Readability: Examining its usefulness in the field of controlled language

Patrick Cadwell
School of Applied Languages and Intercultural Studies
Dublin City University
Dublin 9, Ireland
patrick.cadwell2@mail.dcu.ie

Abstract

This article is based on a survey of text-user attitudes. The survey aimed to discover: whether the concept of readability has merit in the field of controlled language; and whether readability is increased by applying controlled-language rules to a sample of technical texts. The article attempts to provide much-needed empirical data to a neglected area of controlled-language research, and to examine the concept of readability that appears to be misunderstood, undervalued and misused. In particular, the paper examines issues concerning reader preference, the predictability of readability formulas, and the variables that impact on readability as a whole. Moreover, though the participant samples in the study were too small to be generalised to larger populations, the trends identified here indicate useful directions for future research.

Keywords: *readability; controlled language; reader preference; empirical data.*

1 Introduction

In this article, the results of a survey carried out to examine whether readability has merit in the field of controlled language (CL), and whether it is increased by applying CL rules to texts is described. The paper is based on a Masters-level dissertation completed earlier this year by the author: the full dissertation can be viewed at http://www.localisation.ie /resources/Awards/Theses/Theses.htm.

This article focuses foremost on the theoretical issues that informed the study; only a very brief outline of the methodology is given, while placing focus on how questions raised in the review of the literature were answered by empirical data.

The literature, in this case, is made up of three main groups: Group 1 consists of papers submitted to the various International Workshops on Controlled Language Applications (CLAW), workshops that have been running since 1998; Group 2 is made up of scholarly works on the concept of readability, in particular Klare (1963, 1974, 2000), Flesch (1948), Dale and Chall (1948), Fry (1958), Gunning (1952); Group 3 is comprised of articles from journals that specialise in technical writing, for example Journal of Technical Writing and Communication or ACM Journal of Computer Documentation.

The three groups speak to very different audiences and represent different trends in research into document production: Group 1 sees readability as interesting, but focuses its efforts on matters more directly linked to machine translation; Group 2 defines and helps us understand the concept of readability, but is weak at exemplifying its practical applications; Group 3 tends to be highly critical of readability as a concept and focuses research on other areas of document analysis. In short, the literature is broad and does not generally examine a link between a clear understanding of what readability is and how it can be practically and beneficially applied to the field of CL. This is the gap that my study hoped to fill. In particular, I had three main motivations for carrying out my work:

1.1 Motivation: need for empirical research

Eight years ago, Knops (2000, p.134) called out for more empirical data in the field of CL. He said:

> *Generally speaking, there is an urgent need for facts and figures obtained in experimental situations and real-life production environments and relating to the effects of particular CL standards, rules and rule sets on readability and translatability.*

Since that time, researchers have answered that call and several empirical studies have been published in the field of CL rules. However, these studies have focused largely on more machine-oriented topics, such as translatability, comprehensibility, proof-read-

ing effort, etc. More human-oriented analyses - such as readability, usability, etc. - have been neglected. In fact, since Knops call I have found no major work that outlines how readability and controlled language interact in a contemporary, commercial situation. According to Hayes, Maxwell and Schmandt (1996, pp.84-85), this may be because readability advantages are harder to quantify than those of translatability, comprehensibility, etc. I believe, however, that this may also be because few people have an understanding of how readability is correctly defined; without such an understanding, how can anyone determine the advantages that applying readability guidelines may bring?

1.2 Motivation: assumptions held about readability

Several under-tested and unchallenged assumptions about readability now hold firm. For example, "reducing the complexity of syntactic structures of a text increases its readability" (Spaggiiari, Beaujard and Canneson 2003, p.152). Similarly, Reuther (1998, p.174) claims that:

> *It is a well known and indisputable fact within the CL community that the use of a Controlled Language (CL) in technical documentation leads to quality improvement with respect to readability, consistency and translatability.*

In the literature, it is difficult enough to find a consistent definition for readability, let alone empirical evidence for such strong statements. My study intended to show the ultimate value of readability in the field of CL: regardless of how easy a CL makes a document to translate or comprehend, these benefits will be for nothing if the text is written in a way that causes the reader to discard it.

1.3 Motivation: lack of terminological rigour

In preparation for this study, I came across an abundance of terms used to describe how CL texts can be analysed and evaluated (hereafter referred to as metrics). Metrics in the field of CL include: readability; comprehensibility; translatability; usability; post-editing effort; consistency, legibility; acceptability; accessibility; learnability. Depending on the author, these terms can be treated as sharing many, all or no characteristics. Clearly, there is a need for a more systematic and rigorous definition and treatment of these concepts.

2 Brief outline of the survey

For this study, I was given access to various naturally-occurring technical texts by the software publisher, Symantec. An internal training document was deemed to be the most appropriate for use in the survey. This type of document was chosen as it could be claimed that the Symantec participants would all be expert in, or at least familiar with, its contents. Moreover, it seemed long and varied enough to provide different examples of writing styles and readability.

This training text was divided up into short, similar passages and popular readability formulas were applied to the passages. Over 50 readability formulas have been developed over the years, but, of these, six formulas are particularly influential: Flesch; Gunning Fog; Dale-Chall; Fry Graph; SMOG; and Automated Readability Index (ARI). The formulas used in this study were Flesch Reading Ease, Flesch Kincaid Grade Level and Gunning Fog; they were chosen for their speed and ease of calculation, and for the large number of times that they are cited in works in the field of CL.

Of the short, similar passages, three individual passages were finally selected to be experimented on. One received an extremely unfavourable readability prediction from all three formulas. It will henceforth be referred to as the 'Norton' passage. One received an extremely favourable readability prediction from all three formulas. It will be referred to as the 'Shared' passage. The last passage received a prediction score midway between these two from all the formulas and will be referred to as the 'HijackThis' passage. Collectively these three naturally-occurring passages with varied readability scores were termed 'NCL' passages in the dissertation.

Next, these three passages were controlled for readability. Identifying the rules to apply to control for readability was a long and difficult task, and will be described in great detail in 3.3. Here, it will suffice to say that once the author had edited the passages (termed "CL" passages in the study) so that they conformed to the readability rules, the same readability formulas were applied to them: all three scores improved marginally. However, the relationship between the three passages remained the same (i.e. the most difficult to read remained the most difficult, etc.).

With NCL and CL versions prepared, it was time to carry out the survey. The survey used two samples of respondents: one sample was made up of Symantec staff that had knowledge and practical experience of

the technical domain from which the texts were taken (Symantec); the other sample was made up of participants without such domain knowledge and experience (Control). The two groups, with 12 participants in each group, were largely balanced for English ability, gender, and educational profile.

The survey was done on a test / retest basis. There was a survey in Stage 1 and another similar survey in Stage 2 to examine attitude variance depending on whether a CL or NCL version was read. To counter any inherent bias, participants did not know when they received a CL version: for both groups, a table of random numbers was used to randomly distribute a CL version to half the participants in Stage 1 and half in Stage 2. At both stages, the subjects read the passages and then filled out a questionnaire that examined their opinions as to the readability of the texts.

3 Theoretical overview

The survey outlined in Section 2 set out to discover whether readability has merit in the field of CL, and whether it is increased by applying CL rules to the training texts under experimentation. However, to decide how to move from such broad goals to specific survey questions required a lot of theoretical preparation. This preparation involved examining readability in isolation - how we define, measure, predict and produce readability - and examining readability in relation to other metrics used in the field of CL.

3.1 Defining and measuring readability

Readability is regularly mentioned in the CL literature, but is rarely defined by the authors that use it. Perhaps this is because it is an idea that is prevalent in general language, and authors assume that readers understand the concept as "the ease with which written language can be read with understanding" (Crystal 1992, p.326). For this experiment, a more detailed definition was needed to show how readability differed from other metrics, such as legibility, comprehensibility or clarity. At first, it was necessary to consult some fairly dated sources: theoretical work on readability began in the US in the 1940s when literacy levels of the general population were still low, but when the government needed to disseminate increasingly complex written documents in the medical, legal and financial fields. Key works by influential scholars at the time include Rudolf Flesch (1948), Edgar Dale and Jeanne Chall (1948), and Robert Gunning (1952). Dale and Chall (1949 in DuBay 2004, p.3) provided the detail lacking in more general definitions. For them, readability is:

> *The sum total (including all the interactions) of all those elements within a given piece of printed material that affect the success a group of readers have with it. The success is the extent to which they understand it, read it at an optimal speed, and find it interesting.*

Klare (1977 in Harkins and Plung 1982, p.149) concurs with Dale and Chall in their definition, and states that when we talk of readable writing, "...we mean that the intended readers are able to read it quickly, understand it clearly, and accept it readily (i.e. persevere in reading it)". In other words, it is the combination of these three elements that differentiates readability from the other metrics. Flesch (1948), Dale and Chall (1948), Gunning (1952), Fry (1968) and Klare (1963) note that we make documents readable to help readers understand them better, and to help them avoid making mistakes that they might otherwise have made. Crucially, though, they emphasise that we also make them readable to save the readers time and effort, and to ensure that they do not give up on reading the document.

These definitions tell us how to go about measuring readability. However, they do not explain how to predict whether one text is more readable than another. Nor do they instruct us how to produce readable text.

3.2 Predicting readability: readability formulas

As was shown in 3.1, measuring what makes a document readable involves the detailed analysis of complex concepts such as the reader's understanding, reading speed, and perseverance. However, such complex analysis may not always be possible. Thus, scholars have tried to develop formulas which use variables in the text to predict how difficult that text may be for a particular audience. Over 50 procedures claiming to compute how difficult a text is to read have been devised over the last 80 years. Of these, six are particularly influential: Flesch; Dale-Chall; Fog; Fry Graph; SMOG; and Automated Readability Index (ARI). Others that are often utilized include FORCAST, Lorge and Spache (Klare 1974, p.68).

Formulas do not define or explain readability; they do not point to all the areas of a text that make it readable or comprehensible (Davison and Kantor 1982, pp.189-190). The formulas are merely intended as indices or predictors of how difficult a text is likely to be for an intended reader. To construct a formula,

the researcher assembles large numbers of 'criterion' passages; these are usually texts taken from the US educational system. Most formulas in use today originated in research projects in the US and are based on the study of American-English data. Moreover, the formulas are compiled with a view to making the English spoken in that part of the world more readable. This is a matter not usually considered by those who use these formulas in other geographic and linguistic settings. Readability formulas are constructed for other languages aside from English, but their features were not considered in this study. Nonetheless, comparing and contrasting the formulas used in different languages would surely be an interesting research theme in the future.

Once the 'criterion passages' have been chosen, language variables from these passages - typically word difficulty and sentence length - are selected. The researcher then sees how these vary with the scores that readers have given the passages in terms of reading speed, reader preference, and comprehension, to name the three most common values. If a language variable and the readers' scores correlate closely, the variable is said to be a characteristic of readable writing and is combined statistically into a formula. These results are then further validated with other scores for reliability (Klare 1977 in Harkins and Plung 1982, p.149).

To use a formula, a passage of at least 100 words is selected; such a length is necessary for the statistical regressions used in most formulas to be valid. Then, a count is made of the language variables that have been identified as being characteristic of readability. These counts are entered into the formula, and an overall score for the passage is given. This score will typically be expressed in different ways: some formulas place the score on a graph (Fry); some express the score as the US grade-school level the reader needs to have completed to be able to read the passage (Flesch); some express the score on a simple scale from 0 to 100, with 0 being the most difficult, and 100 being the easiest (Flesch Reading Ease); while others express it as the number of years of formal education a reader needs to be able to read the passage (Fog).

Crystal (1997, p.254) and DuBay (2004, p.54) both emphasise the increasing significance and popularity of readability formulas in the field of educational research. However, other authors criticise the formulas for being unsophisticated and unsuited to use on any other texts than those intended for children in the US school system: this is because the criterion passages on which they are based have been selected and validated with schoolchildren in mind (Hargis 2000, p.105; Giles and Still 2005, p.66).

Now let us look at how readable writing can be produced.

3.3 Producing readability: rules to conform to

This section describes the CL created by the author for this experiment: the rules outlined here are shown in the literature to have a positive impact on readability. These rules will be divided into four major categories: textual / pragmatic; syntactic; grammatical; lexical.

Textual rules

Have no more than one idea per paragraph.
According to Davison and Kantor (1982, pp.189-191), readability must take into account elements that contribute to a coherent and well-formed text. They emphasise that the inference and cognitive load placed on readers should not be too great.

Each paragraph should start with a topic sentence.
Davison and Kantor (1982, pp.196-197) also found that reading time shortened and readability increased if closely relevant context information was placed at the beginning of each paragraph.

Give old information before new (theme-rheme progression).
Both Farrington (1996, p.16) and Reuther (2003, p.128) assert that human readers process texts better when new and complex information is presented slowly, in a logical progression, and without too many new chunks at one time.

Use headings for paragraphs and leave sufficient 'white space'.
Dayananda (1986 in Crystal 1997, p.383) advises writing for the eye as well as the mind: using white space, combined with headings, subheadings, etc., makes the organisation of ideas in the text clearer

Put long lists in bullet points.
Hargis (2000, p.129) concurs with many CL authors in recommending that long lists should be presented in the form of bullet points.

Syntactic rules

In general, Klare (1977 in Harkins and Plung 1982,

pp.150-151) reminds us that correctly-punctuated 'Simple Active Affirmative Declarative' sentences are the most readable. However, to be more specific, as far as syntax is concerned:

Sentences should not exceed 25 words.
O'Brien (2003, p.110) explains that in CL the maximum number of words allowed in a sentence varies from somewhere between 20 and 25 words. Generally, the lower limits are applied for procedural texts, and the higher limits for descriptive texts.

Have variety in the length of sentences within this 25-word limit.
As was shown in 3.1, perseverance is a key pillar of readability. Klare (1977 in Harkins and Plung 1982, pp.150-151) underlines that if each sentence is uniformly similar, the reader will become bored and give up reading.

Have a maximum of two clauses per sentence.
Bram (1978 in Harkins and Plung 1982, p.146) showed that to increase readability, there should only be one or two statements per sentence, with no additional qualifying or explanatory information. In general, difficult texts have a longer more complex structure and impose a greater cognitive load on the reader.

Grammatical rules

Avoid using ambiguous constructs.
Some linguistic constructs - for example the connectors 'like' and 'or', or the 'slash' - are ambiguous and require resolution by the reader. These increase reading time and complexity, and should be avoided (Nyberg and Mitamura 1996, p.80).

Avoid using the passive voice.
Dayananda (1986 in Crystal 1997, p.383) states that the passive voice is less readable than the active voice as it generates greater cognitive load.

Avoid ellipsis and pronominal reference.
Klare (1977 in Harkins and Plung 1982, pp.150-151) states that leaving out parts of sentences and using pronouns - even when the meaning can be understood without the original noun or ellipted item - creates more difficulty for the reader and should be avoided.

Lexical rules

According to Nyberg and Mitamura (1996, p.77), a pre-approved vocabulary that is consistently used by authors is vital to the success of a CL. However, this predefined word list will vary depending on the domain in which the texts are used: the list for writing a school textbook will be very different to the one used in writing an airplane maintenance manual. No scholar has yet found a better multi-purpose lexical rule for readability than 'the simple word should be favoured over the complex'. This, however, is not very instructive. Without a vocabulary list specific to this experiment, only one other lexical rule was identified by the author. This was:

Ensure that all words are spelt correctly.
Hargis (2000, p.129) reminds us that poor spelling can increase processing time, misunderstanding and frustration. Its impact on readability should not be underestimated.

It must be stressed that the above rules should not be accepted without challenge: many are severely criticised in the literature. For example, though bulleting is intuitively held by many to be easier to read, research done by Garrod (1998 in Grover et al. 2000, p.91) contests whether doing so actually works. Similarly, Davison and Kantor (1982, pp.192-195) claim that shortening sentences can just lead to the dilution of logical relations between clauses and sentences, which in turn leads to mistaken inferences being made by the reader. Moreover, Hargis (2000, p.126) asserts that the break-up of sentences not only interferes with understanding in this way, but also produces a choppy, monotonous style that will bore and frustrate the reader. Despite these criticisms, however, the weight of evidence in the literature at present points to the above rules positively impacting on readability.

So far, only the linguistic variables impacting on readability have been dealt with. However, several extra-linguistic variables also have a strong influence on readability and must not be neglected.

3.4 Extra-linguistic variables
Many variables outside the linguistic realm help or hinder readers in understanding, in reading more quickly, and in persevering with their reading. These include: motivation; reading ability; interest in the topic; relevance of the topic; familiarity; prior knowledge; and testing conditions. DuBay (2004, p.39) points out that many experiments in the field of CL do not achieve the expected results because they fail to control for such variables. It is not difficult to create illustrative examples. Imagine the number of readers that neglect to sign a simple form, even

though the instructions to do so are easy to understand and clearly presented: in such a case, it is likely that motivation or interest is lacking. Similarly, we can think of a document that would be completely unreadable or incomprehensible to the average person, but that would be smoothly read and easily comprehended by an expert with prior knowledge of the topic and familiarity with the text type.

Clearly, then, these variables can have an impact on readability. For example, Klare (1977 in Harkins and Plung 1982, p.150) shows that:

> *Someone who is very highly motivated can read very difficult materials, where the mismatch between reading ability and readability is considerable.*

He gives some examples: low-ability readers are able to successfully complete a tax return or decipher a complex medical chart when failure to do so would result in serious negative consequences for these readers. Even though motivation and the other extralinguistic variables are known to be critical, their subjectivity means that most experiments are unable to control for them.

3.5 Other metrics related to readability

An abundance of metrics in the CL literature has led to conceptual confusion. Many of the metrics do not, in fact, deal with monolingual document production, as is the case with readability; they deal instead with the work of translating a text from a source language into a target language. This is not to say that readability is of no interest to the field of translation; I would argue that an understanding of how to produce readable documents in English is vital to any translator working into or out of English. However, the fact that many people in the field confuse essentially monolingual metrics like readability with translation-oriented metrics makes the need for terminological rigour all the more urgent. Some of the main metrics with which readability can be confused are detailed below.

3.5.1 Comprehensibility

As pointed out by Roturier (2006, p.3), comprehensibility should be defined as the ease with which a translation can be understood by its reader. However, in the literature, comprehensibility is often used synonymously with comprehension, understandability and understanding. Is comprehensibility the same as understanding? Few CL works explicitly define the metrics they use in their research and often use terms interchangeably. The conceptual map of the field becomes still more confused when translatability is introduced.

3.5.2 Translatability

This concept is generally taken to be the extent to which a document is amenable to processing by either a human translator or, more often, a machine translation system. However, Reuther (2003) declares readability to be a subset of translatability, while authors like Hargis (2000) reverse this position entirely and see translatability as being just one level of readability.

3.5.3 Usability

Authors like Redish (2000) and Schriver (2000) see usability as a metric which completely excludes the need for readability and comprehension. For them, the aim is to read to do, or to read to carry out a procedure successfully. They see the reader's level of understanding as unimportant, once the document has been 'used' effectively and the desired result has been achieved.

3.5.4 Others

Aside from these major concepts, other authors introduce even more 'similar-yet-different' ways to look at CL texts. For example, Puurtinen (1995, p.230) defines 'acceptability' as the readability and speakability of a text as well as how well a text receiver accepts a translated text as cohesive, coherent and capable of utilization. She defines 'accessibility' as the ease of comprehension due to the style of writing. Furthermore, Hargis (2000, p.123) sees concepts such as 'learnability' and 'doability' as being merely different levels of readability.

Regardless of whether the above definitions are accurate, it is certain that conceptual organisation is required. How might we introduce such terminological rigour to the field?

3.6 Introducing terminological rigour

The various metrics are highly interrelated: as much as some concepts can appear to contradict each other, others can be shown to be highly complementary. For example, O'Brien and Roturier (2007) were able to show that many of the CL rules used in their separate studies had a high impact on both comprehensibility and post-editing effort, suggesting that these concepts complement each other.

Thus, a way of mapping CL metrics is required that accounts for such interrelations, complementarities

and contrasts. This study proposed the use of a Venn diagram to better understand the conceptual map. Perhaps it will never be possible to draw clear distinctions between what is readable, comprehensible, translatable, etc. Rather than looking to make these concepts entirely distinct, it might be more useful to look at where they have their focus. The metrics in 3.5 focus to a greater or lesser degree on the text itself, the reader of the text, and the outputs of the text. These three elements then become the three circles of a Venn diagram (see Figure 1).

To illustrate with examples from the diagram: take 'interest'. Figure 1 illustrates that this metric tells us more about the reader of the text than anything else. In contrast, 'legibility' tells us more about the text than anything else. It is by no means the author's assertion that this is a perfect mapping of the concepts. It is simply intended to convey the opinion that readability can be shown to have a much lesser focus on the reader than comprehension, or that readability can be shown to have a much lesser focus on the results of the text than usability, and so on.

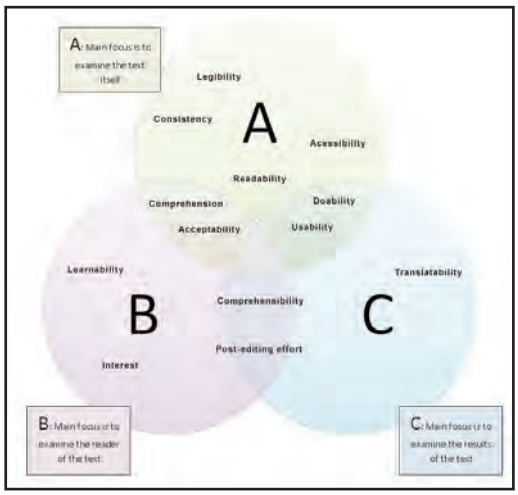

FIGURE 1: VENN DIAGRAM PLOTTING CL METRICS

To summarise, the theory of readability discussed here in Section 2 shows us that the concept is much richer than that which is currently prevalent in the CL literature. When we think of readability, we must consider more than just comprehension and must focus on three key pillars: speed, perseverance and understanding. We must recognise, too, other important variables outside of the linguistic realm that impact on readability. Moreover, it is important to differentiate between the prediction and production of readability when it comes to analysing texts, and especially how this relates to the many popular formulas now in widespread use. With these theoretical issues in mind, let us look at the data produced by this experiment.

4 Summary of the empirical data

The survey carried out for this dissertation produced a large amount of raw data that can be interpreted in many ways. However, the theoretical issues discussed in Section 3 of this article lead us to ask four main questions. Here is how the data seemed to answer these questions:

4.1 Would the CL version be preferred by readers?

At Stage 2, after having read both CL and NCL versions of the text, participants were asked which one they found easier to read. A majority of participants (see Figure 2) in both groups said that they found the CL version of the texts easier to read. To this extent, it can be claimed that the CL versions were preferred by these readers.

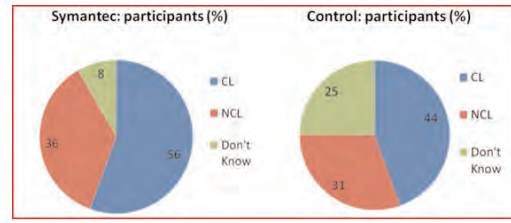

FIGURE 2: WHICH TEXTS WERE EASIER TO READ - WHAT % OF PARTICIPANTS CHOSE WHAT

If a version is preferred by a reader, they are more likely to persevere with reading it than another version. So, on this level, we can say that CL versions appear to be more readable. But what do the questionnaires tell us about the other two pillars of readability?

4.2 Would the other two pillars of readability be altered in the CL version?

According to Klare (1977 in Harkins and Plung 1982, p.149) the key elements of readability - efficiency and understanding - can be tested by analysing reading speed and retention of key vocabulary respectively. The first point we will examine is the retention of key vocabulary. After reading all the passages, participants were asked to identify keywords from a list including synonyms that did not appear in the passages. Figure 3 breaks down the number of correctly retained keywords - less incorrectly selected synonyms (noise) - for each group at each stage.

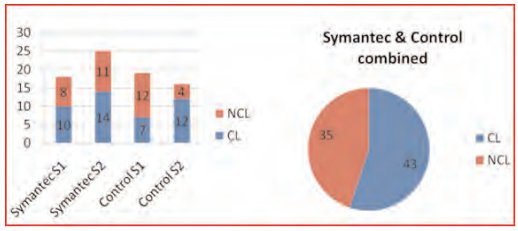

FIGURE 3: NUMBER OF KEYWORDS CORRECTLY RETAINED MINUS NOISE

Overall, this shows that more key vocabulary was correctly retained with less noise when the CL version was read. This result indicates that retention is better when a CL version is used, and that, on this level, the CL texts were more readable. However, as we know from 3.1, readability is about more than just retention and perseverance. The last element to be tested was reading speed.

Figures 4 and 5 show that in the majority of texts, for both groups and at both stages, the CL versions actually took longer to read. This was an unexpected result and was probably due to the fact that the method of timing used was crude and inaccurate; in the study, participants were asked to use a regular wall clock and note starting and finishing times on the questionnaire sheet. This might be sufficient for very long time periods, but in this study, with such short passages, accurate counting of seconds and even milliseconds would have provided much richer data. In this way, the use of eye-tracking software in future work, with its accurate time measurement and complex reading-pattern display, would bring great benefits.

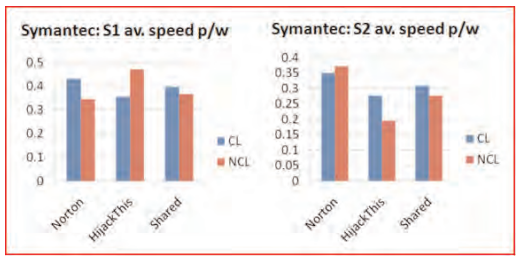

FIGURE 4: BREAKDOWN OF AVERAGE READING SPEED PER WORD PER TEXT FOR SYMANTEC (SECONDS)

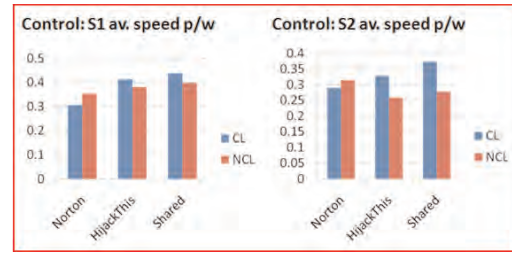

FIGURE 5: BREAKDOWN OF AVERAGE READING SPEED PER WORD PER TEXT FOR CONTROL (SECONDS)

In summary, speed was not positively impacted by applying CL rules, though retention and perseverance were: therefore, it would seem that applying CL rules to technical documents does increase readability.

4.3 Would the formulas' predictions correspond to readers' opinions?

Overall, it seemed that the predictions made by the formulas did not correspond to what readers thought. Figure 6 illustrates that for both groups the majority of readers' ratings did not correctly correspond to the readability formula predictions.

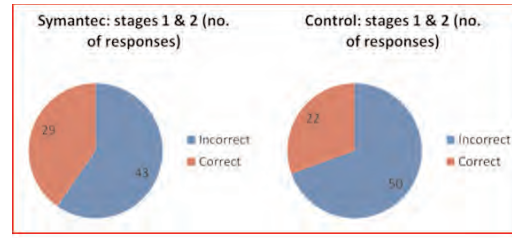

FIGURE 6: NO. OF TIMES READERS' RATINGS CORRECTLY CORRESPONDED TO FORMULA PREDICTIONS

The formulas chosen do not appear to be useful predictive tools. However, there was another group of data that told a different story: participants were asked in the questionnaire to underline any parts of the text that they found difficult to read. By very crudely counting the number of texts in which participants underlined something for reading difficulty, it can be shown that the formulas actually predicted the relationship between the three passages correctly. Thus, in Table 1 we see that in the naturally-occurring NCL versions, 'Norton' had the most responses with underlined sections (17), 'Shared' had the least (11), and 'HijackThis' came in the middle (13). This corresponds exactly to the ranking for difficulty that the formulas predicted in Section 2.

Total no. of texts with underlining for difficulty				
		Symantec	Control	TOTAL
NCL	Norton	11	6	17
	HijackThis	6	7	13
	Shared	5	6	11
CL	Norton	6	5	11
	HijackThis	6	7	13
	Shared	6	7	13

TABLE 1: NO. OF TEXTS WITH UNDERLINING FOR DIFFICULTY

Of course there is a 'black-box' problem with the validity of this data. That is, the underlined sections can only be said to represent areas that participants were less satisfied with; we do not know whether people were truly underlining for problems of readability, comprehension, or some other objection (type-face, legibility, subject matter, etc.). Despite this validity issue, the underlined sections still raise important issues about the way of testing reading difficulty: asking an opinion or rating is subjective and value-laden. Perhaps task-related testing, such as the underlining task, generates more objective data.

4.4 Would extra-linguistic variables impact on readability?

In 3.4, prior knowledge of a domain (or technical expertise) was identified as an extra-linguistic variable that is shown to increase readability in the minds of readers. By this hypothesis, then, the Symantec group should have given more favourable ratings than the Control group. However, this was not the case. Table 2 shows that Symantec underlined more for difficulty than Control, again pointing to the fact that prior knowledge of the domain was not positively impacting on readability.

Symantec: breakdown of underlining (no. of texts)					
	Stage 1		Stage 2		TOTAL
NCL	Norton	6	Norton	5	11
	HijackThis	4	HijackThis	2	6
	Shared	4	Shared	1	5
					22
CL	Norton	1	Norton	5	6
	HijackThis	4	HijackThis	2	6
	Shared	3	Shared	3	6
					18

Control: breakdown of underlining (no. of texts)					
	Stage 1		Stage 2		TOTAL
NCL	Norton	4	Norton	2	6
	HijackThis	4	HijackThis	3	7
	Shared	3	Shared	3	6
					19
CL	Norton	2	Norton	3	5
	HijackThis	5	HijackThis	2	7
	Shared	3	Shared	4	7
					19

TABLE 2: NO. OF TEXTS UNDERLINED FOR DIFFICULTY - SYMANTEC VS. CONTROL

In contrast, the opposite effect was shown for familiarity. Table 3 illustrates that, in both groups, from Stage 1 to Stage 2 underlining decreased, favourability improved, and speed decreased. This would seem to suggest that - because it occurred equally in Control and Symantec - just becoming familiar with a text, even if it was not comprehended or used effectively, makes that text seem more readable.

Familiarity had an impact on readability					
		Symantec		Control	
		Stage 1	Stage 2	Stage 1	Stage 2
Underlining decreased		22	18	21	17
Favourability scores improved		54	49	53	47
Reading speed improved	CL	0.39395	0.31128	0.38513	0.33023
	NCL	0.39483	0.28092	0.37756	0.28336

TABLE 3: FAMILIARITY APPEARED TO HAVE A STRONGLY POSITIVE IMPACT ON READABILITY

The final extra-linguistic variable that was tested by this study was participant profile. In particular, native-English ability was shown to have a strong impact on views of readability. By comparing Figures 7 and 8, it can be seen that non-native speakers found the CL versions easier to read, while native speakers tended to find the NCL versions easier to read. This is not a criticism of the non-native speakers' English ability: most rated themselves 8, 9 or 10, where 10 represented native-level fluency. This difference most likely comes about because native speakers are much more familiar with, and tolerant of, the eccentricities and exceptions of naturally-occurring language.

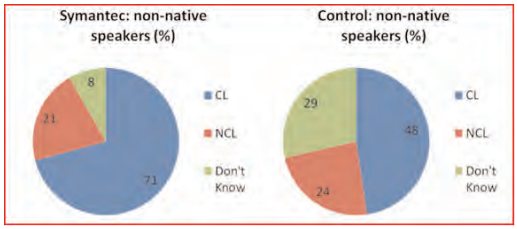

FIGURE 7: NON-NATIVE SPEAKERS FOUND CL MORE READABLE

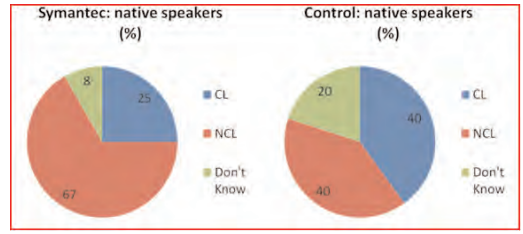

FIGURE 8: NATIVE SPEAKERS FOUND NCL MORE READABLE

5 Conclusions

In conclusion, therefore, this work has provided new empirical data, albeit limited in scope, to show that CL versions are thought to be easier to read; are viewed more favourably; and encourage better retention of keywords. In short, this data seems to suggest that the application of CL rules increases readability. However, these are not the only conclusions that should be drawn from this experiment.

One such additional conclusion concerns readability formulas. Overall, this study appeared to show that the formulas made inaccurate predictions. Most criticisms of formulas probably arise because of people incorrectly using them in ways for which they were not intended: as guidelines for writing or as tools for correction. If the formulas are used within their recognised limitations - as rough predictors of difficulty to prompt more detailed textual analysis - then

they may have more merit.

An important conclusion that needs to be made, too, is that much work remains to be done on clarifying the many metrics that are used in the field of CL. The ideas presented in this study are a tentative first step at disentangling the web. Rather than thinking that one form of analysis is better or worse, any new approach should encourage people just to consider the appropriateness of each metric to their situation.

The final conclusion concerns recommendations for how future studies might proceed. Clearly, more widely-differing texts will produce more noticeable trends in the data: the documents presented in this study were all from the same original document and were selected only based on the predictions of readability formulas. Future studies should incorporate not only formulas, but also semantic assessments, expert advice from the users or authors of the texts, and other criteria to decide the difficulty of the passages to be experimented on. Similarly, such tests should try to incorporate extra-linguistic variables into their methodologies.

Hopefully this article has shown that analysing texts for readability is a useful exercise in the field of CL. By making us consider key elements like reading speed, reader perseverance and reader understanding, as well as influential external factors like motivation and familiarity, the study of readability can promote a comprehensive approach to the theory of document production.

References

Crystal, D. 1992. An encyclopedic dictionary of language and languages. Oxford, UK; Cambridge, Mass., USA: Blackwell.

Crystal, D. 1997. The Cambridge encyclopedia of language. 2nd ed. Cambridge; New York: Cambridge University Press.

Dale, E. and Chall, J.S. 1948. A formula for predicting readability: Instructions. Educational research bulletin. 27 (2). pp.37-54.

Davison, A. and Kantor, R.N. 1982. On the failure of readability formulas to define readable texts. Reading research quarterly. 17 (2), pp.187-209.

DuBay, W.H. 2004. The principles of readability [Online]. Available from: http://www.impact-information.com/impactinfo/readability02.pdf [Accessed 2 September 2008].

Farrington, G. 1996. AECMA Simplified English: an overview of the international aircraft maintenance language. IN: Proceedings of the first controlled language application workshop, March 1996. Leuven: Centre for Computational Linguistics, pp.1-21.

Flesch, R. 1948. A new readability yardstick. Journal of applied psychology. 32 (3), pp.221-233.

Fry, E.B. 1968. A readability formula that saves time. Journal of reading. 11, pp.513-516 and pp.575-578.

Giles, T.D. and Still, B. 2005. A syntactic approach to readability. Journal of technical writing and communication. 35 (1), pp.47-70.

Grover, C., Holt, A., Klein, E. and Moens, M. 2000. Designing a controlled language for interactive model checking. IN: Proceedings of the third international workshop on controlled language applications. April 2000. Seattle: ANLP/NACLA, pp.90-104.

Gunning, R. 1952. The technique of clear writing. New York: McGraw-Hill.

Hargis, G. 2000. Readability and computer documentation. ACM journal of computer documentation. 24 (3), pp.122-131.

Harkins, C. and Plung, D.L. 1982. A guide for writing better technical papers. New York: IEEE Press.

Hayes, P., Maxwell, S. and Schmandt, L. 1996. Controlled English advantages for translated and original English documents. IN: Proceedings of the first controlled language application workshop, March 1996. Leuven: Centre for Computational Linguistics, pp.84-92.

Klare, G.R. 1963. The measurement of readability. Ames, Iowa: Iowa State University Press.

Klare, G.R. 1974. Assessing readability. Reading research quarterly. 10 (1), pp.62-102.

Klare, G.R. 2000. Readable computer documentation. ACM journal of computer documentation. 24 (3), pp.148-168.

Knops, U. 2000. Efficient roll-in and roll-out of controlled language applications. IN: Proceedings of the third international workshop on controlled language

applications. April 2000. Seattle: ANLP/NACLA, pp.134-135.

Nyberg, E.H. and Mitamura, T. 1996. Controlled language and knowledge-based machine translation: principles and practice. IN: Proceedings of the first controlled language application workshop, March 1996. Leuven: Centre for Computational Linguistics, pp.74-83.

O'Brien, S. 2003. Controlling controlled English: an analysis of several controlled language rule sets. IN: European association for machine translation: international workshop, May 2003. Dublin: Dublin City University, pp.105-114.

O'Brien, S. and Roturier, J. 2007. How portable are controlled languages rules: a comparison of two empirical MT studies [Online]. Available from: http://www.mt-archive.info/MTS-2007-OBrien.pdf [Accessed 2 September 2008].

Puurtinen, T. 1995. Linguistic acceptability in translated children's literature. Joensuu: University of Joensuu.

Redish, J. 2000. Usability testing reveals more than readability formulas reveal. ACM journal of computer documentation. 24 (3), pp.132-137.

Reuther, U. 1998. Controlling language in an industrial application. IN: Proceedings of the second international workshop on controlled language applications. May 1998. Pennsylvania: Carnegie Mellon University, pp.174-184.

Reuther, U. 2003. Two in one - can it work? Readability and translatability by means of controlled language. IN: European association for machine translation: international workshop, May 2003. Dublin: Dublin City University, pp.124-132.

Roturier, J. 2006. An investigation into the impact of controlled English rules on the comprehensibility, usefulness and acceptability of machine-translated technical documentation for French and German users. PhD Thesis. Dublin City University.

Schriver, K. 2000. Readability formulas in the new millennium: what's the use? ACM journal of computer documentation. 24 (3), pp.138-140.

Spaggiiari, L., Beaujard, F. and Canneson, E. 2003. A controlled language at airbus. IN: European association for machine translation: international workshop, May 2003. Dublin: Dublin City University, pp.151-159.