

Högskolan i Skövde

School of Communication and Informatics

English

Complete vs Abridged:
A Readability Study of Charlotte Brontë's *Jane Eyre*

Jessica Åkerhage

English C-Course

Spring 2008

Tutor: Ingalill Söderqvist

TABLE OF CONTENTS

INTRODUCTION.....	1
1. BACKGROUND.....	2
1.1. Definition of Readability.....	2
1.2. Early Studies.....	3
1.3. Later Studies.....	5
1.4. Difficult and Easy Language.....	10
2. METHOD.....	12
2.1. The Noun Phrase.....	12
2.2. The Verb Phrase.....	14
2.3 The Clause.....	15
2.4. The Fog Index.....	17
3. MATERIAL.....	17
4. ANALYSIS.....	18
4.1. The Noun Phrase.....	18
4.1.1. The Complete version.....	18
4.1.2. The Abridged Version.....	20
4.1.3. Comparative Discussion.....	21
4.2. The Verb Phrase.....	22
4.2.1. The Complete version.....	22
4.2.2. The Abridged version.....	23
4.2.3. Comparative Discussion.....	23
4.3. The Clause.....	24
4.3.1. The Complete Version.....	24
4.3.2. The abridged Version.....	26
4.3.3. Comparative Discussion.....	27
4.4. The Fog Index.....	29
CONCLUSION.....	30

WORKS CITED.....32

Introduction

This essay deals with the issue of readability, the term *readability* referring to what it is that makes a reader perceive a text as difficult or easy. Some factors are related to the reader but there are also those which depend on the text as such, one such factor being style which is the one that will be focused on in this essay.

The investigation is based on the analysis and comparison of a complete version and an abridged version of Charlotte Brontë's *Jane Eyre*, and the questions to be investigated are whether the author of the abridged version has succeeded in making it less complicated, and if he or she has done so by considering stylistic features said to be affecting readability.

Further, this essay is divided into four chapters. The first chapter contains the background for the analysis and is divided into 4 parts dealing with the following aspects: the definition of readability, early research on readability, later research on readability, and difficult and easy language. Chapter two describes the limitations made and the method used for the analysis which involves looking at the noun phrase, the verb phrase, and the clause. Chapter three gives a detailed description of the corpus investigated. Moving on to chapter four, this is where the results of the investigation are presented. This is done by dividing it into four different subchapters, each of them dealing with issues related to the different areas described in the method. Each of the subchapters then begins with the presentation of the results for each edition which is then followed by a comparative discussion. The essay ends with a conclusion part where conclusions regarding the four areas presented in the analysis are made along with the answering of research questions.

1. Background

1.1. Definition of Readability

Why are some texts considered to be more complex than others? What is it that makes a reader perceive a text as difficult or easy? This is what the term *readability* implies.

Readability refers to all those factors that have an impact on the reading of a text. On one hand there are factors related to the reader, and on the other, factors connected to the text as such. According to Christer Platzack (1971 2-3) elements depending on the reader which may facilitate the reading process, or complicate it, are: reading skill, prior knowledge, and motivation. He states that reading skill varies between individuals but often corresponds to the reader's level of education, prior knowledge implies that the reader is already, at some level, familiar with the subject treated in the text. He also states that motivation is important because the more motivated a reader is to read a certain text the higher difficulty it may have without causing problems to the reading process. Regarding features depending on the text as such it is stated, in William H. DuBay's article *The Principles of Readability*, that they can be divided into four main categories: content, style, design (also described as format), and structure (also described as features of organization), all referring to different aspects of the text. Content refers to its propositions, organization, and coherence, style to the semantic and syntactic elements of the text, design to such things as the typography, format, and illustrations, and structure to the chapters, headings, and navigation (17-18).

1.2. Early Studies

Platzack (1971 4) states that readability can be traced back so long as to the 9th century when Talmudists made calculations of the density of words and ideas in a number of texts. He also points out that it was not until in the 19th century that serious research on readability began, and that the main part of the research in this field has been conducted in the United States (4). According to DuBay (4), before the mid- 19th century, students in the United States were not divided into groups according to grade but they learned to read from books that their families owned, which were often hornbooks and bibles. These books did not constitute the best material for readers, the hornbooks being too easy and *The Bible* too difficult. The first investigations on readability were conducted on children and made it possible for pedagogues to group students according to grade, and to supply the students of each grade with proper reading material corresponding to their reading ability (4). The readability issue began to

interest not only educators, but also journalists, publishers, and librarians realised that it could be advantageous to be able to decide how easy or difficult a text was (4), and therefore the investigations on readability came to include adults as well as children (Platzack 1971 4).

DuBay states that the aim of the first readability studies was to create practical methods which would make it possible to easily match texts with the reading abilities of adults and students, and that this principally resulted in the development of easily applied readability formulas which educators, and other interested parties, could use (10). He further states that during the years preceding the 1950s you can discern three important issues within readability research (10).

The first of these issues is the statistical analysis of literature, introduced in the late 19th century by L.A Sherman (DuBay 10). He compared prose writers from different periods and noticed that sentences were progressively shortened over time. Sherman decided to see if this could be statistically demonstrated and started by counting the average sentence length per 100 periods, and found that it was possible to statistically show how average sentence length had shortened over time (10). For example, he found that the average sentence length in Victorian times was 29 words per sentence, while in his own time it had decreased to 23 words per sentence (10). According to DuBay, Sherman's work was a landmark which set the direction for the research in reading that was to come by suggesting that: literature is a subject for statistical analysis, shorter sentences and concrete terms increase readability, spoken language is more efficient than written language, and, over time, written language becomes more efficient by becoming more like spoken language (10-11). Sherman also showed that writers, to a great extent, are consistent in their average sentence length, and this discovery later became the basis for the validity of the use of text samples instead of a whole text in order to predict readability (11). He also discovered that, over time, sentences were not only made shorter but also simpler and less abstract (11).

The second of the three landmark issues is the publication of vocabulary-frequency lists, which were books which listed words in English by frequency (DuBay 12). In 1921 Edward L. Thorndike published *The Teacher's Word Book*, which was the first elaborate listing of English words by frequency. It consisted of 10 000 terms (12). Thorndike later published two even more extensive listings of this sort, *The Teachers Word Book of 20,000 Words* and *The Teachers Word Book of 30,000 Words*. The vocabulary-frequency lists made it possible for educators to objectively measure the difficulty of words in texts (12). They found that the

more frequent a word is, the more familiar it is, and the more familiar a word is, the easier it is to use (12).

The third significant issue in early readability study, discussed in DuBay's article, is the first creations of readability formulas. The first formula was created by Bertha Lively and S. L. Pressey in 1923 and it was used in schools in order to predict readability of childrens' texts (14). It measured the number of different words in parts of 1,000 words and the number of words not on the first vocabulary-frequency list published by Thorndike (14). The Lively and Pressey method was tested on 700 textbooks and reached a correlation coefficient¹ of 0.80 (14). A few years later, a study was conducted by Mabel Vogel and Carleton Washburne who were the first to put focus on the text's structural characteristics by investigating ten different features, for example, kinds of sentences and prepositional phrases, as well as word difficulty and sentence length. Their formula, which was called the Winnetka formula, was the first formula that predicted difficulty by grade levels. The Winnetka formula had a high correlation with the test scores achieved in reading tests, and became the model of modern readability formulas (14).

In 1931 a study made of adult reading interests was published and one of its conclusions was that many people are limited in their reading because of the lack of adequate material, that is, the reading materials which interest them are too difficult (DuBay 15). This study caused researchers to pay attention also to the reading ability of adults, not only that of children, and so the creation of adult readability formulas began (15). Ralph Tyler and Edgar Dale published a formula based on three factors: the number of different technical words, the number of different hard non-technical words, and the number of indeterminate clauses (16). Another study is that of William S. Gray and Bernice Leary (16-19). The aim of this study was to discover what makes a book readable for adults of limited reading ability. They reached the conclusion that elements could be grouped under these four headings: content, style, design, and structure (17). Gray and Leary found that; content was most important of the four; style was almost as important as content; third in importance was design, with a slight margin over structure. The authors also found that the only group they could measure statistically was style so they concentrated on a number of style variables and created a formula based on five different features: average sentence length in words, number of

¹ “a statistic representing how closely two variables co-vary; it can vary from -1 (perfect negative correlation) through 0 (no correlation) to +1 (perfect positive correlation)” (www.sensagent.com).

different “hard” words, number of first, second, and third-person pronouns, percentage of different words, and number of prepositional phrases (17-18). In 1939 the researcher Irving Lorge created a formula for children’s reading which also became widely used for adult material as well. Lorge’s formula included three elements: average sentence length in words, number of prepositional phrases per 100 words, and number of words not on the Dale list of 769 words (20).

Another spokesperson for the need of readability was Rudolf Flesch who produced two readability formulas. One of the variables used in his first formula was affixes and another was personal references, meaning, for example, names and personal pronouns. Flesch’s second formula consisted of two parts. The first part was called the Reading Ease formula and it made use of two variables: the number of syllables and the number of words for each sample of 100 words (DuBay 20). The Reading Ease formula uses a scale from 1 to 100, where a score of 30 means that the text is very difficult, and a score of 70 means that the text is easy (21). The second part of the Flesch-formula predicts human interest through counting personal words like pronouns and names, and personal sentences like quotes, exclamations and incomplete sentences. The Reading Ease formula created by Flesch became one of the most tested and reliable formulas and it was used by many (21).

In 1952, Robert Gunning also published a formula which is known as the Fog Index (DuBay 24). The Fog Index uses two variables, average sentence length and number of hard words, that is, words with more than two syllables for each 100 words and, due to the fact that it is easy to use, it became a popular method used to establish the grade level of a text. The Formula for Fog Index is as follows:

$$FI = 0.4 * (\text{average sentence length} + \text{percentage of hard words})$$

The process of calculating the Fog Index can be described as follows: 1) count at least 100 words, never excluding any sentences, 2) calculate the average sentence length by dividing the total number of words counted by the total number of sentences, 3) calculate the percentage of hard words (words with more than two syllables) by dividing the total number of hard words by the total number of words counted, 4) multiply the percentage of hard words with 100, 5) add the average sentence length to the percentage of hard words and multiply by 0.4.

1.3. Later Studies

According to DuBay, the research from about 1950 was characterized by consolidation and deeper study: “Investigators sought to learn more about how the formulas work and how to improve them” (DuBay 25). In 1953, Wilson Taylor published a work called *Cloze Procedure: A New Tool for Measuring Readability* (27). He created this new tool because he was not satisfied with the existing readability formulas. He noticed that some works measured much easier on the readability scales than he had expected (27). Taylor also claimed that words are not the best indicators of difficulty but how they relate to one another is more important, and therefore he suggested that tests with regularly omitted words should be used, so called cloze tests (27). The cloze testing is based on the thought that readers improve their ability to fill in the missing words as their reading skills improve (27). The test score is the percentage of words filled in correctly and the lower the score, the more difficult the text was.

While earlier studies on readability only sporadically took in consideration the human interest factors affecting readability (in the Flesch formula personal words and sentences were counted), the new research established that, along with vocabulary and sentence structure, the reader’s reading ability, prior knowledge, and motivation are important contributors to the readability of a text (DuBay 28). In the military, studies on how prior knowledge, as well as text variables, affects readability were conducted. The investigators found that: “. . . while style difficulty appears to affect immediate retention of subjects who are naïve regarding material, subjects who have considerable knowledge of the material may profit little if any from an easier style of material” (qtd. in DuBay 28). In other words, out of two individuals with the same reading ability, the one who has prior knowledge of the subject treated in the text will find it easier to read than does the one who has no knowledge at all. The factor of motivation seems to have an effect of similar character on readability (29), that is, the higher the motivation is, the easier it will be to read the text (30).

Another difference between the earlier studies and the later ones was how they used different measures of readability (DuBay 30). While the first readability studies used reader comprehension as a measure (the reading tests consisted of questions regarding the content) the new ones used other measures like reader persistence and reading efficiency (30). Several researchers showed that there exists a significant relationship between reading persistence, the tendency to keep reading a text, and readability (30). For example Charles E. Swanson made an experiment where he created one easy and one more difficult version of a story and

distributed each to 125 families, and his results were quite astonishing: “A survey of readers taken 30 hours after distribution showed a gain in the easier version over the hard version of 93% of total paragraphs read, 83% in mean number of paragraphs read, and 82% in the number of correspondents reading every paragraph” (30). Regarding reading efficiency, a study was made of 120 male aviators attending a mechanics course at Chanute Air Force Base in Illinois. The aviators received two versions of the same text, one easy and one difficult, and the investigators found that the easy text, in a significant way, improved reading speed (31).

As stated before, the second wave of readability studies was of a more profound character. It sought to find out why the different factors used in early readability formulas were such good predictors of text difficulty, that is, what effect the interpretation and thinking process of the reader have on readability (Platzack 1971 13). This quest led to the collaboration between psychologists and linguists creating a new branch of research called psycholinguistics (Platzack 1971 13). The psychological theorists and linguists promoted the idea that reading was mainly an act of thinking, and, conscious of the limitations of the readability formulas, they attempted to complete them with ways to measure the coherence, organization, and content, and although they failed to do so, their studies strengthened the importance of these variables for comprehension (DuBay 31-32). In 1977 the role of coherence was investigated by Walter Kintsch who suggested measuring readability by measuring the number of propositions in a text. He found that a deficiency in coherence has greater effects on lower-grade readers than on upper-grade ones (32). According to DuBay, an attempt was made, in 1982, to use the organization of larger units of texts as a possible measure for readability. It was claimed that a text that follows a topical scheme, that is, has a logical ordering of its topic, is easier to read and generates more results:

That is, people remember more and read faster information which is logically organized with a topical plan than they do when the same information is presented in a disorganized, random fashion. . . . Thus the plan of discourse can be considered apart from content, and deserves separate consideration from researchers, as from those who are planning a composition. (qtd in DuBay 33)

Two years later, another study directed its attention towards textual coherence and found that this type of coherence is the most important feature for learning and comprehension, and that there are two types: global coherence, which is the cohesion of an idea across an entire section, chapter, or book, and local coherence, which implies the linking of ideas between and

within sentences by using connectives (DuBay 34). For local coherence, the linking that carries meanings from one phrase, clause, or sentence to another, is particularly important, for example, pronouns referring to previous nouns, conjunctions, connectives, and substitutions for a previously used phrase or clause (34).

The collaboration between psychologists and linguists also resulted in a renewed interest in deciding the level of a text through a subjective and qualitative evaluation, which was to complement the objective measurement realized by the readability formulas (Dubay 35). This subjective analysis consisted in examining vocabulary, format, content, length, illustrations, repetition of words, and curriculum, curriculum meaning educational plan. Researchers proposed some new methods of this subjective way of leveling texts. For example, R.P. Carver suggested that specific persons trained for this task, so called qualified raters, should be used (DuBay 36). Another way was introduced by Jeanne Chall and her associates who, in 1996, published *Qualitative Assessment of Text Difficulty, A Practical Guide for Teachers and Writers* which provided examples of graded passages along with layouts and illustrations to which texts could be compared (36). Besides this renewed interest of a subjective analysis for text leveling, several studies also looked at how effective it was to use formula variables not only for the purpose of deciding the level of a written text, but also in order to compose and re-write texts for different levels (37). The first studies showed that using formula variables for the transformation of a text was not a very successful method. However, when, in a later study, George R. Klare surveyed 36 studies dealing with this issue he found that: in 19 of the studies the readability variables had a significant effect on comprehension; in 6 of the studies there were mixed results; in 11 of the studies the variables had no measurable effect on comprehension (38). Klare also came to the conclusion that in the 19 studies, where comprehension was affected, the researchers had, maybe without realizing it, changed more than word and sentence length: “The best assumption, it seems to me, is that the research workers, probably with considerable effort, managed to change basic underlying causes of difficulty in producing readable versions” (qtd. in DuBay 38). The same author established that the following word factors affect comprehension: the number of content words, the frequency, familiarity, and length of these words, their concreteness or abstractness, their association value, and the number of nominalised verbs. Regarding the sentence he found the following sentence factors to be affecting comprehension: length, whether it is active or passive, if it is affirmative or negative, if it is embedded, and if it has a great depth (38).

Extensive readability surveys also looked at the solidity of numerous text variables and new formulas were created. Formulas were now often validated on the cloze test instead of on a multiple choice test, which was used for the early formulas, and the first to use the cloze test for validation was Edmund B. Coleman (DuBay 43). He produced four formulas using these variables: number of one-syllable words per 100 words, number of sentences per 100 words, number of pronouns per 100 words, and number of prepositions per 100 words. The first formula comprised only the word variable, the second both the word and the sentence variable, the third extended to include the pronoun variable as well, and the fourth formula included all of the variables. John Bormuth conducted several studies and developed a number of new readability formulas, some of them using as many as 14 to 20 variables (44). But Bormuth soon discovered that the new variables contributed little to the rigour of the two classic predictors, the word and sentence length, and so the new variables were ultimately abandoned (44). DuBay states that Bormuth's observation about the solidity of the traditional formula variables was later confirmed by other researchers:

The findings of Bormuth about the reliability of the classic variables were later confirmed by MacGintie and Tretiak (1971) who said that the newer syntactic variables proposed by the cognitive theorists correlated so highly with sentence length that they added little accuracy to the measurement. They concluded that average sentence length is the best predictor of syntactic difficulty. (45)

One example of a formula created by Bormuth is the Bormuth Mean Cloze formula which includes three variables: number of words on the Dale-Chall frequency list of 3000 words, average sentence length in words, and average word length in letters (DuBay 44). Other examples of formulas, created through later research on readability, are the Hull formula and the Advantage TASA Open Standard (ATOS) formula. The Hull formula was produced in 1979 by Leon C. Hull and it used two variables: sentence length and the density of modifiers. He claimed that an increase in the number of adjectives and adverbs before a noun lowers comprehension, and this suggestion was confirmed by the test results in his study which showed that modifier density has a negative effect on comprehension, in other words, it increases the textual complexity. According to DuBay, Hull's advice to writers is that they use shorter sentences and eliminate strings of nouns, adjectives, and adverbs as pre-modifiers placing them after the verb rather than before the noun (51). The ATOS formula is a readability formula designed for books created by researchers at School Renaissance Institute and Touchstone Applied Science Associates. The developers of this formula found that the

combination of the following three variables gives the best description of text difficulty: words per sentence, the average grade-level of words, and letters per word (54).

1.4. Difficult and Easy Language

Since English sources on this subject have been hard to find, the following Swedish sources that discusses stylistics in general, that is, do not refer specifically to Swedish stylistics, are used.

According to Osmo Wiio difficult language includes: a relatively large number of long words, quite a lot of uncommon words, many adjectives and adverbs, and many long sentences of more than 15 words. His definition of easy language is based on the opposition of difficult language, in other words it involves: many short words, many common words, a low number of adjectives and adverbs, and short sentences (109).

In 1973 Platzack published *Språket och läsbarheten: en studie i samspelet mellan läsare och text*, in which he investigates the effect a number of different linguistic factors have on readability, namely if they complicate the reading of a text. Through the testing of a series of hypotheses Platzack found several linguistic features that affected text difficulty. For example, he found that a sentence in which subordinate clauses are placed between the finite verb and the object is harder to read than a sentence where subordinate clauses are placed in final position (1973 88-90). He could also establish that a text becomes more complicated if it contains relative clauses in which the relative pronoun has been omitted (1973 64-71). Another of his investigations showed that, although long sentences complicate the reading of a text, a text with too short sentences will have a similar effect on readability. His experiment showed that a text with an average sentence length of about 13 words is easier to read than one which has an average length of 9 words (1973 113-117). Another of the investigations made by Platzack emphasized that a sentence which does not follow the weight principle is more difficult to read than one which complies with this rule (1973 94-98). Platzack (1973) explains the weight principle in the following manner: "A line which consists of a single word is generally lighter than a line which consists of a group of words....Light lines are generally placed before heavy lines.... This applies to the sentence as well as to [...] noun -, adjective-

and adverb phrases... ”² (”Ett led som består av ett enda ord är i regel lättare än ett led som består av en ordgrupp.... Lätta led placeras som regel före tunga led.... Detta gäller såväl satsen [...] som substantiv-, adjektiv- och adverbfraser...”) (qtd. in Platzack 1973 90).

Another survey which gives an account of linguistic features complicating a text is the study *Åtta texter ur Dagens Nyheter: en läsbarhetsstudie*, published by Ulf Teleman et al in 1973. Teleman and his associates discovered, for example, that a text with many common words is more easily read than a text containing many uncommon words (45-46), and that a text which contains a great variety of lexical words is more difficult than a text with a small variety of lexical words (49-50). Regarding the noun phrase, Teleman et al (61) state that pre-modifications make the noun phrase more laborious to read, which is a view also communicated by Leon C.Hull (DuBay 51). Teleman et al also found that a passive construction is more difficult than an active one (63), and that: “a text is more difficult to read if its sentences are long, if its sentences consist of many clauses, if the subordinate clauses are long and/or many (70).”

In the *Longman Grammar of Spoken and Written English*, the noun phrase is referred to as often having a complex construction: “Noun phrases...frequently have a very complex structure and can contain several layers of embedding. . . . As phrases can also contain clauses (e.g. relative clauses modifying noun phrases and degree clauses modifying adjective phrases), and as clauses are in their turn built up of phrases, the complexity can be very great indeed” (113). An adequate example is provided of this embedding in noun phrases:

the direct result of the continuing loss of yet another typical feature of the English countryside. (113)

From this sequence it is possible to discern 4 different noun phrases. The first NP includes the whole sequence, the nominal head being *result*. The second NP, which starts at *the continuing* and extends to the full stop, has the word *loss* as a head. The third NP, in which *feature* functions as a nominal head, begins at *yet another* and continues to the full stop. The fourth and final NP begins at *the English* and extends to the full stop and its head being *countryside*.

² Quotations extracted from sources written in other languages than English have been translated by the author of this essay.

2. Method

For this study on readability it is necessary to establish some restrictions in its outline. A readability investigation discusses factors related to the text such as content, style, design, and structure, and factors which are connected to the reader reading ability, prior knowledge, and motivation. Out of the four main groups of textual characteristics influencing readability, mentioned in DuBay's article, only the category referred to as style is being analysed here. This limitation to style is based on the conclusions made by William S. Gray and Bernice Leary in their study *What Makes a Book Readable*, where they discovered that style, only preceded with a slight margin by content, was the second most important category, and that it was the only one of the four categories which could be analysed objectively and statistically. Further, the author of this essay, although aware of the importance semantic elements have on readability, chooses to investigate only syntactic elements that affect readability. This is done by looking at the following syntactic constituents: the noun phrase, the verb phrase, and the clause. In addition to the analysis of these three syntactic constituents the readability formula referred to as the Fog Index is applied to both primary texts.

2.1. The Noun Phrase

According to Sidney Greenbaum the basic structure of the noun phrase is as follows:

- Determiner
- Pre- modification
- Head
- Post- modification

Based on their order of appearance in relation to each other the determiners can be divided into three groups: pre-determiners, central determiners, and post-determiners. The pre-determiners consist of the words *all*, *half*, *both*, and multipliers like, for example, *once*, *double*, and *two times* (Greenbaum 75). Central determiners include the definite and indefinite articles, demonstratives like, for example, *this* and *that*, possessives like, for example, *her* and *our*, wh- words like *which* and *what*, the universal *every* and *each*, the non-assertive *either*, the negative dual *neither*, the general assertive *some*, the general non-assertive *any*, the quantitative *enough*, and the negative *no* (Greenbaum 73-74) . Post-determiners are ordinals like *first* and *other*, and quantifiers like *six* and *plenty of* (Greenbaum 77).

Pre-modifiers can be adjectives like in *an evil plan*, participles like in *a singing girl*, genitive constructions like in *the boy's cap*, nouns like in *a country girl*, and adverbs of different kinds like in *a far-away cottage* (Greenbaum 383).

According to Biber et al the head of a noun phrase can be a noun, a pronoun, or a proper noun but that they differ in frequency and ability to take modifiers:

Common nouns are the most frequent type of noun phrase head, and they are also the most productive head type freely occurring with both premodifiers and postmodifiers. . . . However, [pronouns] represent the opposite end of the continuum from nouns in the extent to which they take modifiers: examples of a . . . pronoun with either a premodifier or a postmodifier are quite rare. . . . Proper nouns and other naming expressions usually do not occur with a modifier (581-583)

Reviewing post-modification Greenbaum states that it can appear in the form of finite clauses, non-finite clauses, prepositional phrases, adjectives, adverbs, and appositions, as in:

The girl *who stands outside* is Mary. (finite clause)

The girl *standing outside* is Mary. (non-finite clause)

It is a book *about Mary's life*. (prepositional phrase)

The girl *present* is Mary. (adjective)

The girl *outside* is Mary. (adverb)

Mary, *the dentist*, is rich. (apposition)

Adjectives constitute the most common group of pre-modifiers and the *-ed* and *-ing* participles are relatively uncommon as pre-modifiers (Biber et al 589). Prepositional phrases are the most common post-modifiers (Greenbaum 375) and, according to Biber et al noun phrases with either pre-modification, post-modification, or both are more frequent in language of formal character than informal (578). In this essay, the analysis of the noun phrase is focused on its pre- and post-modifications and, since et al state that pronouns and proper nouns often do not occur with either of these, the current investigation only analyses phrases where a noun acts as head of the phrase. Such noun phrases as exemplified on page 11 in this essay are analysed as one constituent, that is, as one single noun phrase, and not as several independent

phrases. Since post-modification may be embedded they are classified according to their first constituent. Also, Since some determiners could be considered as proper noun phrases, for example, *two of her daughters*, *a lot less than a pound*, *a great deal of money*, it is necessary to inform that they are treated in this essay as determiners and not as noun phrases.

2.2. The Verb Phrase

The structure of a verb phrase is defined in *Longman* (99) as follows: “Verb phrases contain a lexical verb or primary verb as head. . . , either alone or accompanied by one or more auxiliaries.” Verbs can be divided into three main categories: lexical verbs, primary verbs, and modal verbs. Lexical verbs can stand on their own, primary verbs can either stand alone or co-occur with a lexical verb, modal verbs cannot occur without a lexical verb or a primary verb functioning as a lexical verb (et al 358).

According to Greenbaum the verb phrase is important in displaying contrasts as, for example, tense, aspect, finiteness, and voice (42). This investigation considers the aspect of voice, since this is one factor which has been said to complicate the reading of a text (Teleman et al 63).

Voice

A verb phrase where the verb is transitive can express either the active or passive voice, the active voice being the most common form. The typical passive constructions consist of a form of the primary auxiliary *be* followed by an *ed*-participle, or a form of the verb *get* followed by an *ed*-participle (Biber et al 475).

- The girl *was followed* home by a cat. (*be* + *ed*-participle)
- The girl *got caught* by a net. (*get* + *ed*-participle)

The passive form constructed by using *get* is, on the whole, considerably less frequent than the *be*-construction and occurs mostly in informal language, and, in total, the passive construction is much more common in formal language (Biber et al 476). Biber et al also discusses three other forms of passives, namely, three non-finite constructions which occur either as a post-modifier of nouns or as a complement of a verb (935-937).

- It was a present *wrapped* in newspaper (*ed*-participle as postmodifier of noun)

- If unkind words continue *to be said* it will be the end of their friendship. (infinitive as verb complement)
- My brother is having his car *repaired* at the garage. (*ed*-participle as verb complement)

The *ed*-participle as a post-modifier is most frequent in academic prose and passives in non-finite positions are not very frequent (Biber et al 631, 937).

2.3. The Clause

There are two types of clauses: main clauses and subordinate clauses. The basic characteristics of a main clause are that it contains at least a subject and a finite verb (other clause constituents like complement, object, and adverbial not being obligatory) and that it is able to stand on its own still conveying a meaningful expression (Greenbaum 204).

According to Greenbaum “The structural relationship between the reporting clause and direct speech is problematic (298)”. In this essay both the direct speech (*Speak*) and the reporting clause (*he urged*) in the sentence *Speak, he urged* have been classified as main clauses.

In contrast to a main clause, a subordinate clause cannot stand on its own but is dependent on a superordinate clause which is commonly a main clause but which can also be another subordinate clause (Greenbaum 283). As opposed to main clauses, subordinate clauses can, as well as containing finite verbs, contain non-finite verb forms like the *ing*-participle, the *ed*-participle, and the marked or unmarked infinitive, or they can completely leave out the verb (Greenbaum 285-288). This is illustrated by these two sentences:

- Mary predicted that John would be very happy to see the car.
- When possible, he wants to go abroad.

In the first sentence there is a main clause (*Mary predicted*), two subordinate clauses, one finite (*that John would be very happy*), and one non-finite (*to see the car*). The *that*-clause is subordinate to the main clause and the *to*-clause is subordinate to the *that*-clause. In the second sentence there are two clauses: a main clause (*he wants to go abroad*) and a subordinate verbless clause (*When possible*). According to Greenbaum a verbless clause is a clause in which: “it is usually possible to postulate a missing form of the verb *be* and to

recover the subject, when omitted, from the context” (287). According to Teleman et al a text is more difficult if its sentences contain many clauses and if the subordinate clauses are many (70). Platzack (1973 64-71) states that a relative clause in which the pronoun is omitted is more difficult than one in which it is not.

- The car *that* I bought yesterday is red (relative pronoun present)
- The car I bought yesterday is red (relative pronoun omitted)

When a passive construction is applied changes not only occur in the verb phrase but the clause changes too, and regarding the clause, there are two kinds of passives referred to as the *long passive* and the *short passive*. In the long passive the agent is expressed by using a *by*-phrase and in the short passive the agent is omitted (Biber et al 935).

- The necklace was bought *by Mary* (agent expressed through *by*-phrase)
- The necklace was bought for Mary (agent unexpressed)

Biber et al state that, in total, long passives are much less common than short passives, and that the long construction can be described as personal whereas the short one is impersonal:

The short. . . passive makes it possible to eliminate the participant that would have been expressed in the subject of the corresponding active construction, i.e. normally the agent. As the agent is most typically human, it is no doubt correct to describe the short dynamic passive as impersonal. . . . The long passive preserves all the information that would be expressed in the corresponding active construction; therefore it cannot be described as impersonal. (943)

In the following analysis the following punctuation marks are thought of as introducing new clauses: full stop, colon, semi-colon, hyphen, exclamation mark, a question mark, and parenthesis. Other signs of a new clause beginning are: coordinating and subordinating conjunctions, capital letters which are other than the personal pronoun *I* or a proper name which is not preceded by a punctuation mark.

2.4. The Fog Index

The Fog Index is a method used to establish the grade level of a text by estimating how many years of formal education is required in order to read a text with relative ease. The Formula for calculating the Fog Index is as follows:

$$FI = 0.4 * (\text{average sentence length} + \text{percentage of hard words})$$

This investigation applies the Fog Index to both versions of *Jane Eyre* and in each version the Fog Index is calculated for three different passages, namely on the first page, the middle page, and the last page of each text. At least 100 words are counted and since no sentences are excluded, the number of words may vary between passages, depending on if the 100th word is not the word ending a sentence. Out of the three calculations the medial value will be ascribed to respective version. This is done for the purpose of establishing whether the author of the abridged edition has succeeded in making the text less complicated.

3. Material

This investigation is based on the analysis and comparison of a complete version and an abridged version of the novel *Jane Eyre* written by Charlotte Brontë in 1847. The complete version was published by the Penguin Group in 1994 and consists of 438 pages and 38 untitled chapters. The abridged edition, on the other hand, was published in 2008 by Pearson Education Limited, in association with Penguin Books, and it contains 151 pages, and 44 chapters which have all been given titles. The adapted text constitutes one out of six levels of abridged versions of *Jane Eyre*, each version gradually increasing in difficulty. The texts are levelled according to the number of headwords they contain, for example, level 1 contains 300 headwords, while level 6 includes 3000 headwords. Level 1 is recommended to beginners and level six is said to be appropriate for advanced readers. The version which is subjected to analysis in this investigation belongs to level 5, contains 2300 headwords and is defined as suitable for upper-intermediate readers. Due to differences in typography, the complete version having smaller letters and more words per page than the abridged version, it has been considered necessary to investigate more pages in the abridged version than in the complete one. With this in mind, every fifth page, adding up to a total of 30 pages, are examined in the abridged edition while in the unabridged version every eighteenth page, adding up to a total of 24 pages, are analysed.

4. Analysis

This chapter is divided into four subchapters, each related to the four different areas described in the method namely the noun phrase, the verb phrase, the clause and the Fog Index. Each of the four subchapters begins with the presentation of the results attained and is then followed by a comparative discussion. Since the proportion of noun phrases, verb phrases, and clauses found and analysed in each edition does not amount to an equal number, the comparative discussion bases its observations and arguments on the percentage for each phenomenon. In addition, all the percentage numbers have been rounded off, that is, no decimals are given. However, there are some cases where one decimal is given because it is not possible to round off properly and still reach exactly 100 percent.

4.1. The Noun Phrase

4.1.1. The Complete Version

In the complete version a total of 1149 noun phrases have been investigated, and out of these a little more than half have no modification at all, they consist only of one or several determiners and a head or of a head without a determiner. Most of those phrases which are in fact modified appear with either pre-modification or post-modification, the ones appearing with both being much less frequent. For a more clear view of the differences mentioned above, see tables 1 and 2.

Table 1

Occurrence NP

Noun Phrase	Nr	%
Modified	505	44%
Unmodified	644	56%
Total	1149	100%

Table 2

Occurrence Modification

Modification	Nr	%
Pre-modification	226	45%
Post-modification	204	40%
Pre- and Post-modification	75	15%
Total	505	100%

Looking at the type of pre- and post-modifications the following observations have been made. While pre-modification by participle, adverbs, nouns or genitive constructions are not very frequent, the pre-modification by adjectives is very much so. Regarding post-modification, in more than half of the cases it consists of a prepositional phrase, the second most frequent type is the finite clause, the third post-modification is the non-finite clause, and, besides these three types of post-modification, post-modification by appositions, adverbs, and adjectives are not at all frequent. The tables 3 and 4 display the number which has been achieved, respectively, by each of the different types of pre- and post-modification.

Table 3

Occurrence Pre-modification

Pre-modification	Nr	%
Adjectives	277	80%
Participles	40	12%
Nouns	18	5%
Genitives	8	2%
Adverbs	3	1%
Total	346	100%

Table 4

Occurrence Post-modification

Post-modification	Nr	%
Prepositional phrase	180	65%
Finite clause	50	18%
Non-finite clause	29	10%
Apposition	9	3%
Adjectival	7	2%
Adverb	4	1%
Total	279	100%

The noun phrases analysed in the complete edition largely follow the basic structure which is proposed by Greenbaum (see chapter 2.1). However, three different types of deviation from this standard structure have been found.

In one case a strange determiner construction is used: *A third half the contents of my mug of coffee*. The central determiner (*A*) is here followed by a pre-determiner (*third*) and the post-determiner (*half*) which in turn is succeeded by another central determiner (*the*), when, according to Greenbaum, the construction should be: pre-determiner, central determiner, and post-determiner (73-77).

In three cases the pre-modification has been placed before instead of after the determiner: *So haughty a reliance, So low a key, Too sharp a pang*.

In one case the nominal head has been subjected to ellipsis: *I did not like to put my hand into a hireling's, but it is pleasant to feel it circled by Jane's little fingers*. In this sentence the noun phrase *a hireling's* has an absent head in the noun *hand*.

4.1.2. The Abridged Version

The total amount of noun phrases found and investigated in this edition is 959. Most of these 959 noun phrases appear without any modification at all and, among those which are modified in some way, the ones occurring with only post-modification constitute the largest group. The second most common group of modified noun phrases are those that are solely pre-modified, the least frequent group being phrases which appear with both pre- and post-modification at the same time. The total number of noun phrases pertaining to each of the groups mentioned above is shown in tables 5 and 6.

Table 5

Occurrence NP

Noun Phrase	Nr	%
Modified	374	39%
Unmodified	585	61%
Total	959	100%

Table 6

Occurrence Modification

Modification	Nr	%
Pre-modification	142	38%
Post-modification	184	49%
Pre-and Post-modification	48	13%
Total	374	100%

Regarding the character of the pre-modifications found in this part of the material the most common pre-modifying type is undoubtedly the adjective. Among the other types of pre-modifiers the noun is the most frequent and it is followed by the participle, the genitive construction and the adverb which constitute the less frequent groups. Turning to post-modification, the most frequent type is the prepositional phrase and it is followed by the non-finite clause, the finite clause, the apposition, the adverb, and the adjective. Tables 7 and 8 show the number and percentage reached by each group of pre-modification and post-modification.

Table 7

Occurrence Pre-modification

Table 8

Occurrence Post-modification

Pre-modifications	Nr	%
Adjectives	185	87.7%
Nouns	13	6.1%
Participles	8	3.8%
Genitives	4	1.9%
Adverbs	1	0.5%
Total pre-modification	211	100%

Post-modifications	Nr	%
Prepositional phrase	134	58%
Non-finite clause	35	15%
Finite clause	33	14%
Apposition	14	6%
Adverbial	9	4%
Adjectival	7	3%
Total post-modification	232	100%

In the abridged edition all noun phrases but one follow the basic structure described by Greenbaum. The deviation consists in that the pre-modification has been placed before the determiner: *Too unimportant an activity*.

4.1.3. Comparative Discussion

In both versions over half of the noun phrases appear without modification. However, the percentage of these phrases is somewhat higher in the abridged version where 61% percent of the noun phrases belong to this group, compared to 56% in the complete edition. If one considers what has been stated by Biber et al (578), namely, that noun phrases which are modified in some way are most frequent in formal language, this discovery might be an indication that the language in the abridged edition is of a somewhat more informal character, at least when looking at the noun phrase. Out of those noun phrases that were modified there is an interesting observation to be made; while, in the complete edition, noun phrases occurring exclusively with pre-modification (45%) are more frequent than those which appear with only post-modification (40%), the exact opposite applies to the abridged version where pre-modification (38%) is surpassed by post-modification (49%). This observation coincides with the view communicated by both Telemann et al (61) and Hull (DuBay 51) that pre-modifications should be eliminated since they make the noun phrase more difficult to read. Regarding phrases occurring with both pre- and post-modification there is only a slight difference between the two editions, 15% being of this sort in the complete version and in the abridged 13% and, in both versions, they are less frequent than those phrases which occur with either pre- or post-modification.

Looking at the character of the pre-modifications of the noun phrases it is obvious that, in both editions, adjectival pre-modifiers constitute the largest group. Although this is the case the percentage of adjectival pre-modifiers is higher in the abridged version (87.7%) than in the complete (80%). Another difference is observed regarding the use of the participle as pre-

modifier; it is less common in the abridged edition where it is the third most common modifier attaining a percentage of 3.8%, whereas in the complete version it is the second most frequent modification reaching a percentage of 12%. According to Biber et al the participle is a relatively uncommon pre-modification (589). Hence the fact that it has not been used as frequently in the abridged edition as in the complete one might suggest that the author of the easier version has considered this fact when rewriting the text. Comparing the three other groups of pre-modifiers, namely, nouns, genitives, and adverbs, there is no difference between the two versions worth mentioning.

Regarding the character of post-modification it is no surprise to find that post-modification by prepositional phrase is the most frequent one in both editions, since this is stated to be the most common post-modifier (Greenbaum 375). However it was a bit of a surprise to find that, even though the prepositional phrase was the most frequent one in both versions, it was more common in the complete edition (65%) than in the abridged (58%). Another difference between the two editions is how the non-finite clause is allowed more space in the abridged version (15%) than in the complete (10%) since non-finite constructions are considered to be somewhat more difficult than finite constructions.

4.2. The Verb Phrase

4.2.1. The Complete Version

The total amount of verb phrases encountered and analysed in this edition is 1771 of which the great majority is expressing active voice. Of those verb phrases that express passive voice, the passive construction formed with the primary auxiliary *be* followed by an *ed*-participle is the most frequent type. The second most common form used is the one where the auxiliary is absent, namely, where the *ed*- participle alone expresses passive meaning. This group is followed by the infinitive which, in turn, is succeeded by the form composed of *get* acting as an auxiliary and an *ed*- participle. Tables 9 and 10 give the exact number and percentage for each of the forms mentioned above.

Table 9

Occurrence Verb Phrase

Table 10

Occurrence Passive Construction

Voice	Nr	%
Passive	110	6%
Active	1661	94%
Total Verb Phrases	1771	100%

Passive Construction	Nr	%
Be+ participle	65	59%
Get+ participle	1	1%
Ed- participle	34	31%
Infinitive	10	9%
Total Passives	110	100%

4.2.2. The Abridged Version

The number of verb phrases found in the abridged edition is 1638 and the verb phrases that were found to express passive voice are of a considerably smaller amount than those expressing active voice. The most frequent passive construction is that which is formed with the auxiliary *be* succeeded by the *ed-* participle, the second most common structure being the *ed-* participle standing alone without an auxiliary. The passive as expressed through the infinitive is not at all common and the structure where *get*, used as an auxiliary, is followed by an *ed-*participle is non-existent. The number achieved by each group is shown in table 11, dealing with the matter of voice, and table 12, referring to passive constructions.

Table 11

Occurrence Voice

Voice	Nr	%
Passive	95	6%
Active	1543	94%
Total Verb Phrases	1638	100%

Table 12

Occurrence Passive Construction

Passive Construction	Nr	%
Be+ participle	63	66.3%
Get+ participle	0	0%
Ed- participle	26	27.4%
Infinitive	6	6.3%
Total Passives	95	100%

4.2.3. Comparative Discussion

Considering the statement made by Teleman et al (63) that a passive construction is more difficult than an active construction, one should expect to find fewer passive constructions in the abridged version, but this has not been the case here. The distribution between verb phrases expressing active or passive voice is exactly the same in both editions where the active verb phrase (94%) is considerably more frequent than the passive (6%). The fact that nothing seems to have been done in the abridged version, in terms of decreasing or completely eliminating passive constructions, may be the result of the already low number of passives occurring in the complete version; because of their infrequent use in the complete

edition it has not been thought to be of importance to further lower their occurrence. However, one thing worth noting is that in the abridged edition passives formed by the auxiliary *be* followed by an *ed*-participle, which is stated to be the most common form of all (Biber et al 475), are more commonly used (66.3%) than in the complete version (59%). This might be one way in which the author of the easier version has, if not tried to decrease the number of passive constructions, at least tried to make them more readable. Regarding the other structures used to form the passive they follow the exact same order in both versions, that is, the *ed*-participle is the second most frequent type, the infinitive is the third most common, and the least common is the verb *get* followed by an *ed*-participle appearing only once in the complete version (0.9%) and not at all in the abridged (0%). The use of the *ed*-participle and the infinitive are somewhat more frequent in the complete edition where the percentage for respective type is 30.9% and 9.1%, compared to the abridged edition where the percentage is 27.4% and 6.3%, respectively. Since it has been stated that the *ed*-participle is most frequent in formal language and that passives in non-finite forms are not very frequent (Biber et al 631, 937), this observation can be considered as an indication that the language in the complete edition is more complex than that which is used in the abridged.

4.3. The Clause

4.3.1. The Complete Version

In the complete version a total number of 1850 clauses have been found and analysed, and of these a little more than half are found to be main clauses. Regarding the structure of the subordinate clauses the two most frequent constructions are the finite and non-finite constructions, the subordinate clause with an omitted verb phrase not being very common. In table 13 the distribution between main- and subordinate clauses is displayed and in table 14 the frequency of each of the subordinate clause structures is accounted for.

Table 13

Occurrence Clause

Table 14

Occurrence Subordinate Constructions

Clause	Nr	%
Main	1098	59%
Subordinate	752	41%
Total Clauses	1850	100%

Subordinate Constructions	Nr	%
Finite	424	56%
Non-finite	292	39%
Verbless	36	5%
Total Subordinate Clauses	752	100%

Turning to the relative clause and the passive clause construction, the most frequent type of relative clauses is the one where the relative pronoun has been omitted, and, among the short and the long passive, the most common construction is found to be the short passive. Table 15 and 16 show the number and percentage achieved respectively by each group pertaining to the two different phenomena revised above.

Table 15

Occurrence Relative Pronoun

The Relative Clause	Nr	%
Pronoun Present	47	44%
Pronoun Absent	60	56%
Total Relative Clauses	107	100%

Table 16

Occurrence Passive Clause Construction

The Passive	Nr	%
Long Passive	20	18%
Short Passive	90	82%
Total Passive Clauses	110	100%

The number of sentences which constitute the text extracted and investigated in the complete version is counted to 555. The sentences in this version have been found to contain everything from one single clause to as many as 15 clauses. Regarding the number of subordinate clauses included in sentences, it ranges from one to 12 subordinate clauses. The number and percentage referring to the amount of clauses contained by a sentence are displayed in table 17 and the number and percentage which show how many subordinate clauses are included in a sentence are presented in table 18.

Table 17

Clause Frequency in Sentences

Table 18

Subordinate Clause Frequency in Sentences

Clauses	Nr	%
1	149	26.8%
2	122	22%
3	79	14.2%
4	64	11.5%
5	50	9%
6	36	6.5%
7	15	2.7%
8	12	2.2%
9	11	2%
10	6	1.1%
11	3	0.5%
12	3	0.5%
13	2	0.4%
14	2	0.4%
15	1	0.2%
Total Sentences	555	100%

Subordinate Clauses	Nr	%
0	236	42.5%
1	140	25.2%
2	68	12.3%
3	53	9.5%
4	24	4.3%
5	12	2.2%
6	9	1.6%
7	4	0.7%
8	7	1.3%
11	1	0.2%
12	1	0.2%
Total Sentences	555	100%

4.3.2. The Abridged version

In this version the total number of clauses analysed amounts to 1733 and of these clauses the majority are main clauses. Concerning the different structures of the subordinate clauses, the two undoubtedly most common ones are the finite and non-finite structures while the most infrequent one is the verbless construction. The Tables 19 and 20 below account for the number and percentage of main- and subordinate clauses and the three different structures of subordinate clauses, respectively.

Table 19

Occurrence Clause

Clause	Nr	%
Main	1163	67%
Subordinate	570	33%
Total Clauses	1733	100%

Table 20

Occurrence Subordinate Constructions

Subordinate Constructions	Nr	%
Finite	324	57%
Non-finite	217	38%
Verbless	29	5%
Total Subordinate Clauses	570	100%

Referring to the relative clause and the passive clause structure the relative clause where the relative pronoun is present is the most frequent one. Out of the two different passive clause constructions, the short passive surpasses the long passive. In tables 21 and 22 the

differences referred to above are displayed showing the number and percentage for each group related to the relative clause and the passive clause, respectively.

Table 21

Occurrence Relative Pronoun

The Relative Clause	Nr	%
Pronoun Present	59	60%
Pronoun Absent	39	40%
Total Relative Clauses	98	100%

Table 22

Occurrence Passive Clause Construction

The Passive	Nr	%
Long Passive	14	15%
Short Passive	81	85%
Total Passive Clauses	95	100%

The total amount of sentences which compose the text analysed in the abridged edition is counted to 843 and the number of clauses included in these sentences ranges from 1 to 7. The number of subordinate clauses contained within a sentence varies from 1 to 5 clauses. The distribution between the number of clauses constituting a sentence and the number of subordinate clauses included in a sentence is described in numbers and percentages in tables 23 and 24 below.

Table 23

Clause Frequency in Sentences

Clauses	Nr	%
1	353	41.9%
2	263	31.2%
3	121	14.3%
4	58	6.9%
5	31	3.7%
6	15	1.8%
7	2	0.2%
Total Sentences	843	100%

Table 24

Subordinate Clause Frequency in Sentences

Subordinate Clauses	Nr	%
0	497	58.9%
1	206	24.4%
2	81	9.6%
3	37	4.4%
4	19	2.3%
5	3	0.4%
Total Sentences	843	100%

4.3.3. Comparative Discussion

Regarding the issue of main or subordinate clause, the number of main clauses in the abridged version (67%) surpasses the number found in the complete (59%). This observation also, logically, entails the fact that the subordinate clause is less frequent in the abridged version (33%) than it is in the complete (41%). Since the main clause is considered to be less complex in its structure than the subordinate clause, these findings are clear indications that the

language in the abridged version is of a less complex character than that which is used in the complete version.

Looking at the character of the subordinate clauses, that is, whether they are finite, non-finite, or verbless, there are no differences worth mentioning. In both versions the finite subordinate clause is the most frequent of the three, the non-finite construction is the second most common, and the subordinate clause with an omitted verb phrase is, undoubtedly, the least common of all reaching the same percentage (5%) in both versions.

Concerning the issue of an absent or present relative pronoun in relative clauses, the number of clauses where the pronoun is present is greater in the abridged version than in the complete. While in the abridged edition 60% of the relative clauses analysed included the relative pronoun and 40 % excluded it, it was more absent (56%) than present (44%) in the complete version. Since Platzack (1973 64-71) states that relative clauses with an omitted relative pronoun are more difficult to read than those which include it, this result suggests that the author in the abridged edition has indeed considered this fact when composing an easier version.

Moving on to the passive clause construction, namely, the distribution between the long and the short passive, no difference to speak of has been found between the two editions. In both versions the short passive is the most common construction reaching 82% in the complete version and 85% in the abridged. This result is not surprising because Biber et al (1993) state that the short passive is much more common than the long passive. But Biber et al (1993) also state that while the long passive includes all the information that would be included in the corresponding active construction the short passive omits pieces of it. Considering this statement, one might have expected to find the long passive to be more frequent in the abridged version than it is since a clause which does not omit any information ought to be easier to read, but this is not the case: the long passive reaches 18% in the complete version and only 15% in the abridged.

Passing to the two last factors investigated regarding the clause, namely the total number of clauses included in sentences and the number of subordinate clauses occurring in sentences, there are quite obvious differences between the two editions. The highest number of clauses included in sentences in the complete edition is 15 while in the abridged it is only 7. In addition, the percentages of sentences in the abridged version made up by one (41.9%) or two (31.2%) clauses are larger than in the complete where the percentages reach 26.8% and

22%, respectively. Moving on to the number of subordinate clauses included in a sentence, the highest number is found in the complete edition where one sentence has as many as 12. In the abridged version the highest number of included subordinate clauses is 5. Also, sentences without subordinate clauses are more frequent in the abridged version (58.9%) than in the complete (42.5%). No considerable differences are found in the frequency of sentences containing one, two, three, or five subordinate clauses. However, it ought to be noted that the percentage of sentences containing more than three clauses in the abridged edition is 2.7% while in the complete version 10.5% contain more than three. These findings mentioned above coincide with what is stated by Telemann et al (70), namely that a text is more difficult if its sentences contain many clauses and if the subordinate clauses are many. It is obvious that the author of the abridged version has thought it useful to reduce these two factors.

4.4. The Fog Index

As expected the calculated Fog Index for the complete version exceeds the Fog Index for the abridged version in all three passages. The tables 25 and 26 below show the Fog Index achieved by each of the three passages in the two versions.

Table 25

The Complete Version

Passage	Words counted	Sentences	Difficult words	FI
1	122	8	12	10
2	130	14	6	5.5
3	112	8	6	7.7

Table 26

The Abridged Version

Passage	Words counted	Sentences	Difficult words	FI
1	100	11	3	4.8
2	116	13	3	4.6
3	109	19	4	3.7

The medial value of the Fog Index for the complete version is 7.7 and for the abridged it is 4.6. This indicates that to read the text in the complete edition with relative ease between 7 and 8 years of formal education are required while for reading the abridged edition only between 4 and 5 years are needed. It should also be noticed that while the highest Fog Index for the complete edition is 10 it is only 4.8 for the abridged. This result shows that the author

of the abridged edition has succeeded in making the text more readable, at least as concerns average sentence length and number of difficult words which are the two factors used in the Fog Index formula.

Conclusion

This essay investigates differences in readability between a complete and an abridged version of the novel *Jane Eyre*. The analysis is focused on stylistic features and, by analysing the three syntactic constituents the noun phrase, the verb phrase, and the clause, and by calculating the Fog Index for the two versions, it can be concluded that differences do exist.

The noun phrase analysis shows that unmodified noun phrases are more frequent in the abridged version than in the complete, and that noun phrases modified only by pre-modification are less common in the abridged edition than in the complete. Another observation made is that the use of the participle as pre-modifier is less common in the abridged edition.

The verb phrase investigation reveals that there is not much difference between the two editions, both editions having an equal distribution of verb phrases expressing the active and the passive voice. However, one difference, related to the different passive forms, is found and it is that the most common passive form, the auxiliary *be* followed by an *ed*-participle, is more common in the abridged edition.

The clause analysis clearly shows that the text in the abridged version contains a larger number of main clauses, and that the presence of the relative pronoun in relative clauses is greater here than in the complete version where the relative pronoun is more absent than present. Another clear difference between the two editions is how the sentences in the abridged edition consist of a smaller number of clauses and includes fewer subordinate clauses.

The Fog Index calculation resulted in the medial value of 7.7 for the complete edition and 4.6 for the abridged, meaning that to read the complete edition without difficulty you would need more than 7 years of formal education while only more than 4 years are required for the abridged. In addition, it should also be noticed that the Fog Index for each of the three passages was higher in the complete version than in the abridged (see tables 25 and 26).

In conclusion, it is clear that the author of the abridged version has indeed succeeded in making the text in the abridged edition less difficult than the text in the complete version (see chap 4.4). It is also obvious that he or she really has used the stylistic features investigated in this essay, especially those pertaining to the noun phrase and the clause, said to be complicating a text in order to make it less complex. Although some stylistic features referring to the readability are revealed in this essay, it is certain that there remain several others to discover in this complex novel created by Charlotte Brontë.

Works Cited

Primary Sources

Brontë, Charlotte. *Jane Eyre*. London: Penguin, 1994.

Brontë, Charlotte. *Jane Eyre*. Harlow: Pearson Educ. Ltd, 2008.

Secondary Sources

Biber, Douglas et al. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Educ. Ltd., 2000.

DuBay, William H. "The Principles of Readability". Impact Info. 2004. 31 March 2008
<<http://www.impact-information.com/impactinfo/readability02.pdf>>.

Greenbaum, Sidney, and Randolph Quirk. *A Student's Grammar of the English Language*. Harlow: Pearson Educ. Ltd., 1990.

Platzack, Christer. *Läsbarhetsforskning-en översikt*. Lund: Lunds Universitet, 1971.

Platzack, Christer. *Språket och läsbarheten: en studie i samspelet mellan läsare och text*. Lund: Gleerup, 1973.

Teleman, Ulf et al. *Åtta texter ur Dagens Nyheter: en läsbarhetsstudie*. Lund: Lunds Universitet, 1973.

Wiio, Osmo A. *Kommunikation-vad är det?* Trans.Kjell Nowak. Stockholm: Natur och Kultur, 1976.

www.sensagent.com. 22 April 2008.