1-1-2010

The Effects of Linguistic Features and Evaluation Perspective on Perceived Difficulty of Medical Text

Gondy Leroy
Claremont Graduate University

Stephen Helmreich
New Mexico State University - Main Campus

James Cowie
New Mexico State University - Main Campus

The Effects of Linguistic Features and Evaluation Perspective on Perceived Difficulty of Medical Text

Gondy Leroy
Claremont Graduate University
Gondy.Leroy@cgu.edu

Stephen Helmreich
New Mexico State University
shelmreich@psl.nmsu.edu

James R. Cowie
New Mexico State University
jcowie@nmsu.edu

Abstract

Millions of laypersons need more medical information than they are customarily provided during their doctor's visit. Health websites can help fill this knowledge gap, but the text is believed to be too difficult to understand for many laypersons. To help write text that is not perceived as too difficult and leads to better comprehension (actual difficulty), we study how linguistic structures influence text difficulty. Since perceived difficulty has been shown to be a barrier to self-education, evaluating perceived difficulty is an essential first step to take. In this study, we evaluated the impact of noun phrase complexity and of function word density in four sentence structures (active, passive, sentential or extraposed subject). Complex noun phrases significantly increased perceived difficulty while using more function words significantly decreased perceived difficulty. Furthermore, laypersons judge text differently when they perform the evaluation on behalf of themselves compared to evaluating on behalf of other readers.

1 Introduction

Millions of people access health-related websites for information and this number will certainly grow. In addition, clinics and hospitals will often send reading materials to the patient's home. Such information has advantages if understood correctly. Foremost among the advantages is that consumers can be more knowledgeable, which empowers many to ask more informed questions when seeing their caregiver and lessens their fear of the unknown [1]. Better informed consumers can also make healthier choices, such as exercising regularly, or avoid problems, such as interactions between medications. There are also indirect benefits when patients need more detailed information than their healthcare provider can give in a limited amount of time. For example, healthcare providers will frequently order 3 or 4 exams without spending adequate time on explaining why they are ordered or what will happen. When patients understand the nature of diagnostic tests being ordered and their importance, it leads to fewer missed appointments which positively affects the operations of clinics and follow up appointments.

However, incorrect or incomplete comprehension of these texts can aggravate health problems instead of resolving them. Particularly among non-native speakers of English, as well as less-educated native speakers, lack of comprehension of healthcare related materials is a large problem. The Committee on Health Literacy for the Council on Scientific Affairs found that misunderstanding health information increases the risk of making unwise health decisions, leading to poorer health and higher healthcare costs [2]. In the U.S., an estimated 89 million people do not have sufficient health literacy to understand treatments or preventive care [3] and costs associated with limited health literacy are estimated to be as high as $50 to $73 billion per year [4].

Many researchers and clinicians alike have looked at improving health literacy by educating patients or by making it easier for them to self-educate. One particularly effective approach would be to facilitate learning from text. Two aspects need to come together for this to happen. The first aspect is that consumers need correct information. Since the Internet is not regulated, there is no guarantee that the information provided is correct and trustworthy (though there are attempts to rate web information [5]). Certainly, consumers should be educated in usage of this information but this is not the problem we address. The second problem, the one we tackle, is that consumers have to be able to learn from the text and comprehend the information they need in their individual situation.

Our goal is to follow a systematic approach to refine, improve, and expand existing research that looks at text readability for today's consumers. We aim to verify the association between readability formula outcome and understanding. To this end, we have started a systematic review of linguistic characteristics and consumer characteristics and their relation to text understanding. We are conducting a series of tests that

measure perceived and actual difficulty of text (an important difference ignored by most), evaluate with representative users and not just experts, and start taking characteristics of intended users into account. Much ongoing research takes only one characteristic into account while others may be as important and even interact with the first. For example, highly educated elderly have different skills and needs than less educated teenagers and focusing on age while disregarding education would be an over simplification of the problem.

We discuss here one aspect of our research agenda. We present the effects on perceived difficulty of simplifying texts based on particular linguistic structures and the influence of different evaluation perspectives.

2 Background

2.1 Health Information

Today, medical information on almost any topic can be found on the Internet. Although the information may not always be correct or objective, it often leads to very positive outcomes when correct information is used toward the education of laypersons. Many laypersons, health information consumers, are willing to learn about diseases and treatments, or how to improve their lifestyles, but this process is hindered by three problems. The first is that there are consumers who do not have access to the Internet. The second problem is that not all information is correct. The third problem is that many consumers do not understand the information they gather online. The first and second problems are not the topic of this paper. We focus on the third problem and our goal is to help laypersons understand medical text.

There are many potential solutions, but to make any solution feasible, it should be cost-effective and scalable. We have proposed in previous work, building on a model by Soergel [6], that an information technology layer can help interpret, understand and personalize information [7]. For example, it could be used to provide help with terminology, to add overviews, summaries or more detailed explanation and images.

According to applications of readability formulas, millions of documents would need to be rewritten because they are currently written at a too high grade level. This problem is difficult to solve and the solution needs to be scalable. Manual rewriting of millions of pages is too expensive, impossible to enforce on the Internet, and still not necessarily a good solution because there is no 'best level' that fits all. Enforcing and verifying compliance would drive up costs even more. On top of everything else, most health professionals are already rushed and do not have the time to rewrite all information. Their staff does not have the linguistic or medical expertise to rewrite the information. An alternative solution, currently adopted by some hospitals, is to buy the materials written at a low grade level. For example, the CareNotes collection, provided by Thomson Reuters, contains about 5,000 documents specifically written for easy comprehension. However, when choosing this solution, clinicians are dependent on the available information and do not have any tools to help with ad hoc communication. Moreover, many clinicians, especially those not part of large hospital systems, would not have the resources to acquire all necessary materials.

2.2 Writing Guidelines and Readability Formulas

Writing guidelines are available for clinicians who write for a lay audience. Most guidelines provide advice on word choice and sentence construction. For example, the use of active voice, short paragraphs and one- or two-syllable words is advised [8]. The guidelines also tend to include referrals to a specific readability grade level to aim for: 6^{th} or 8^{th} grade. The number refers to the grade the reader should have completed in school to understand the text. The readability levels are calculated with fairly simple formulas and most are based on syllable and word counts. The Flesch-Kincaid formulas are available in MS Word. Reviews of online text, including our own [9, 10], reveal that most online text is written at a 10^{th} grade level or higher.

Although the writing guidelines and readability formulas form an excellent start, there are several problems with them. The first problem is that the guidelines treat all patients as having little education or background knowledge. As a result, 6^{th} grade text is recommended regardless of who the patient is: a native English speaker or not, a highly educated person or not, a newly diagnosed patient or an expert patient. For example, patients who have suffered from chronic diseases have often become "expert" patients and are very much up to date on treatments and options. In other cases, patients have complex information needs that are often not met by their healthcare providers [11]. Given the complex nature of many treatments, e.g., genetics-based treatments, oversimplification may dilute the information. A second problem is that even following advice on readability formulas may not simplify the text as intended. It is fairly straightforward to lower readability grade levels by using shorter sentences though this does not guarantee an easier-to-

read text. A final problem is that the available tools and guidelines have not been tuned for today's culture and medical text. For example, in general it may be true that shorter words are easier to understand than longer words, this is not always the case in medicine where words such as "apnea" are difficult, while "diabetes" or "menopause" are easier. Moreover, using many abbreviations would lower the readability grade level while most people find a text full of abbreviations difficult to read.

Tools are needed that are better attuned to readers with different skills. Kim et al [12] developed potential new metrics which may be more sensitive. While the results matched the outcomes from readability formulas well, follow up studies should clarify the effects with representative consumers. Rosemblat et al. [13] worked with experts and included 24 different text characteristics. Four experts evaluated 15 characteristics and the results indicated that only 2 characteristics, vocabulary and the main information in the document, contributed to the experts' decision of suitability of the document for health consumers.

3 The Perceived Difficulty Barrier

We believe it is essential to distinguish between perceived and actual difficulty of text. Although this is not commonly done in medical text readability research, psychological models of human behaviors support this distinction. The Theory of Reasoned Action (TRA) and its extension the Theory of Planned Behavior (TPB), have been put forward as models to explain behaviors and what determines them [14]. The TPB includes perceived behavioral control as an additional factor to the original model. This factor has further been shown to consist of two distinct components: perceived difficulty and perceived control. The distinction was supported by Trafimow et al. [15] in 5 experiments. In a more medical context, Liu et al. [16] evaluated the information search behavior of patients using a questionnaire with 1000 osteoarthritis patients. From their factor analysis, they concluded that perceived difficulty and self-efficacy played a role in drug information seeking behaviors. Depending on the type of optimization used in the analysis, the two loaded on a single or on two factors. Controllability also played a role in perceived difficulty and self-efficacy: with high controllability (as reported by the patients), the dimensions were again distinct.

A second model, the Health Belief Model (HBM), which is better known in medicine, proposes a factor similar to perceived difficulty. The model contains four dimensions: perceived susceptibility, perceived severity, perceived benefits, and perceived barriers. Support for the overall model has been mixed, but a review study in 1984 [17] showed that the perceived-barriers dimension was the most significant of all four in explaining health behavior.

Besides research in medicine, studies focusing on education also support the notion that perceived difficulty matters. When comparing different presentation media, e.g., text only versus text with graphics, Velayo [18] found that both media type and perceived difficulty influenced understanding independently.

Readability research studies provide a snapshot of ongoing cognitive processes. The HBM and TPB can help lead toward a systematic approach in tackling the readability problem. Based on these models, we believe that perceived difficulty of text is a barrier encountered by many consumers who are expected to read text and educate themselves. It is a barrier that can be lowered.

4 Methods

Our goal is to systematically evaluate linguistic features that can be discovered in text during the writing process and that are associated with text difficulty. These features should be more specific than an overall score for a sentence or text. For this reason, we focus on word sequences in a sentence - called features from here on - that can be recognized by parsers.

4.1 Features and Hypotheses

When evaluating the impact of sentence structures, different options exist. We chose to focus on specific sentence structures that are commonly found in online text available to health information consumers. By focusing on a few structures, in contrast to entire paragraphs with many different structures, we believe we can systematically evaluate difficulty.

We chose four different sentence structures to work with. Active and passive sentences are well-known structures that require little explanation. We want to note, however, that writing guidelines advise against the use of passive voice in text written for health information consumers. In additional to these two, we also included sentences where the subject is a sentential subject, for example, Identifying molecules involved in the immunologic response will help … researchers design better … ". Sentential subject sentences are regularly seen in clinical trials documents. The extraposed version can be used to communicate the same information using a different structure, for example "It will help researchers to identify …"

At the phrase level, we evaluate the effect of different noun phrases. Noun phrases are the referential units of sentences. It can be expected that more complex noun phrases make it increasingly difficult to identify the referent and thus understand the sentence. First of all, a difficult noun as the base of the phrase can increase the overall complexity. For example, "apnea" or "islets" are difficult nouns, not because of their length, but because of their relative rarity in common usage. Writing guidelines indirectly address this by advising the use of simple words. Furthermore, the use of compound nouns can increase difficulty. These noun phrases consist of a sequence of nouns that form a larger referring expression, e.g. "diabetes risk" or "common prostate cancer treatment." These structures are often found in medical texts, as opposed to "common treatment of prostate cancer" or "risk of diabetes." The compound phrases pose understanding difficulties both in determining how the nouns are related (does "common" modify "prostate" or "cancer" or "treatment"?) and in the exact relationship between the nouns (does "risk" refer to a risk in contracting diabetes or to the risk of having diabetes?). We hypothesize that sentences with complex noun phrases will be perceived by more people as difficult sentences.

At the word level, we evaluate the effect of function words in a sentence. We use "function words" to refer to prepositions, wh-words, modals, auxiliaries, and determiners, e.g., "of," "what," "should," "be", "a." To the best of our knowledge, function word density, i.e., the proportion of function words in a sentence, has not been evaluated by others for its effects on readability. In earlier work [9], we found that patient blogs contained many more function words than formal documents. The occurrence of function words was twice as high in blogs as in formal documents. In the small, follow-up pilot study [9] laypersons judged sentences to be easier when they contained a higher proportion of function words. A high proportion of function words leads to a different cadence closer to spoken language. It may also help space out individual concepts in text to facilitate assimilation. We hypothesize that sentences with a lower proportion of function words will be perceived by more people as difficult.

Finally, in earlier work, we also noticed differences between laypersons and experts when they evaluated the vocabulary and style of a document [10]. The layperson was asked to judge whether a document was difficult; the expert was asked to judge if a document was difficult for an average health information consumer (layperson). We incorporate an evaluation in this study to verify if judging from different perspectives affects the outcome. We hypothesize there will be a difference between perceived difficulty for the two perspectives – difficulty for oneself and difficulty for others.

4.2 Study Design

Students at community college, undergraduate, and graduate institutions in New Mexico and California were invited to participate. We selected this group because they are representative of consumers who look online for information and include different education levels. Other large consumer groups, such as the aged, will be invited later.

The study reported here was part of a larger study that measured actual difficulty and perceived difficulty of sentences. Basic demographic questions about gender, age, native language, and education were also included. Actual difficulty was measured in a separate, stand-alone module using paragraphs of text and by asking content questions. Perceived difficulty, reported here, was measured by showing sentences and requesting participants to choose the most difficult and easiest version among the different options. We report here on the difficulty scores only since they led to the same conclusions as scores based on choosing the easiest sentences.

The sentences used in this study were selected from online documents available to consumers. Our goal was to study the impact of the overall sentence structure (active, passive, sentential subject, extraposed subject), the noun phrase complexity (high, low) and the function word density (low, high). Ideally, 16 different versions should be shown to participants: four sentence structures, each with two different noun phrases complexities and two function word densities (4x2x2). Unfortunately, showing this many different versions makes it impossible for participants to choose. We noted during pilot studies that the results become nonsensical. Similarly, showing eight different versions was still too confusing. Therefore, practical constraints forced us to show only foursomes of sentences. We therefore constructed 4 versions (noun phrases complexity x function word density) for each sentence structure. Table 1 shows an active sentence with its four versions. Subjects are asked to choose the most difficulty version in each set of 4. Each participant received 4 active sentence examples, 4 passive sentence examples, 4 sentential subject sentence examples, and 4 extraposed subject sentence examples.

Table 1: Active sentence examples

Noun Phrase Complexity	Function Word Density	Example
Simple	Low	Fortunately, changes in personal habits can prevent more damage to arteries supplying the heart.
	High	Fortunately, a few changes in your personal habits can prevent any more damage to the arteries supplying the heart.
Complex	Low	Fortunately, lifestyle changes can prevent further damage to coronary arteries.
	High	Fortunately, a few lifestyle changes can prevent further damage of your coronary arteries.

Sentences were randomized once and this order was used for all subjects and examples. The order used was the same as in Table 1, which is different from the one shown in the results (see Figures 1-4). This constant order may have resulted in stronger effects compared to complete randomization and is a limitation of the study. Follow-up experiments will be conducted with completely randomized ordering. However, the results show that the order itself, e.g., the first or last as the most difficult, is not responsible for these results. The evaluations on behalf of oneself (Self) or on behalf of others (Other) are not affected by this ordering.

5 Results

The study was designed as a within-subjects design to compare variations within each sentence structure (active, passive, sentential subject, extraposed subject). For each structure, we showed four sentence versions, which differed for noun phrase complexity (complex versus simple NP) and the function word density (low and high density). We requested participants to choose from two perspectives: an evaluation on behalf of themselves and one on behalf of other consumers (Self versus Other). To provide a clear overview of the results, we have ordered the conditions in the output (Figures 1-4) according to our hypotheses (not the order shown to users) with easier sentence to the left and more difficult sentences to the right.

To evaluate if sentence versions were perceived as of different difficulty levels, we conducted a repeated-measures ANOVA per sentence structure, with noun phrase complexity and function word density as the independent variables. Repeated-measures was used since each participant viewed each sentence version. The dependent variable is the percentage of participants who selected a particular sentence version as the most difficulty for themselves (Self) or for others (Other). We include partial eta-squared (η^2) information to indicate the proportion of total variability attributable to a factor. Since we are interested in seeing if evaluations differ when judging for oneself (Self) or for others (Other), we conducted paired-samples T-tests for each condition and report the statistically significant differences.

5.1 Demographics

Ninety-seven subjects participated. Eleven did not complete the survey as intended, e.g., did not choose the most difficult sentence in each condition, and their responses were removed from the dataset. Of the remaining 86 participants in the study, 57% were female and 43% male. The average age was 26 years old, with a range between 17 and 72 years old. More than half of the group (59%) had not yet earned a bachelor's degree but 35% had a bachelor's degree or better. Due to recruiting students at local colleges, our sample is slightly more educated than the estimates for the U.S. population based on the U.S. Census 2008 data of those 18 years and older, where only 27% of the population has achieved a bachelor's degree. Several of our participants may yet earn a bachelor's degree.

Table 2: Highest education achieved by participants

Highest Education Level Achieved N = 86	Percentage
High School	6
Some Community College	20
Community College Associate Degree	13
Some College	26
Bachelor's Degree	19
Master's Degree	11
Ph.D. Degree	5

5.2 Active Sentences

Figure 1 shows the results for active sentences. The most difficult versions were the sentences with complex noun phrases and few function words; they were selected by 33% of participants as the most difficult for themselves and by 37% as the most difficult for others. The easiest sentences were those with simple noun phrases and many function words;

only 19% of participants selected this as difficult for themselves and only 11% thought it would be difficult for others. The evaluations on behalf of oneself or of others are very similar.

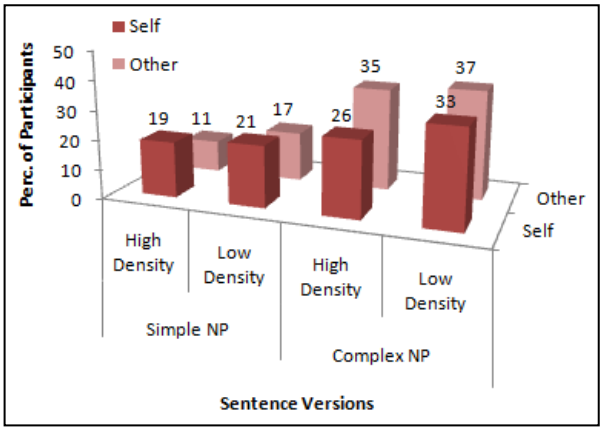

Figure 1: Percentage of participants selecting the most difficult version among active sentences

The repeated-measures ANOVA for Self indicates a significant main effect of noun phrase complexity. When selecting the most difficult sentences for oneself, the sentence with complex noun phrases was chosen more often, ($F(1,85) = 17.277$, $p = <.001$), which explained 17% of the variability ($\eta^2 = .17$).

The repeated-measures ANOVA for Other showed a similar main of effect of noun phrase complexity, ($F(1,85) = 66.908$, $p = <.001$), which explained 44% of the variability ($\eta^2 = .44$).

Paired-samples T-test indicated that in two conditions the differences between Self and Other were significant. The sentences with simple noun phrases and high density function words were more often chosen as difficult for oneself (19%) than for others (11%), $p < .001$. In contrast, the sentences with complex noun phrases and high function word density were less often chosen as the most difficult for oneself (26%) than for others (35%), $p < .005$.

5.3 Passive Sentences

Figure 2 shows the results for passive sentences. The most difficult version is clearly the structure with complex noun phrases and low density of function words. It was chosen as the most difficult by 37% of participants when choosing for themselves and by 47% when choosing on behalf of others. When choosing for oneself, there is no distinct easier version; however, when choosing for others, the sentence with simple noun phrases and high function word density was chosen the least often (12%) as the difficult sentence.

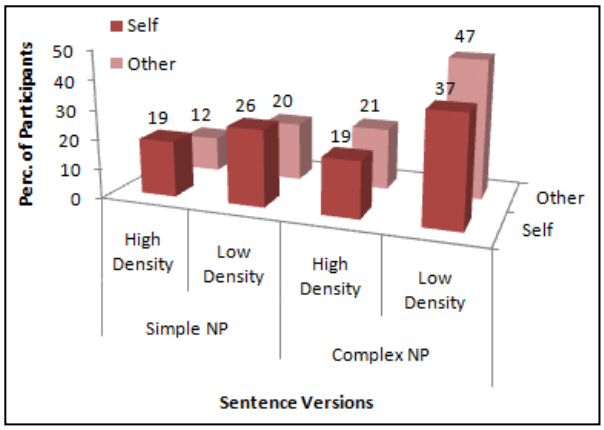

Figure 2: Percentage of participants selecting the most difficult version among passive sentences

The repeated-measures ANOVA for Self indicated a significant main effect of noun phrase complexity ($F(1,85) = 4.483$, $p < .05$) which explained 5 % of the variability ($\eta^2 = .005$), with more complex noun phrases being selected more often as complex. There was a second main effect of function word density, ($F(1,85) = 17.603$, $p < .001$), with sentences with fewer function words being seen as more difficult, an effect which accounted for 17% of the variability ($\eta^2 = .172$). Finally, there was a significant interaction effect between the two variables ($F(1,85) = 4.379$, $p < .05$), which accounted for 5% of the variability ($\eta^2 = .049$).

The repeated-measures ANOVA for Other showed similar effects: a main effect of noun phrase complexity ($F(1,85) = 40.112$, $p < .001$, $\eta^2 = .321$) and of function word density ($F(1,85) = 27.964$, $p < .001$, $\eta^2 = .248$). The interaction effect was also significant ($F(1,85) = 9.286$, $p < .001$, $\eta^2 = .098$).

Paired samples T-test indicated that sentences with simple noun phrases and high function word density ($p < .01$) or low function word density ($p < .05$) were more often chosen as difficult for oneself than for others: 19% versus 12% and 26% versus 20%. In contrast, sentences with complex noun phrases and low function word density were chosen less often as difficult for oneself (37%) than for others (47%), $p < .005$.

5.4 Extraposed Subject Sentences

Figure 3 shows the results for the extraposed subject sentences. In comparison to active and passive

sentences, the impact of noun phrase complexity seems more striking for extraposed and sentential subjects, which was unexpected. The sentence chosen as the easiest, was the sentence with simple noun phrases and high function word density, which only 11% of participants considered difficult. The sentences with complex noun phrases and low function word density were chosen by almost half of the participants, 44%, as the most difficult. The patterns are even more extreme when evaluating on behalf of others.

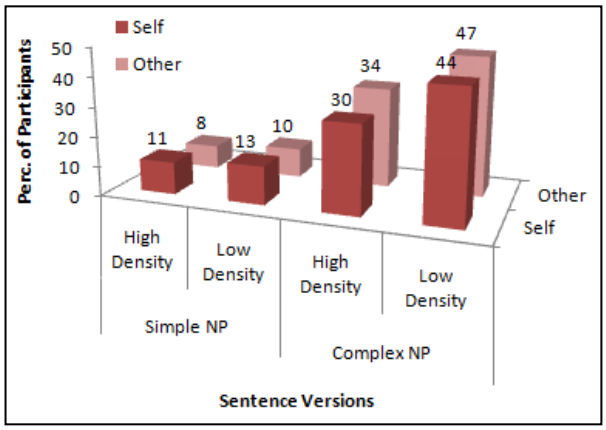

Figure 3: Percentage of participants selecting the most difficult version among extraposed subject sentences

The repeated-measures ANOVA for Self indicated two significant main effects and a significant interaction. Sentences with simpler noun phrases were considered simpler ($F(1,85) = 107.4$, $p < .001$), which explained most of the variability: 56% ($\eta^2 = .558$). There was a second, weaker, main effect of function word density ($F(1,85) = 6.647$, $p < .05$), with lower function word density being considered more often as difficult, an effect which accounted for 7% of the variability ($\eta 2 = .073$). The interaction was also significant. Function word density effects mattered especially with complex noun phrases ($F(1,85) = 4.449$, $p < .05$), which accounted for 5% of the variability ($\eta 2 = .050$).

The repeated-measures ANOVA for Others indicated main effects for noun phrase complexity ($F(1,85) = 192.6$, $p < .001$, $\eta^2 = .694$) and for function word density ($F(1,85) = 6.639$, $p < .05$, $\eta 2 = .072$). The interaction between the two was not significant ($p = .064$).

Paired samples T-tests showed only one significant effect. Sentences with simple noun phrases and high function word density were more often considered as difficult for oneself (11%) than for others (8%), $p < .05$.

5.5 Sentential Subject Sentences

Figure 4 shows the results for the sentential subject sentences. The patterns of perceived difficulty are similar for these sentential subject sentences compared to the extraposed subject sentences. The easiest conditions, where the fewest participants indicated a sentence as difficult, contained simple noun phrases and high function word density (11%). The most difficult condition was again the set of sentences with complex noun phrases and low function word density, which 40% of participants indicated to be the most difficult version. Differences between evaluations on behalf of oneself or others show the same, but somewhat more pronounced, pattern.

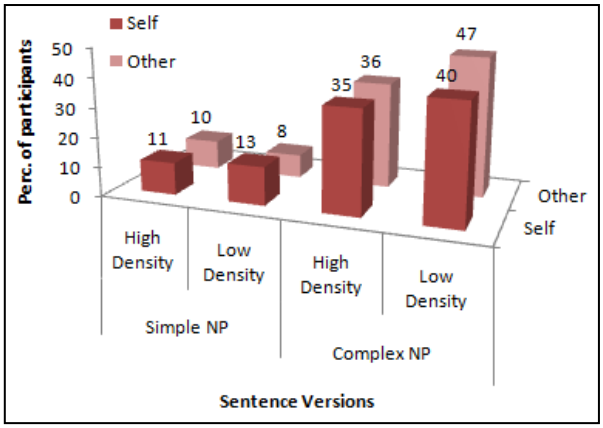

Figure 4: Percentage of participants who selected a sentence as difficult among sentential subject sentences

The repeated-measures ANOVA for Self indicates one very strong main effect of noun phrase complexity ($F(1,85) = 108.1$, $p < .001$) with sentences with more complex noun phrases being chosen more often as the most difficult. This variable accounted for 56% of the variability ($\eta^2 = .560$). There were no other significant effects.

The repeated-measures ANOVA for Other similarly indicated only a significant main effect for noun phrase complexity ($F(1,85) = 187.8$, $p < .011$, $\eta^2 = .688$).

Paired samples T-tests indicated that the differences between evaluations on behalf of oneself and others are significant only when the function word density is low. For sentences with simple noun phrases and low function word density, more participants indicated the

sentence as most difficult for themselves (13%) than for others (8%), $p < .05$. In contrast, sentences with complex noun phrases and low function word density were more often considered difficult for others (47%) than for oneself (40%), $p < .05$.

5.6 Readability Grade Levels

To complete our analysis, we compared the perceived difficulty with the commonly calculated readability grade levels. Table 3 provides an overview of the average Flesch-Kincaid Readability Grade Level for our examples in each condition.

Table 3: Highest average Flesch-Kincaid readability grade levels per condition

Structure	Noun Phrase Complexity	Function Word Density	Average Flesch-Kincaid Grade Level (N=4)
Active	Simple	High	12.9
		Low	12.8
	Complex	High	16.0
		Low	15.7
Passive	Simple	High	12.7
		Low	12.9
	Complex	High	15.0
		Low	14.8
Extraposed Subject	Simple	High	11.3
		Low	11.0
	Complex	High	13.1
		Low	12.1
Sentential Subject	Simple	High	12.2
		Low	11.7
	Complex	High	14.6
		Low	12.8

Compared to the writing guidelines for medical text for laypersons, all our sentences are considered too difficult. The recommended level is 6^{th} or 8^{th} grade. The numbers do not show as dramatic differences between conditions as our perceived difficulty evaluations did. We calculated the Pearson's correlation coefficient, which evaluates if there is a linear relationship. Although there are very few sample points and this analysis should only be seen as an indication of relations, we found one significant correlation for active sentences between the readability grade level and the evaluation on behalf of others ($p < .05$). There was no significant linear relation for any sentence structure for the evaluation on behalf of oneself or others and the grade levels.

It is noteworthy that more function words led to slightly higher readability grade levels but lower perceived difficulty according to our subjects.

6 Discussion

As hypothesized, we found strong effects of noun phrase complexity. Simpler noun phrases are recognized by most people as simple. We also hypothesized that function word density would play a role, with more function words leading to sentences that are perceived as easier. The results point in that direction, although the effects were not significant for every sentence structure. These combined results show that more sensitive measures can be developed that are associated with perceived difficulty, a first barrier in text understanding.

We also hypothesized that evaluating for oneself or on behalf of someone else would lead to different results. This assumption was based on previous work where we compared expert and layperson evaluations. Where we expected that people would overestimate their own knowledge and underestimate other's knowledge, the results pointed in a different direction. The overall scores on behalf of others were in the same direction but more extreme: a sentence considered easy for oneself was considered even easier for others, a sentence considered difficult for oneself was considered even more difficult for others.

Finally, readability grade levels cannot explain the results associated with function word density levels or the different patterns for different sentence structures. The readability grade levels are associated with word syllable and word. It was expected that sentences received higher grade levels when they have complex noun phrases. However, the main and interaction effects shown here indicate there is more at play then readability formulas can currently capture.

7 Conclusion

Although generalizations about sentence structure are limited since we did not randomize the sentence order per person, the strong effects, which do not coincide with the presentation order, suggest that noun phrase complexity and function word density have an enormous impact on what is perceived as difficult text or not. Simpler noun phrases and more function words lead to text that is perceived as simpler. Overall sentence structure also matters. Especially with sentences that have an extraposed or sentential subject structure, the difference related to noun phrase

complexity and function word density stand out. Finally, there evaluations done for oneself and on behalf of others differ.

This study is among the first to study different linguistic structures in comparison to each other. There are limitations that need to be taken into account. First and foremost is that sentences were not randomized per person. Although the ordering cannot explain the results, follow-up verifications studies will be conducted to exclude any possible irrelevant effects of ordering. There are also several strengths. The approach used is tuned to leverage information technology and automate any processes necessary to simplify. We also worked with representative consumers, not experts.

Text readability is an important aspect of human computer interaction (HCI). Many have looked at font types, sizes, and colors especially when websites, which provide many opportunities for different text representation. Positioning and attracting attention online have also been studied extensively with a focus on banners, pop-ups, and other attention tracking methods. The readability aspects, however, have usually been treated as a separate aspect. We believe that both perceived and actual text difficulty will be influenced by many factors for which a complete model does not yet exist. We evaluated one aspect in this study, the perceived readability of text, which has been shown to influence how readers interact with text. Other aspects, such as actual difficulty, text length, or text style, need to be combined and evaluated. Moreover, psychological research on modalities and memory should also be consulted. For example, text that is read by the person (written text) or heard (spoken text) has different characteristics and different effects on memory and understanding. Depending on the situation, one is preferred over the other. For example, short spoken instructions are often more effective in emergency directions than a written message. Finally, personal characteristics such as language skills, memory skills, literacy skills, health literacy skills, and cognitive skills will influence how difficult a text seems. These text, personal, and situation factors need to be integrated before a complete model of text understanding can be achieved.

In general, we believe we have made a first significant step toward more sensitive measures of text difficulty evaluation that are based on data with representative consumers. Future work will include more characteristics of texts and a focus on understanding and retention of information.

8 Acknowledgements

We would like to thank our many study participants who volunteered their time and made this work possible. We would also like to thank David W. Hindman, Ph.D. (Charles R. Drew University of Medicine and Science), and Charles V. Brown, Jr., MD, ABN, for their examples, materials, and support for this project.

This work was made possible by a grant from the National Science Foundation, SGER, 0742223, “U3 – Understanding User Understanding”.

9 References

[1] S. Fox and D. Fallows, "Internet Health Resources - Health searches and email have become more commonplace, but there is room for improvement in searches and overall Internet access.," Pew Internet & American Life Project, Washington D.C. July 16 2003.

[2] Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs - American Medical Association, "Health Literacy: Report of the Council on Scientific Affairs," JAMA, vol. 281,, pp. 552-557, 1999.

[3] Committee on Health Literacy - Institute of Medicine of the National Academies, "Health Literacy: A Prescription to End Confusion," N. Nielsen-Bohlman, A. M. Panzer, and D. A. Kindig, Eds. Washington, DC: The National Academies Press, 2004.

[4] R. B. Friedland, Understanding Health Literacy: New Estimates of the Costs of Inadequate Health Literacy. Washington, DC, 1998.

[5] M. J. Martin, "Reliability and Verification of Natural Language Text on the World Wide Web," in ACM-SIGIR Doctoral Consortium, Sheffield, England, 2004.

[6] D. Soergel, T. Tse, and L. Slaughter, "Helping Healthcare Consumers Understand: An "Interpretative Layer" for Finding and making Sense of Medical Information," in MEDINFO, San Francisco, USA, 2004, pp. 931-935.

[7] G. Leroy, E. Eryilmaz, and B. T. Laroya, "Health Information Text Characteristics," in American Medical Informatics Association (AMIA) Annual Symposium, Washington DC, 2006.

[8] B. D. Weis, Health Literacy and patient safety: Help patients understand. Manual for Clinicians, Second Edition ed.: AMA and AMA Foundation, 2007.

[9] G. Leroy, S. Helmreich, J. R. Cowie, T. Miller, and W. Zheng, "Evaluating Online Health Information: Beyond Readability Formulas," in American Medical Informatics

(AMIA) Association Annual Symposium, Washington DC, 2008.

[10] G. Leroy, T. Miller, G. Rosemblat, and A. Browne, "A Balanced Approach to Health Information Evaluation: A Vocabulary-based Naïve Bayes Classifier and Readability Formulas," Journal of the American Society for Information Science, vol. 59, pp. 1409-1419, Published Online: Apr 28 2008 2008.

[11] L. Rosmovits and S. Ziebland, "What do Patients with Prostate or Breast Cancer want from an Internet Site? A Qualitative Study of Information Needs," Patient Education and Counseling, vol. 53, pp. 57-64, 2004.

[12] H. Kim, S. Goryachev, G. Rosemblat, A. Browne, A. Keselman, and Q. Zeng-Treitler, "Beyond Surface Characteristics: A New Health Text-Specific Readability Measurement," in American Medical Informatics (AMIA) Annual Symposium, 2007, pp. 418-422.

[13] G. Rosemblat, R. Logan, T. Tse, and L. Graham, "Text Features and Readability: Expert Evaluation of Consumer, Health Text," in Mednet 2006.

[14] I. Ajzen, "The Theory of Planned Behavior," Organizational Behavior and Human Decision Processes, vol. 50, pp. 179-211, 1988.

[15] D. Trafimow, P. Sheeran, M. Conner, and K. A. Finlay, "Evidence that Perceived Behavioral Control is a Multidimensional Construct: Perceived Control and Perceived Difficulty," British Journal of Social Psychology, vol. 41, pp. 101-121, 2002.

[16] Y. Liu, W. R. Doucette, and K. B. Farris, "Perceived Difficulty and Self-Efficacy in the Factor Structure of Perceived Behavioral Control to Seek Drug Information from Physicians and Pharmacists," Research in Social and Administrative Pharmacy, vol. 3, pp. 145-159, 2007.

[17] N. K. Janz and M. H. Becker, "The Health Belief Model: A Decade Later," Health Education Quarterly, vol. 11, pp. 1-47, 1984.

[18] R. S. Velayo, "Retention of Content as a Function of Presentation Mode and Perceived Difficulty," Reading Improvement, vol. 30, pp. 216-227, 1993.