Journal of the American Medical Informatics Association, 29(11), 2022, 1976–1988
https://doi.org/10.1093/jamia/ocac149
Advance Access Publication Date: 9 September 2022
Review

Review

A survey of automated methods for biomedical text simplification

Brian Ondov, **Kush Attal**, and **Dina Demner-Fushman**

Computational Health Research Branch, National Library of Medicine, Bethesda, Maryland, USA

Corresponding Author: Brian Ondov, PhD, Computational Health Research Branch, National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894, USA; brian.ondov@nih.gov

Received 27 May 2022; Revised 26 July 2022; Editorial Decision 7 August 2022; Accepted 16 August 2022

ABSTRACT

Objective: Plain language in medicine has long been advocated as a way to improve patient understanding and engagement. As the field of Natural Language Processing has progressed, increasingly sophisticated methods have been explored for the automatic simplification of existing biomedical text for consumers. We survey the literature in this area with the goals of characterizing approaches and applications, summarizing existing resources, and identifying remaining challenges.
Materials and Methods: We search English language literature using lists of synonyms for both the task (eg, "text simplification") and the domain (eg, "biomedical"), and searching for all pairs of these synonyms using Google Scholar, Semantic Scholar, PubMed, ACL Anthology, and DBLP. We expand search terms based on results and further include any pertinent papers not in the search results but cited by those that are.
Results: We find 45 papers that we deem relevant to the automatic simplification of biomedical text, with data spanning 7 natural languages. Of these (nonexclusively), 32 describe tools or methods, 13 present data sets or resources, and 9 describe impacts on human comprehension. Of the tools or methods, 22 are chiefly procedural and 10 are chiefly neural.
Conclusions: Though neural methods hold promise for this task, scarcity of parallel data has led to continued development of procedural methods. Various low-resource mitigations have been proposed to advance neural methods, including paragraph-level and unsupervised models and augmentation of neural models with procedural elements drawing from knowledge bases. However, high-quality parallel data will likely be crucial for developing fully automated biomedical text simplification.

Key words: natural language processing, machine learning, text simplification, plain language, health literacy

INTRODUCTION

The ability of patients to understand health issues has been shown to have a significant impact on outcomes.[1] The preponderance of information now available to the general public may have the potential improve health literacy, which could increase patients' autonomy, engagement, and compliance when it comes to their own healthcare.[2–4] Resources along these lines range from medical literature and online encyclopedias to clinical notes, which are increasingly made available to patients through healthcare portals. Utilizing these resources, however, comes with major barriers.[5] Naturally, patients and caretakers will not always have the educational background needed to grasp all the concepts they encounter in such documents. However, much of the impenetrability of these types of documents is more akin to a language barrier, created by the distinctive lexicon of medicine,[6] which often parallels general terminology. Further, literature and encyclopedias often employ long sentences with complex syntax, making their general reading levels higher than much of the public's.[7] Both of these potential hurdles may be conquerable in light of recent methods in Machine Translation and Text Simplification. This has led researchers to explore the possibility of automatically adapting biomedical text to plain language for a consumer audience. Figure 1 illustrates an example of a manual simplification that highlights some of the difficulties that may be faced by automated systems. The original passage is from a biomedical abstract.[8] In this survey, we provide relevant background, review progress that has been made in biomedical text simplification, and discuss current challenges.

BACKGROUND

Biomedical text simplification is an intersection of a computational task and a language domain. We will present background relating to both areas.

Plain language in medicine

As a remedy for low health literacy, the use of Plain Language is increasingly advocated for by practitioners and researchers.[9–14] Many consumer-oriented online knowledge bases, such as the Merck Manuals[15] and MedlinePlus,[16] incorporate Plain Language. However, the time and resources needed to survey and adapt the latest literature for the public means they are not likely to contain very recent research findings. This is especially important during emerging health crises, when consumers are looking to the Internet to find such knowledge.[17] Plain Language Summaries (PLS), which are short, consumer-oriented summaries published alongside original research, may be a solution. There has been considerable effort to encourage the publishing of PLS along with research articles, systematic reviews (such as from Cochrane), and clinical trials. However, PLS have nonetheless been reported to be difficult to find and inconsistently written.[18–20]

Automatic text simplification

Since the general task of automatically simplifying text is a broad area and beyond the scope of this review, we will direct the reader to other surveys, such as those by Shardlow et al[21] and Al-Thanyyan and Azmi,[22] for a fuller understanding of methods and resources. However, here we will briefly cover the major developments and practices as they may relate to biomedical text simplification.

Procedural methods

Earlier automatic simplification methods focused on procedurally addressing the 2 major components of simplification: lexical and syntactic.[23] Procedural (or "rule-based") lexical simplification uses knowledge bases (such as plain language thesauri) to identify rare or difficult terms and substitute them for more common or shorter synonyms.[24] Procedural syntactic simplification generally uses parse trees to perform operations like pruning, splitting, or pattern-based paraphrasing.[25,26]

Statistical methods

Driven by availability of parallel data and improvements in computing power (among other factors), Statistical Machine Translation (SMT)[27] drove much of the early progress in Machine Translation (the use of computers to automatically translate from one natural language to another).[28] By posing the simplification problem as one of "translating" from one sublanguage (ie, the more complex register) to another (ie, the simplified register), SMT is also applicable to this task. However, there are challenges in translating within a language, the foremost being data acquisition. For translation between languages, data are largely found in existing parallel corpora, since textual resources often need to be read in different languages. These corpora have typically already been translated sentence by sentence, preserving as much meaning as possible. Individual pairs of sentences simply need to be aligned from such "bitexts," typically using fairly straightforward heuristics and dynamic programs. Bitexts for different audiences in the same language, however, are rarer. Those that exist are typically adapted at the document level, potentially splitting, rearranging, or omitting sentences and phrases. These types of sentence-level operations, though rare in translation,[29] are much more common in simplification, requiring more advanced strategies for alignment. Much of the progress in data-driven simplification (ie, statistical and neural approaches) has thus been enabled by the "mining" of training data from comparable corpora, using various heuristics and measures of semantic similarity.[30–32] Using

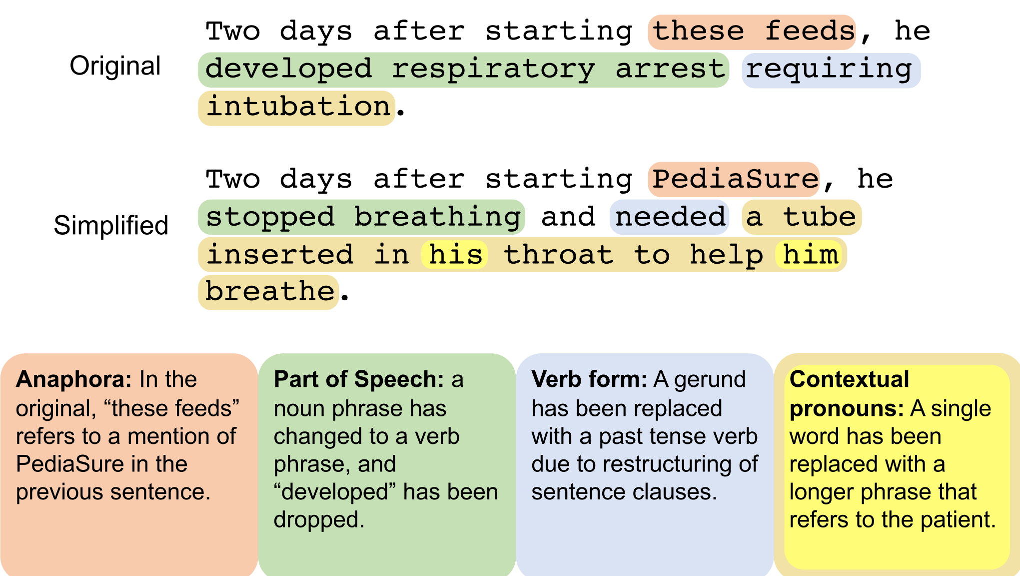

Figure 1. An example of biomedical text simplification highlighting some of the difficulties that may be encountered.

such mined data, the SMT framework and simplification-specific variants of it have been explored.[32–35] However, simplification tends to involve open-ended paraphrasing to a greater extent than translation, with many options for lexical substitution. It has been noted that this poses a problem for SMT, which relies on consistently observed parallel words and phrases.[36]

Neural methods

Advances in Deep Learning methods, along with further advances in hardware and data sources, led to the dominance of the neural sequence-to-sequence model[37] over statistical models for Machine Translation and many other linguistic tasks. Since then, much of the progress in simplification has naturally focused on neural models as well. Parallel data sets can be used to directly train sequence-to-sequence models.[38,39] Deviations from this basic paradigm have included teaching neural models to make particular types of edits[40] or output text with desired characteristics, such as parse tree depth,[41] and using reinforcement learning to reward simplicity, relevance, and fluency of the output.[42] Since neural models, like statistical models, require large amounts of training data, some work in the neural era has also focused on using auxiliary neural networks to improve the mining of training pairs.[43–45] The lack of truly parallel training data, however, has also generated interest in hybrids of procedural and neural methods,[46] unsupervised methods,[47] and zero-shot methods that leverage parallel texts across languages.[48]

Evaluation

As with other natural language tasks, automated evaluation metrics are crucial for the development of increasingly advanced methods, especially those that are data-driven. However, the difficulty of rating simplifications automatically has also made human evaluation important. Automatic metrics can be broadly grouped into readability, reference-based, and reference-free:

- Readability: systems can in theory be evaluated with traditional measures of readability, such as the Flesch-Kincaid Grade Level (FKGL),[49,50] SMOG (Simple Measure of Gobbledygook),[51] LIX/RIX/OVIX,[52–55]which use attributes like syllable counts and sentence length to calculate text difficulty. However, these metrics are easily manipulated[56] and do not necessarily correspond to reader comprehension, which is ultimately the goal of simplification.[53] Further, word complexity may not correspond to familiarity when considering domain-specific language.[57,58] Additionally, inline explanations of terms are often required for comprehension but make sentences longer and more complex.[59] These phenomena can make standard readability metrics misleading for automatic text simplification, leading to alternatives being proposed.[60]
- Reference based: For data-driven approaches that are framed as translation problems, it is intuitive to use reference-based metrics, such as BLEU (BiLingual Evaluation Understudy),[61] which measures the fraction of common substrings between a system-generated output and a gold-standard reference translation (or set of translations). However, BLEU can be misleading when applied to simplification because of the specific operations of this task.[62] Xu et al[35] thus introduce a new, simplification-specific reference-based metric, SARI, that incorporates the original input in addition to reference simplifications. This allows SARI to characterize n-grams that were kept, added or deleted compared to the original, and balance these 3 operations in rating system outputs. Additionally, other translation-based metrics such as METEOR and TER (Translation Edit Rate) have been used.[63] Paraphrase in n-gram changes (PINC) was developed specifically for rating paraphrases and rewards novelty, essentially making it the inverse of BLEU.[64] BERTscore[65] relaxes the exact token (or stem) matching requirement of other metrics by comparing contextual embeddings using BERT.[66]
- Reference free: Sulem et al[67] introduce SAMSA, a metric that compares syntactic content of a generated output to its input, forgoing the need for reference outputs.

Figure 2 shows examples of scoring with BLEU and SARI, which are the most commonly used automatic, reference-based evaluations for simplification.[68] Alva-Manchego et al,[68] after analyzing correlations of various automated metrics with human judgments, recommend using the precision component of BERTscore foremost, and then SARI (especially if the system performs lexical paraphrasing) or SAMSA (especially if the system performs sentence splitting) if BERTscore is low.

Human evaluations are typically performed using a small subset of system output due to the time and labor involved. They can be divided into several basic types:

- Qualitative: human evaluators may be tasked with providing judgments about attributes like simplicity, grammaticality, and faithfulness of the simplification to the original. These judgments are usually captured quantitatively using Likert scales or forced-choice responses.
- Comprehension: since the end goal of simplification is for readers to better comprehend text, human readers can be tested on how well they comprehend a simplified version of a passage versus the original. There are 2 main methods used to determine comprehension:
  - Multiple choice questions (MCQ): After reading a passage of either original or simplified text, the subject is presented with a crafted question with a multiple-choice response, similar to a standardized test. The question is created such that the subject would need to have comprehended the text to know the answer. These are reliable but require labor and expertise to generate.
  - Cloze procedure: Due originally to Taylor,[69] the cloze procedure involves masking words throughout the text and tasking the subject with predicting the missing word from context. This has been shown to correlate well with other measures of comprehension and has the benefit of being mostly automatic. The basic cloze procedure is often modified, for example by choosing which words to mask by their importance ("rational deletion") or adding multiple choice distractors to each blank.

MATERIALS AND METHODS

Since we are concerned with not only a particular language task, but one applied to a specific domain, we approach our literature search using pairwise sets of synonyms. For the task, we begin with the terms "sentence simplification" and "text simplification" and, following initial search results, also include "text style transfer" and "text adaptation." For the domain, we use the terms "biomedical," "medical," and "clinical." We search for all pairs of these synonyms using Google Scholar, Semantic Scholar, PubMed, ACL Anthology, and DBLP. From the results, we include papers that are both chiefly concerned with the biomedical domain and either: (1) describe a

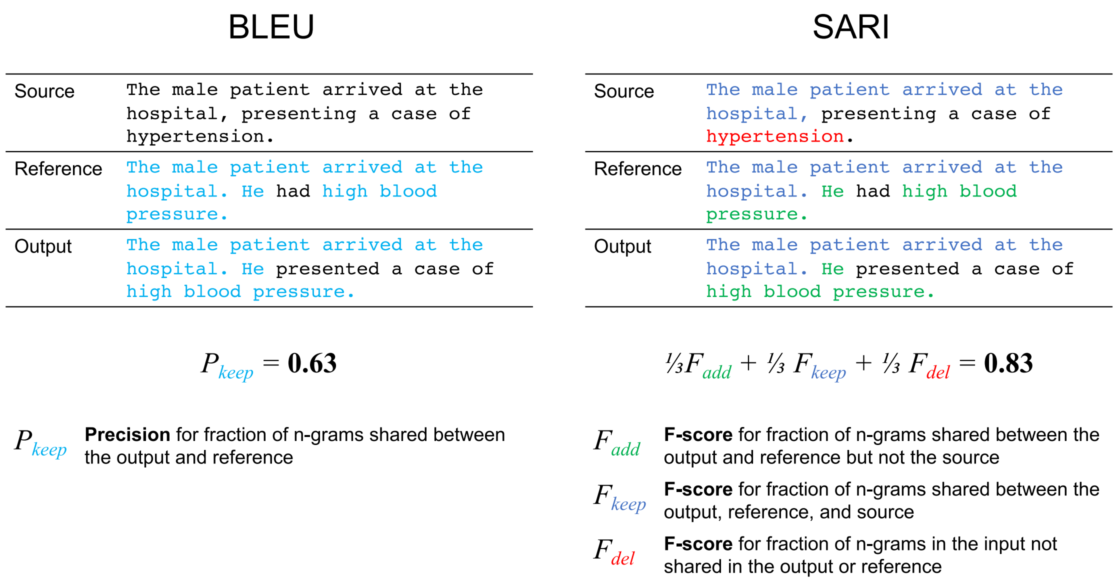

Figure 2. Automatic reference-based and evaluation with BLEU and SARI. Note that colors are applied to 1-grams (single words) though both metrics include up to 4-grams (overlapping stretches of 4 words) in counts when computing precision and F-score.

method for simplifying or adapting text to a consumer audience, (2) empirically evaluated simplification or its effects on downstream tasks, or (3) describe a data set or other resource valuable for developing simplification systems. We exclude papers that are not chiefly concerned with the biomedical domain or do not fall into at least one of the latter 3 categories. We further exclude publications that are not peer reviewed (eg, preprints and theses) or that only evaluate proprietary systems. For semantic searches (ie, Google Scholar and Semantic Scholar), whose results are not limited by the keywords, we end the search when an entire page of results does not meet the inclusion criteria. We categorize papers by natural language, subdomain within the biomedical domain, and computational paradigm.

RESULTS

Our search and inclusion criteria yielded 45 papers. Of these (nonexclusively), 32 describe tools or methods, 13 present data sets or resources, and 12 describe impacts on human comprehension or downstream applications. A complete list of papers and their inclusion criteria can be found in Table A1 in the Supplementary Material.

Categorization

First, we categorize papers by basic factors, choosing to sort them by natural language and subdomain within the biomedical domain.

Natural languages

The majority of papers (32) were concerned with only English text data. Either in addition to or instead of English, 8 papers dealt with French,[43,70–76] 2 with Spanish,[77,78] 1 with Portuguese,[79] 1 with Swedish,[80] 1 with Italian,[81] and 1 with Romanian.[82]

Subdomains

The biomedical domain encompasses a range of textual resources and applications (Figure 3). Friedman et al[6] identify 2 main sublanguages of biomedical text: that of the biomolecular domain (corresponding to literature) and that of the clinical domain (corresponding to patient reports). Each subdomain poses unique challenges for simplification. Literature contains technical terms and complex language. Clinical notes, on the other hand, use abbreviations and short, often incomplete sentences with informal grammar. Literature can further be divided into journal articles and reference texts such as books and encyclopedias. The distinction is noteworthy since articles are more likely to describe experiments and results, employing elements such as the first person, the passive voice, and statistical analyses. Clinical notes also contain further subdomains, according to the medical specialty. We find 35 papers dealing primarily with literature and 15 primarily with clinical notes. Of the latter, 2 papers focused specifically on radiology,[73,83] which has a fairly restricted lexicon based around anatomical descriptions, appearance descriptors, and diagnostic terms, and 2 focused on prescription instructions,[84,85] which have a particular shorthand lexicon for aspects like medication dosages, routes, and frequencies.

Data sources

Regardless of the automated approach (procedural, statistical, or neural), robust sources of parallel data are required. Procedural methods largely depend on thesauri for making word-level substitutions. Data-driven statistical and neural methods, however, require corpora that are parallel at least at the phrase or sentence level, which are relatively rare within languages (see Automatic Text Simplification). The problem of finding such data sets is only exacerbated for a specialized domain like the biomedical text, with instances of individual sentences or phrases being explicitly rewritten by experts restricted to small data sets or narrow subdomains.[84,85] In place of parallel corpora, the 2 major solutions are comparable corpora and pseudoparallel corpora. A summary of available data sets for biomedical text simplification can be seen in Table 1.

Plain language thesauri

The most basic parallel resource for simplification is a plain language thesaurus, which links technical terms (words or short

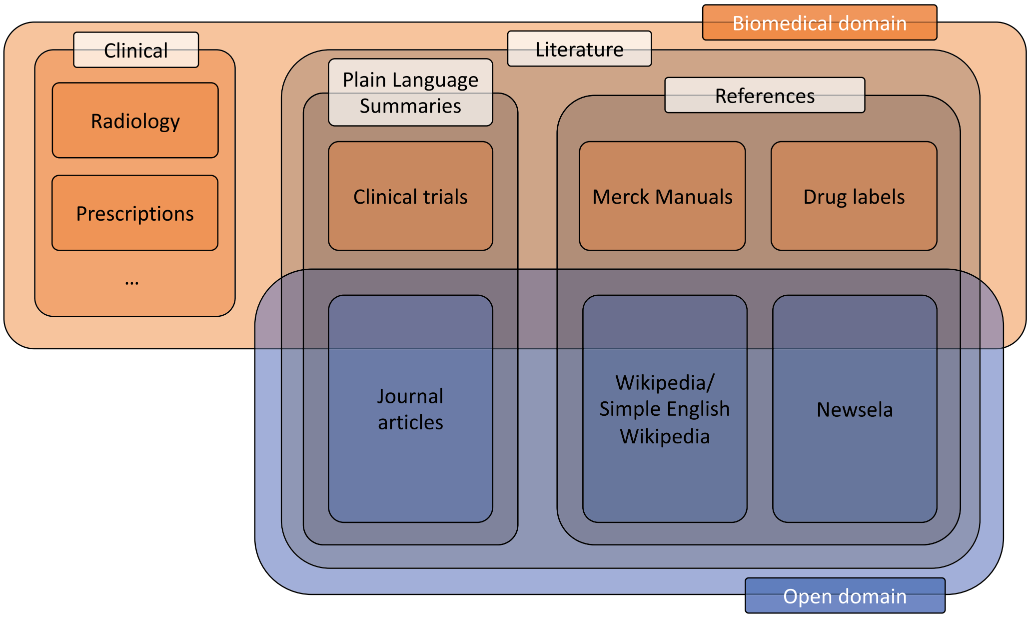

Figure 3. Domains and subdomains of current and potential sources of simplification data. Size is arbitrary and not meant to represent amount of data available.

Table 1. Published corpora for biomedical text simplification, listed in reverse chronological order

Name	Year	Author	Type	Size	Language
CochranePLS[a]	2021	Devaraj et al	Comparable	4495 document pairs	English
CDSR[b]	2021	Guo et al	Comparable	7805 document pairs	English
MSD Train	2020	Cao et al	Nonparallel	130 349 sentences (professional); 114 674 sentences (consumer)	English
MSD test	2020	Cao et al	Parallel[c]	675 passage pairs[d]	English
CLEAR (thesaurus)	2020	Koptient and Grabar	Thesaurus	11 272 terms	French
Medical Paper-Blog[b]	2020	Pattisapu et al	Parallel[c]	706 sentence pairs	English
DBF[a]	2020	Sakakini	Parallel	4554 sentence pairs	English
AutoMeTS[a]	2020	Van et al	Pseudoparallel	3300 sentence pairs	English
WikiMed[a]	2019	van den Bercken et al	Pseudoparallel	9212 sentence pairs	English
WikiSWiki[b]	2018	Adduru et al	Pseudoparallel	2493 sentence pairs	English
CLEAR	2018	Grabar and Cardon	Comparable	16 193 document pairs	French
CLEAR (aligned)	2018	Grabar and Cardon	Pseudoparallel	663 sentence pairs	French
OAC CHV	2011	Doing-Harris et al	Thesaurus	88 529 terms	English

[a]Name assigned here.
[b]As referred to within paper.
[c]Mined from existing text but screened by human annotators to ensure parallel content.
[d]Passages have on average 1.38 and 1.55 sentences on the professional and consumer sides, respectively.

phrases) to plain language synonyms, such as those from a Consumer Health Vocabulary (CHV).[86] The linkage between technical and familiar terms can be partially automated using statistical methods, as done by Koptient and Grabar, Elhadad and Sutaria, and Doing-Harris and Zeng-Treitler.[76,87–89] The Unified Medical Language System (UMLS)[90] Metathesaurus[91] contains such relationships for the Open Access Collaborative (OAC) CHV.[88]

Parallel and pseudoparallel corpora

The pair-mining approach, as used by many in the open domain (see Automatic Text Simplification),[30–32,44,45] can provide aligned subsets of sentences from comparable biomedical corpora. Statistical methods like those used for linking thesauri can also be used to find short phrases revolving around specific noun/verb constructs, as done by Deléger and Zweigenbaum[70] and Tchami,[71] and other statistics and heuristics can be used to find semantically similar pairs of sentences or longer passages. However, it is rare for complete sentences written in the contexts of different documents to have identical meanings. Even allowing for one-to-many alignment, authors note that many such pairs have extra clauses on one side or make different points. We thus refer herein to corpora generated with this method as *pseudoparallel*. Adduru et al[63] and Cardon and Grabar[72] execute this approach by training classifiers to distinguish known parallel sentences from randomly paired sentences, then using the trained classifier to identify potentially parallel sentences from comparable corpora. Van den Bercken et al,[92] taking advantage of the biomedical articles included in Wikipedia, filter an open domain pseudoparallel Wikipedia/Simple-English-Wikipedia data set[31] to create a biomedical domain-specific subset. Van et al[93] take a similar approach beginning from another data set derived from Wikipedia/Simple-English-Wikipedia.[30] Cao et al[94] employ expert annotators to identify highly similar passages (typically one to two sentences on each side) from the online Merck Manuals, which feature many articles on similar topics written for both professional and consumer audiences. Pattisapu et al[95] also use expert annotators to similarly select pairs from research papers and corresponding medical blogs about them. Human filtering makes data sets from both Cao et al and Pattisapu et al closer to true parallel corpora. Though these data sets relatively small because of the labor involved, the higher quality makes them valuable for testing.

Comparable corpora

Comparable corpora consist of pairs of documents that convey roughly the same information. For example, Guo et al extract abstracts from the Cochrane Database of Systematic Reviews (CDSR)[96] paired with their corresponding Plain Language Summaries (PLS),[97] with the aim of training lay summarization systems.[98] The pairs are filtered for a source length of 300 to 1000 words. Devaraj et al[99] build on this approach by using heuristics to align sections of CDSR abstracts, which are often structured into sections, with analogous sections of corresponding PLS. This results in passages of text that describe roughly the same experimental results in both the professional and consumer registers. Similarly to Guo et al, passages retained are limited to 1024 tokens, which is the size that the largest current models can ingest.[100,101] Grabar and Cardon[72] compile a corpus of comparable document pairs from several sources: French Wikipedia versus Vikidia (an online encyclopedia targeted toward children, featuring mainly articles in French), professional versus consumer drug leaflets, and French versions of Cochrane abstracts versus their French Plain Language Summaries.

Elements of comprehension

Biomedical text simplification is important chiefly as a means to the end of improved consumer understanding. Consequently, there has been a large amount of interest in identifying which of the many possible linguistic operations captured by simplification are the most effective in achieving this goal. This knowledge is especially valuable for procedural approaches but can also inform data set creation, evaluation, and error analysis for data-driven approaches.

- Corporal analysis: Kauchak et al[102] use classical Machine Learning to weight features such as part-of-speech, lexical frequency, and semantic ambiguity in terms of their predictive value for the difficulty of biomedical text. Interestingly, they find that more difficult text contains more specific terms (such as the technical term *myocardium*), while simpler text has more ambiguous ones (such as the more common but polysemous word *heart*). In another study characterizing the accessibility of text, Kauchak et al[78] identify transition words, which are known to impact comprehension.
- Empirical studies: In empirical studies of comprehension with nonexpert readers by Leroy et al, substitution of less familiar words (as measured by frequency in the Google Web Corpus)[103] by more familiar ones showed mixed results, improving perception of text difficulty but only improving comprehension in some cases.[104–106] Splitting up longer noun phrases into short ones was also hypothesized by Leroy et al[107] to improve comprehension but when studied only improved perceived difficulty. In contrast, interventions that were shown by Leroy et al[106] to improve comprehension were increasing coherence using anaphora (ie, using a pronoun to refer to a recently used term rather than repeating it) and emphasizing ideas using typographical coherence (ie, indented enumerations). When a difficult term has no simpler substitution, an explanation may be inserted. Gu et al[108] examine the effects of these explanations on comprehension, finding that optimal placement (after the term or at the end of the sentence) depends on the type of corpus.
- Qualitative measurements: To facilitate comparison of papers that describe methods, we identify 4 major types of qualitative human judgments used for evaluation in these papers. From descriptions of what annotators were asked to judge, we map evaluations onto these 4 categories:
  - Grammaticality: how well the output conforms to grammatical rules or is fluent
  - Simplicity: how simple or comprehensible the output is, either absolutely or compared to the original
  - Preservation: how much of the original semantic content is preserved in the output
  - Accuracy: how correct the information in the output is

Computational paradigms

Here, we describe work based on the computational paradigm (Figure 4). Of the 32 tools or methods described in the papers included, 22 are chiefly procedural and 10 are chiefly neural. Note that we did not find any papers employing the statistical paradigm within the biomedical domain. We also discuss 2 special cases: human-in-the-loop computing and the use of simplification as preprocessing for further computational tasks.

Procedural

Despite the widespread adoption of statistical, and then neural, methods over procedural ones in the open domain (see Automatic Text Simplification), work on procedural methods for simplifying biomedical text has continued. This is due to: (1) the robustness of in-domain knowledge bases, (2) the lack of in-domain parallel training data, and (3) the unpredictability of neural models. The latter is especially important when health information is involved, as both spurious substitutions and omissions of potentially crucial context could mislead users and potentially cause inappropriate actions. Like in the open domain, the 2 most common elements of procedural methods are lexical simplification methods and syntactic simplification.

- *Lexical* methods typically use plain language thesauri (see Data Sources), such as the UMLS Metathesaurus,[91] to identify and replace difficult medical terms.[59,73,79–81,83,89,104,109] Open-domain knowledge bases that naturally contain scientific language, such as WordNet[110] and Wiktionary, have also been used, as by Leroy et al[104] Several improvements to this general paradigm have been explored:
  - Explanation generation: The hierarchical relationships between words or concepts in the UMLS[90] and other knowledge bases can also be used to generate explanations for words that do not have simpler synonyms. For example, "*Pulmonary atresia*" can be explained through its hyponymy with "*birth defect*," creating the phrase "*Pulmonary atresia* (a type of *birth defect*)." This approach is taken by Kandula et al,[59] Leroy et al,[104] and Zeng-Treitler et al.[109]
  - Subwords: Kloehn et al[77] further expand on the basic word-substitution approach by breaking words into roots and affixes and providing substitutions for each piece of complex words. Abrahamsson et al[80] similarly adapt the approach to

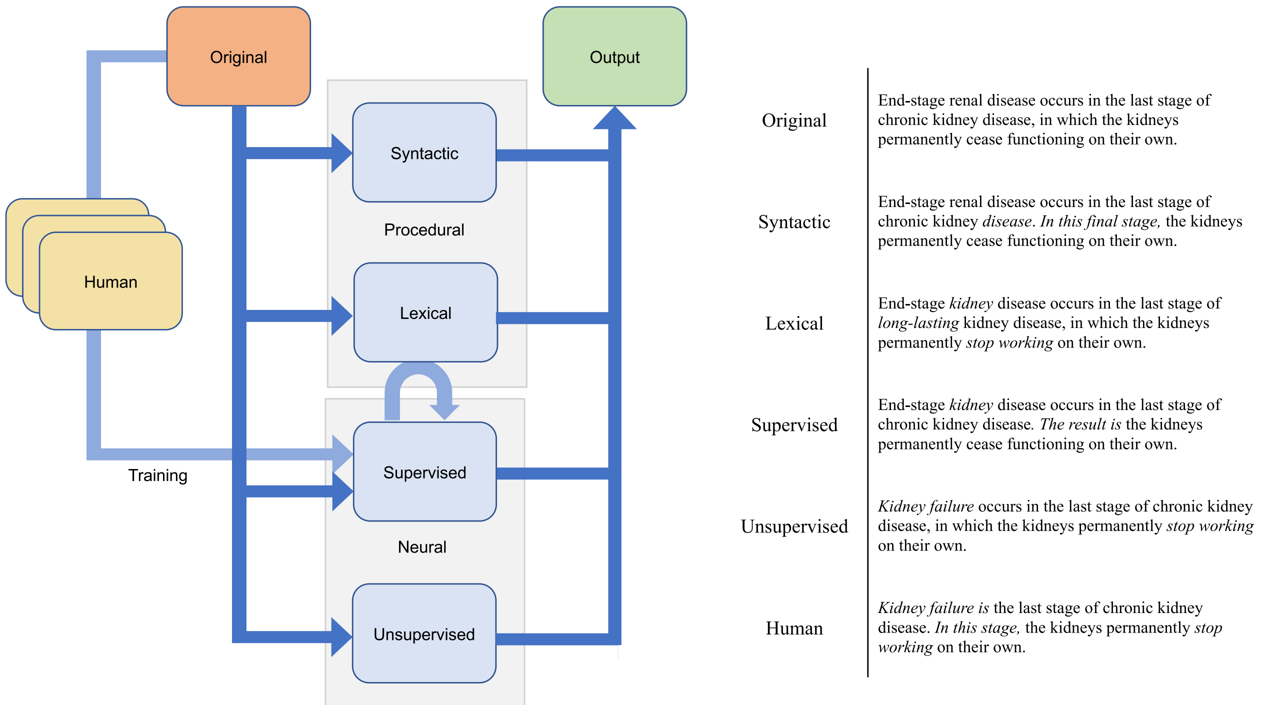

Figure 4. Computational paradigms. Procedural approaches include syntactic and lexical simplifications. Neural approaches simplify holistically, based on training data. Supervised neural approaches are trained with human-created examples (either simplified by sentence or mined from a larger, simple corpus). Unsupervised neural approaches learn internal representations and output distributions, requiring only unaligned expert and simple corpora. Supervised neural models may be augmented with lexical simplification the via a phrase table, combining neural and procedural paradigms.

Swedish, a compounding language, by examining substrings of words.

- Language models: Sakakini et al[84] improve on the lexical substitution approach by using a language model to help disambiguate possible substitutions using their surrounding context. Tran et al[111] use a neural Masked Language Model to directly generate context-aware candidates for substitutions.

- *Syntactic* approaches have included those common in the open domain. Koptient and Grabar and Kandula et al employ sentence splitting methods,[59,74] and Kauchak and Leroy[112] use pattern-based paraphrasing. Peng et al[113] use finite state machines to identify coordination, relative clauses, and appositions (adjacent noun phrases referring to the same thing). Double negation, where words with negative prefixes are also negated in the sentence (such as “not abnormal”), can be simplified by removing both negations. This method is implemented by Mukherjee et al,[114] and this implementation is used as part of a larger system by Kauchak and Leroy.[112]

Procedural approaches have the advantage of predictability, preservation of content, and generally accurate lexical substitutions. However, lexical substitutions are typically made without considering context, causing ambiguity (eg, “no” interpreted as “nitrous oxide,” or “ultrasound” meaning either a treatment or a diagnostic tool) and grammatical issues (eg, “renal cortex” becoming “kidney outer layer of an organ”). While procedural methods also have the advantage that parallel training data are not needed, this also means they are typically evaluated using bespoke human evaluation methods, making it difficult to compare the performance of different systems. In Table 2, we provide an overview of publications presenting procedural methods.

Neural

Scarcity of parallel simplification resources is only exacerbated within domains. Perhaps the biggest exception is Li et al,[85] who are able to obtain a large set (around 530 000) of prescription directions and their manually translated equivalents for training a sequence-to-sequence model. However, these data are from a highly specific subdomain and are not publicly available, leaving the problem open in the broader biomedical domain, especially the literature subdomain. Nonetheless, research on neural models has been pushed forward by several innovations:

- Procedural elements: One way to mitigate data scarcity is to augment the standard, supervised sequence-to-sequence model with procedural elements that draw from knowledge bases, such as phrase tables. This approach is taken by Koptient and Grabar, Van den Bercken et al, and Shardlow and Nawaz.[75,92,115]
- Pseudoparallel corpora: Pairs of sentences similar in meaning can be mined from comparable corpora (see Parallel and Pseudoparallel Corpora). This allows standard sequence-to-sequence models to be used, following the Machine Translation paradigm. This approach is taken by Van et al and van den Bercken et al.[92,93]
- Document-level simplification: Another strategy is to take advantage of recent models that are able to ingest and output entire documents. This allows the use of comparable corpora for training (see “Comparable Corpora”). Devaraj et al and Guo et al train BART[101] models using Cochrane abstracts along with their Plain Language Summaries.[98,99] Limited human evaluation by Guo et al indicated that, though outputs lagged in simplicity versus reference PLS, they were at least as fluent and accurate. However, an issue for both systems was the “hallucination” (spontaneous introduction into output) of passages common in

Table 2. Publications describing procedural methods for biomedical text simplification, in reverse chronological order

Author	Year	Language	Subdomain	Evaluation			
				Readability	Reference	Human	Comprehension
Tran et al[a]	2021	English	Literature	–	–	gram., simp., acc.	–
Alfano et al	2020	English, Italian	Literature	Familiarity	–	simp.,	–
Kauchak and Leroy	2020	English	Literature	–	–	–	
Koptient and Grabar	2020	French	Literature	–	–	gram., simp., pres.	–
Sakakini et al	2020	English	Clinical	–	PINC, BLEU	–	MCQ
Zilio et al	2020	Portuguese	Literature	–	–	simp., acc.	–
Kloehn	2018	English, Spanish	Literature	–	–	simp., acc.	–
Ramadier et al	2018	French	Clinical	–	–	–	Cloze
Mukherjee et al	2017	English	Literature	–	–	–	–
Qenam et al	2017	English	Clinical	–	–	simp., acc.	–
Abrahamsson et al	2014	Swedish	Literature	LIX, OVIX	–	–	–
Leroy et al	2013 (a)	English	Literature	–	–	–	MCQ
Topac and Stoicu-Tivadar	2013	English, Romanian	Literature	–	–	simp.	–
Leroy et al	2012	English	Literature	–	–	–	MCQ
Peng et al[b]	2012	English	Literature	–	–	–	–
Evans[b]	2011	English	Clinical				
Jonnalagadda and Gonzalez[b]	2010 (a)	English	Literature	–	–	–	–
Kandula et al	2010	English	Literature	FKGL, SMOG	–	–	Cloze
Jonnalagadda et al[b]	2009	English	Literature	–	–	–	–
Ong et al	2007	English	Literature	FKGL	–	–	–
Zeng-Treitler et al	2007	English	Clinical	–	–	simp., acc.	Cloze
Elhadad et al	2006	English	Literature	–	–	simp.	–

Note: Human evaluations are coded as grammaticality, simplicity, preservation, and accuracy.
[a]Contains neural elements.
[b]Evaluated using performance of downstream automated tasks.

training targets, such as "the aim of this study..." or "the evidence is up-to-date as of..."

- Unsupervised learning: As in the open domain, the lack of parallel data has also motivated unsupervised techniques. Cao et al[94] pose the simplification problem as Style Transfer,[79] training various architectures for this task using nonparallel professional and consumer data to represent source and target styles, respectively. Pattisapu et al[95] explore a denoising autoencoder model for unsupervised simplification. In this method, to create training data, medical text in the consumer register is altered (or "noised") by replacing simple medical terms with their more technical counterparts from knowledge bases. By trying to recover the original (simple) sentence, the model learns to substitute terms while also retaining fluency in the consumer register. The latter denoising approach fares much better than the former style transfer approach in human evaluations. However, both approaches suffer from unpredictable hallucinations and repeated syllables, words, or phrases.

In addition to the fact that no single automatic evaluation metric fully captures the quality of simplification, it has been difficult to directly compare systems because of a lack of standardized data sets. However, van den Bercken et al[92] put forth a pseudoparallel corpus as a test set and use it to compare their system to a reimplemented baseline system. Pattisapu et al,[95] build on this, using the same data set to compare their system to that of van den Bercken et al and Shardlow and Nawaz, and claiming a new state-of-the-art.[21] A summary of publications describing neural methods can be seen in Table 3.

Human-in-the-loop

While complete automation is the goal of much of the work in biomedical text simplification, computational approaches may also have value in assisting human writers. A human-in-the-loop model allows progress in the area to benefit medical communicators before fully automated models are viable and protects against harms that could be caused by incorrect system output. Procedural simplification in this vein can consist of suggesting lexical substitutions and syntactic rearrangements to a user, either within an editor, as implemented by Leroy et al and Kauchak and Leroy[104,112] or a reader, as implemented by Alfano et al and Zilio et al.[79,81] On the neural side, Van et al[93] propose adapting the sequence-to-sequence model to interactively provide words likely to come next in the simplified version of the text, creating a specialized "auto-complete."

Simplification as preprocessing

Though consumers may ultimately be the greatest beneficiaries of automatically simplified biomedical text, human readers are not always the intended audience. Research by Jonnalagadda et al has shown that both lexical and syntactic simplification can aid in parsing[116] and automatic identification of protein-protein interactions.[117] Similarly, Peng et al[113] show that syntactic simplification, specifically targeting coordinations, relative clauses, and appositions, can improve extraction of protein phosphorylation and ranking of sentences by gene actions. Evans et al use syntactic simplification to improve the extraction of data from clinical patient information.[118,119] Finally, simplification in the form of substitution of expert biomedical terms by more common ones

Table 3. Publications describing neural methods for biomedical text simplification, in reverse chronological order

Author	Year	Language	Subdomain	Evaluation		
				Readability	Reference	Human
Devaraj et al	2021	English	Literature	FKGL, ARI	SARI, BLEU, ROUGE	–
Guo	2021	English	Literature	FKGL, GFI, CLI	ROUGE	gram., simp., pres., acc.
Cardon and Grabar[a]	2020	French	Literature	Kandel	SARI, BLEU	–
Cao et al	2020	English	Literature	–	SARI, BLEU	pres.
Li et al	2020	English	Clinical	–	BLEU, METEOR	acc.
Pattisapu et al	2020	English	Literature	–	SARI, BLEU, ROUGE, METEOR	gram., simp., pres.
Van et al	2020	English	Literature	–	–	–
Shardlow and Nawaz[a]	2019	English	Clinical	FKGL, GFI, CLI	–	simp.
van den Bercken et al[a]	2019	English	Literature	–	SARI, BLEU	gram., simp., pres.
Adduru et al	2018	English	Literature	–	BLEU, METEOR, TER	–

Note: Human evaluations are coded as grammaticality, simplicity, preservation, and accuracy.
[a]Procedural elements are used.

has also been proposed by Chen et al as a remedy for unseen phrases in Statistical Machine Translation.[120]

DISCUSSION AND CONCLUSION

Adaptation of biomedical text for the consumer is not one specific problem, but rather a branch of study spanning various topics, registers, and applications. The idiosyncrasies of biomedical language (and its subdomains) provide unique methodological challenges beyond the task of open-domain automatic text simplification. At the same time, the high stakes of medical information, the availability of knowledge bases, and the lack of parallel text data, have likely led to the persistence of procedural methods (as opposed to neural methods) within this domain, while neural methods have dominated in the open domain. Procedural methods have made progress in substituting difficult terms, providing explanations, and some forms of syntactic simplification. However, rule-based methods require maintenance of knowledge bases, and they typically fail to capture the complexity of language, requiring human editing for grammaticality.[112] Deep Learning models hold a lot of promise for incorporating both the lexical knowledge and the transformational rules needed in a purely data-driven, and thus ultimately more sustainable, way. However, limitations in training data currently cause neural models to be worse than procedural methods at preserving meaning and critical information, despite improving grammaticality.[84,85,111] In addition to training data, a smaller, but still substantial, amount of high-quality reference data is also needed for measuring progress of methods development. In our reading of the literature, obtaining both these types of data has been a sticking point for the progress of the task. The most common approaches to this lack to date have been the use of pseudoparallel and comparable corpora and the manual annotation of small data sets, or those within specific subdomains. For training, future circumventions of the parallel data issue may also include unsupervised methods and zero-shot multitask models. For testing, however, the most likely way forward will be further manual creation of truly parallel data sets.

As fully automated systems of any paradigm become more feasible, ethical concerns are likely to arise surrounding the potential of mistranslations to do real-world harm. These must be taken seriously to avoid bad outcomes and erosion of trust in health information systems. However, concerns must also be weighed against the fact that consumers are already seeking out information they may not understand.[5] If systems are deployed, it is prudent to provide disclaimers to users about their experimental nature, if appropriate, and to emphasize that they are not substitutes for professional health advice. Additionally, semiautomated systems and human editing of system output may provide a bridge, ensuring veracity until fully automated systems can be trusted. This would be in keeping with the approach of "progressive caution."[121] It is also crucial to consider privacy when collecting training data or deploying a system that accepts user queries. These data could contain information about personal health conditions that, if leaked, could affect patients financially via insurance rates and erode trust, which would be counterproductive for engagement.[121]

The task of fully automating the simplification of biomedical text remains a daunting one, with issues to be resolved in methods, data collection and ethical concerns. However, each bit of progress brings us closer to realizing the potential for this application to improve patient understanding, health literacy, and, ultimately, health outcomes.

FUNDING

This research was supported by the Intramural Research Program of the National Library of Medicine (NLM), National Institutes of Health.

AUTHOR CONTRIBUTIONS

BO performed database searches. BO and KA read papers and contributed to the manuscript. DDF edited the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

No new data were generated or analyzed in support of this research.

REFERENCES

1. Berkman ND, Sheridan SL, Donahue KE, *et al.* Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011; 155 (2): 97–107.
2. Safeer RS, Keenan J. Health literacy: the gap between physicians and patients. *Am Fam Physician* 2005; 72 (3): 463–8.
3. Wicklund K, Ramos K. Plain language: effective communication in the health care setting. *J Hosp Librariansh* 2009; 9 (2): 177–85.
4. Dees RH. Health literacy and autonomy. *Am J Bioeth* 2007; 7 (11): 22–3.
5. Aydın GÖ, Kaya N, Turan N. The role of health literacy in access to online health information. *Procedia Soc Behav Sci* 2015; 195: 1683–7.
6. Friedman C, Kra P, Rzhetsky A. Two biomedical sublanguages: a description based on the theories of Zellig Harris. *J Biomed Inform* 2002; 35 (4): 222–35.
7. Snow CE. Academic language and the challenge of reading for learning about science. *Science* 2010; 328 (5977): 450–2.
8. Gupta N, Rutledge C. Pyruvate dehydrogenase complex deficiency: an unusual cause of recurrent lactic acidosis in a paediatric critical care unit. *J Crit Care Med (Targu Mures)* 2019; 5 (2): 71–5.
9. Converse JM. On "MEDICALESE" a plea for proper English usage. *Plast Reconstr Surg* 1974; 54: 206–8.
10. Stableford S, Mettger W. Plain language: a strategic response to the health literacy challenge. *J Public Health Policy* 2007; 28 (1): 71–93.
11. Quesenberry AC. Plain language for patient education. *J Consum Health Internet* 2017; 21 (2): 209–15.
12. Warde F, Papadakos JK, Papadakos T, *et al.* Plain language communication as a priority competency for medical professionals in a globalized world. *Can Med Ed J* 2018; 9 (2): e52–9.
13. Hadden KB, Coleman CA, Scott AJ. The bilingual physician: seamless switching from Medicalese to plain language. *J Grad Med Educ* 2018; 10 (2): 130–3.
14. Dormer L, Walker J. Plain language summary of publication articles: helping disseminate published scientific articles to patients. *Future Oncol* 2020; 16 (25):1873–4.
15. Bullers K. Merck manuals. *J Med Libr Assoc* 2016; 104 (4): 369–71.
16. Miller N, Lacroix E-M, Backus JE. MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. *Bull Med Libr Assoc* 2000; 88 (1): 11–7.
17. SeyyedHosseini S, Lund BD, BasirianJahromi R. Comparing health consumer search behavior and scientific research productivity related to COVID-19 vaccines in the USA: an infodemiology study [published online ahead of print 2022]. *Glob Knowl Mem Commun* 2022. doi: 10.1108/GKMC-09-2021-0155.
18. Jelicic Kadic A, Fidahic M, Vujcic M, *et al.* Cochrane plain language summaries are highly heterogeneous with low adherence to the standards. *BMC Med Res Methodol* 2016; 16 (1): 4.
19. FitzGibbon H, King K, Piano C, *et al.* Where are biomedical research plain-language summaries? *Health Sci Rep* 2020; 3 (3): e175.
20. Penlington M, Silverman H, Vasudevan A, *et al.* Plain language summaries of clinical trial results: a preliminary study to assess availability of easy-to-understand summaries and approaches to improving public engagement. *Pharmaceut Med* 2020; 34 (6): 401–6.
21. Shardlow M. A survey of automated text simplification. *Int J Adv Comput Sci Appl* 2014; 4: 58–70.
22. Al-Thanyyan SS, Azmi AM. Automated text simplification: a survey. *ACM Comput Surv CSUR* 2021; 54: 1–36.
23. Siddharthan A. An architecture for a text simplification system. In: language engineering conference. Los Alamitos, CA: IEEE Computer Society; 2002: 64–71.
24. Carroll J, Minnen G, Canning Y, *et al.* Practical simplification of English newspaper text to assist aphasic readers. In: proceedings of the AAAI-98 workshop on integrating artificial intelligence and assistive technology. Menlo Park, CA: Association for the Advancement of Artificial Intelligence; 1998: 7–10.
25. Filippova K, Strube M. Dependency tree based sentence compression. In: proceedings of the fifth international natural language generation conference. Stroudsburg, PA: Associationfor Computational Linguistics; 2008: 25–32.
26. Filippova K, Strube M. Sentence fusion via dependency graph compression. In: proceedings of the 2008 conference on empirical methods in natural language processing. Stroudsburg, PA: Association for Computational Linguistics; 2008: 177–85.
27. Lopez A. Statistical machine translation. *ACM Comput Surv* 2008; 40 (3): 1–49.
28. Hutchins J. Machine translation: a concise history. *Comput Aided Transl Theory Pract* 2007; 13: 11.
29. Frankenberg-Garcia A. A corpus study of splitting and joining sentences in translation. *Corpora* 2019; 14 (1): 1–30.
30. Coster W, Kauchak D. Simple English Wikipedia: a new text simplification task. In: proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics; 2011: 665–9.
31. Hwang W, Hajishirzi H, Ostendorf M, *et al.* Aligning sentences from standard Wikipedia to simple Wikipedia. In: proceedings of the 2015 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Denver, Colorado: Association for Computational Linguistics; 2015: 211–7.
32. Zhu Z, Bernhard D, Gurevych I. A monolingual tree-based translation model for sentence simplification. In: proceedings of the 23rd international conference on computational linguistics (Coling 2010). Beijing, China: Tsinghua University Press; 2010: 1353–61.
33. Coster W, Kauchak D. Learning to simplify sentences using Wikipedia. In: proceedings of the workshop on monolingual text-to-text generation. Stroudsburg, PA: Association for Computational Linguistics; 2011: 1–9.
34. Wubben S, Van Den Bosch A, Krahmer E. Sentence simplification by monolingual machine translation. In: proceedings of the 50th annual meeting of the Association for Computational Linguistics (volume 1: long papers). Stroudsburg, PA: Association for Computational Linguistics; 2012: 1015–24.
35. Xu W, Napoles C, Pavlick E, *et al.* Optimizing statistical machine translation for text simplification. *Trans Assoc Comput Linguist* 2016; 4: 401–15.
36. Štajner S, Béchara H, Saggion H. A deeper exploration of the standard PB-SMT approach to text simplification and its evaluation. In: proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing (volume 2: short papers). Stroudsburg, PA: Association for Computational Linguistics; 2015: 823–8.
37. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. *Adv Neural Inf Process Syst* 2014; 27: 3104–12.
38. Wang T, Chen P, Rochford J, *et al.* Text simplification using neural machine translation. In: proceedings of the AAAI conference on artificial intelligence. Menlo Park, CA: Association for the Advancement of Artificial Intelligence; 2016.
39. Nisioi S, Štajner S, Ponzetto SP, *et al.* Exploring neural text simplification models. In: proceedings of the 55th annual meeting of the Association for Computational Linguistics (volume 2: short papers). Stroudsburg, PA: Association for Computational Linguistics; 2017: 85–91.
40. Dong Y, Li Z, Rezagholizadeh M, *et al.* EditNTS: an neural programmer-interpreter model for sentence simplification through explicit editing. In: proceedings of the 57th annual meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics; 2019: 3393–402.
41. Martin L, De La Clergerie ÉV, Sagot B, *et al.* Controllable sentence simplification. In: proceedings of the 12th language resources and evaluation conference. Paris, France: European Language Resources Association; 2020: 4689–98.

42. Zhang X, Lapata M. Sentence simplification with deep reinforcement learning. In: proceedings of the 2017 conference on empirical methods in natural language processing. Stroudsburg, PA: Association for Computational Linguistics; 2017: 584–94.
43. Cardon R, Grabar N. Parallel sentence retrieval from comparable corpora for biomedical text simplification. In: RANLP 2019. Shoumen, Bulgaria: INCOMA, Ltd.; 2019.
44. Jiang C, Maddela M, Lan W, *et al.* Neural CRF model for sentence alignment in text simplification. In: proceedings of the 58th annual meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics; 2020: 7943–60.
45. Martin L, Fan A, de la Clergerie É, *et al.* MUSS: multilingual unsupervised sentence simplification by mining paraphrases. ArXiv:200500352, 2020.
46. Sulem E, Abend O, Rappoport A. Simple and effective text simplification using semantic and neural methods. In: proceedings of the 56th annual meeting of the Association for Computational Linguistics (volume 1: long papers). Stroudsburg, PA: Association for Computational Linguistics; 2018: 162–73.
47. Surya S, Mishra A, Laha A, *et al.* Unsupervised neural text simplification. In: proceedings of the 57th annual meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics; 2019: 2058–68.
48. Mallinson J, Sennrich R, Lapata M. Zero-shot crosslingual sentence simplification. In: proceedings of the 2020 conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: Association for Computational Linguistics; 2020: 5109–26. doi:10.18653/v1/2020.emnlp-main.415
49. Flesch R. A new readability yardstick. *J Appl Psychol* 1948; 32 (3): 221–33.
50. Kincaid JP, Fishburne RP Jr, Rogers RL, *et al.* Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel. *Naval Technical Training Command Millington TN Research Branch*. Springfield, VA: National Technical Information Service; 1975.
51. Mc Laughlin GH. SMOG grading-a new readability formula. *J Read* 1969; 12: 639–46.
52. Björnsson C. *Läsbarhet*. Stockholm, Sweden: Liber; 1968.
53. Smith M, Taffler R. Readability and understandability: different measures of the textual complexity of accounting narrative. *Account Audit Account J* 1992; 5: 84–98.
54. Mühlenbock K, Kokkinakis SJ. LIX 68 revisited-An extended readability measure. In: proceedings of corpus linguistics. Lancaster, UK: University Centre for Computer Corpus Research on Language; 2009.
55. Coleman M, Liau TL. A computer readability formula designed for machine scoring. *J Appl Psychol* 1975; 60 (2): 283–4.
56. Tanprasert T, Kauchak D. Flesch-Kincaid is not a text simplification evaluation metric. In: proceedings of the 1st workshop on natural language generation, evaluation, and metrics (GEM 2021). Stroudsburg, PA: Association for Computational Linguistics; 2021: 1–14.
57. Kauchak D, Leroy G. Moving beyond readability metrics for health-related text simplification. *IT Prof* 2016; 18 (3): 45–51.
58. Leroy G, Kauchak D. The effect of word familiarity on actual and perceived text difficulty. *J Am Med Inform Assoc* 2014; 21 (e1): e169–72.
59. Kandula S, Curtis D, Zeng-Treitler Q. A semantic and syntactic text simplification tool for health content. In: AMIA annual symposium proceedings. Rockville, MD: American Medical Informatics Association; 2010: 366.
60. Yan X, Song D, Li X. Concept-based document readability in domain specific information retrieval. In: proceedings of the 15th ACM international conference on information and knowledge management. New York, NY: Association for Computing Machinery; 2006: 540–9.
61. Papineni K, Roukos S, Ward T, *et al.* BLEU: a method for automatic evaluation of machine translation. In: proceedings of the 40th annual meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics; 2002: 311–8.
62. Sulem E, Abend O, Rappoport A. Semantic structural evaluation for text simplification. In: proceedings of NAACL-HLT. Stroudsburg, PA: Association for Computational Linguistics; 2018: 685–96.
63. Adduru V, Hasan SA, Liu J, *et al.* Towards dataset creation and establishing baselines for sentence-level neural clinical paraphrase generation and simplification. In: *KHD@ IJCAI*. Aachen, Germany: CEUR-WS; 2018.
64. Chen D, Dolan WB. Collecting highly parallel data for paraphrase evaluation. In: proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics; 2011: 190–200.
65. Zhang T, Kishore V, Wu F, *et al.* BERTscore: evaluating text generation with BERT. ArXiv:190409675, 2019.
66. Devlin J, Chang M-W, Lee K, *et al.* BERT: pre-training of deep bidirectional transformers for language understanding. In: proceedings of NAACL-HLT. Stroudsburg, PA: Association for Computational Linguistics; 2018: 4171–86.
67. Sulem E, Abend O, Rappoport A. BLEU is not suitable for the evaluation of text simplification. In: proceedings of the 2018 conference on empirical methods in natural language processing. Stroudsburg, PA: Association for Computational Linguistics; 2018: 738–44.
68. Alva-Manchego F, Scarton C, Specia L. The (un) suitability of automatic evaluation metrics for text simplification. *Comput Linguist* 2021; 47 (4): 861–89.
69. Taylor WL. Cloze procedure: a new tool for measuring readability. *Journal Q* 1953; 30 (4): 415–33.
70. Deléger L, Zweigenbaum P. Paraphrase acquisition from comparable medical corpora of specialized and lay texts. In: AMIA annual symposium proceedings. Rockville, MD: American Medical Informatics Association; 2008: 146.
71. Tchami OW. Acquiring verb frames for a text simplification lexicon in the medical domain. In: *Term Bases Linguist Linked Open Data*. Frederiksberg, Denmark: Copenhagen BusinessSchool; 2016: 101.
72. Grabar N, Cardon R. Clear-simple corpus for medical French. In: proceedings of the 1st Workshop on Automatic Text Adaptation (ATA). Stroudsburg, PA: Association for Computational Linguistics; 2018: 3–9.
73. Ramadier L, Lafourcade M. Radiological text simplification using a general knowledge base. In: international conference on computational linguistics and intelligent text processing. New York, NY: Springer; 2017: 617–27.
74. Koptient A, Grabar N. Fine-grained text simplification in French: steps towards a better grammaticality. In: ISHIMR proceedings of the 18th international symposium on health information management research. Enschede, Netherlands: University of Twente; 2020.
75. Cardon R, Grabar N. French biomedical text simplification: when small and precise helps. In: the 28th international conference on computational linguistics. Prague, Czech Republic: International Committee on Computational Linguistics; 2020: 710–6.
76. Koptient A, Grabar N. Rated lexicon for the simplification of medical texts. In: the fifth international conference on informatics and assistive technologies for health-care, medical support and wellbeing HEALTHINFO 2020. Hannover, Germany: Leibniz University; 2020.
77. Kloehn N, Leroy G, Kauchak D, *et al.* Improving consumer understanding of medical text: development and validation of a new subsimplify algorithm to automatically generate term explanations in English and Spanish. *J Med Internet Res* 2018; 20 (8): e10779.
78. Kauchak D, Leroy G, Pei M, *et al.* Predicting transition words between sentence for English and Spanish medical text. In: AMIA annual symposium proceedings. Rockville, MD: *American Medical Informatics Association*; 2019: 523.
79. Zilio L, Paraguassu LB, Hercules LAL, *et al.* A lexical simplification tool for promoting health literacy. In: proceedings of the 1st workshop on tools and resources to empower people with REAding DIfficulties (READI). Paris, France: European Language Resources Association; 2020: 70–6.

80. Abrahamsson E, Forni T, Skeppstedt M, *et al.* Medical text simplification using synonym replacement: adapting assessment of word difficulty to a compounding language. In: proceedings of the 3rd workshop on predicting and improving text readability for target reader populations (PITR). Stroudsburg, PA: Association for Computational Linguistics; 2014: 57–65.
81. Alfano M, Lenzitti B, Bosco GL, *et al.* Design, development and validation of a system for automatic help to medical text understanding. *Int J Med Inform* 2020; 138: 104109.
82. Topac V, Stoicu-Tivadar V. Patient empowerment by increasing the understanding of medical language for lay users. *Methods Inf Med* 2013; 52 (5): 454–62.
83. Qenam B, Kim TY, Carroll MJ, *et al.* Text simplification using consumer health vocabulary to generate patient-centered radiology reporting: translation and evaluation. *J Med Internet Res* 2017; 19 (12): e8536.
84. Sakakini T, Lee JY, Duri A, *et al.* Context-aware automatic text simplification of health materials in low-resource domains. In: proceedings of the 11th international workshop on health text mining and information analysis. Stroudsburg, PA: Association for Computational Linguistics; 2020: 115–26.
85. Li J, Lester C, Zhao X, *et al.* PharmMT: a neural machine translation approach to simplify prescription directions. In: findings of the Association for Computational Linguistics: EMNLP 2020. Stroudsburg, PA: Association for Computational Linguistics; 2020: 2785–96.
86. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. *J Am Med Inform Assoc* 2006; 13 (1): 24–9.
87. Elhadad N, Sutaria K. Mining a lexicon of technical terms and lay equivalents. In: *Biological, Translational, and Clinical Language Processing*. Stroudsburg, PA: Association for Computational Linguistics; 2007: 49–56.
88. Doing-Harris KM, Zeng-Treitler Q. Computer-assisted update of a consumer health vocabulary through mining of social network data. *J Med Internet Res* 2011; 13 (2): e37.
89. Elhadad N. Comprehending technical texts: predicting and defining unfamiliar terms. In: AMIA annual symposium proceedings. Rockville, MD: American Medical Informatics Association; 2006: 239.
90. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004; 32 (Database Issue): D267–70.
91. Schuyler PL, Hole WT, Tuttle MS, *et al.* The UMLS Metathesaurus: representing different views of biomedical concepts. *Bull Med Libr Assoc* 1993; 81 (2): 217–22.
92. Van den Bercken L, Sips R-J, Lofi C. Evaluating neural text simplification in the medical domain. In: the World Wide Web conference. New York, NY: Association for Computing Machinery; 2019: 3286–92.
93. Van H, Kauchak D, Leroy G. AutoMeTS: the autocomplete for medical text simplification. In: *proceedings of the 28th international conference on computational linguistics*. Barcelona, Spain: International Committee on Computational Linguistics; 2020: 1424–34.
94. Cao Y, Shui R, Pan L, *et al.* Expertise style transfer: a new task towards better communication between experts and laymen. In: proceedings of the 58th annual meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics; 2020: 1061–71.
95. Pattisapu N, Prabhu N, Bhati S, *et al.* Leveraging social media for medical text simplification. In: Proceedings of the 43rd international ACM SIGIR conference on research and development in information. New York, NY: Association for Computing Machinery; 2020: 851–60.
96. Starr M, Chalmers I, Clarke M, *et al.* The origins, evolution, and future of the Cochrane database of systematic reviews. *Int J Technol Assess Health Care* 2009; 25 (S1): 182–95.
97. Langendam MW, Akl EA, Dahm P, *et al.* Assessing and presenting summaries of evidence in Cochrane reviews. *Syst Rev* 2013; 2 (1): 81.
98. Guo Y, Qiu W, Wang Y, *et al.* Automated lay language summarization of biomedical scientific reviews. In: proceedings of the AAAI conference on artificial intelligence. Menlo Park, CA: Association for the Advancement of Artificial Intelligence; 2021: 160–8.
99. Devaraj A, Marshall I, Wallace BC, *et al.* Paragraph-level simplification of medical texts. In: proceedings of the 2021 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics; 2021: 4972–84.
100. Liu Y, Ott M, Goyal N, *et al.* Roberta: a robustly optimized BERT pretraining approach. ArXiv:190711692, 2019.
101. Lewis M, Liu Y, Goyal N, *et al.* BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: proceedings of the 58th annual meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics; 2020: 7871–80.
102. Kauchak D, Mouradi O, Pentoney C, *et al.* Text simplification tools: using machine learning to discover features that identify difficult text. In: 2014 47th Hawaii international conference on system sciences. Los Alamitos, CA: IEEE Computer Society; 2014: 2616–25.
103. Brants T, Franz A. *Web 1T 5-Gram Version 1*. Philadelphia, PA: Linguistic Data Consortium; 2006.
104. Leroy G, Endicott JE, Mouradi O, *et al.* Improving perceived and actual text difficulty for health information consumers using semi-automated methods. In: AMIA annual symposium proceedings. Rockville, MD: American Medical Informatics Association; 2012: 522.
105. Leroy G, Endicott JE, Kauchak D, *et al.* User evaluation of the effects of a text simplification algorithm using term familiarity on perception, understanding, learning, and information retention. *J Med Internet Res* 2013; 15 (7): e2569.
106. Leroy G, Kauchak D, Mouradi O. A user-study measuring the effects of lexical simplification and coherence enhancement on perceived and actual text difficulty. *Int J Med Inf* 2013; 82 (8): 717–30.
107. Leroy G, Kauchak D, Hogue A. Effects on text simplification: evaluation of splitting up noun phrases. *J Health Commun* 2016; 21 (Suppl 1): 18–26.
108. Gu Y, Leroy G, Kauchak D. When synonyms are not enough: optimal parenthetical insertion for text simplification. In: AMIA annual symposium proceedings. Rockville, MD: American Medical Informatics Association; 2017: 810.
109. Zeng-Treitler Q, Goryachev S, Kim H, *et al.* Making texts in electronic health records comprehensible to consumers: a prototype translator. In: AMIA annual symposium proceedings. Rockville, MD: American Medical Informatics Association; 2007: 846.
110. Miller GA. WordNet: a lexical database for English. *Commun ACM* 1995; 38 (11): 39–41.
111. Tran L, Velazquez E, Sips R-J, *et al.* Evaluating medical lexical simplification: rule-based vs. BERT. *Stud Health Technol Inform* 2021; 281: 1023–4.
112. Kauchak D, Leroy G. A web-based medical text simplification tool. In: 53rd annual Hawaii international conference on system sciences, HICSS 2020. Los Alamitos, CA: IEEE Computer Society; 2020: 3749–57.
113. Peng Y, Tudor CO, Torii M, *et al.* iSimp: a sentence simplification system for biomedical text. In: 2012 IEEE international conference on bioinformatics and biomedicine. Los Alamitos, CA: IEEE; 2012: 1–6.
114. Mukherjee P, Leroy G, Kauchak D, *et al.* NegAIT: a new parser for medical text simplification using morphological, sentential and double negation. *J Biomed Inform* 2017; 69: 55–62.
115. Shardlow M, Nawaz R. Neural text simplification of clinical letters with a domain specific phrase table. In: *proceedings of the 57th annual meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics; 2019: 380–9.
116. Jonnalagadda S, Tari L, Hakenberg J, *et al.* Towards effective sentence simplification for automatic processing of biomedical text. In: Proceedings of human language technologies: the 2009 annual conference of the North American Chapter of the Association for Computational Linguistics, companion volume: short papers. Stroudsburg, PA: Association for Computational Linguistics; 2009: 177–80.
117. Jonnalagadda S, Gonzalez G. BioSimplify: an open source sentence simplification engine to improve recall in automatic biomedical information

extraction. In: AMIA annual symposium proceedings. Rockville, MD: American Medical Informatics Association; 2010: 351.
118. Evans R, Orăsan C. Sentence simplification for semantic role labelling and information extraction. In: proceedings of the international conference on recent advances in natural language processing (RANLP 2019). Shoumen, Bulgaria: INCOMA, Ltd.; 2019: 285–94.
119. Evans RJ. Comparing methods for the syntactic simplification of sentences in information extraction. *Lit Linguist Comput* 2011; 26 (4): 371–88.
120. Chen H-B, Huang H-H, Chen H-H, *et al.* A simplification-translation-restoration framework for cross-domain SMT applications. In: proceedings of COLING 2012. Bombay, India: IndianInstitute of Technology Bombay; 2012: 545–60.
121. Goodman KW, Miller RA. *Ethics in Biomedical and Health Informatics: Users, Standards, and Outcomes*. New York, NY: Springer; 2021: 391–423.