When synonyms are not enough: Optimal parenthetical insertion for text simplification

Article *in* AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium · April 2018

CITATION
1

READS
83

3 authors:

Yang Gu
The University of Arizona
9 PUBLICATIONS **60** CITATIONS
SEE PROFILE

Gondy Leroy
The University of Arizona
165 PUBLICATIONS **2,535** CITATIONS
SEE PROFILE

David Kauchak
Middlebury College
48 PUBLICATIONS **1,338** CITATIONS
SEE PROFILE

When synonyms are not enough: Optimal parenthetical insertion for text simplification

Yang Gu[1], Gondy Leroy, PhD MS[1], David Kauchak, PhD[2]
[1]University of Arizona, Tucson, AZ; [2]Pomona College, Claremont, CA

Abstract

As more patients use the Internet to answer health-related queries, simplifying medical information is becoming increasingly important. To simplify medical terms when synonyms are unavailable, we must add multi-word explanations. Following a data-driven approach, we conducted two user studies to determine the best formulation for adding explanatory content as parenthetical expressions. Study 1 focused on text with a single difficult term (N=260). We examined the effects of different types of text, types of content in parentheses, difficulty of the explanatory content, and position of the term in the sentence on actual difficulty, perceived difficulty, and reading time. We found significant support that enclosing the difficult term in parentheses is best for difficult text and enclosing the explanation in parentheses is best for simple text. Study 2 (N=116) focused on lists with multiple difficult terms. The same interaction is present although statistically insignificant, but parenthetical insertion can still significantly simplify text.

Introduction

Text-based information plays an important role in patient education. Doctors send leaflets and instructions home with patients so they can adequately care for themselves. More than half of American Internet users independently search for health-related information online[1]. Therefore, it is essential that medical information be written in language simple enough to be correctly interpreted by patients. Changing the words and grammar of a text can achieve text simplification, and appropriate modifications will enhance readability and comprehension.

In our prior research, we found that term familiarity, as approximated by term frequency, is a reliable indicator of text difficulty[2]. Term familiarity can identify and rank terms by difficulty. By substituting familiar synonyms for difficult terms, we were able to decrease both actual and perceived difficulty. However, many medical concepts do not have simple synonyms. For example, a single word is not enough to understand "reactogenecity"; the reader needs to make the connection between three concepts: a medical product, its adverse reaction, and the anticipated nature of the adverse reaction. For such concepts, we must insert an explanation in the text. Enclosing the new text in parentheses is one of the most natural ways to insert relevant but discontinuous information into text. Previous research has found that adding explanatory content enclosed in parentheses simplifies text[3, 4]; however, the best way to formulate and insert the new information has not been investigated.

This research contributes to a semi-automated text simplification system for writers, which will automatically identify difficult text and suggest options for simplification. The human writer can choose to accept, refine or reject system suggestions to create coherent text. We examine four factors that may affect simplification using parentheticals: the difficulty of the original text, whether to place a term or an explanation in the parentheses, the difficulty of the explanatory content, and the position of the difficult term in the sentence. We conducted one study to investigate those four factors in sentences that contain a single difficult term. In a second study, we examine the extent to which parallelism in the formulation of parenthetical expressions will impact simplification of texts with multiple difficult terms in a list.

Background

Useful Existing Models to Guide Simplification

The fields of psychology and linguistics have explored the reading and discourse comprehension process, and there are many models of text comprehension. Some of the more well-known models include the Construction-Integration Model[5], the Landscape Model[6], and the Resonance Model[7]. Every model differs in its assumptions and computational model, but they agree on how the comprehension process works on a high level. During reading, the text provides the reader with cues, such as words and punctuation. Readers search their long-term memory to give meanings to these cues and construct propositions. Comprehension occurs when the reader is able to make connections between the propositions. In text simplification, we can facilitate comprehension by modifying the cues, such as words and syntactic markers, which will help readers make connections.

Explaining Difficult Concepts

Discovering and collecting a consumer-friendly medical vocabulary has been an ongoing effort in the health and medical informatics community[8]. For example, there have been several efforts to collect vocabulary through web mining or automatically generating simple explanations based on relationships in the UMLS Metathesaurus[3, 9, 10].

Synonym replacement is an intuitive, effective, and popular approach for lexical simplification, but it is not always possible. When no synonyms are available, explanations need to be inserted. Kandula et al. made the distinction between "definitions" and "explanations"[3] for terms. A "definition" is a description of the terminology that aims to be correct, accurate, and precise. For example, the Wikipedia page for "reactogenecity" provides a definition from the NIH (National Institute of Health), which states *"Reactogenicity events are adverse events that are common and known to occur for the intervention/investigational product being studied...*[11]*"* An "explanation" is a description that is more abstract but also more understandable to the reader. It should not introduce more difficult terminology. In a sentence about anthrax vaccines, we can explain "reactogenecity" as "*known side effects from vaccines*."

Some automated text simplification systems employ explanation generation[12]. For example, Eom, Dickinson, and Sachs designed a system for second language learners that allows users to upload text in one panel and show the sense-specific definition for a selected vocabulary in a side panel[13]. Damay et al.'s SimText system first uses a thesaurus to look for synonym replacements for difficult terms and then appends a definition when synonyms are unavailable[14]. Kandula et al.'s system specializes in the medical domain and also first searches for consumer-friendly synonyms to medical terms in the Open Access and Collaborative Consumer Health Vocabulary[3]. When that is unavailable, they generated simple explanations using high-level relationships in the UMLS[3].

Simplification through Parenthetical Expressions

A natural way to introduce new content into text is with parenthetical expressions. A parenthetical expression is any expression embedded in a host expression that makes no contribution to the structure of the host and makes the host expression discontinuous[15]. In written text, the expression can be enclosed using a variety of punctuation characters, such as brackets, dashes, or commas. For example, consider "He told John – his best friend from college – all about it." While it introduces more information, the parenthetical expression also disrupts the structure of the original sentence. We are interested in appositive parenthetical expressions, which reformulate parts of the original host expression[15]. The reformulation provides additional information about the concept in the host expression. In the context of text simplification, we only consider appositive parenthetical expressions enclosed by literal parentheses in written texts.

We have two options for formulating parenthetical expressions to simplify difficult words. The common approach is to insert the explanatory content as a parenthetical expression behind the target term, as in "*local and general reactogenicity (known side effects from vaccines) are expected*." This can be easily automated. The reader expects the parentheses to signal an interruption to the flow and structure of the entire sentence, so the language of the explanatory content need not blend perfectly with the host expression. However, the reader needs to read the parenthetical expression to understand the concept.

Alternatively, we can incorporate the explanatory content into the text and state the obscure medical term in a parenthetical expression, as in "*local and general known side effects from vaccines (reactogenicity) are expected*." In this formulation, it becomes apparent immediately after the modifier clause that the clause describes the side effects of medicine. The reader can comprehend this information faster with less strain on their memory. Most reading comprehension models agree that readers keep snippets of information in their working memory, which has an approximate limit of seven plus/minus two units of information[16]. Therefore, making information comprehensible as early as possible, and avoiding long interruptions and divergences in the text that must be stored in working memory, can be advantageous for comprehension. However, directly inserting definitions into text will create cohesion issues, especially when definitions are long, like those of many medical terms[14].

Because adding definitions within parentheses is widely used in written English, this formulation has been adopted as the default for parenthetical insertions[3]. Nonetheless, there is little empirical support that one formulation is superior to the other.

Study Objectives

The goal of our studies is to discover combinations of features that would influence and optimize simplification when parenthetical insertion is needed. The first study focuses on simplifying sentences with one difficult "target" term. We investigate the impact of the difficulty of the source material, the placement of the parentheses, the position

in the sentence, and the difficulty of the explanatory content. The second study focuses on simplifying sentences with multiple target terms in a single list. We investigate the impact of the difficulty of the source material and the parallelism of the parentheses. The target terms always remain in the text, so readers can learn the difficult vocabularies in their medical texts.

Methods

We conducted two user studies on the effects of parenthetical insertion on simplification. The subjects were asked to read short passages and answer questions about the actual and perceived difficulty of the text. Subjects were recruited on Amazon Mechanical Turk (MTurk), an online service that allows human workers to complete small tasks for monetary compensation. We restricted the workers to being in the United States with 95% approval ratings on previous tasks. MTurk has been used for tasks ranging from survey completion, data annotation or tagging, to user studies; study results have been shown to be as reliable as those from traditional approaches[17, 18].

Study Instrument. Each study used 16 passages identified from 16 documents. Eight documents are Wikipedia articles about medical topics and eight are PubMed abstracts of studies on various diseases. To select the passages, sentences were randomly selected and manually checked for suitability based on the following criteria:

1. The sentence does not contain difficult terms other than the target term, which we will augment with parentheses. The target term can be defined sufficiently in one sentence or less, and a reasonable, simple explanation can be created based on the target term.
2. The sentence can be edited to satisfy all experiment conditions, i.e., the definition and explanation for the term can be incorporated into the text of the sentence with minimal editing. Study 1 also requires that the target term can be moved to the end of the sentence with only minor changes to the words in the sentence.
3. The sentence is preceded and followed by one or two sentences from the same document that provide context for the passage. The context does not contain difficult terms or difficult terms can be removed.

The following is a passage from the Wikipedia article on Anthrax with the target word "reactogenicity":

> *"Vaccines against anthrax for use in livestock and humans have had a prominent place in the history of medicine. The French scientist Louis [....] All currently used anthrax vaccines show considerable local and general reactogenicity (erythema, induration, soreness, fever) and serious adverse reactions occur in about 1% of recipients."*

We removed or simplified other difficult words besides the target word and split up sentences to make the context easier to understand. We also ensured that we did not introduce misinformation about the topic in our modifications. The edited passage used in our study is:

> *"Vaccines against anthrax in humans and livestock have had a prominent place in the history of medicine. Considerable local and general reactogenicity are expected from all currently used anthrax vaccines for humans. Common reactions include soreness and fever, and serious adverse reactions occur in about 1% of recipients."*

We conducted all statistical analyses with SPSS Statistics 24 with standard settings.

Study 1: Single Parenthetical Insertion

Design

This study focuses on simplifying sentences with one difficult target term. We use a 2x2x2x2 full-factorial within-subjects experimental design with the following factors:

1. Types of text: Wikipedia and PubMed. We use the source of the text as a proxy for difficulty. The Wikipedia articles represent easier text, while the PubMed research abstracts represent more difficult text. Wikipedia articles may contain difficult vocabulary and content, but they are meant to be descriptive and informative. Research articles have a different writing style and assumption of the reader's knowledge, but they are also an important resource for expert patients and caregivers. We want to investigate how localized simplification that targets one single difficult concept may impact a generally difficult piece of text differently from a text that is only difficult at that single point.

2. *Placement of Explanatory Content: Inside Parentheses and Outside Parentheses.* Explanatory content placed inside parentheses is commonly used and easy to automate, but the alternative is less disruptive to the flow of the sentence and the reading process. Table 1 provides an example of how factors 2, 3, and 4 combine to form sentences.

3. Difficulty of the Explanatory Content: Simple (explanation) and Difficult (definition). Definitions can be extracted from established domain resources and provide more comprehensive information. Explanations tend to contain simpler language. For each target term, we create a "definition" using definitions from the term's Wikipedia page or from a dictionary. Then, the definition is shortened and simplified manually to create an explanation. We did not use a single resource such as UMLS or Wikipedia for definitions because some terms are not defined concisely or understandably to a layperson in any one place. Since our system is designed for writers, they can intervene to refine existing definitions.

4. Position in Sentence: Middle and End. Since parenthetical expressions introduce an interruption in the flow of the text, we consider where this interruption occurs. If we insert the parenthetical expression at the end of the sentence, the interruption may be less disruptive than if we insert it in the middle.

Table 1. Demonstration of Study 1 Experiment Conditions

Placement of Explanatory Content	Difficulty of Explanatory Content	Position in Sentence	Sentence
Inside Parentheses	Simple (Explanation)	Middle	Considerable local and general reactogenicity (known side effects from vaccines) are expected from all currently used anthrax vaccines for humans.
		End	All currently used anthrax vaccines for humans are expected to produce considerable local and general reactogenicity (known side effects from vaccines).
	Difficult (Definition)	Middle	Considerable local and general reactogenicity (adverse events that are common and known to occur for a medical intervention or investigational product) are expected from all currently used anthrax vaccines for humans.
		End	All currently used anthrax vaccines for humans are expected to produce considerable local and general reactogenicity (adverse events that are common and known to occur for a medical intervention or investigational product).
Outside Parentheses	Simple	Middle	Considerable local and general known side effects from vaccines (reactogenicity) are expected from all currently used anthrax vaccines for humans.
		End	All currently used anthrax vaccines for humans are expected to produce considerable local and general known side effects from vaccines (reactogenicity).
	Difficult	Middle	Considerable local and general adverse events that are common and known to occur for a medical intervention or investigational product (reactogenicity) are expected from all currently used anthrax vaccines for humans.
		End	All currently used anthrax vaccines for humans are expected to produce considerable local and general adverse events that are common and known to occur for a medical intervention or investigational product (reactogenicity).

We evaluate simplification using the following metrics:

Actual Difficulty. Average accuracy on three multiple-choice questions about the content of the passage, which the subjects answer without the text present. One question specifically focuses on the meaning of the target term, which is most directly related to the parenthetical expression.

Perceived Difficulty. A 5-point Likert Scale (1 = Very Easy, 5 = Very Difficult) response to "How difficult would this passage look in a text" immediately after reading.

Reading Time. Time (in seconds) the respondents spent on the page containing the passage.

Procedure

Subjects recruited on MTurk were redirected to a survey on Qualtrics. We used Qualtrics' randomization features to assign 8 passages to each subject. The subjects do not see more than one version of any single passage.

After reading each passage, subjects were taken to a new page with six questions. The first question asked about perceived difficulty. Then there were three multiple-choice questions about the content of the passage and an attention question. The final question always asks about the subject's prior knowledge or experience with the condition: "How familiar are you with the topic of [topic of passage]." Response choices were: "I have never heard of it before," "I have only heard of it in passing," "I have studied it in detail for educational purposes," "I know of someone with this condition," or "I have (or have had) this condition." We interpret scores of 1 or 2 (first two

options) as minimal or no knowledge of the condition and scores of 3 to 5 (last three options) as high knowledge, and use this metric to control for the impact of background knowledge on the results.

We also use an attention question to filter out respondents who have not read the passage. For example, for the earlier passage about Anthrax vaccines, we asked "Which word appeared in the passage?" To someone who read the passage, the correct answer is obvious among the choices: Anthrax, Arthur, Anthropology, and Antonym.

Results

We invited 56 MTurk workers to read and answer questions on 8 passages for compensation of $3 USD. We removed three subjects for having missed two or more attention questions and six subjects for spending an unreasonably long or short amount of time reading or answering the questions. Prior knowledge of the topic is significantly correlated with reading time ($r = 0.102$, $p < 0.046$), perceived difficulty ($r = -0.212$, $p < 0.000$), and actual difficulty ($r = 0.1$, $p < 0.038$), so we removed all responses from someone with high knowledge of the medical condition.

Table 2 summarizes the demographic characteristics of subjects whose responses were included in the analysis. The majority are under 40 years old. More than half have a bachelor's degree and about one-quarter have a high school diploma. About three-quarters are white and most of the respondents speak only English at home. Females make up just above half of our respondents.

Table 2. Demographics of Study 1 Subjects (N=47)

Age	**Count (%)**
<30	14 (0.298)
31 to 40	16 (0.340)
41 to 50	9 (0.192)
51 to 60	8 (0.170)
61 to 70	0 (0.0)
>71	0 (0.0)
Gender	
Female	27 (0.574)
Male	20 (0.426)
Race (multiple answers allowed)	**(N=49)**
American Indian/Alaska Native	1 (0.020)
Asian	7 (0.143)
Black	2 (0.041)
Native Hawaiian/Pacific Islander	1 (0.020)
White	38 (0.776)
More than one race	1
Education	
Less than high school	0 (0.0)
High school diploma	13 (0.277)
Associate's degree	6 (0.128)
Bachelor's degree	26 (0.553)
Masters	1 (0.021)
Doctorate	1 (0.021)
Language Spoken at Home	
Never/Rarely English	0 (0.0)
Half English	1 (0.021)
Mostly English	2 (0.043)
Only English	44 (0.936)

Table 3. Means for Actual Difficulty, Perceived Difficulty, and Reading Time for each Experimental Condition

Source of Passage	Placement of Explanatory Content	Difficulty of Explanatory Content	Position in Sentence	N	Actual Difficulty (all)		Actual Difficulty (term)		Perceived Difficulty		Reading Time	
					Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Wikipedia	**Baseline**	**-**	**-**	**13**	**0.641**	**0.214**	**0.462**	**0.519**	**2.923**	**1.115**	**115.700**	**169.519**
	Inside Parentheses	Simple (explanation)	Middle	16	0.792	0.269	0.688	0.479	2.875	0.957	53.977	54.054
			End	15	0.822	0.213	0.733	0.458	2.733	1.163	56.160	31.194
		Difficult (Definition)	Middle	14	0.714	0.410	0.786	0.426	3.071	0.997	59.401	54.057
			End	17	0.667	0.312	0.588	0.507	3.529	1.179	56.022	46.248
	Outside parentheses	Simple	Middle	17	0.627	0.309	0.647	0.493	3.176	1.131	130.004	385.895
			End	18	0.667	0.280	0.833	0.383	3.000	1.372	44.541	42.111
		Difficult	Middle	18	0.611	0.308	0.556	0.511	3.389	1.195	47.502	34.802
			End	14	0.690	0.243	0.429	0.514	2.786	1.188	55.126	50.646
PubMed	**Baseline**	**-**	**-**	**15**	**0.556**	**0.371**	**0.600**	**0.507**	**3.067**	**0.961**	**43.417**	**52.232**
	Inside Parentheses	Simple	Middle	16	0.750	0.285	0.625	0.500	3.063	1.063	51.646	49.439
			End	19	0.684	0.304	0.684	0.478	2.895	1.100	48.933	25.946
		Difficult	Middle	16	0.500	0.344	0.438	0.512	2.938	1.124	88.809	89.867
			End	18	0.500	0.383	0.500	0.514	3.111	1.278	53.344	39.877
	Outside Parentheses	Simple	Middle	13	0.744	0.277	0.615	0.506	2.692	1.251	84.045	105.127
			End	18	0.685	0.267	0.667	0.485	2.444	0.856	51.518	44.647
		Difficult	Middle	17	0.706	0.232	0.647	0.493	3.176	1.074	53.871	44.675
			End	14	0.619	0.288	0.571	0.514	3.143	0.949	48.320	22.109

Table 3 summarizes the means for all experimental conditions. Actual difficulty is measured by accuracy, so lower scores denote more difficult texts. Consistent with conventional wisdom, inserting any explanatory material in any formulation reduces actual difficulty compared to the baseline condition, with a few exceptions for when difficult definitions are inserted into the text. For Wikipedia texts, the improvement from baseline is the largest when we insert simple explanations in parentheses at the end of the sentence ($t_{26} = 2.241$, $p < 0.034$). Other configurations do show improvement as well, though not statistically significant, though this may be due to the small number of responses we have in each condition.

When we aggregate data based on features of interest, we do observe significant interactions, though, some parenthetical formulations are more successful at simplification than others. The best formulation for Wikipedia passages (simple explanation inside parentheses at the end of sentences) yields significantly higher accuracy (i.e. easier documents) than the worst formulation (difficult explanation outside parentheses the middle of the sentence; $t_{31} = 2.242$, $p < 0.032$). Similarly, the best configuration for PubMed (explanation inside parentheses in the middle) is significantly different than the worst configuration (definition inside parentheses in the middle; $t_{30} = 2.236$, $p < 0.033$). It is important that we understand which combinations of features can affect the effectiveness of each formulation.

We performed a 4-way ANOVA on the experimental condition for actual difficulty, perceived difficulty, and reading time (N = 260). We exclude the baseline from this discussion because it does not contain any explanatory content.

Actual Difficulty. We found a main effect for the difficulty of explanatory content on the accuracy of the content questions ($f_{1,244} = 6.501$, $p < 0.011$). Readers presented with a technical definition correctly answered 62.6% of the content questions, while readers presented with the explanation achieved higher accuracy and were able to correctly answer 72.1% of the time. The difficulty of the explanatory content also has a significant effect ($f_{1,244} = 4.056$, $p < 0.045$) on the accuracy of the questions focusing on the meaning of the target term. Readers presented with a technical definition were able to correctly answer the question about meaning of the target term 56.4% of the time, while readers presented with the explanation were able to correctly answer 68.7% of the time.

Figure 1.A shows the significant interaction between placement of the parentheses and the source material of the text ($f_{1,244} = 5.759$, $p < 0.017$). If the text is from Wikipedia, adding explanatory content inside the parentheses (with the difficult term outside the parentheses) results in better accuracy than adding the new content outside the parentheses (with the difficult term inside the parentheses). However, if the text is from PubMed, which is more difficult and less familiar to the general audience, the results are reversed but the magnitude of change is much smaller. This suggests that the best approach for inserting new explanatory content into text depends on the difficulty, or at least type of text. Using the wrong approach could reduce comprehension and increase difficulty.

We also found a nearly statistically significant trend ($f_{1,244} = 3.633$, $p < 0.058$) showing a similar interaction between the placement of parentheses based on type of explanatory content. As seen in Figure1.B, given a simple explanation, adding explanatory content inside parentheses resulted in a high retention of content compared to adding the explanatory content outside. However, given a technical definition, adding it inside parentheses resulted in lower retention compared to keeping the explanatory content outside of parentheses and integrated with the text.

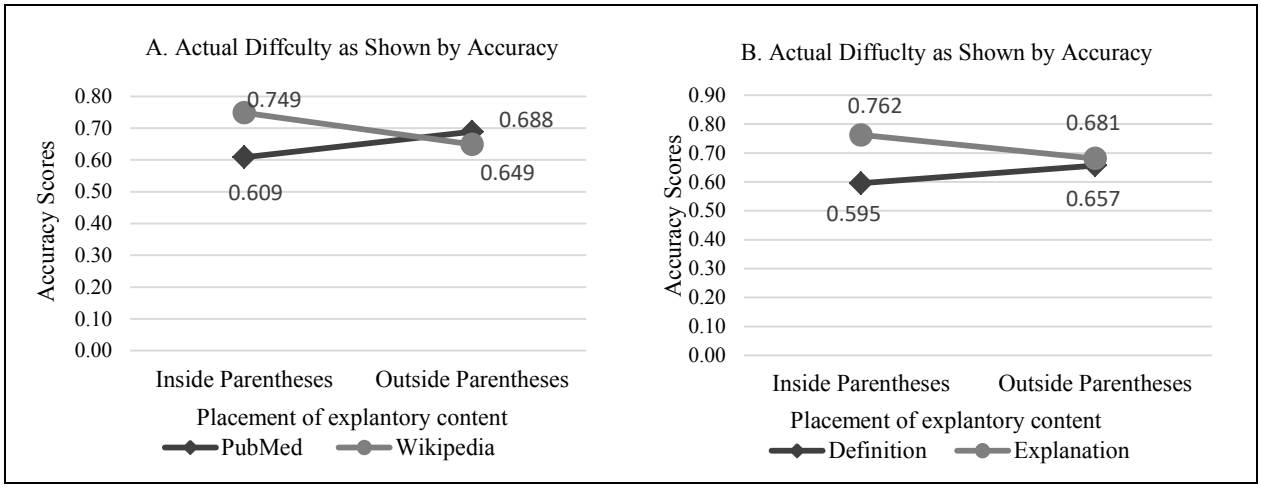

Figure 1. Interaction Effects from Placement of Parentheses and Difficulty of Source or Explanatory Content

Perceived Difficulty. While perceived difficulty is significantly correlated with actual difficulty ($r = -.243$, $p < 0.001$), we do not see a significant difference between individual experiment conditions and the baseline due to the small sample size. However, if we aggregate the data based on specific features, we find that the difficulty of the explanatory content has a significant effect ($f_{1,244} = 4.053$, $p < 0.045$) on perceived difficulty. Readers presented with a technical definition report the average difficulty of the text as 3.143, while readers presented with a simple explanation report a lower difficulty of 2.860.

Reading Time. One concern is that retention of the text material would be influenced by the time spent reading it. In our experiment, we kept wording consistent between conditions as much as possible, only changing the position and placement of the target term and associated vocabulary. We ran correlations between our difficulty scores and reading time. We found that reading time is not significantly correlated with perceived difficulty ($r = -0.027$, $p < 0.666$), term question ($r = 0.083$, $p < 0.185$), and actual difficulty ($r = 0.042$, $p < 0.490$), which rules out reading time as an alternative explanation to the effects we observe in our data.

Table 3 shows two outliers for mean reading time of Wikipedia passages. Three subjects in the baseline condition took about six minutes each. Only the baseline condition had multiple subjects needing more than two minutes. This behavior can be explained by the subjects becoming confused by passages that were lexically simple (Wikipedia text) but incomprehensible without the explanatory content. The other outlier condition is caused by one single subject whose time pattern suggests he/she took a break during the experiment and returned later. This reflected plausible behavior of a subject who was committed to the task, so we retained the data. (We removed all data from subjects who took multiple long breaks). Without this subject, the mean reading time is 36.547 seconds.

Study 2: Multiple Parenthetical Insertions

Design

Study 2 focuses on the cases where there are several target terms that form a list in a sentence to understand whether using parallel parentheses insertion, which would provide structure to the text, simplifies the text. Based on the results from Study 1, position of target terms in the sentence is not considered, and we use only explanations. We also included a baseline condition containing the original sentence. We use a 2x3 full-factorial within-subjects experimental design with the following factors and levels:

1. Types of text: Wikipedia and PubMed. Identical to Study 1.

2. *Parallelism of Parentheses: All Inside, Mismatched, and All Outside.* We expect parallelism to provide more cues about the structure and organization of the information in the passage, which should help with simplification. Table 4 provides an example for each level.

Table 4. Demonstration of Study 2 Experiment Conditions

Parallelism of Parentheses	Sentence
All inside	Heart murmurs are most frequently categorized into systolic (when the heart is contracting) heart murmurs and diastolic (when the heart is expanding) heart murmurs, differing in the part of the heartbeat on which they can be heard.
Mismatched	Heart murmurs are most frequently categorized into systolic (when the heart is contracting) heart murmurs and heart murmurs produced when the heart is expanding (diastolic), differing in the part of the heartbeat on which they can be heard.
All outside	Heart murmurs are most frequently categorized into heart murmurs produced when the heart is contracting (systolic) and heart murmurs produced when the heart is expanding (diastolic), differing in the part of the heartbeat on which they can be heard.

As in Study 1, we evaluate simplification using actual difficulty, perceived difficulty, and reading time using the same scales.

In this study, each subject was assigned 4 passages (since we have fewer conditions).

Results

We recruited 48 workers for our study for $2 each. Similar to Study 1, we removed 6 workers for spending an unreasonably long or short amount of time reading or answering the questions and those with high background knowledge.

The majority of our respondents are under 40 years old (see Table 5). Just over half hold a bachelor's or associate's degree. Over three-quarters are white and most speak only English at home. Females make up just above half of our respondents.

Table 5. Demographics of Study 2 Subjects (N = 42)

Age	**Count (%)**
<30	13 (0.309)
31 to 40	17 (0.405)
41 to 50	9 (0.214)
51 to 60	1 (0.024)
61 to 70	2 (0.048)
>71	0 (0.0)
Gender	
Female	22 (0.524)
Male	20 (0.476)
Race (multiple answers allowed)	**(N=43)**
American Indian/Alaska Native	0 (0.0)
Asian	5 (0.116)
Black	3 (0.070)
Native Hawaiian/Pacific Islander	0 (0.0)
White	35 (0.814)
More than one race	1
Education	
Less than high school	0 (0.0)
High school diploma	12 (0.286)
Associate's degree	9 (0.214)
Bachelor's degree	15 (0.357)
Masters	5 (0.119)
Doctorate	1 (0.024)
Language Spoken at Home	
Never/Rarely English	0 (0.0)
Half English	0 (0.0)
Mostly English	2 (0.048)
Only English	40 (0.952)

Table 6 summarizes the mean values of our major metrics for each experiment condition. The reduction in actual difficulty of Wikipedia text is significant when we insert all explanations in parentheses ($t_{34} = 3.647$, $p < .001$) or insert some explanations in a mismatched manner ($t_{33} = 3.647$, $p < .010$). In general, the baseline condition is most difficult. Mismatched parentheses are an improvement over the baseline, but are still more difficult than any parallel formulation. Curiously, for PubMed passages, inserting explanatory content increased perceived and actual difficulty (decreased content question accuracy). However, results for Wikipedia passages appear consistent with our expectation, though not statistically significant.

We performed 2-way ANOVA for actual difficulty, perceived difficulty, and reading time (N = 116 without baseline condition).

Actual Difficulty. Readers scored much higher on Wikipedia (81.6%) than on PubMed passages (66.4%). The difference in difficulty is significant ($f_{1,92} = 4.222$, $p < 0.043$). No significant effects were observed for term-specific questions.

Contrary to our expectations, actual difficulty is not significantly affected by the misalignment of parentheses. However, the interaction effect between different source material and placement of parentheses is still evident, though not statistically significant. Figure 2 illustrates this. For passages from PubMed, integrating all explanatory content into the text yields higher accuracy scores, but for passages from Wikipedia, keeping explanatory content in parentheses yielded higher accuracy scores.

Perceived Difficulty. PubMed passages are perceived to be more difficult than Wikipedia passages ($f_{1,92} = 7.333$, $p < 0.008$). PubMed passages had an average perceived difficulty of 3.572, while Wikipedia passages' was 3.106.

Reading Time. None of our experiment variables were significantly related to reading time. As in Study 1, reading time did not impact our results and was not correlated with actual difficulty ($r = 0.107$, $p < 0.254$), term comprehension ($r = 0.017$, $p < 0.858$), or perceived difficulty ($r = 0.098$, $p < 0.294$).

Table 6. Means for Actual Difficulty, Perceived Difficulty, and Reading Time for each Experimental Condition

Document Source	Placement of Parentheses	N	Actual Difficulty (all)		Actual Difficulty (term)		Perceived Difficulty		Reading Time	
			Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
Wikipedia	**Baseline**	**15**	**0.556**	**0.272**	**0.600**	**0.507**	**3.133**	**0.990**	**59.293**	**33.517**
	All inside	21	0.873	0.247	0.857	0.359	3.238	1.136	67.541	51.772
	Mismatched	23	0.710	0.307	0.739	0.449	3.130	0.920	41.355	27.415
	All outside	20	0.783	0.224	0.850	0.366	3.150	0.988	40.047	27.922
PubMed	**Baseline**	**17**	**0.686**	**0.322**	**0.647**	**0.493**	**3.529**	**1.007**	**46.319**	**20.155**
	All inside	18	0.593	0.334	0.611	0.502	3.389	1.092	67.296	67.817
	Mismatched	18	0.685	0.242	0.833	0.383	3.222	1.060	66.362	39.256
	All outside	16	0.771	0.315	0.750	0.447	3.813	0.655	67.864	41.717

Although we were able to significantly reduce the actual difficulty of the text using optimal strategies identified in Study 1, Study 2 did not provide conclusive results about the effect of parallel parentheses for simplifying a list of difficult terms. Therefore, in a post hoc analysis, we examined the properties of the lists in our study. Six of the eight Wikipedia passages included lists of similar terms (terms are semantically related, such as "systolic" and "diastolic"), while five of the PubMed passages were lists with dissimilar terms. Parallel parentheses may highlight the similarity between related terms and make reading easier. The significant main effects we observed in Study 2, in which Wikipedia passages are actually and perceived to be easier than PubMed material, may be partially attributed to the similarity between the terms in the lists. However, in this study, similarity is also confounded with the difficulty of the source material.

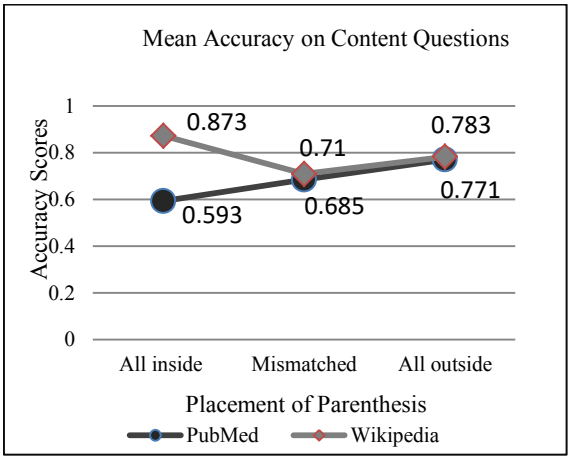

Figure 2. Non-significant Interaction between Parenthesis and Difficulty of text

Limitations

The small scale of the study, with only 16 passages in each experiment, is a limitation. We selected and edited the passages to restrict the number of difficult words and the position in the sentence. This allowed us to precisely control different parameters in the experiment. To draw more generalizable conclusions, future research should use passages that were originally published with difficult terms in positions of interest in the sentence. Additionally, we used Wikipedia articles and PubMed abstracts to represent simple and difficult texts, respectively. However, in addition to lexical difficulty, these texts also differ in writing style, which is a nuanced concept we can't address in the scope of this study.

Conclusion and Lessons Learned

Through our studies, we found that parenthetical expressions can simplify text, but only under the correct conditions. Augmenting the text with simple explanations for difficult terms can reduce the difficulty of the text. However, there exists a significant interaction between difficulty of the text overall and whether to enclose the difficult term or its explanation in the text. Modifying the text without taking these factors into account will be ineffective or even counter-productive to text simplification.

The following text simplification lessons result from our two studies:

1. Difficult content benefited from putting the target term (i.e., difficult term) inside parentheses and incorporating the explanatory content into the text. In Study 1, PubMed passages with the term inside the parentheses had lower actual difficulty, while Wikipedia passages with an explanation in the parentheses had lower actual difficulty. The same pattern is observed in Study 2, though it is not statistically significant. Also in Study 1, in difficult texts, , the texts were easier when a definition (i.e., difficult text) was incorporated in the text rather than in parentheses. A possible explanation is that parentheses provide a clear boundary for snippets of text. When the text is more difficult, the reader becomes unwilling to read the difficult material and will simply skip to the end of the parentheses. By incorporating explanatory content into difficult text, or difficult definitions into any text, readers are encouraged to read the difficult content and benefited from absorbing the information.

2. If we want to add explanatory content to a text, adding a simple explanation helps more than a detailed, technical explanation. Practically, we need to generate more resources with consumer-friendly explanations of medical terminology, which is part of our ongoing research.

3. The position of the target term does not significantly affect actual or perceived difficulty of the text. Even though parenthetical expressions interrupt the flow of a sentence, the location of the interruption is not significant. Its impact on the reader's processing of the sentence is overshadowed by the difficulty of the content.

4. We have preliminary but non-significant evidence suggesting that parallelism in a list may play a role in simplification. We can consider parallelism as the similarity between difficult concepts in a list, which indicates a consistency in the theme or semantic types. Another aspect of parallelism is using parentheses and other cohesion markers in a consistent manner. This is a topic for future research.

5. In general, simplification strategies are inconsistent across different types of text. For example, PubMed passages in Study 2 became more difficult after employing parenthetical insertion in non-optimal configurations. From patients to caregivers to health professionals, the medical domain produces many types of text, ranging from electronic health records to scientific research to general guidebooks. Simplification strategies need to be tailored to the task at hand to be effective.

Acknowledgements

Research reported in this publication was approved by the Institutional Review Board at the University of Arizona, and supported by the National Library of Medicine of the National Institutes of Health, under Award Number R01LM011975. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Hansberry DR, Agarwal N, Shah R, Schmitt PJ, Baredes S, Setzen M, et al. Analysis of the readability of patient education materials from surgical subspecialties. The Laryngoscope. 2014;124(2):405-12.
2. Leroy G, Kauchak D. The effect of word familiarity on actual and perceived text difficulty. Journal of the American Medical Informatics Association. 2014;21(e1):e169-e72.
3. Kandula S, Curtis D, Zeng-Treitler Q. A semantic and syntactic text simplification tool for health content. AMIA Annual Symposium Proceedings. 2010;2010:366-70.
4. Choi YK, Kirchhoff K, Turner AM, editors. Medical Text Simplification by Medical Trainees: A Feasibility Study. 2016 IEEE International Conference on Healthcare Informatics (ICHI); 2016 4-7 Oct. 2016: IEEE.
5. Kintsch W. The role of knowledge in discourse comprehension: a construction-integration model. Psychological review. 1988;95(2):163.
6. Van den Broek P, Risden K, Fletcher CR, Thurlow R. A "landscape" view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. Models of understanding text. 1996:165-87.
7. Myers JL, O'Brien EJ. Accessing the discourse representation during reading. Discourse processes. 1998;26(2-3):131-57.
8. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. Journal of the American Medical Informatics Association. 2006;13(1):24-9.
9. Doing-Harris KM, Zeng-Treitler Q. Computer-assisted update of a consumer health vocabulary through mining of social network data. Journal of medical Internet research. 2011;13(2):e37.
10. Vydiswaran VV, Mei Q, Hanauer DA, Zheng K, editors. Mining consumer health vocabulary from community-generated text. AMIA Annual Symposium Proceedings; 2014: American Medical Informatics Association.
11. Reactogenicity: Wikipedia, The Free Encyclopedia; 2015 [updated 10 September 2015]. Available from: https://en.wikipedia.org/w/index.php?title=Reactogenicity&oldid=680444801.
12. Shardlow M. A survey of automated text simplification. International Journal of Advanced Computer Science and Applications. 2014;4(1).
13. Eom S, Dickinson M, Sachs R, editors. Sense-specific lexical information for reading assistance. Proceedings of the Seventh Workshop on Building Educational Applications Using NLP; 2012; Montreal, Canada: Association for Computational Linguistics.
14. Damay J, Lojico G, Lu K, Tarantan D, Ong E, editors. SIMTEXT: Text Simplification of Medical Literature. Proceedings of the 3rd National Natural Language Processing Symposium-Building Language Tools and Ressources; 2006.
15. Burton-Roberts N. Parentheticals. In: Brown K, editor. Encyclopedia of Language and Linguistics 2ed. Amsterdam: Elsevier; 2006. p. 179-82.
16. Peterson L, Peterson MJ. Short-term retention of individual verbal items. Journal of experimental psychology. 1959;58(3):193.
17. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? Perspectives on psychological science. 2011;6(1):3-5.
18. Paolacci G, Chandler J, Ipeirotis PG. Running experiments on amazon mechanical turk. Judgment and Decision Making. 2010;5(5).