	29.48%	5.78%
8	Breast thermography	11.43%	1.41%
9	Brain CT-scan	12.96%	0.46%
10	Cervicothoracic CT-scan	20.63%	6.00%
11	Colonoscopy and Esophagoscopy	60.50%	6.72%
12	GI Endoscopy	52.71%	2.32%
13	Echocardiogram	17.73%	21.28%
14	GI Motility test	13.27%	2.65%
15	Right knee MRI	22.74%	51.89%
16	Abdominal CT-scan	16.24%	4.31%
17	Spine MRI	18.64%	10.75%
18	Thyroid ultrasound	26.70%	0%
19	Thoracic and abdominal CT-scan	27.70%	8.82%
20	Colon endoscopic biopsy	59.78%	7.25%
21	Retrograde cystography	36.50%	8.03%
22	Fever and chills	34.81%	7.51%
23	Seizures	30.59%	3.23%
24	Stomach, duodenum, and sigmoid colon biopsy	96.30%	10.18%

25	Amenorrhea	50.27%	21.86%
26	Vaginal ultrasound scan	26.58%	8.33%
27	Bilateral COVID-19 Pneumonia	44.56%	8.16%
28	Human Papilloma Virus	45.13%	26.67%
29	SARS-CoV-2 PCR test	22.67%	4.31%
30	AKI and hypokalaemia	21.87%	6.94%
31	Systemic Erythematosus Lupus	54.47%	8.38%
32	Tendinitis	31.39%	7.30%
33	Low-grade glioma	13.89%	4.26%
34	Right hemiparesis	10.91%	12.38%
35	Retinal detachment	15.48%	1.21%
36	Cervical root disorder	32.87%	4.40%
37	Intramuscular hematoma	20.30%	4.31%
38	Mechanical lumbago	50.75%	14.92%
39	Polytrauma	17.14%	5.27%
40	Sacrococcygeal contusion	27.91%	8.14%
41	De Quervain's disease	17.29%	0.75%
42	Hip fracture	21.48%	8.11%
43	Tibial eminence injury	13.65%	3.69%
44	Cranioencephalic trauma	29.90%	3.86%
45	Catatonia and loss of consciousness	32.71%	4.59%
46	Acute Respiratory Failure	38.77%	9.79%
47	Acute Respiratory Distress Syndrome	46.26%	9.48%

48	Thoracotomy	24.93%	6.43%
49	Genitofemoral neuralgia	26.21%	5.34%
50	Varicocele	65.57%	18.03%

As is evident from Table 10, TSE resulted in a lower percentage change than that determined by SIMPLE® in forty eight out of fifty reports (i.e. in 96% of cases). Furthermore, in the majority of cases (thirty nine reports, i.e. 78% of cases) the percentage change produced by TSE was lower than 10%. In contrast, SIMPLE® resulted in a percentage change greater than 20% in thirty eight reports (i.e. in 76% of cases).

Therefore, the fact that SIMPLE® performed worse than TSE with the Fog index can be attributed to the fact that sentence length plays a decisive role in calculating this specific index.

A primary cause of the production of longer simplified texts by SIMPLE® in our experiment is certainly the fact that we had to enter all the relevant listed definitions up to the first useful one, which in turn led to adding more words than with TSE (see Section 4). But this was the only possible way to replicate the behaviour of a lay reader who is obliged to read through several definitions before finding one that suits them.

5.3 A simplification proposal based on the results of the analysis

Based on the results obtained from the comparison of these two tools, it is evident that the replacement of technical terms with the equivalent consumer terms has a positive impact on the readability of the text, whereas it would appear that the syntactic simplification of the texts does not have a significant influence on it. Accordingly, in order to most effectively simplify a medical report and thus make it easier for the lay reader to read and understand it, it would be appropriate to focus on lexical simplification. Furthermore, as we ascertained with SIMPLE, replacing a technical term with a long definition is not the best way to improve the readability of the report. Therefore, it would be appropriate to work on the corpus of definitions that the software draws from in order to provide the user with short and simple definitions that have a readability index suitable for the lay reader.

In order to further improve lexical substitution, the match between each technical term and the consumer term should be verified by an expert, so that a single technical term does not correspond to several consumer terms, which could mislead the lay reader who would not be able to select the correct alternative.

As far as the software interface is concerned, that of TSE would seem to be better since it allows the readers to obtain a new text on the basis of the simplifications they have chosen to make. An interface like that of SIMPLE®, on the other hand, could further hinder the reading of the report since it is not possible to apply the simplifications in the source text.

To sum up, it appears to us that effective simplification software should focus on lexical simplification only and, only in the absence of consumer terms, provide a single short definition through an interface that allows the source text to be edited thus obtaining a simplified equivalent text.

One approach to obtain clear and straightforward medical reports is to intervene early during their composition, using a commonly agreed reference standard. As we mentioned in Section 3.1, the ASD STE specification guidelines were originally conceived for the aerospace industry. However, considering the results of this study on the impact of sentence length and structure on the readability of texts, the following ASD STE specification rules could possibly be used to successfully simplify medical report writing ([5]):

- Rule 5.1: Use a maximum of 20 words in each sentence (for procedural writing).
- Rule 6.3: Use a maximum of 25 words in each sentence (for descriptive writing).
- Rule 6.5: Make sure that each paragraph has only one topic.
- Rule 6.6: Make sure that no paragraph has more than six sentences.

By following these rules, it is possible to adjust both the length of the sentences that make up each medical report, by working upstream on the number of words that make up each sentence, and the complexity of the texts as a whole, by reducing the number of sentences that make up each paragraph of the report.

To the best of our knowledge, no research has been done so far on how these rules affect the readability of medical reports. However, the impact these guidelines have had on other types of technical texts ([36]; [46]; [17]) bodes well in this respect.

6. Conclusions and future work

In this paper, we analysed the performance of two different text simplification tools on a sample of fifty MRs. To this aim, we compared the tools' output using two different readability indexes.

On the whole, we can say that both tools succeeded in producing some degree of simplification; however, their performance varied depending on text and on the readability formula considered. When we applied the SMOG score, SIMPLE® performed better than TSE, with 68% (vs. 50%) of simplified texts showing better readability scores than the corresponding non-simplified texts, and 32% (vs. 48%) showing worse readability. On the contrary, when our reference was the Fog index, TSE performed significantly better than SIMPLE®, with 95% (vs. 66%) of simplified texts showing better readability scores than the corresponding non-simplified texts, and 5% texts (vs. 43%) showing worse readability. At the same time, the target texts obtained with SIMPLE® showed similar scores regardless of the readability formula (with a positive result in about two thirds of the cases), while the target texts obtained with TSE showed scores corresponding to improved readability in 95% cases when using the Fog formula, compared to only 50% when using the SMOG formula.

Such results are largely due to the different types of lexical suggestions provided by each tool matched to the fact that the SMOG score considers the length of the sentences that make up the text. In fact, when no consumer term is available for a technical term, SIMPLE® provides a long list of definitions in random order, thus obliging the patient to read several, often long, definitions before they can find a suitable one. As shown in our calculations, sentence length impacts on readability scores, and in particular on the Fog score. But this should not be considered a purely mathematical issue, in so far as having to read long definitions certainly increases the time and effort a patient puts into reading their MRs, and if excessive, may even put them off from using the simplification tool. To solve this issue, simplification tools should ideally find a way to display only one definition, selected among the many available ones according to readability as well as information criteria.

Syntactic simplification as offered in TSE, on the other hand, has produced very little improvement, if any, on the readability of our MRs. However, we believe this to be due to the fact that the types of syntactic reformulation suggested by TSE (see Section 4), though certainly suitable in other contexts, are not sufficient when it comes to MRs. As seen in Section 2, besides nominalization and passive forms, MRs are characterised by other syntactic and stylistic features which hamper understanding by patients, such as use of pre-modification instead of relative clauses, and omission of phrasal elements. Thus, further research should focus on assessing the impact of reformulation of these other syntactic features on readability, how to automatically identify the parts of a text that need simplifying and what kind of suggestions the software tool should provide.

Furthermore, some texts consistently obtained worst readability indexes in SIMPLE® regardless of the index considered (texts 7, 11, 20, 23, 27, 31, 33, 37, 43, 47). In the TSE, on the other hand, only five texts obtained worst readability scores with both indexes (texts 26, 27, 28, 35, 36). Only text 27 is common to both lists. Detailed linguistic analysis of these specific texts and of their corresponding simplified versions has not been performed in the current study, but it may provide further insight into the features that proved most problematic for these simplification tools and lead to new suggestions on how simplification tools should work for this kind of medical texts.

Finally, text simplification should be tested on a much wider variety of genres, including other text types produced by doctors or healthcare providers and intended for lay readers (such as drug leaflets, information brochures, websites, etc.). Moreover, considering that numerous readability formulas exist based on different criteria, further methods of evaluating the effectiveness of these types of tools are needed. It would be appropriate, for example, to have several users run the tools on a sample of texts and then submit a questionnaire to the users to check their actual understanding of the source and target texts. The participants should be lay readers and have no specific knowledge in the medical domain. Considering that any readability formula can be associated with a level of education, the sample of participants should be divided not only on the basis of demographic variables, native language, and occupational sector, but also on the basis of the highest level of education achieved. In addition, the study could also involve non-native speakers (such as immigrants or occasional visitors) categorised on the basis of their proficiency. This could prove the actual usefulness of this type of tool for real users, allowing developers to introduce new features or combine elements from different tools.

References

[1] Adkins, A. D., and Singh, N. N. 2001. "Reading Level and Readability of Patient Education Materials in Mental Health." *Journal of Child and Family Studies* 10 (1): 1-8. https://doi.org/10.1023/A:1016617115729

[2] Ainsworth-Vaughn, N. 2003. "The Discourse of Medical Encounters". In *The Handbook of Discourse Analysis*, edited by D. Schiffrin, D. Tannen, and H. E. Hamilton, 453-469. Malden: Blackwell Publishing. https://doi.org/10.1002/9780470753460.ch24

[3] Alfano, M., Lenzitti, B., Bosco, G. L., Muriana, C., Piazza, T., and Vizzini, G. 2020. "Design, development and validation of a system for automatic help to medical text understanding." *International Journal of Medical Informatics* 138 (1): 1-8. https://doi.org/10.1016/j.ijmedinf.2020.104109

[4] Aripova, D. A., & Bashmakova, I. S. (2018). Towards Phraseologism Formation and Terminologization in Scientific and Technical Texts. *SHS Web Conf. The International Scientific and Practical Conference "Current Issues of Linguistics and Didactics: The Interdisciplinary Approach in Humanities and Social Sciences" (CILDIAH-2018)*, 50, 1-8. https://doi.org/10.1051/shsconf/20185001028

[5] ASD, 2021. "Simplified Technical English Specification ASD-STE100". *International specification for the preparation of technical documentation in a controlled language* 1 (8): 81-93.

[6] Askehave, I., and Korning Zethsen, K. 2000. "Medical texts made simple – dream or reality?". *Hermes, Journal of Linguistics* 1 (25): 63-73. https://doi.org/10.7146/hjlcb.v13i25.25585

[7] Carlino, O., 2021. "Can medical reports be made SIMPLE? Improving consumer understanding of medical reports by testing the SIMPLE® software on English and Italian texts." Unpublished Master's thesis. Università del Salento.

[8] Bailin, A., and Grafstein, A. 2001. "The linguistic assumptions underlying readability formulae: A critique." *Language and Communication* 21 (3): 285-301. https://doi.org/10.1016/S0271-5309(01)00005-2

[9] Cabré, M. T. 1999. *Terminology: Theory, Methods, and Applications*. Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/tlrp.1

[10] Cappelletti, P. 2004. "Il "referto" in Medicina di Laboratorio." *Rivista Italiana di Medicina di Laboratorio* 5 (3): 197-208. https://www.sipmel.it/it/riviste/articolopdf.php/1253

[11] Cappuzzo, B. 2020. "Medical English Translation into Italian. The Case of Netter's Work." *ESP ACROSS CULTURES* 17 (1): 31-48. https://edipuglia.it/wp-content/uploads/2021/07/Cappuzzo.pdf

Compton, Donald L., Amanda C. Appleton and Michelle K. Hosp. 2004. "Exploring the Relationship Between Text-Leveling Systems and Reading Accuracy and Fluency in Second-Grade Students Who Are Average and Poor Decoders." *Learning Disabilities Research and Practice* 19 (1): 176-184. https://doi.org/10.1111/j.1540-5826.2004.00102.x

[13] Connatser, B., and Peac, E. 1999. "Last rites for readability formulas in technical communication." *Journal of Technical Writing and Communication* 29 (1): 271-287. https://doi.org/10.2190/6EWH-J5C5-AV1X-KDGJ

[14] Dale, E., and Chall, J. 1948. "A Formula for Predicting Readability." *Educational Research Bulletin* 27 (1): 11-28. http://garfield.library.upenn.edu/classics1982/A1982NY36000001.pdf

[15] Danes, F. 1974. "Functional sentence perspective and the organisation of the text." In *Papers in Functional Sentence Perspective*, edited by F. Danes, 106-128. Prague: Academia. https://doi.org/10.1515/9783111676524

[16] Derguech, W., Sana E Zainab, S., & D'Aquin, M. 2018. "Assessing the Readability of Policy Documents: The Case of Terms of Use of Online Services." *ICEGOV '18: Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance,* 247-256. New York: Association for Computing Machinery. https://doi.org/10.1145/3209415.3209498

[17] Díaz-Galán, A., and Fumero-Pérez, M. C. (2020). "Construcciones en ASD-STE100: Una formalizacion del significado no proposicional en los textos instructivos empleados en la aviacion." *RaeL Revista Electronica de Linguistica Aplicada,* 1 (19): 24-34. https://link.gale.com/apps/doc/A632853600/IFME?u=anon~4c9d9aeb&sid=googleScholar&xid=76ef55e6.

[18] Flesch, R. 1948. "A new readability yardstick." *Journal of Applied Psychology* 32 (3): 221-233. https://doi.org/10.1037/h0057532

[19] Gatto, M. 2011. "The 'body' and the 'web'. The web as corpus ten years on." *ICAME Journal 35*, 35-58. http://icame.uib.no/ij35/Maristella_Gatto.pdf

[20] Gotti, M. 2003. *Specialized Discourse. Linguistic Features and Changing Conventions*. Bern: Peter Lang. https://doi.org/10.1016/j.pragma.2004.12.009

[21] Gunning, R. 1952. *The Technique of Clear Writing*. New York: McGraw-Hill Book Company.

[22] Ha, J., & Longnecker, N. (2010). Doctor-patient communication: a review. *The Ochsner Journal, 10*(1), 38-43. https://pubmed.ncbi.nlm.nih.gov/21603354/

[23] Harrison, C. 1980. *Readability in the classroom*. New York: Cambridge University.

[24] Heydari, P. 2012. "The Validity of Some Popular Readability Formulas." *Mediterranean Journal of Social Sciences* 3 (2), 423-435. https://doi.org/10.5901/mjss.2012.v3n2.423

[25] Huckin, T. N., and Olsen, L. A. 1983. *English for Science and Technology: A Handbook for Nonnative Speakers*. McGraw-Hill College.

[26] Kauchak, D., and Leroy, G. 2020. "A Web-Based Medical Text Simplification Tool." *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 3749-3757. https://core.ac.uk/download/pdf/286030516.pdf

[27] Klare, G. 1963. *The measurement of readability*. Ames: Iowa State University Press.

[28] Kilgariff, A. 2001. "Web as a corpus." *Proceedings of Corpus Linguistics 2001 conference*, 342-344. Lancaster: Lancaster University. https://www.kilgarriff.co.uk/Publications/2003-KilgGrefenstette-WACIntro.pdf

[29] Kincaid, J., Fishburne, R., Rodgers, R., and Chissom, B. 1975. *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel.* Millington: Institute for Simulation and Training, University of Central Florida. https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary

[30] Koch-Weser, S., DeJong, W., and Rudd, R. E. 2009. "Medical word use in clinical encounters." *Health Expectations* 12 (4): 371-382. https://doi.org/10.1111/j.1369-7625.2009.00555.x

[31] Koptient, A., and Grabar, N. 2020. "Rated Lexicon for the Simplification of Medical Texts." *The Fifth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing HEALTHINFO 2020*, 1-8. https://hal.archives-ouvertes.fr/hal-03095275/document

[32] Kvashnina, O. S., and Sumtsova, O. V. 2018. "Using Google to Search Language Patterns in Web-Corpus: EFL Writing Pedagogy." *Italian Journal of Educational Technology* 13 (3): 173-179. https://doi.org/10.3991/ijet.v13i03.7712

[33] Kuzmina, O. D., Fominykh, A. D., and Abrosimova, N. A. 2015. "Problems of the English Abbreviations in Medical Translation." i 199 (1): 548-554. https://doi.org/10.1016/j.sbspro.2015.07.545

[34] Leroy, G., and Kauchak, D. 2014. "The effect of word familiarity on actual and perceived text difficulty." *Journal of the American Medical Informatics Association* 21 (1): 169-172. https://doi.org/10.1136/amiajnl-2013-002172

[35] Leroy, G., Endicott, J. E., Mouradi, O., Kauchak, D., and Just, M. L. 2012. "Improving Perceived and Actual Text Difficulty for Health Information Consumers using Semi-Automated Methods." *AMIA 2012 Annual Symposium*, 522-531. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540563/

[36] Mader, S., and Urban, B. (2010). "Creating Instructional Content for Augmented Reality based on Controlled Natural Language Concepts." *20th International Conference on Artificial Reality and Telexistence*, 9-13. https://icat.vrsj.org/ICAT2010_Proceedings/paper/paper-002.pdf

[37] Madera, M., Tomoń, R., and Lorenc, P. 2016. "Medical reporting in web-based applications designed to meet regulatory and industry standards." *Position Papers of the Federated Conference on Computer Science and Information Systems*, 151-156. https://annals-csis.org/proceedings/2016/pliks/574.pdf

[38] Maglie, R. 2009. *Understanding the language of medicine*. Rome: Aracne.

[39] McCall, W., and Schroeder, L. C. 1961. *McCall-Crabbs Standard Test Lessons in Reading*. New York: Teachers College Press.

[40] McLaughlin, G. 1969. "SMOG grading: a new readability formula." *Journal of Reading* 12 (8): 639-646. https://eric.ed.gov/?id=EJ004016

[41] Meyer, I., & Mackintosh, K. (2000). When terms move into our everyday lives: An overview of de-terminologization. *Terminology, 6*, 111-138. https://doi.org/10.1075/term.6.1.07mey

[42] Ong, E., Damay, J., Lojico, G., Lu, K., and Tarantan, D. 2006. "Simplifying Text in Medical Literature." *Journal of Research in Science, Computing and Engineering* 4 (1): 1-27. https://ejournals.ph/article.php?id=1754

[43] Roter, D., and Hall, J. 1992. *Doctors talking with patients/patients talking with doctors*. Westport: Auburn House. https://psycnet.apa.org/record/1992-98766-000

[44] Sager, J. C., Dungworth, D., and McDonald, P. F. 1980. *English Special Language: Principles and Practice in Science and Technology*. Amsterdam: John Benjamins Pub Co. https://doi:10.1017/S0047404500009143

[45] Salager-Meyer, F. 1990. *Discoursal flaws in Medical English abstracts: A genre analysis per research- and text-type*. Berlin: Walter de Gruyter. https://doi.org/10.1515/text.1.1990.10.4.365

[46] Vaskic, L., and Paetzold, K. (2019). "A Critical Review of the Integrated Logistics Support Suite for Aerospace and Defence Programmes." *International Conference on Engineering Designs, ICED19*, 3541-3550. https://doi:10.1017/dsi.2019.361

[47] Vihla, M. 1999. *Medical Writing: Modality in Focus*. Leiden: Brill (1999). https://doi.org/10.1075/jhp.2.1.09bib

[48] Wang, L., Miller, M. J., Schmitt, M. R., and Wen, F. K. 2013. "Assessing readability formulas differences with written health information materials: Application, results, and recommendations." *Research in Social and Administrative Pharmacy* 9 (5): 503-516. https://doi.org/10.1016/j.sapharm.2012.05.009

[49] Wermuth, M. C., and Verplaetse, H. 2019. *Medical terminology in the Western world*. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/hot.2.05wer

[50] Zamanian, M., and Heydari, P. 2012. "Readability of Texts: State of the Art." *Theory and Practice in Language Studies* 2 (5): 43-53. https://doi:10.4304/tpls.2.1.43-53