5 ICLEHI 2017-076 Eva Tuckyta Sari Sujatna

The Readability Test of the English Children Short Stories

Eva Tuckyta Sari Sujatna*[a], Yuyu Yohana Risagarniwa[b], Rahmat Sopian[c]
[ab]Department of Linguistics, Universitas Padjadjaran, Bandung Indonesia
[c]Department of Historical Science and Philology, Universitas Padjadjaran, Bandung, Indonesia
*Corresponding Author: eva.tuckyta@unpad.ac.id

ABSTRACT

Many researchers do their research on grammatical intricacy or lexical density as the main elements of readability of texts but it is very limited research on English children short stories. The readability test of texts is different from one text to another. The texts can be classified into spoken or written text. The classification can be based on some elements, such as the total word count or the hard words involved in the texts. This preliminary research tries to figure out the readability test based on the grammatical intricacy or lexical density of the English children short stories by applying the free application called text analyser as the text content analysis tool found in http://www.usingenglish.com/resources/text-statistics.php. The research based on the two aims of the research, they are, (1) To identify the average of hard words found in the English children short story, and (2) To identify the average of grammatical intricacy or lexical density of the English children short stories. The method applied in this research is descriptive analysis. The writers describe the data based on the data which is described by the text analyser. Based on the analysis, the result shows that (1) the average of the hard words found in the English children short story is 3.06% and (2) the average of the hard words found in the English children short story is 33.62%.

Keywords: Readability test, English children short stories, grammatical intricacy, lexical density

Introduction

Every text has a specific readability level. It is found in every language, including English. The readability level, in fact, depends on the grammatical intricacy or lexical density as the main elements of readability of texts. Commonly, people argue that a text is easier or harder to understand based on their intuition. Of course, it is hard to prove that their arguments are accurate since they have different experience and ability in understanding the text so that it cannot be proven scientifically.

The texts can be classified into spoken or written text. The classification can be based on some elements, such as the total word count or the hard words involved in the texts. This preliminary research tries to figure out the readability test based on the grammatical intricacy or lexical density of the English children short stories.

To describe the grammatical intricacy and lexical density of the text can be analysed manually based on the formula introduced by Halliday and Ure. Besides the manual way, the grammatical intricacy and lexical density of the text can be analysed by applying the free application called text analyser as the text content analysis tool found in http://www.usingenglish.com/resources/text-statistics.php. The research based on the two aims of the research, they are, (1) To identify the average of hard words found in the

English children short stories, and (2) To identify the average of grammatical intricacy or lexical density of the English children short stories.

Previous Research

This research is based on the previous research. To identify whether the text is spoken or written, the writers refer to the theories from Halliday (2004) in Presnyakova (2011) which is supported by Gerot and Wignell (1995). Besides the three theories above, the writers also refer to Mode Continuum which is applied by Cruickshank (2008) in Hertzberg (2012). To describe the grammatical intricacy and lexical density of the text the writers refer to Ure's theory (1971) and Halliday (1989) and also Eggins (1994).

Mode Continuum

Mode Continuum is a term offered by Cruickshank (2008) to define written or spoken text. According to Gerot and Wignell (1995),

> "Spoken and written languages are both complex but in different ways. Spoken language tends to be complex grammatically and written language tends to be complex lexically. Spoken language tends to be grammatically intricate whereas written language tends to be lexically dense."

This argument is in line with Halliday (2004) in Presnyakova (2011) argue that

> "Typically, written language becomes complex by being lexically dense: it packs a large number of lexical items into each clause; whereas spoken language becomes complex by being grammatically intricate: it builds up elaborate clause complexes out of parataxis and hypotaxis."

From the arguments above, it can be concluded that the text can be classified as the spoken text or the written text is depended on the grammatical intricacy and the lexical density of the texts.

The following is the Mode Continuum described by Cruickshank (2008) in Hertzberg (2012) in Figure 1.

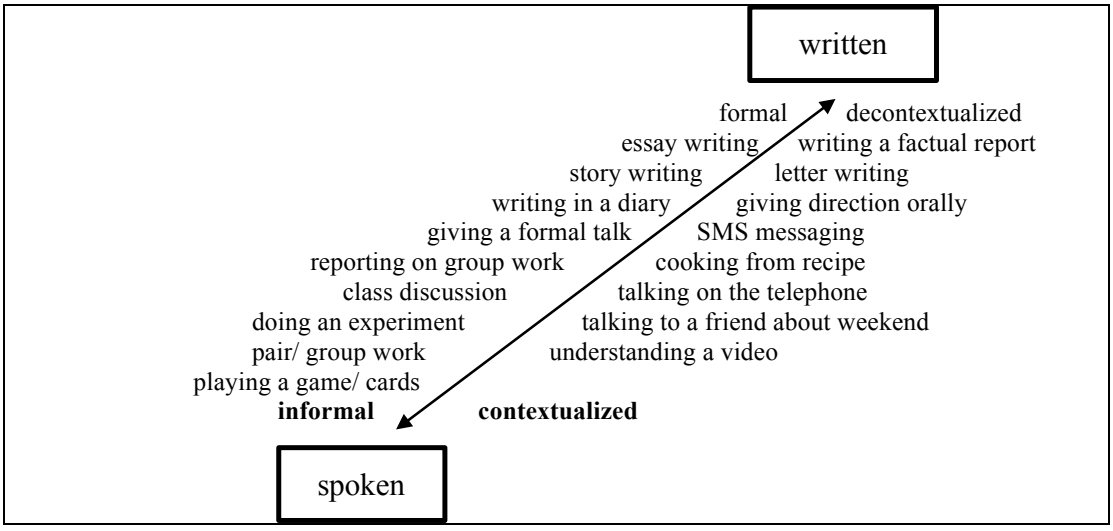

Figure 1. Mode Continuum by Cruickshank (2008) in Hertzberg (2012)

Besides the mode of continuum described by Cruickshank (2008) in Hertzberg (2012), Gerot and Wignell (1995) have introduced the scale of spoken and written text as described in the figure 3.

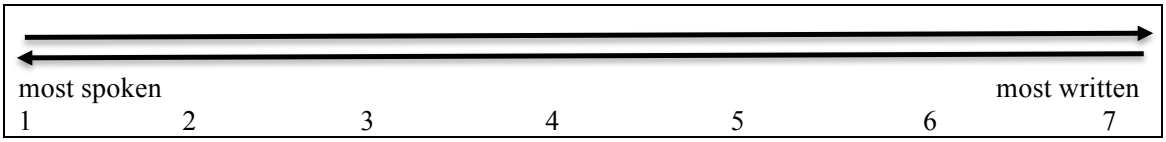

Figure 2. The Scale of Spoken and Written Text

Grammatical Intricacy and Lexical Density

Grammatical intricacy and the lexical density of the texts is being influenced each other. They are very close to the types of the text, both spoken and written text. The grammatical intricacy and lexical density can define the text as the spoken or written text. As Eggins (1994) describes in the following figure:

Spoken language	Written language
Low lexical density - few content-carrying words as a proportion of all words	High lexical density - many content-carrying words as a proportion of all words
High grammatical intricacy - many clauses per sentence	Low grammatical intricacy - few clauses per sentence

Figure 3. Spoken and Written Text (Eggins, 1994: 61)

To count the index of Grammatical Intricacy (GI), Halliday (1985) describes the total number of ranking clauses divided by the total number of clauses complexes of the texts as explained in the following formulae:

$$\text{GI (index)} = \frac{\text{Total number of ranking clauses}}{\text{Total number of clause complexes}}$$

Lexical Density (LD) or sometimes defined as complexity factor. LD is used to scale the difficultness of a text. Texts with a higher density are more difficult understood. In this paper, the writers call it as LD. Johansson (2008) in his article titled *Lexical diversity and lexical density in speech and writing: a developmental perspective* mentioned that

> "*By investigating this (lexical density), we receive a notion of information packaging; a text with a high proportion of content words contains more information than a text with a high proportion of function words (prepositions, interjections, pronouns, conjunctions and count words*)". To count the LD of a text, Ure (1971) in Johansson (2008) described in the following formulae:

$$Ld\ (\%) = \frac{total\ number\ of\ words\ with\ lexical\ properties}{total\ number\ of\ orthographic\ words}\ x\ 100$$

Figure 4. Lexical Density of Ure's in Johansson

Relating to the definition used in the application used in this research (**https://www.usingenglish.com/resources/text-statistics.php**) *words with lexical*

properties mentioned by Ure means different words while *orthographic words* means words as described in the following formulae:

Lexical Density = (Number of different words / Total number of words) x 100

Figure 5. Lexical Density of https://www.usingenglish.com/resources/text-statistics.php

Example 1:

*People like Mamma Mia. Kids like **Mamma Mia**. Teenagers love **Mamma Mia**. Parents love **Mamma Mia.** Grandparents love **Mamma Mia**.*

$$Ld\ (\%) = \frac{total\ number\ of\ words\ with\ lexical\ properties\ (9)}{total\ number\ of\ orthographic\ words\ (20)}\ x\ 100 = 45\%$$

General Statistics

Total Word Count:	**20**
Total Unique Words:	**9**
Number of Sentences:	**5**
Average Sentence Length:	**4.00**
Number of Paragraphs:	**1**
Hard Words:	**2** (10.00%) (what's this?)
Lexical Density:	**45.00%** (what's this?)
Fog Index:	**5.60** (what's this?)

Word Length Breakdown

Length	Count	Graph
3 letter words	5	25.0%
4 letter words	6	30.0%
5 letter words	5	25.0%
6 letter words	1	5.0%
7 letter words	1	5.0%
9 letter words	1	5.0%
12 letter words	1	5.0%

Figure 6. Text statistics http://www.usingenglish.com/resources/text-statistics.php.

From the text statistics of the example 1, it can be seen that that total word count is 20 words, the total unique words is nine words, it means that there are nine words (*People, like, Mamma, Mia, Kids, Teenagers, love, parents, grandparents*) in the text that appears for the first time, for the second and the third times are not counted as unique words.

Research Methods

This research employs descriptive method. It describes the phenomenon being studied relating to the data. The data obtained from children short stories written in

English. The writers took four texts to be analysed. After collecting the data, the writers analyse them by applying a free ware taken from http://www.usingenglish.com/resources/text-statistics.php. The free ware helps the writers in counting the total word, the unique words, the number of sentences, average sentence length, the number of paragraph, hard words, lexical density, and Fog Index. Later, the present writers count the average of each element to answer the research questions. Finally, from the test, the present writers can analyse the readability of each text taken from children short stories.

Result and Discussion

In this paper, the writers apply four example texts taken from four children short stories in English. After doing the analysis, the writers found the text statistics of each as described in the following figure.

General Statistics

Total Word Count:	**1119**
Total Unique Words:	**343**
Number of Sentences:	**132**
Average Sentence Length:	**8.48**
Number of Paragraphs:	**1**
Hard Words:	**21** (1.88%) (what's this?)
Lexical Density:	**30.65%** (what's this?)
Fog Index:	**4.14** (what's this?)

Word Length Breakdown

Length	Count	Graph
1 letter words	31	2.8%
2 letter words	152	13.6%
3 letter words	321	28.7%
4 letter words	275	24.6%
5 letter words	155	13.9%
6 letter words	77	6.9%
7 letter words	55	4.9%
8 letter words	13	1.2%
9 letter words	15	1.3%
10 letter words	3	0.3%
11 letter words	3	0.3%
12 letter words	2	0.2%
13 letter words	1	0.1%

Figure 7. Statistics Test Story 1

General Statistics

Total Word Count:	**888**
Total Unique Words:	**300**
Number of Sentences:	**93**
Average Sentence Length:	**9.55**
Number of Paragraphs:	**1**
Hard Words:	**26** (2.93%) (what's this?)
Lexical Density:	**33.78%** (what's this?)
Fog Index:	**4.99** (what's this?)

Word Length Breakdown

Length	Count	Graph
1 letter words	28	3.2%
2 letter words	140	15.8%
3 letter words	247	27.8%
4 letter words	179	20.2%
5 letter words	119	13.4%
6 letter words	72	8.1%
7 letter words	51	5.7%
8 letter words	15	1.7%
9 letter words	13	1.5%
10 letter words	3	0.3%
11 letter words	2	0.2%
12 letter words	1	0.1%

Figure 8. Statistics Test Story 2

General Statistics

Total Word Count:	**669**
Total Unique Words:	**250**
Number of Sentences:	**74**
Average Sentence Length:	**9.04**
Number of Paragraphs:	**1**
Hard Words:	**29** (4.33%) (what's this?)
Lexical Density:	**37.37%** (what's this?)
Fog Index:	**5.35** (what's this?)

Word Length Breakdown

Length	Count	Graph
1 letter words	13	1.9%
2 letter words	71	10.6%
3 letter words	185	27.7%
4 letter words	156	23.3%
5 letter words	66	9.9%
6 letter words	55	8.2%
7 letter words	50	7.5%
8 letter words	36	5.4%
9 letter words	22	3.3%
10 letter words	9	1.3%
11 letter words	1	0.1%
13 letter words	1	0.1%

Figure 9. Statistics Test Story 3

General Statistics

Total Word Count:	**964**
Total Unique Words:	**315**
Number of Sentences:	**92**
Average Sentence Length:	**10.48**
Number of Paragraphs:	**1**
Hard Words:	**30** (3.11%) (what's this?)
Lexical Density:	**32.68%** (what's this?)
Fog Index:	**5.44** (what's this?)

Word Length Breakdown

Length	Count	Graph
1 letter words	25	2.6%
2 letter words	124	12.9%
3 letter words	262	27.2%
4 letter words	263	27.3%
5 letter words	93	9.6%
6 letter words	66	6.8%
7 letter words	56	5.8%
8 letter words	27	2.8%
9 letter words	33	3.4%
10 letter words	3	0.3%
13 letter words	2	0.2%

Figure 10. Statistics Test Story 4

From the collected data in figure 7-10, it can be found that the words total number of the text is 669 words as the shortest and 1119 words as the longest text or the average of the words total number is 910, so that they can be classified into short stories. The average of the unique words is 302 or 33%. In the story 1, for example, there are 343 words that appear for the first time from 1119 words as the total number (or 776 words appear for the second, the third, or the fourth times in the story).

The data shows that the average sentence length is around 9.4, it is concluded that every sentence has around nine words in each text of the short stories. Besides the words total number, the average of unique words and the average of sentence length, the data shows the hard words, lexical density and fog index. Hard words mentioned in this paper are defined as words with three or more syllables. The data described that the average of

the hard words is 3.06%, in other words, the words with three or more syllables in each text is not too many.

The average lexical density of the four stories is 33.6%, it is under 40%. The last element of the statistics test is Fog Index. Fog Index is a test of readability of a text; it is designed to show how easy or difficult a text is to read. It is mentioned in http://www.usingenglish.com/resources/text-statistics.php that

> "The Gunning Fog Index gives the number of years of education that your reader hypothetically needs to understand the paragraph or text. The Gunning Fog Index formula implies that short sentences written in plain English achieve a better score than long sentences written in complicated language. For reference, the New York Times has an average Fog Index of 11-12, Time magazine about 11. Typically, technical documentation has a Fog Index between 10 and 15, and professional prose almost never exceeds 18."

From the analysis, it is shown that the average of fog index is 4.98, it is can be categorized that the text is children short stories.

Conclusion

The result of the research shows that the statistics test of the four children short stories mentioned in figure 7- 10 are summarized in the following table 1.

	Story 1	Story 2	Story 3	Story 4	average
Total Word Count:	1119	888	669	964	910
Total Unique Words:	343	300	250	315	302
Number of Sentences:	132	93	74	92	97.7
Average Sentence Length:	8.48	9.55	9.04	10.48	9.4
Number of Paragraphs:	1	1	1	1	1
Hard Words:	21 (1.88%)	26 (2.93%)	29 (4.33%)	30 (3.11%)	26.5 (3.06%)
Lexical Density:	30.65%	33.78%	37.37%	32.68%	33.6%
Fog Index:	4.14	4.99	5.35	5.44	4.98

Table 1. The Four English Children Short Stories Texts Statistics

The table 1 describes that the average lexical density of the four stories is 33.6%, it is under 40%. It can be categorized into easy text (children short stories) since a lexically dense text has a lexical density of around 60-70% and those which are not dense have a lower lexical density measure of around 40-50%.

References

Bhattacherjee, A. (2012). Social Science Research: Principles, Methods, and Practices.

Castello, E. (2008). Text Complexity and Reading Comprehension Tests. Bern: International Academic Publisers.

Eggins, S. (1994). An Introduction to Systemic Functional Linguistics. London: Continuum.

Gerot, L., & Wignell, P. (1995). Making Sense of Functional Grammar. Sydney: Gerd Stabler.

Haliday, M. (1985). Spoken and Written Language. Geelong Vict: Deakin University.

Halliday, M., & Matthiessen, C. (2004). Introduction to Functional Grammar (Third Edition). London: Edward Arnold.

Hertzberg, M. (2012). Teaching English language Learners in mainstream Classes. Newtown: Primary English Teaching Association Australia.

Johansson, V. (2008). Lexical Diversity and and Lexical Density in Speech and Writing: A Develop-mental Perspective. Dept. of Linguistics and Phonetics , pp. 61-79.

Presnyakova, I. (2011). Systemic Functional Analysis of Elementary School Language Arts Textbooks. Theses.