Intralingual Translation and Cascading Crises:

Evaluating the Impact of Semi-Automation on the Readability and Comprehensibility of Health Content

Alessandra Rossetti

https://orcid.org/0000-0002-2162-9639

alessandra.rossetti@dcu.ie

ADAPT Centre and School of Applied Language and Intercultural Studies, Dublin City University, Dublin, Ireland

Intralingual Translation and Cascading Crises:
Evaluating the Impact of Semi-Automation on the Readability and Comprehensibility of Health Content

Abstract

During crises, intralingual translation (or simplification) of medical content can facilitate comprehension among lay readers and foster their compliance with instructions aimed to avoid or mitigate the cascading effects of crises. The onus of simplifying health-related texts often falls on medical experts, and the task of intralingual translation tends to be non-automated. Medical authors are asked to check and remember different sets of plain language guidelines, while also relying on their interpretation of how and when to implement these guidelines. Accordingly, even simplified health-related texts present characteristics that make them difficult to read and comprehend, particularly for an audience with low (health) literacy. Against this background, this chapter describes an experimental study aimed at testing the impact that using a controlled language (CL) checker to semi-automate intralingual translation has on the readability and comprehensibility of medical content. The study focused on the plain language summaries and abstracts produced by the non-profit organisation Cochrane. Using Coh-Metrix and recall, this investigation found that the introduction of a CL checker influenced some readability features, but not lay readers' comprehension, regardless of their native language. Finally, strategies to enhance the comprehensibility of health content and reduce the vulnerability of readers in crises are discussed.

Keywords: intralingual translation; plain language; health content; readability; comprehensibility; cascading crises

1. Introduction and related work

Communication during crises and disasters often revolves around the health and well-being of the affected communities (Reynolds and Seeger, 2007). Disease outbreaks are followed by health messages on the recommended course of action to avoid the spread of infections (Bults *et al.*, 2011; Holmes, 2008); authorities might declare that drinking tap water represents a health threat after an earthquake (Yoshii *et al*., 2014); and industrial accidents often lead respondents to consider the risks to health represented by toxic waste (Cutchin *et al.*, 2008). The production of effective health messages can therefore ensure that crises and disasters do not propagate with cascading effects, which are not unlikely in today's interconnected and complex environmental, health, and financial systems (Helbing, 2013).

Glik (2007, p. 38) argues that, for a message to be effective,

> the recipient of threat information must (a) receive the information, (b) understand the information, (c) understand that the message applies to them, (d) understand that they are at risk if they do not take protective action, (e) decide that they need to act on the information, (f) understand which actions need to be taken, and (g) be able to take action.

Comprehension (or understanding) of a message is therefore key. Interlingual translation of health messages (e.g. from English to Kiswahili following the Ebola outbreak) has already been shown to facilitate comprehension (O'Brien and Cadwell, 2017). However, the importance of intralingual translation (or simplification) of medical[1] content has also been widely recognized, since health-related texts are often characterized by specialized vocabulary, complex syntactic structures, and cohesion gaps that make it difficult for lay people[2] (even native speakers) to read and understand them (Lachance *et al.*, 2010; Mićić, 2013). Individuals with low levels of health literacy—i.e. with reduced ability to process and understand health-related information (Berkman *et al.*, 2010)[3]—represent a particularly vulnerable group, having lower sense of empowerment and decreased ability to manage their own health (Dumenci *et al.*, 2014; Sørensen *et al.*, 2012).

Intralingual translation can be defined as '[t]he translation of scientific knowledge into useful constructs or concepts that nontechnical audiences can understand' (Glik, 2007, p. 36). Muñoz-Miquel (2012, 187) describes intralingual translation in terms of expert-to-layman reformulation. This reformulation often involves the use of plain language. As Redish (2000, p. 165) remarks, ‘a document in plain language is one that works for its users’. Warde et al. (2018, p. e54) provide a definition of plain language applicable to the health content, whereby

> [p]lain language is defined as communication that can be understood the first time it is seen or heard, that uses succinct active-voiced grammatically correct complete sentences to better enable patients and caregivers to engage with information, using a more informal tone and common terms whenever possible.

At the World Health Organization, communicators are encouraged to translate specialized information into messages that a lay audience could understand, e.g. by explaining technical terms or by breaking long portions of content into chunks (World Health Organization, 2019). The U.S. Department of Health and Human Services adopted a plain-language approach for their website *healthfinder.gov* (Schriver, 2017), and expanded its efforts across its various divisions, including the Centers for Disease Control and Prevention, the Centers for Medicare and Medicaid Services, and the National Institutes of Health, where training on plain language writing is provided to staff members (see U.S. Department of Health and Human Services, 2018). Moreover, guidelines are made available which mainly focus on vocabulary (e.g. avoidance of medical jargon); syntax (e.g. avoidance of passive voice); content organisation; and, only occasionally, cohesion (Centers for Disease Control and Prevention, 2016).

Intralingual translation at these organisations is often a manual/non-automated task for which no technological assistance is provided to guide authors in the consistent application of plain language guidelines. Accordingly, authors might forget to implement a guideline, or interpret it in a different way from other authors within the same organisation (Harper and Zimmerman, 2009). Unsurprisingly, research has shown that medical content tends to require reading skills associated with a ninth- or higher US grade level (Carbone and Zoellner, 2012). In the context of a crisis, the difficulty of health-related texts is further enhanced by the trauma of the event and by the resulting anxiety, which has a detrimental effect on comprehension (Keselman *et al.*, 2005). It is therefore important that plain language guidelines are applied consistently and uniformly across a text so as to mitigate the effects of anxiety and facilitate comprehension of heath content in a crisis.

Providing authors of health content (and in particular medical experts) with some form of technological assistance to semi-automate the traditionally manual intralingual translation tasks has been proposed as a solution to the inconsistent application of plain language guidelines and, in turn, to the low readability and comprehensibility of medical texts. Leroy *et al.* (2013) describe an algorithm whose goal is to semi-automate lexical simplification and coherence enhancement. The authors (ibid.) argue that semi-automating the intralingual translation/text simplification task (e.g. by means of a Microsoft Word plugin) could help health domain experts become more effective in the production of readable and comprehensible texts. Similarly, Smith *et al.* (2011, n.p.) state that 'given the difficulty of engineering comprehensibility of clinical text, the most useful informatics tools will be those that can support the physicians, nurses, and patient educators tasked with making clinical information understandable to patients.' Despite these remarks, empirical evidence on the effect of introducing technological assistance into a manual intralingual translation task for health-related texts is scarce—the study described in this chapter sought to help fill this research gap.

2. The case of Cochrane

The study focused on the English medical content produced by Cochrane, an international non-profit organisation mainly relying on volunteer health professionals to systematically gather and review medical evidence on the effectiveness of treatments and interventions (Higgins and

Green, 2011). Cochrane contributors produce systematic reviews that address specific research questions and are then published on the Cochrane Library website[4]. Examples of systematic reviews are *Interventions to Increase Tuberculosis Case Detection at Primary Healthcare or Community-Level Services* (Mhimbira *et al.*, 2017)—produced in an effort to fight tuberculosis—or *Polymer-Based Oral Rehydration Solution for Treating Acute Watery Diarrhoea* (Gregorio *et al.*, 2016)—written as part of a special collection on interventions for the prevention and treatment of water-related diseases after natural disasters. Cochrane has been collaborating with, among others, the World Health Organization to communicate the results of medical research to a wide audience, from policy-makers and clinicians, to the lay public in low- and middle-income countries (Cochrane Infectious Diseases, 2019).

Since systematic reviews tend to be long and inaccessible for lay readers due to their specialized language (Harvey, 2018), each systematic review is preceded by: 1) a plain language summary (PLS) that summarizes and simplifies the content of the systematic review for lay readers; 2) an abstract, i.e. a non-simplified summary targeting health professionals. Similar to other organisations sharing the same mission, Cochrane adopts a manual intralingual translation approach for the production of PLS, whereby authors are asked to check, remember, interpret, and manually apply sets of plain language guidelines. These guidelines are spread across different documents, show contradictions, and lack examples and specificity (see e.g. Cochrane Norway, 2017 or The Cochrane Collaboration, 2013). Accordingly, PLS have been found to have low readability and comprehensibility (Karačić *et al.*, 2017; Maguire and Clarke, 2014; Santesso *et al.*, 2015).

The project tested if and how asking Cochrane authors to use the Acrolinx controlled languages[5] (CL) checker to edit their PLS would enhance the readability and comprehensibility of these texts. The Acrolinx CL checker is a tool that automatically and consistently flags readability issues in a text (e.g. in relation to sentence length, style, vocabulary), while also providing: 1) examples/suggestions on how to address them; and 2) an overall score accounting for style, spelling, grammar, tone of voice, and terminology (Rodríguez Vázquez, 2016). The use of this CL checker represents a semi-automated approach to intralingual translation since the author or editor needs to manually select/apply the recommended changes even though these are automatically flagged (Schwitter, 2015).

Concretely, the project addressed the following research question:

> *Does semi-automating a manual approach for the intralingual translation of health content increase text readability and comprehensibility?*

For the purpose of this study, it was necessary to make a distinction between readability and comprehensibility, since the former is determined by text characteristics and is measured to predict the difficulty of a text for a group of readers, while the latter is determined by the reader's characteristics (such as native language, reading skills, prior knowledge, or disabilities), and involves testing actual comprehension (Collins-Thompson, 2014; Shardlow, 2014). Therefore, these two aspects were analysed separately. Furthermore, this investigation expanded its analysis to also include abstracts—namely, the non-simplified summaries of Cochrane Systematic Reviews—as these texts would act as benchmark, i.e. they were expected to allow the investigator to test the impact of intralingual translation/simplification (whether manual or semi-automated) against the lack of simplification efforts.

3. Materials, methods, design, and procedure

3.1. The Acrolinx tool

Twelve Cochrane authors used the Acrolinx CL checker (provided as a Microsoft Word plugin) to edit their PLS, previously produced with the manual intralingual translation approach and written in English. The authors were also instructed to use their common sense in deciding whether to apply a change recommended by Acrolinx. The Acrolinx plugin for Microsoft Word displays the identified readability issues in a sidebar, along with suggestions and examples. In other words, two different types of information are made available to authors: 1) linguistic content (e.g. style or grammar rules that have been contravened); and 2) didactic content (i.e. explanation of the readability issues and assistance for the authors) (Reuther and Schmidt-Wigger, 2000). When a readability issue is selected in the sidebar, it is also flagged in the document. Figure 1 shows how the Acrolinx tool worked on a sample PLS.

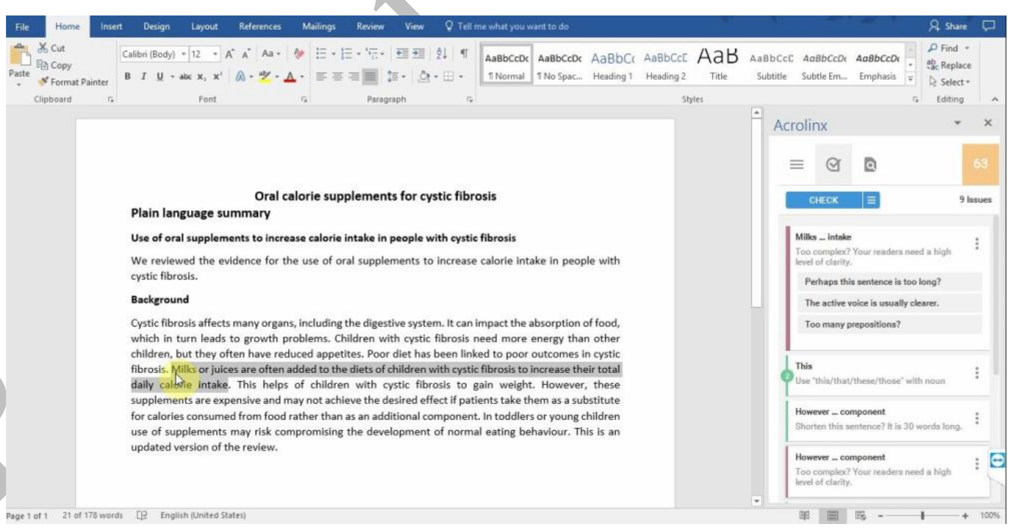

Figure 1. Running a readability check with Acrolinx on a sample PLS.

Regarding spelling, the texts were checked against US English spelling—the set of Acrolinx CL rules called *Standard_US* was adopted. In the Cochrane Style Manual (The Cochrane

Collaboration, 2016), it is specified that both British and American spelling are allowed as long as they are used consistently.

The Acrolinx CL rules that would contravene Cochrane PLS guidelines, such as the Acrolinx rules on hyphens, possessives, or en dashes were deselected, as permitted by the local server that was used for this study. More importantly, the Acrolinx rule on avoiding modal verbs was deactivated since modality is widely used in Cochrane texts, particularly when evidence of the benefits of a treatment or intervention is uncertain (Cochrane Norway, 2017).

3.2. Experimental materials

Three sets of texts were available for this study. They dealt with a variety of health-related topics, from strabismus, to stroke, to mental health problems, to aneurysm:

(1) 12 PLS produced with the manual simplification approach (henceforth pre-Acrolinx PLS);

(2) the same 12 PLS edited with the Acrolinx CL checker (henceforth post-Acrolinx PLS);

(3) 12 abstracts (i.e. non-simplified texts) produced by the same authors.

3.3. Coh-Metrix

In order to measure and compare the readability of these three sets of texts, Coh-Metrix was used. This is a computational tool that automatically analyses text difficulty along a variety of dimensions, including vocabulary, syntax, and cohesion (Dowell *et al.*, 2016). For the present study, this tool was preferred to traditional readability formulas (such as the Flesch Reading Ease or the SMOG) since these formulas only consider shallow text characteristics—mainly word length and sentence length—while ignoring other text characteristics that have been shown to be predictors of reading comprehension, such as cohesion (Liu and Rawl, 2012). This limitation of traditional readability formulas is highlighted in Leroy *et al.* (2013, p. 719), who argue:

> many formulas equate long words with difficult words. However, in medicine, this relationship may not hold true, e.g., 'apnea' would be considered more difficult than 'diabetes' or 'obesity' by most readers.

Furthermore, previous studies had already used Coh-Metrix to investigate differences between texts simplified for readers with different skills (Crossley *et al.*, 2012), and between simplified and non-simplified texts (Crossley and McNamara, 2008).

The Coh-Metrix measures that were selected for this study are: narrativity; syntactic simplicity; word concreteness; referential cohesion; deep cohesion; Coh-Metrix L2 (second language) Reading Index; and the Flesch-Kincaid Grade Level, a traditional readability formula introduced for comparison with the other multidimensional measures of readability provided by Coh-Metrix. Narrativity is determined by the extent to which the language in a text is frequent and familiar (McNamara *et al.*, 2014). Syntactic simplicity is influenced by the length of the sentences and by their syntactic structures (ibid.). Word concreteness is linked with the presence of concrete words that facilitate the evoking of mental images (ibid.). Referential cohesion is determined by the extent to which the same words and ideas are repeated across sentences and the entire text (McNamara *et al.*, 2011). Deep cohesion is linked with the presence of causal and intentional connectives signalling these relationships in the text (McNamara *et al.* 2014). The Coh-Metrix L2 Reading Index measures three variables that have been shown to predict the difficulty of L2 reading more accurately than traditional readability formulas, namely lexical coreferentiality, syntactic similarity, and word frequency (Crossley *et al.*, 2008). Finally, the Flesch-Kincaid Grade Level is determined by word length and sentence length, and it indicates the reading ability (in terms of US grade-school level) required to be able to read a text (McNamara *et al.*, 2014).

Scores for narrativity, syntactic simplicity, word concreteness, referential cohesion, deep cohesion, and the Flesch-Kincaid Grade Level were obtained through Coh-Metrix Common Core Text Ease and Readability Assessor (T.E.R.A.)[6], while the Coh-Metrix L2 Reading Index was obtained by Coh-Metrix 3.0[7]. With the exception of the Flesch-Kincaid Grade Level, for each of the selected measures, the higher the score the higher the readability of the texts (McNamara *et al.*, 2014).

3.4. Text-retelling

Text-retelling was adopted in order to assess comprehensibility. In this method, participants are prompted to read a text and, subsequently, to tell or write what they can recall from it—including inferences and elaborations—in their own words, and without the possibility of going back to the text (Crossley and McNamara, 2016; Reed and Vaughn, 2012). Data were elicited by means of both free recall protocols (i.e. participants were asked to write everything they could remember about an entire text) and cued recall protocols (i.e. participants were asked to write everything they could remember about specific sections of a text) (McNamara *et al.*, 2011). Text-retelling was preferred to other comprehension tests (e.g. cloze tests or multiple-choice questions) since, unlike other tests, text-retelling ensures that the format of the questions is equivalent across texts (as in 'write everything you can remember about…') (Reed and Vaughn, 2012); allows for the avoidance of clues in the questions (Hansen, 1978); does not expose participants to incorrect statements (Roediger and Marsh, 2005); and does not allow them to guess (Crossley and McNamara, 2016). Furthermore, text-retelling is theoretically supported by Kintsch's (1998) Construction-Integration model of reading comprehension, according to which an idea is more likely to be recalled when it is integrated in the mental representation of a text through its connections with the other ideas.

For the text-retelling experiment, a random sampling technique was used to recruit participants from the pool of students at Arizona State University (ASU) who were not enrolled in health-related courses or training—this requirement enabled testing comprehensibility of Cochrane/medical content among lay readers. Despite their young age (Section 4.2), these participants represented a realistic target audience for Cochrane, whose research revolves around a broad range of health issues (from mental health, to skin disorders, to allergies).

Fifty-nine native speakers of English and 23 non-native speakers of English were involved. The study was conducted in a laboratory setting. Each participant was asked to conduct the following tasks: (i) complete a short background questionnaire on their gender, age, ASU college/school being attended, year of college, native language, types of texts generally read in English, and English ability (for non-native speakers of English only); (ii) complete the Gates-MacGinitie Reading Test assessing their reading skills; (iii) read three texts—a pre-Acrolinx PLS, a post-Acrolinx PLS, and an abstract—and answer one free recall question and two cued recall questions on each of them; and (iv) answer nine multiple-choice questions assessing their knowledge of the health topics discussed in the texts.

In other words, a within-subjects design was adopted, whereby each participant read the three texts representing the three experimental conditions (i.e. manual intralingual translation, semi-automated intralingual translation, and lack of intralingual translation efforts) while being blinded to the design (Buljan *et al.*, 2018). A within-subjects design was selected to isolate the impact of individual differences (Lazar *et al.*, 2010). To compensate for order effects, namely fatigue and learning effect, the order in which texts from different sets were presented to readers was counterbalanced, and each reader was assigned texts on three different health-related topics.

The questions on topic knowledge were asked because, even though the participants involved were lay readers with no systematic knowledge of the medical domain, they might have had isolated knowledge about a specific topic (e.g. if they had undergone one of the treatments described in the texts or knew somebody who had) (Alexander *et al.*, 1994). The questions on topic knowledge were submitted to participants after the experimental tasks so as not to influence their reading behaviours. The texts read were also checked to ensure that they did not contain answers to the topic-knowledge questions (Ozuru *et al.*, 2009).

Participants could spend as much time as they needed reading the three texts assigned to them, which contained between 290 and 460 words. However, in line with a similar study described in Crossley and McNamara (2016), a time limit was set for the for text-retelling—four minutes for each free recall question and one and a half minute for each cued recall question. Participants were also instructed not to worry about spelling mistakes.

4. Data analysis and results

4.1. Readability

For each of the selected readability measures (namely, narrativity, syntactic simplicity, word concreteness, referential cohesion, deep cohesion, L2 Reading Index, and Flesch-Kincaid Grade Level), descriptive statistics were first calculated for each set of texts. Subsequently, a series of repeated measures analyses of variance (ANOVA) were conducted, where the readability measures represented the dependent variable, and the three sets of texts represented the levels of the independent variable (manual intralingual translation, semi-automated intralingual translation, and lack of intralingual translation efforts). Regarding syntactic simplicity scores, the Shapiro-Wilk test showed that the data for one set of texts did not meet the assumption of normality (p=0.01). Therefore, a Friedman test was used instead. When the repeated measures ANOVA and the Friedman test indicated the presence of at least one significant difference, post hoc tests (i.e. Tukey Honestly Significant Difference post hoc test and Wilcoxon signed-rank tests with Bonferroni adjustment) were conducted to find out where the significant differences lay. Descriptive statistics—means, standard deviations (SD), and medians—for the scores produced by Coh-Metrix for the selected readability measures are reported in Table 1, divided by set of texts. The asterisks indicate which differences were found to be statistically significant. For instance, in the case of narrativity, a significant difference was found between pre-Acrolinx PLS and abstracts (signalled with one asterisk), and between post-Acrolinx PLS and abstracts (signalled with double asterisk).

Measures	Pre-Acrolinx PLS (N=12)	Post-Acrolinx PLS (N=12)	Abstracts (N=12)
	Means (SD)	***Means (SD)***	***Means (SD)***
Narrativity	21.75 (14.15) (*)	20.66 (10.33) (**)	8.25 (2.45) (*)(**)
Word concreteness	25.41 (15.10)	21.91 (14.38)	22.75 (13.83)
Referential cohesion	44.08 (15.58) (*)	41.08 (16.92) (**)	20.33 (7.86) (*)(**)
Deep cohesion	54.83 (24.65) (*)	48.58 (23)	34.33 (16) (*)
Coh-Metrix L2 Reading Index	12.25 (3.5) (*)	14.01 (3.72) (**)	6.2 (2.17) (*)(**)
Flesch-Kincaid Grade Level	12 (1.59) (*)(**)	10.58 (0.9) (*)(***)	13.83 (1.11) (**)(***)

	Median	*Median*	*Median*
Syntactic simplicity	61.5(*)	78.5(*)	70

Table 1. Descriptive and inferential statistics for readability measures analysed, per corpus.

It can be observed that introducing the CL checker into the manual simplification approach resulted in a statistically significant increase in the syntactic simplicity of Cochrane PLS, as well as in a statistically significant decrease in their word length and sentence length, as indicated by the Flesch-Kincaid Grade Level. Furthermore, compared with the abstracts—where no simplification attempts were made—simplified texts showed significantly higher levels of narrativity, referential cohesion, and L2 readability, while also being associated with a significantly lower Flesch-Kincaid Grade Level, both before and after the introduction of Acrolinx. In other words, regardless of being non-automated or semi-automated, intralingual translation led to texts scoring higher on a variety of readability measures.

4.2. Comprehensibility

In terms of demographic profile of the participants/lay readers, Figures 2 and 3 show a good gender balance among both groups of native and non-native speakers of English. In both groups, the majority of participants was between 18 and 19 years old.

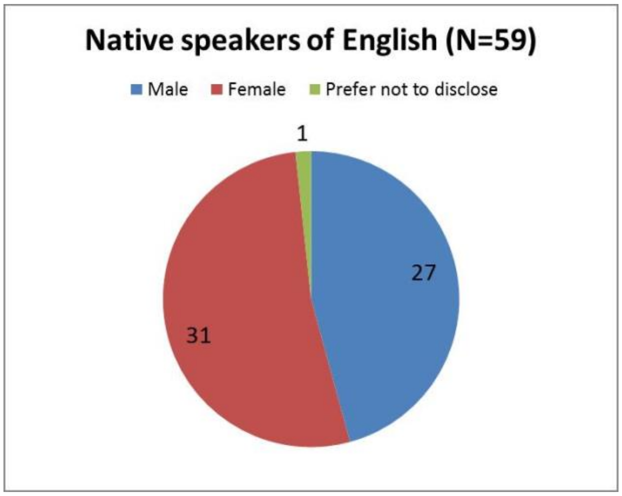

Figure 2. Gender distribution of native participants.

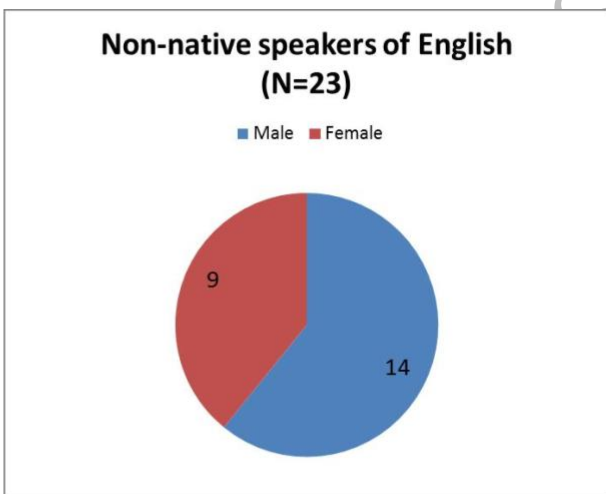

Figure 3. Gender distribution of non-native participants.

Five participants reported being enrolled in health-related courses at ASU. However, drawing upon previous studies (Boshuizen and Schmidt, 1992; Shapiro, 2004), for the purpose of this investigation, they could be treated as lay readers since they were all between their first and third year of study.

With regard to the types of texts generally read in English, both native and non-native participants stated that they read mainly emails, followed by essays and notes. Within the non-native group, participants had different first languages, mainly Chinese and Spanish (see Figure 4). The majority of non-native participants (n=17) reported having spent seven years or more speaking English. However, only two of them spoke English at home. When asked to answer the question ‘How well do you speak English?’ (Vickstrom *et al.*, 2015), most non-native participants answered either ‘very well’ (n=10) or ‘well’ (n=10).

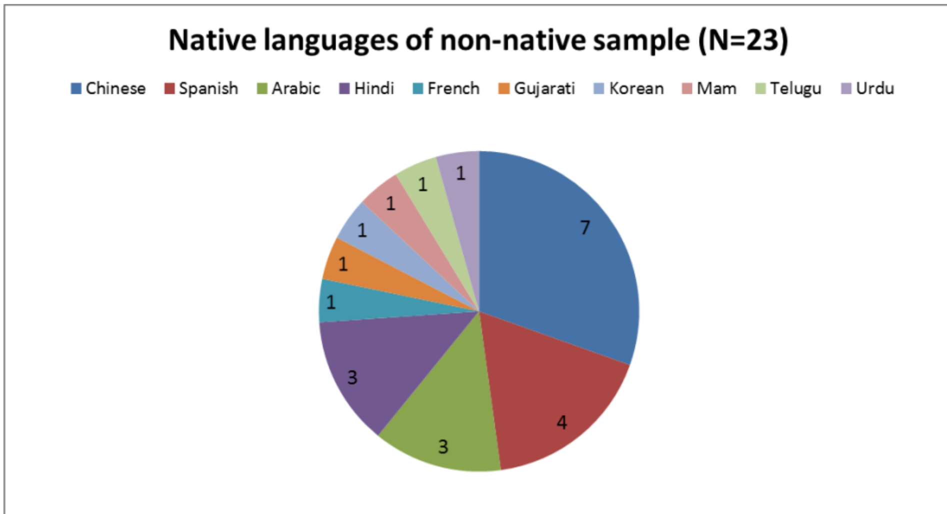

Figure 4. Native languages of non-native sample.

Results for participants’ reading skills, assessed through 48 multiple-choice questions in the Gates-MacGinitie Reading Test, are reported in Table 2, where it can be observed (somewhat unsurprisingly) that native English speakers had higher reading skills—for native readers, the scores on reading skills ranged between 13 and 46 (out of 48). For non-native readers, they ranged between 10 and 36 (out of 48). Regarding topic knowledge, Table 3 shows that, overall, the lay readers were not very familiar with the health-related topics discussed in the three texts assigned to them.

Participants/lay readers	**Scores (out of 48)** ***Mean (SD)***
Entire sample (N=82)	29.84 (10.94)
Native English speakers (N=59)	33.72 (9.04)
Non-native English speakers (N=23)	19.86 (8.97)

Table 2. Descriptive statistics for participants’ reading skills.

Descriptive statistics on topic knowledge of lay readers	Scores (out of 9)
Mean (SD)	4.91 (1.55)
Min	1
Max	7

Table 3. Descriptive statistics for lay readers' topic knowledge.

Before analysing the free and cued recall protocols produced by the participants, a rubric was developed to:

(1) segment both the texts and the recall protocols into idea units, defined as phrases (Richards *et al.*, 2016);

(2) assign a score to each idea unit/phrase in the recall protocols based on its accuracy and completeness compared with the corresponding idea unit in the text.

Furthermore, the author of this chapter and a second annotator segmented a set of sample sentences. Subsequently, segmentation was compared, and disagreements were resolved through discussion. However, only the author of this chapter conducted the segmentation and the scoring of all the recall protocols. Concretely, each phrase in the recall protocol was assigned: a score of 1 if it contained accurate and complete information when checked against the corresponding phrase in the text; a score of 0.5 if the information was incomplete or partially inaccurate; and a score of 0 if all the information was inaccurate (Best *et al.*, 2008). Correct inferences (produced by two participants only) were assigned a score of 2. Participants were not penalized for typos or for using synonyms (e.g. for using 'fixing' instead of 'treating') (Bovair and Kieras, 1981; Diao and Sweller, 2007). Unintelligible and irrelevant content in the recall protocols (e.g. 'do not remember') was excluded from the analysis, along with repetitions.

Free recall and cued recall scores were calculated separately. For each recall protocol, both a raw score and a percentage score (out of the maximum score that a participant might have obtained) were calculated. Percentage scores were used for the subsequent analysis. Since each participant was assigned two cued recall questions per text, the mean of their cued recall percentage scores was calculated prior to the analysis.

For each group of participants (i.e. native vs non-native speakers of English), a series of repeated measures ANOVA/Friedman tests[8] were run, followed by post hoc tests, where free recall and cued recall represented the dependent variables, and the three sets of texts represented the levels of the independent variable (manual simplification, semi-automated simplification, lack of simplification efforts). Table 4 reports descriptive statistics for recall scores, with the asterisks indicating significant differences. Different numbers of asterisks are used to indicate where the significant differences lie.

Type of recall	Participants	Abstracts	Pre-Acrolinx PLS	Post-Acrolinx PLS
		Mean (SD)	*Mean (SD)*	*Mean (SD)*
Free recall	*Native English speakers*	9.08 (5.21) (*)(**)	13.39 (7.59) (*)	12.87 (7.28) (**)
	Non-native English speakers	5.26 (4.05) (*)(**)	8.36 (5.09) (*)	8.26 (5.38) (**)
Cued recall	*Native English speakers*	29.77 (21.4) (*)(**)	14.9 (9.09) (*)	15.72 (12.58) (**)
	Non-native English speakers	26.45 (22.88) (*)(**)	6.3 (7.09) (*)	7.68 (7.07) (**)

Table 4. Descriptive and inferential statistics for recall scores, per corpus and sample of participants.

It emerges that: (i) the introduction of semi-automation during simplification did not have a significant impact on lay readers' ability to recall Cochrane PLS; (ii) free recall of PLS was significantly higher than free recall of abstracts; and (iii) cued recall of PLS was significantly lower than cued recall of abstracts.

Several within-subjects analyses of covariance (ANCOVA) were conducted, where reading skills (as assessed through the Gates-MacGinitie Reading Test) were treated as covariate. Reading skills are an important component of (health) literacy (Simonds *et al.*, 2017). A separate within-subjects ANCOVA for each type of recall (free vs cued) and sample of participants (native vs non-native) was conducted. Here the partial eta squared (η_{p2}) is also given as a measure of effect size. After ensuring that the assumption of homogeneity of regression slopes was met, the study found that:

(1) for native speakers, the within-subjects ANCOVA on free recall including reading skills as covariate was not significant, $F(1.584, 85.527)=0.461$, $p=0.587$, $\eta_{p2}=0.008$. A significant effect of reading skills on free recall of native readers was observed, $F(1, 54)=5.535$, $p=0.022$, $\eta p2=0.093$. Therefore, by including reading skills as a covariate, the significant differences that had been observed between free recall of abstracts and free recall of PLS were no longer significant;

(2) for native speakers, the within-subjects ANCOVA on cued recall including reading skills as covariate was not significant, $F(1.569, 86.315)=0.6$, $p=0.512$, $\eta_{p2}=0.011$. A significant effect of reading skills on cued recall of native readers was observed, $F(1, 55)=4.628$, $p=0.036$, $\eta_{p2}=0.078$. Therefore, by including reading skills as a covariate, the significant differences that had been observed between cued recall of abstracts and cued recall of PLS were no longer significant;

(3) for non-native speakers, the within-subjects ANCOVA on free recall including reading skills as covariate was not significant, $F(2, 42)=0.02$, $p=0.980$, $\eta_{p2}=0.001$. Moreover, reading skills did not have a significant effect on free recall of non-native readers, $F(1, 21)=3.457$, $p=0.077$. However, reading skills were shown to account for about 14% of

the variance in free recall. When the sample size is small—as in the case of the non-native speakers of English in this study (n=23)—'strong and important effects can be nonsignificant' (Levine and Hullett 2002, p. 614);

(4) for non-native speakers, the within-subjects ANCOVA on cued recall including reading skills as covariate was not significant, $F(1.143, 22.852)=0.039$, $p=0.874$, $\eta_{p}^{2}=0.002$. Moreover, reading skills did not have a significant effect on cued recall of non-native readers, $F(1, 20)=3.697$, $p=0.069$. However, reading skills were shown to account for about 15% of the variance in cued recall.

5. Discussion and conclusions

Adopted to avoid or mitigate the cascading effects of crises by facilitating comprehension among readers with low health literacy and no medical knowledge, intralingual translation (or simplification) of health content is often conducted by authors with a medical background, and its effectiveness is influenced by the authors' memory and subjective interpretation of how and when plain language guidelines should be applied. Furthermore, as in the case of Cochrane, guidelines can be contradictory and vague.

This chapter described an experimental investigation of the impact that semi-automating a manual intralingual translation approach has on the readability and comprehensibility of health-related texts—having a CL checker that automatically and consistently flags issues in a text was assumed to be beneficial for these two aspects. A similar study is reported in Leroy *et al.* (2013). However, this investigation is the first to: (i) focus on the non-profit international organisation Cochrane and the Acrolinx CL checker; and (ii) adopt Coh-Metrix and recall to examine, respectively, the readability and comprehensibility of health-related texts. If Cheek's (2010) classification of plain language into three categories were to be applied, it could be argued that this study relies on: (i) a formula-based approach to plain language (readability assessment); (ii) an outcomes-focused approach (i.e. comprehension testing); and (iii) an elements-focused approach (consideration of the impact of guidelines and rules).

With regard to readability, it was observed that using the Acrolinx CL checker led to Cochrane PLS with significantly higher syntactic simplicity. This result might be due to the higher specificity of Acrolinx CL rules, compared with Cochrane guidelines. For instance, while Cochrane guidelines contain the vague '[l]imit sentences to one key point' (The Cochrane Collaboration, 2013, p. 4), Acrolinx specifies the maximum number of words allowed in a sentence. In addition, even when the same syntactic issue (such as the avoidance of passive voice) is specifically addressed both in Cochrane guidelines and Acrolinx rules, authors might be more likely to apply the recommended modification in the semi-automated approach because the readability issue is automatically flagged, i.e. they do not need to rely on their memory of the guideline. Compared with pre-Acrolinx PLS, post-Acrolinx PLS were also associated with a significantly lower US grade level, i.e. they had significantly lower word and sentence length. Both Cochrane guidelines and Acrolinx rules address these readability issues—see e.g. Cochrane's reliance on the SMOG index that calculates word and sentence length (Mc Laughlin 1969; The Cochrane Collaboration, 2013, p. 4). Again, the significant increase in readability observed after revising the PLS with Acrolinx might be due to its automatic flagging of the issues.

Despite these readability improvements, there was no significant impact of semi-automation on comprehensibility, as assessed though recall. In other words, using the Acrolinx CL checker was not beneficial for the comprehension of lay readers with different language backgrounds and relatively low knowledge of the health topics discussed. One explanation for this finding could be that, while having an impact at the word- and syntax-level, the CL checker did not increase cohesion, a text characteristic that facilitates the development of connections between the ideas in the text, and, in turn, their recall (McNamara *et al.*, 2014). Furthermore, several studies showed that, compared with interventions at other text levels (e.g. vocabulary), increasing cohesion is more beneficial in terms of recall of medical texts (Smith *et al.*, 2011).

To summarize the main findings of this study, it is useful to reconsider the research question here:

> *Does semi-automating a manual approach for the intralingual translation of health content increase text readability and comprehensibility?*

The findings discussed here, based on the impact of the Acrolinx CL checker, show that semi-automating a manual intralingual translation approach for health content can increase some readability measures (particularly at the word and sentence level), but does not prove beneficial in terms of comprehension among native or non-native speakers of English.

Regarding practical implications, this study points to the need for developing technological assistance for intralingual translation that also accounts for cohesion, especially considering that this aspect is rarely mentioned in the plain language guidelines provided to authors. An example of a CL rule aimed at increasing cohesion (for Spanish) is reported in Cascales (2002, p. 55): 'Avoid the use of referring expressions such as pronouns and deictic determiners, instead repeat the concept'. Additionally, authors should be trained to search for cohesion issues at the macro-level of a text, e.g. by ensuring that the presentation of arguments follows a logical order, and that each paragraph begins with a topic sentence (Kools *et al.*, 2004).

Additional results emerged from this investigation. In line with previous studies showing the beneficial effect of intralingual translation or simplification on comprehension (Kurtzman and Greene, 2016), this study also found that comprehension of abstracts/non-simplified texts (measured through *free* recall) was significantly lower than comprehension of PLS/simplified texts, among both native and non-native speakers of English. This result is not surprising when considering that the readability analysis had shown that abstracts were characterized by significantly lower levels of narrativity, referential cohesion, lexical coreferentiality, syntactic similarity, and word frequency, among others.

Yet, it also emerged that *cued* recall of abstracts was significantly higher than cued recall of simplified Cochrane texts. This finding is surprising, especially considering that free recall and cued recall are usually highly correlated (McNamara *et al.*, 2011). To explain this result, it was useful to consider characteristics other than language in both abstracts and PLS. By looking at these, it was observed that: (i) sections in the abstracts were usually shorter than sections in the PLS; and (ii) while abstracts always had headings in bold formatting to separate the text into different coherent sections, some PLS either were not divided into sections, or had section headings that did not coherently match the content of the sections. For instance, it was not uncommon for PLS to contain a section with a broad heading such as *Background*, under which unrelated types of information were reported. This claim would need to be tested in future studies. However, it is worth pointing out that previous works have already underlined the beneficial effects (on comprehension) of using (bold) headings, and short sections (Kurtzman and Greene, 2016; Rusko *et al.*, 2012). See, for example, this quote from Kools *et al.* (2004, p. 723, emphasis added):

> [h]eadings and subheadings clarify overall text structure and can serve as "anchors" for the reader […]. In general, it may be assumed that headings influence cognitive processing by (a) acting as cues for activating related prior knowledge, (b) accentuating the relationship

among important concepts in a text, and (c) providing retrieval cues for subsequent recall of a text.

It should be remembered that, with cued recall, participants wrote everything they could remember about a specific section. These results show that, while beneficial, text simplification might not be enough, especially if specific sections of a text (e.g. on prevention of an infection) need to be comprehended and remembered by the target audience. Formatting and text segmentation can all enhance the effectiveness of the health message. Future studies might examine the impact of semi-automation on these extra-linguistic aspects. Acrolinx, for example, allows users to configure the maximum number of sentences allowed in a paragraph—this functionality might in turn reduce the length of the section (Carter, 2018).

This study also found that, for both native and non-native readers, reading skills tend to drive comprehension, often more than the type of text being read. In other words, individuals with low literacy will show poorer comprehension than individuals with higher literacy, regardless of text characteristics such as plain language or formatting. This result indicates a need for disseminating health content into a variety of formats, such as audio and images. For instance, as reported in a paper by Brown *et al.* (2014, p. 270) on the role of literacy in disasters, 'pictographs or simple drawings of figures demonstrating explicit, detailed actions are an effective way to communicate health care instructions to adults with low-literacy skills.' It is worth mentioning that Cochrane is currently producing podcasts of their systematic reviews (Maguire and Clarke, 2014).

To conclude, this study highlighted the impact (and lack thereof) of semi-automating the intralingual translation of medical content. It also suggested ways in which technological assistance in the form of a CL checker (or similar) could be developed with a view to enhancing text comprehension. The importance of considering aspects other than language, and formats other than texts, was also discussed. There are, however, several limitations and areas for future testing. A small number of abstracts and PLS were available when setting up this study. Moreover, the text-retelling experiment involved a limited number of non-native speakers of English. This limited availability of texts and participants might have resulted in the inability to observe a true effect in the statistical results. The findings of the statistical tests should therefore be confirmed in larger studies. Furthermore, having additional raters segmenting and scoring the recall protocols would have enhanced the reliability of the procedure. In addition to measuring generic reading skills through the Gates-MacGinitie Reading Test, it might also be necessary to specifically measure participants' health literacy, for example by means of the Test of Functional Health Literacy in Adults (Parker *et al.*, 1995).

Finally, the CL checker used had not been tailored to Cochrane/medical content—future research with a customized tool might show stronger impact on readability and comprehensibility. Despite these limitations, findings from this investigation can inform and build on the work that Cochrane and other organisations (e.g. the Centers for Disease Control and Prevention) are conducting to make health content more comprehensible for lay readers with low health literacy, and to reduce their vulnerability during crises.

Acknowledgement

This work has received funding from the Irish Research Council (GOIPG/2017/1409), and from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 734211. The author would like to thank Professor Danielle McNamara and her team at the Science of Learning and Educational Technology Lab at Arizona State University for their technical assistance with participants' recruitment and with the setup of the study.

Disclosure statement and ethical approval

The author has no financial interest or benefit arising from the direct applications of her research. This study received approval from the Research Ethics Committee at Dublin City University (DCUREC/2017/066), and from the Institutional Review Board at Arizona State University (STUDY00006514).

References

Alexander, Patricia, Jonna Kulikowich, and Sharon Schulze. 1994. "The Influence of Topic Knowledge, Domain Knowledge, and Interest on the Comprehension of Scientific Exposition." *Learning and Individual Differences* 4: 379–397.

Berkman, Nancy, Terry Davis, and Lauren McCormack. 2010. "Health Literacy: What Is It?" *Journal of Health Communication* 15: 9–19.

Best, Rachel, Randy Floyd, and Danielle McNamara. 2008. "Differential Competencies Contributing to Children's Comprehension of Narrative and Expository Texts." *Reading Psychology* 29 (2): 137–164.

Boshuizen, Henny, and Henk Schmidt. 1992. "On the Role of Biomedical Knowledge in Clinical Reasoning by Experts." *Cognitive Science* 16: 153–184.

Bovair, Susan, and David Kieras. 1981. *A Guide to Propositional Analysis for Research on Technical Prose (Technical Report No. 8).* Tucson: Department of Psychology, University of Arizona.

Brown, Lisa, Jolie Haun, and Lindsay Peterson. 2014. "A Proposed Disaster Literacy Model." *Disaster Medicine and Public Health Preparedness* 8 (3): 267–275.

Buljan, Ivan, Mario Malički, Elizabeth Wager, Livia Puljak, Darko Hren, Frances Kellie, Helen West, Žarko Alfirević, and Ana Marušić. 2018. "No Difference in Knowledge Obtained from Infographic or Plain Language Summary of a Cochrane Systematic Review: Three Randomized Controlled Trials." *Journal of Clinical Epidemiology* 97: 86–94.

Bults, Marloes, Desirée Beaujean, Onno de Zwart, Gerjo Kok, Pepijn van Empelen, Jim van Steenbergen, Jan Richardus, and Hélène Voeten. 2011. "Perceived Risk, Anxiety, and Behavioural Responses of the General Public during the Early Phase of the Influenza A (H1N1) Pandemic in the Netherlands: Results of Three Consecutive Online Surveys." *BMC Public Health* 11(1): no page number.

Carbone, Elena, and Jamie Zoellner. 2012. "Nutrition and Health Literacy: A Systematic Review to Inform Nutrition Research and Practice." *Journal of the Academy of Nutrition and Dietetics* 112 (2): 254–265.

Carter, Merlin. 2018. *Configuring Guidelines that Regulate Paragraph and Sentence Length*. Accessed April 2, 2019. https://bit.ly/2E3lNTe.

Cascales, Remedios Ruiz. 2002. "A Specification and Validating Parser for Simplified Technical Spanish." MSc thesis, *University of Limerick*. Accessed April 2, 2019. https://bit.ly/2JXlH40.

Centers for Disease Control and Prevention. 2016. *Everyday Words for Public Health Communication*. Accessed April 3, 2019. https://bit.ly/2Du1bnK.

Cheek, Annetta. 2010. "Defining Plain Language." *Clarity* 64: 5–15.

Collins-Thompson, Kevyn. 2014. "Computational Assessment of Text Readability: A Survey of Current and Future Research." *International Journal of Applied Linguistics* 165 (2): 97–135.

Cochrane Infectious Diseases. 2019. *WHO Collaborating Centre for Evidence Synthesis for Infectious and Tropical Diseases.* Accessed April 3, 2019. https://bit.ly/2CUkcOe.

Cochrane Norway. 2017. *How to Write a Plain Language Summary of a Cochrane Intervention Review*. Accessed April 3, 2019. https://bit.ly/2BG9FWv.

Crossley, Scott, and Danielle McNamara. 2008. "Assessing L2 Reading Texts at the Intermediate Level: An Approximate Replication of Crossley, Louwerse, McCarthy and McNamara (2007)." *Language Teaching* 41 (3): 409–429.

Crossley, Scott, and Danielle McNamara. 2016. "Text-Based Recall and Extra-Textual Generations Resulting from Simplified and Authentic Texts." *Reading in a Foreign Language* 28 (1): 1–19.

Crossley, Scott, David Allen, and Danielle McNamara. 2012. "Text Simplification and Comprehensible Input: A Case for an Intuitive Approach." *Language Teaching Research* 16 (1): 89–108.

Crossley, Scott, Jerry Greenfield, and Danielle McNamara. 2008. "Assessing Text Readability Using Cognitively Based Indices." *TESOL Quarterly* 42 (3): 475–493.

Cutchin, Malcolm, Kathryn Martin, Steven Owen, and James Goodwin. 2008. "Concern about Petrochemical Health Risk before and after a Refinery Explosion." *Risk Analysis: An International Journal* 28 (3): 589–601.

Diao, Yali, and John Sweller. 2007. "Redundancy in Foreign Language Reading Comprehension Instruction: Concurrent Written and Spoken Presentations." *Learning and Instruction* 17 (1): 78–88.

Dowell, Nia, Arthur Graesser, and Zhiqiang Cai. 2016. "Language and Discourse Analysis with Coh-Metrix: Applications from Educational Material to Learning Environments at Scale." *Journal of Learning Analytics* 3 (3): 72–95.

Dumenci, Levent, Robin Matsuyama, Daniel Riddle, Laura Cartwright, Robert Perera, Harold Chung, and Laura Siminoff. 2014. "Measurement of Cancer Health Literacy and Identification of Patients with Limited Cancer Health Literacy." *Journal of Health Communication* 19(0 2): 205–224.

Glik, Deborah. 2007. "Risk Communication for Public Health Emergencies." *Annual Review of Public Health* 28: 33-54.

Gregorio, Germana, Maria Liza Gonzales, Leonila Dans, and Elizabeth Martinez. 2016. "Polymer-Based Oral Rehydration Solution for Treating Acute Watery Diarrhoea." *Cochrane Database of Systematic Reviews* 12 (CD006519): no page number.

Hansen, Cheryl. 1978. "Story Retelling Used with Average and Learning Disabled Readers as a Measure of Reading Comprehension." *Learning Disability Quarterly* 1 (3): 62–69.

Harper, Raquel, and Don Zimmerman. 2009. "Exploring Plain Language Guidelines." In *Proceedings of the IEEE International Professional Communication Conference (IPCC 2009)*. Accessed April 3, 2019. https://bit.ly/2OwZfeU.

Harvey, Lisa. 2018. "Summaries of Cochrane Systematic Reviews: Making High-Quality Evidence Accessible." *Spinal Cord* 56 (185): no page number.

Helbing, Dirk. 2013. "Globally Networked Risks and How to Respond." *Nature* 497: 51–59.

Higgins, Julian, and Sally Green. (editors) 2011. *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.1.0). Accessed April 3, 2019. https://bit.ly/2t90rNM.

Holmes, Bev. 2008. "Communicating about Emerging Infectious Disease: The Importance of Research." *Health, Risk & Society* 10 (4): 349–360.

Karačić, Jasna, Ivan Buljan, Darko Hren, Pierpaolo Dondi, and Ana Marušić. 2017. "Readability of Different Formats of Information about Cochrane Systematic Reviews: A Cross Sectional Study." In *Abstracts of the Ninth Croatian Cochrane Symposium*. Accessed April 3, 2019. https://bit.ly/2ShUmJa.

Keselman, Alla, Laura Slaughter, and Vimla Patel. 2005. "Toward a Framework for Understanding Lay Public's Comprehension of Disaster and Bioterrorism Information." *Journal of Biomedical Informatics* 38 (4): 331–344.

Kintsch, Walter. 1998. *Comprehension: A Paradigm for Cognition.* Cambridge, United Kingdom: Cambridge University Press.

Kools, Marieke, Robert Ruiter, Margaretha Van de Wiel, and Gerjo Kok. 2004. "Increasing Readers' Comprehension of Health Education Brochures: A Qualitative Study into How Professional Writers Make Texts Coherent." *Health Education & Behavior* 31 (6): 720–740.

Kuhn, Tobias. 2014. "A Survey and Classification of Controlled Natural Languages." *Computational Linguistics* 40 (1): 121–170.

Kurtzman, Ellen, and Jessica Greene. 2016. "Effective Presentation of Health Care Performance Information for Consumer Decision Making: A Systematic Review." *Patient Education and Counseling* 99 (1): 36–43.

Lachance, Christina, Lori Erby, Beth Ford, Vincent Allen, and Kimberly Kaphingst. 2010. "Informational Content, Literacy Demands, and Usability of Websites Offering Health-Related Genetic Tests Directly to Consumers." *Genetics in Medicine* 12 (5): 304–312.

Lazar, Jonathan, Jinjuan Feng, and Harry Hochheiser. 2010. *Research Methods in Human-Computer Interaction*. Chichester, England: John Wiley & Sons Ltd.

Leroy, Gondy, David Kauchak, and Obay Mouradi. 2013. "A User-Study Measuring the Effects of Lexical Simplification and Coherence Enhancement on Perceived and Actual Text Difficulty." *International Journal of Medical Informatics* 82 (8): 717–730.

Levine, Timothy, and Craig Hullett. 2002. “Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research.” *Human Communication Research* 28 (4): 612–625.

Liu, Chiung-Ju, and Susan Rawl. 2012. “Effects of Text Cohesion on Comprehension and Retention of Colorectal Cancer Screening Information: A Preliminary Study.” *Journal of Health Communication* 17: 222–240.

Maguire, Lisa, and Mike Clarke. 2014. “How Much do You Need: A Randomised Experiment of whether Readers Can Understand the Key Messages from Summaries of Cochrane Reviews without Reading the Full Review.” *Journal of the Royal Society of Medicine* 107 (11): 444–449.

Mc Laughlin, Harry. 1969. “SMOG Grading - A New Readability Formula.” *Journal of Reading* 12 (8): 639–646.

McNamara, Danielle, Arthur Graesser, Philip McCarthy, and Zhiqiang Cai. 2014. *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge, MA: Cambridge University Press.

McNamara, Danielle, Arthur Graesser, Zhiqiang Cai, and Jonna Kulikowich. 2011. “Coh-Metrix Easability Components: Aligning Text Difficulty with Theories of Text Comprehension.” In *Annual Meeting of the American Educational Research Association*. Accessed April 3, 2019. https://bit.ly/2Rp2Tdo.

McNamara, Danielle, Yasuhiro Ozuru, and Randy Floyd. 2011. “Comprehension Challenges in the Fourth Grade: The Roles of Text Cohesion, Text Genre, and Readers’ Prior Knowledge.” *International Electronic Journal of Elementary Education* 4 (1): 229–257.

Mhimbira, Francis, Luis Cuevas, Russell Dacombe, Abdallah Mkopi, and David Sinclair. 2017. “Interventions to Increase Tuberculosis Case Detection at Primary Healthcare or Community-Level Services.” *Cochrane Database of Systematic Reviews* 11 (CD011432): no page numbers.

Mićić, Sofija. 2013. “Languages of Medicine — Present and Future.” *JAHR* 4 (7): 217–233.

Muñoz-Miquel, Ana. 2012. "From the Original Article to the Summary for Patients: Reformulation Procedures in Intralingual Translation." *Linguistica Antverpiensia, New Series–Themes in Translation Studies* 11: no page number.

O'Brien, Sharon, and Patrick Cadwell. 2017. "Translation Facilitates Comprehension of Health-Related Crisis Information: Kenya as an Example." *The Journal of Specialised Translation* 28: 23-51.

Ozuru, Yasuhiro, Kyle Dempsey, and Danielle McNamara. 2009. "Prior Knowledge, Reading Skill, and Text Cohesion in the Comprehension of Science Texts." *Learning and Instruction* 19 (3): 228–242.

Parker, Ruth, David Baker, Mark Williams, and Joanne Nurss. 1995. "The Test of Functional Health Literacy in Adults." *Journal of General Internal Medicine* 10 (10): 537–541.

Patel, Vimla, and David Kaufman. 2006. "Cognitive Science and Biomedical Informatics." In *Biomedical Informatics: Computer Applications in Health Care and Biomedicine,* 3rd edition, edited by Edward Shortliffe, and James Cimino, 133–185. New York: Springer.

Redish, Janice. 2000. "What Is Information Design?". *Technical Communication* (47) 2: 163–166.

Reed, Deborah, and Sharon Vaughn. 2012. "Retell as an Indicator of Reading Comprehension." *Scientific Studies of Reading* 16 (3): 187–217.

Reuther, Ursula, and Antje Schmidt-Wigger. 2000. "Designing a Multi-Purpose CL Application." In *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW 2000).* Accessed April 3, 2019. https://bit.ly/2QtPEKU.

Reynolds, Barbara, and Matthew Seeger. 2007. "Crisis and Emergency Risk Communication as an Integrative Model." *Journal of Health Communication* 10 (1): 43–55.

Richards, Todd, Stephen Peverly, Amie Wolf, Robert Abbott, Steven Tanimoto, Rob Thompson, William Nagy, and Virginia Berninger. 2016. "Idea Units in Notes and Summaries for Read Texts by Keyboard and Pencil in Middle Childhood Students with Specific Learning Disabilities: Cognitive and Brain Findings. *Trends in Neuroscience and Education* 5 (3): 146–155.

Rodríguez Vázquez , Silvia. 2016. “Assuring Accessibility during Web Localisation: An Empirical Investigation on the Achievement of Appropriate Text Alternatives for Images.” PhD diss., University of Geneva.

Roediger, Henry, and Elizabeth Marsh. 2005. “The Positive and Negative Consequences of Multiple-Choice Testing.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31 (5): 1155–1159.

Rusko, Elina, Karel Van der Waarde, and Raija-Liisa Heiniö. 2012. “Challenges to Read and Understand Information on Pharmaceutical Packages.” In *Proceedings of the 18th IAPRI World Packaging Conference,* edited by Jay Singh, 79–85. Lancaster, Pennsylvania: DEStech Publication.

Santesso, Nancy, Tamara Rader, Elin Nilsen, Claire Glenton, Sarah Rosenbaum, Agustín Ciapponi, Lorenzo Moja, Jordi Pardo, Qi Zhou, and Holger Schünemann. 2015. “A Summary to Communicate Evidence from Systematic Reviews to the Public Improved Understanding and Accessibility of Information: A Randomized Controlled Trial.” *Journal of Clinical Epidemiology* 68 (2): 182–190.

Schriver, Karen. 2017. “Plain Language in the US Gains Momentum: 1940-2015.” *IEEE Transactions on Professional Communication* 60 (4): 343–366.

Schwitter, Rolf. 2015. “Controlled Language.” In T*he Routledge Encyclopedia of Translation Technology*, edited by Chan Sin-Wai, 450–464. London and New York: Routledge.

Shapiro, Amy. 2004. “How Including Prior Knowledge as a Subject Variable May Change Outcomes of Learning Research.” *American Educational Research Journal* 41 (1): 159–189.

Shardlow, Matthew. 2014. “A Survey of Automated Text Simplification.” *International Journal of Advanced Computer Science and Applications, Special Issue on Natural Language Processing* 4 (1): 58–70.

Simonds, Vanessa, Eva Garroutte, and Debra Buchwald. 2017. “Health Literacy and Informed Consent Materials: Designed for Documentation, Not Comprehension of Health Research.” *Journal of Health Communication* 22 (8): 682–691.

Smith, Catherine, Scott Hetzel, Prudence Dalrymple, Alla Keselman. 2011. "Beyond Readability: Investigating Coherence of Clinical Text for Consumers." *Journal of Medical Internet Research* 13 (4): no page numbers.

Sørensen, Kristine, Stephan Van den Broucke, James Fullam, Gerardine Doyle, Jürgen Pelikan, Zofia Slonska, and Helmut Brand. 2012. "Health Literacy and Public Health: A Systematic Review and Integration of Definitions and Models." *BMC Public Health* 12 (80): no page number.

The Cochrane Collaboration 2013. *Standards for the Reporting of Plain Language Summaries in New Cochrane Intervention Reviews (PLEACS).* Accessed April 3, 2019. https://bit.ly/2QsT2pv.

The Cochrane Collaboration. 2016. *Cochrane Style Manual.* Accessed April 3, 2019. https://bit.ly/2re0wP8.

U.S. Department of Health and Human Services. 2018. *Plain Writing Act Compliance Report*. Accessed April 3, 2019. https://bit.ly/2SWD2dJ.

Vickstrom, Erik, Hyon Shin, Sonia Collazo, and Kurt Bauman. 2015. *How Well—Still Good? Assessing the Validity of the American Community Survey English-Ability Question*. Accessed April 2, 2019. https://bit.ly/2vv00Az.

Warde, Fiona, Janet Papadakos, Tina Papadakos, Danielle Rodin, Mohammad Salhia, and Meredith Giuliani. 2018. "Plain Language Communication as a Priority Competency for Medical Professionals in a Globalized World." *Canadian Medical Education Journal* 9 (2): e52–e59.

World Health Organization. 2019. *Communicating for Health.* Accessed April 3, 2019. https://bit.ly/2HVeCPG.

Yoshii, Hatsumi, Hidemitsu Saito, Saya Kikuchi, Takashi Ueno, and Kineko Sato. 2014. "Report on Maternal Anxiety 16 Months after the Great East Japan Earthquake Disaster: Anxiety over Radioactivity." *Global Journal of Health Science* 6 (6): 1–10.

[1] In this chapter we use the terms 'health/health-related' and 'medicine/medical' interchangeably.

2 We define a lay person as a person having 'only common sense or everyday knowledge of a domain' (Patel and Kaufman, 2006, p. 152).

3 For the purposes of this paper, we adopt the following definition of health literacy, from Berkman *et al.* (2010, p. 16): '[t]he degree to which individuals can obtain, process, understand, and communicate about health-related information needed to make informed health decisions'.

4 The Cochrane Library website is available at: https://bit.ly/2CrMUX3 (Accessed March 5, 2019).

5 A controlled language is 'a constructed language that is based on a certain natural language, being more restrictive concerning lexicon, syntax, and/or semantics, while preserving most of its natural properties' (Kuhn, 2014, p. 123). Plain language is regarded as a type of controlled language (ibid.). The name assigned by the Acrolinx company to their software is *content optimisation software*/*platform.* However, in line with scholarly tradition (e.g. Rodríguez Vázquez 2016), we call this tool *CL checker*.

6 Coh-Metrix T.E.R.A is available at: http://tea.cohmetrix.com (Accessed March 7, 2019).

7 Coh-Metrix 3.0 is available at: http://tool.cohmetrix.com (Accessed March 7, 2019).

8 Friedman tests were conducted for cued recall scores, which did not meet the assumption of normality according to the Shapiro-Wilk test ($p<0.05$).