LEXICAL DENSITY AND READABILITY OF NON-ENGLISH MAJORED FRESHMEN'S WRITING IN VIETNAMESE CONTEXT

Nguyen Ngoc To Ngan - KOTO Foundation

nguyenngoctongan@gmail.com

Nguyen Thi Thu Thao - HNUE

thaok57a@gmail.com

Abstract

The principal objective of this investigation is to evaluate the lexical density and readability in the writings of first-year Mathematics-majored students with dual major in English at a pedagogical university in Hanoi. The data were collected from 26 written products, using two methods in calculating lexical density and readability proposed by Ure (1971) and Flesh (1994) respectively to with the aid of some online text analyzers. The study shows that students can only achieve the average level of both lexical density and readability, which suggests that they need to enhance their writing skills with more complex grammar and vocabulary.

Key words: *lexical density, readability, writing, text analyzers*

1. INTRODUCTION

This investigation mainly concerns the evaluation of two descriptive parameters: lexical density and readability in written products of non-English majored students in Vietnamese language teaching and learning context. To meet the practical teaching and researching demands of scientific subjects in the modern era, Vietnamese pedagical students are required to study English at the very begining of their training. Despite receiving enthusiastic guidance of instructors from Faculty of English, those students still cannot acquire a proficiency level to pursue their future career. Recently, English has been integrated into the curriculum to equip Mathematic majored students with English language skills for their teaching Maths in English later. The two indices metioned above are not comprehensive assessment of one's language competence, but can show the quantified notion of text complexity, which can provide a very first and relative look at their productive ability.

Objectives of the study

This study examines the lexical density and readability of written products of Mathematic majors who are trained to teach Maths in English in the future, from which an outlook of their writing competence could be envisioned. This study does not cover every aspect of written work, but the quantitative one.

Research questions

The study mainly addresses the following two questions:

a. What are the lexical density and readability of non-English majored first-year students' writings at a pedagogical university in Vietnamese context?
b. What can be inferred from those factors?

2. LITERATURE REVIEW

Lexical density

In discourse analysis, the concept "lexical density" suggested by Ure (1971) is used to describe the proportion of lexical words (content words) to the total number of words in either spoken or written form of language. Johansson (2008) stated "a notion of information package" could be extracted by examining this aspect, which means the more lexical words the text content, the more information we can exploit from it.

In order to actually calculate the lexical density of a text, the term "lexical words" should be clarified. Along with the introduction of the concept "lexical density", Ure (1971) gave out the distinction of lexical words and non-lexical words. Languages comprise lexical words which are the primary carriers of meaning, and non-lexical words which do not have lexical function, but "purely in terms of grammar". According to Ure (1971), a word is only orthographic, and a lexical item such as "turn out" is counted as two seperated words: "turn" is a lexical word, while "out" is a non-lexical word. It can be regarded that lexical words belong to the open class and non-lexical words belong to the close class.

Later, the development of the concept "lexical density" was marked when Halliday further refined Ure's formulas as his first approximation to measure lexical density. Halliday (1985) also identified grammar items, or function words as a close system of determiners such as articles, pronouns, most of prepositions, conjunctions, some classes of adverbs and finite verbs, and lexical items or content words as an open system to which new words can be added. However, in contrary with Ure, Halliday referred to the term "items" rather than "words" when it came to "lexical density" since he considered that it might take more than a word to represent a sense, for example, phrasal verbs like "turn out" are regarded as lexical items.

The formula proposed by Ure (1985)

Lexical density (%) = (Number of lexical words/ Total number of words) x 100

Readability

Another parameter mentioned in this study is "readability" which is defined as "the level of ease or difficulty with which text material can be understood by a particular reader who is reading that text for a specific purpose" (Pikulski, 2002). Readability concerns the possibility that the content of a text can be comprehended by the reader, which should be differentiated from legibility, a measure of how easily individual letters or characters can be distinguighed from each other and how they are presented in the form of a text. Dubay (2004) estimated that there were about more than 200 formulas developed to measure readability of written texts. However, due to the time constraint, only Flesch–Kincaid readability tests is applied in this study to calculate readability. It consists of two parts, the Flesch Reading Ease, and the Flesch–Kincaid Grade Level, both of which concern the same core measures (word length and sentence length); however, they have different weighting factors. The results of the two tests correlate quite inversely: a text with a comparatively high score on the Reading Ease test should have a lower score on the Grade Level test.

Flesch Reading Ease test rates text on a 100-point scale. The higher the score, the easier it is to understand the document. The formula for the Flesch Reading Ease score is:

$$RB = 206.835 - (1.015 \times ASL) - (84.6 \times ASW)$$

Where: ASL = average sentence length (the number of words divided by the number of sentences)

ASW = average number of syllables per word (the number of syllables divided by the number of words)

The table below shows the relevance of the Flesch Reading Ease score with the education level of the reader. It is expected that a piece of writing should reach the readability score of at least 60 or it would be too easy for average adult readers.

0-30	Very difficult	Postgraduate
30-50	Difficult	Undergraduate
50-60	Fairly difficult	Grade 10-12
60-70	Standard	Grade 8-9
70-80	Fairly easy	Grade 7
80-90	Easy	Grade 6
90-100	Very easy	Grade 5

Table 1. Original Flesch reading eases description of style Educational Attainment Level (US) (Courtis et al. 2002) (Vinh To et al. 2013)

3. METHODOLOGY

Participants

The subjects of our study were 25 non-English majored freshmen at Faculty of Mathematics, including 2 males, 21 females. They studied English at high school and had to take English test as a part of university admission process. During their university, they study English with the purpose of teaching Maths in English in the future. Right after

entering university, they started to learn and practise all four English skills and writing paragraphs is considered compulsory.

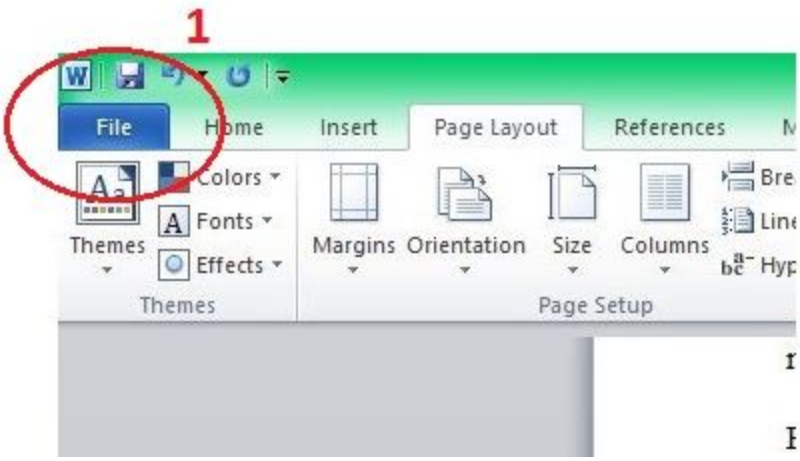
Procedure

a. Calculate readability

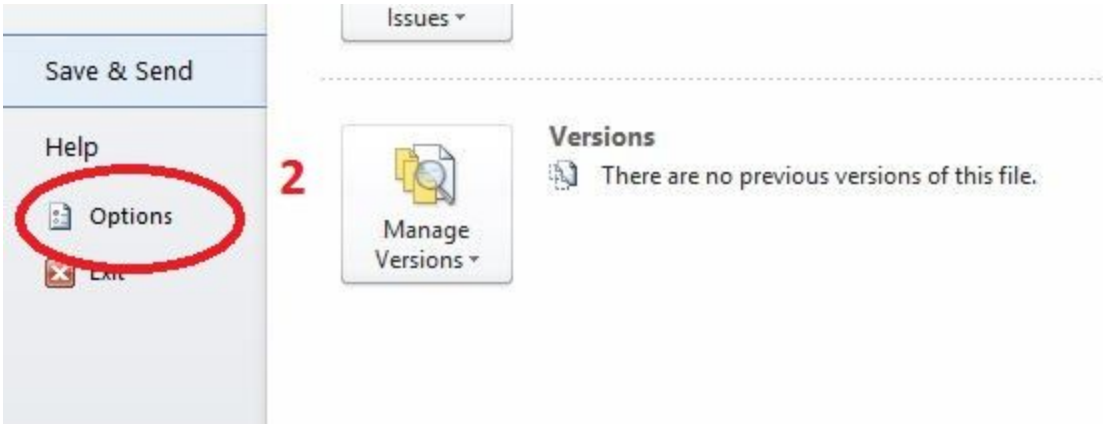
25 writing samples of the 25 participants were collected then transferred into digitized texts with Microsoft Office 2010.

To analyze the readability of the texts, the readability score of a text is obtained by using the readability testing tool in Microsoft Office 2010. The procedure followed these steps:

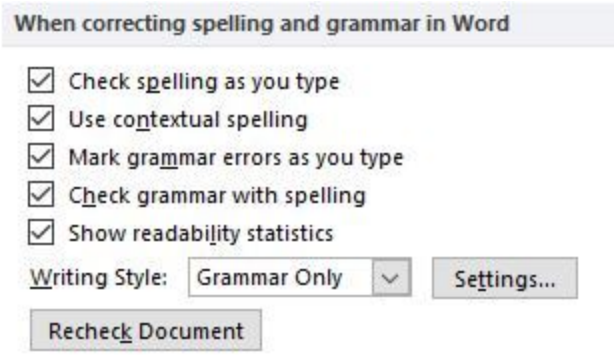
Step 1: Click on "**File**" in the upper left corner of the window



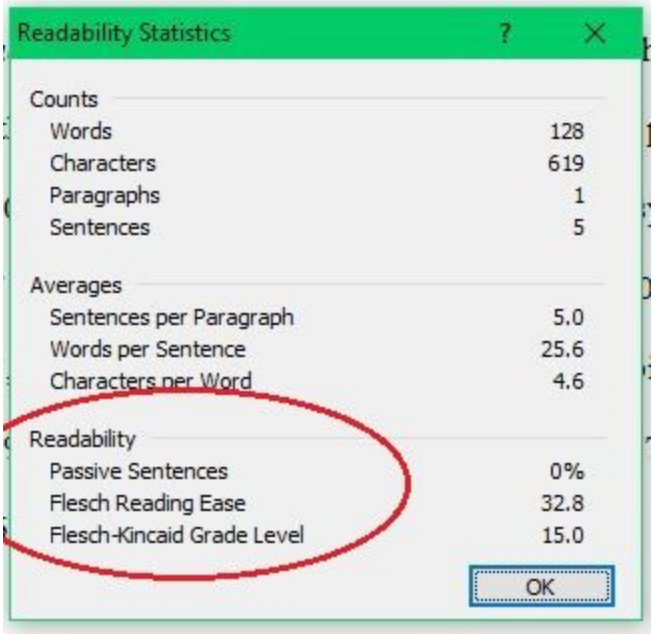
Step 2: Choose "**Option**"



Step 3: Click on "**Proofing"** and "**Check grammar with spelling"**. Under "**When correcting grammar in Word"**, select the "**Show readability statistics"** check box. Finally click "**Ok**"



Step 4: Open each text and select "**Spelling and Grammar**" or press F7 on the key board to get the test result.



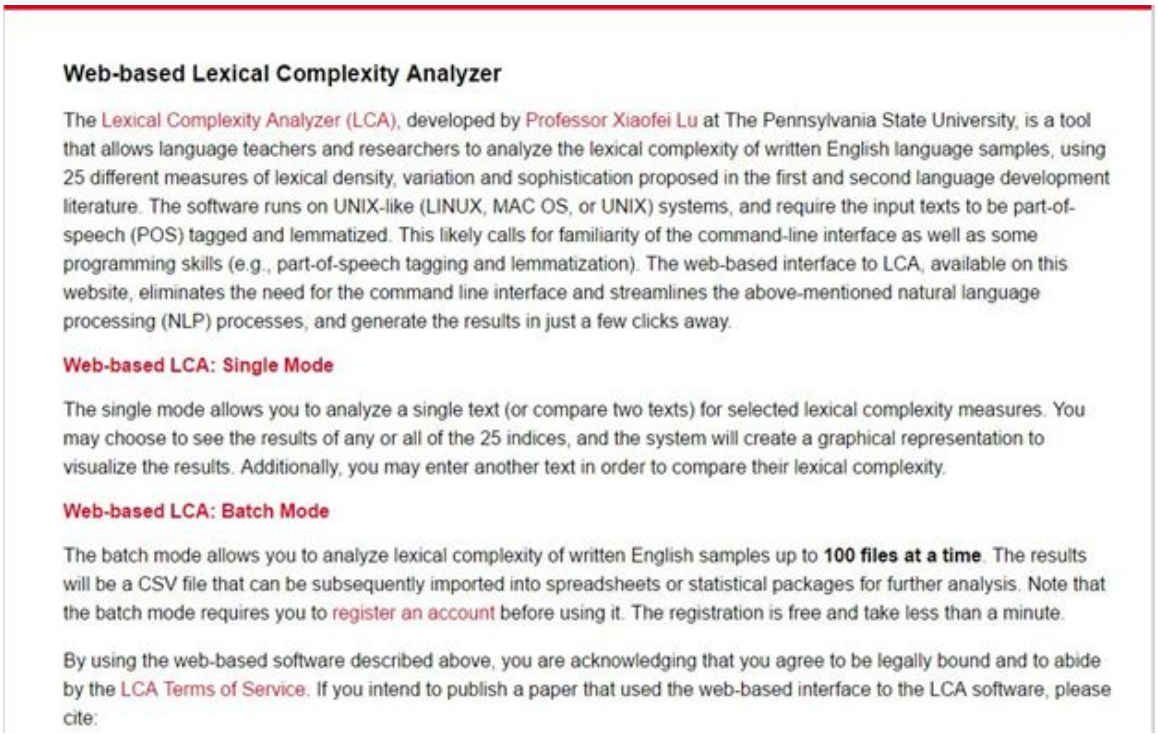
This tool is integrated into Microsoft Office 2010 providing readability result based on the two formulas of Flesch in measuring readability

b. Calculating lexical density

The lexical density calculation is relied on an online tool called "Lexical Complexity Analyzer" developed by Lu (2010). This tool enables language teachers and researchers to analyze the lexical complexity of written English language samples, using 25 different measures of lexical density, variation and sophistication proposed in the first and second language development literature. This tool is provided in an online version for trial and a downloadable version for futher utility. The lexical data of this study is collected by the online tool and following these steps belows.

Step 1: Browsed the website http://aihaiyang.com/software/ . Regist with email and confirm the registation.

Step 2: When the web interference appears, choose **Web-based LCA: Single Mode**



Web-based Lexical Complexity Analyzer

The Lexical Complexity Analyzer (LCA), developed by Professor Xiaofei Lu at The Pennsylvania State University, is a tool that allows language teachers and researchers to analyze the lexical complexity of written English language samples, using 25 different measures of lexical density, variation and sophistication proposed in the first and second language development literature. The software runs on UNIX-like (LINUX, MAC OS, or UNIX) systems, and require the input texts to be part-of-speech (POS) tagged and lemmatized. This likely calls for familiarity of the command-line interface as well as some programming skills (e.g., part-of-speech tagging and lemmatization). The web-based interface to LCA, available on this website, eliminates the need for the command line interface and streamlines the above-mentioned natural language processing (NLP) processes, and generate the results in just a few clicks away.

Web-based LCA: Single Mode

The single mode allows you to analyze a single text (or compare two texts) for selected lexical complexity measures. You may choose to see the results of any or all of the 25 indices, and the system will create a graphical representation to visualize the results. Additionally, you may enter another text in order to compare their lexical complexity.

Web-based LCA: Batch Mode

The batch mode allows you to analyze lexical complexity of written English samples up to **100 files at a time**. The results will be a CSV file that can be subsequently imported into spreadsheets or statistical packages for further analysis. Note that the batch mode requires you to register an account before using it. The registration is free and take less than a minute.

By using the web-based software described above, you are acknowledging that you agree to be legally bound and to abide by the LCA Terms of Service. If you intend to publish a paper that used the web-based interface to the LCA software, please cite:

Step 3: Follow the instructions. First, copy and paste the text into the first text box. This tool allows the analysis of two texts at once so that the calculation of these two texts can be compared to each other. Then, choose the indice(s) needed analysing. Choose the language: British or American English. Finally, click on "**Submit**"

Step 2: Select indice(s)

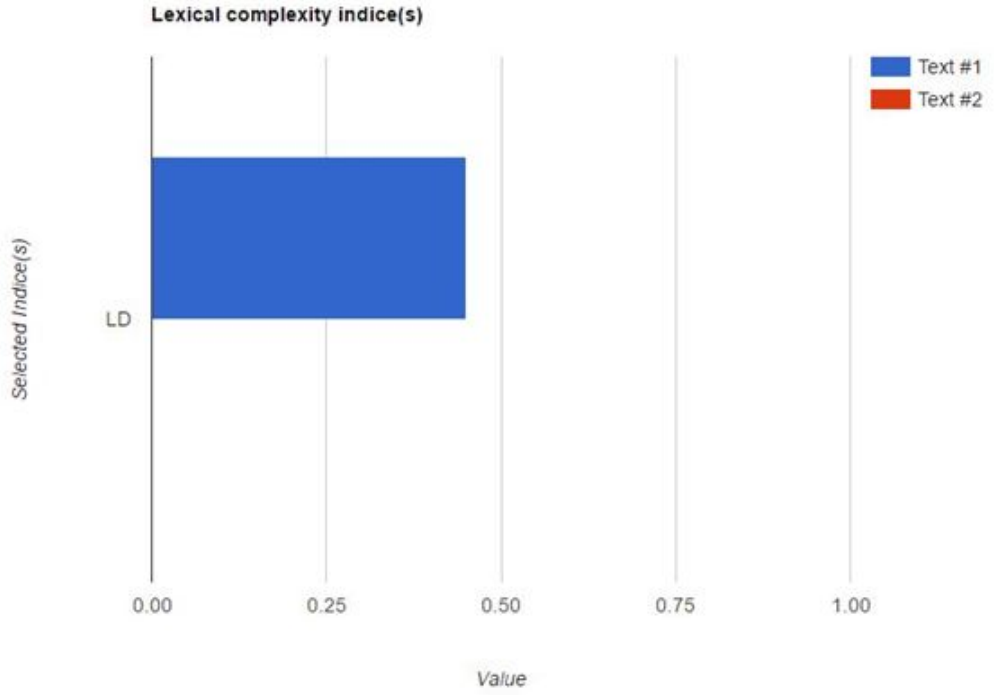
- Lexical density (LD)
- **Lexical Sophistication**
  - Lexical sophistication-I (LS1)
  - Lexical sophistication-II (LS2)
  - Verb sophistication-I (VS1)
  - Verb sophistication-II (VS2)
  - Corrected VS1 (CVS1)
- **Lexical Variation**
  - **NDW**
    - Number of different words (NDW)
    - NDW (first 50 words) (NDWZ-50)
    - NDW (expected random 50) (NDW-ER50)
    - NDW (expected sequence 50) (NDW-ES50)
  - **TTR**
    - Type/Token ratio (TTR)
    - Mean Segmental TTR (50) (MSTTR-50)
    - Corrected TTR (CTTR)
    - Root TTR (RTTR)
    - Bilogarithmic TTR (logTTR)
    - Uber Index (Uber)
  - **Verb diversity**
    - Verb variation-I (VV1)
    - Squared VV1 (SVV1)
    - Corrected VV1 (CVV1)
  - **Lexical word diversity**
    - Lexical word variation (LV)
    - Verb variation-II (VV2)
    - Noun variation (NV)
    - Adjective variation (AdjV)
    - Adverb variation (AdvV)
    - Modifier variation (ModV)

Step 3: Select English Variety:

British English

Submit

Visualization



4. DATA ANALYSIS AND RESULTS

Data analysis

All figures are presented in a table to analyze.

Table 2: Total words, readability scores and lexical density of each student

STT	Student	Total words	Readability	Lexical density (%)
1	No.1	157	59,2	52,9
2	No.2	92	57,5	59,8
3	No.3	98	58,5	56,7
4	No.4	135	51,11	57,8
5	No.5	115	68.2	66,1
6	No.6	150	65,6	61,7
7	No.7	166	79,1	53,6
8	No.8	150	64,8	56
9	No.9	116	70,7	54,3
10	No.10	151	65,7	61,6
11	No.11	126	57,5	55,1
12	No.12	92	75,4	68,5
13	No.13	138	58,3	58,9
14	No.14	123	72,0	59,3
15	No.15	121	62,7	62
16	No.16	153	61,2	60,4
17	No.17	123	61,5	62,6
18	No.18	154	76,2	61
19	No.19	170	60,7	57,3
20	No.20	81	51,4	58,5
21	No.21	90	66,5	66,7
22	No.22	144	67,7	59,03
23	No.23	192	51	66,84
24	No24	167	63,8	53,89
25	No 25	101	52,7	63,11

The researchers used Microsoft Excel to sort data into order according to Total words, Readability and Lexical density to find out the range of each parameter, then compared to the criteria to divide data into smaller groups

The total of words range from 81 to 192 words, including 3 works under 100 words.

Readability scores range from 51 to 79,1. The writing with the largest number of words (192 words) also has lowest score of readability (51). Comparing readability scores with

Table 1 which presents the educational attainment levels, we can perceive that 5 out of 25 works which have readability scores in the 70- 80 range are at the "fairly easy" level, accounting for 20%. 10 out of 25 works which have the readability scores in the 60-70 range are at the "standard" level, making up 40%. Works which have readability scores in the 50-60 range comprise 40% with 10 works at the "fairly hard" level. The average readability score is about 59, 6 at the "standard" level.

Lexical density falls in the range of 52,9% to 68,5%. Overall, indices of lexical density on average are required for written language, according to Ure (1971), which is over 40%.

Results

Divide the participants into three groups according to the readability score then we have

Table 3. Lexical density range of each student group divided according to readability scores

Groups	Readability	Lexical density
1	51-59,2	52,9-66,84
2	60,7-68,2	56-66,7
3	70,7-79,1	53,6-68,,5

From the table we can see that the readability score and the lexical density are not really in a corresponding relation. Lexical density can be varied in each range of readability score. Moreover, even with higher total number of words, it does not mean an article can get higher lexical density and readability scores because of the less complexity in structure of sentences or words.

Overall, both lexical density and readability scores of students are just in the average range, resulting from the fact that they have not yet mastered the technique of using complex or compound words and sentences, with a minor exception. Their writings mostly rely on using simple words and simple sentences, which lowers the variables in those formulas.

This study shows that students need to improve their ability to produce complex and compound sentences paralleling with ability to use complex and compound words in their writing, then lexical density and readability score can be higher in value. This requires more intensive and regular practice to master writing skills.

CONCLUSION

Summary

This study concentrated on investigating lexical density and readability of non-English majored students' writings at a pedagogical university in Vietnamese context. First of all, this study statistically calculated lexical density and readability of the subject's written tasks with the help of some available computing programs, then pointed out that most of the students only reach an average level. Thus, students need more training to enhance writing competence as well as using vocabulary.

Implication for language teaching and learning

In order to do so, students can arrange group study to exchange writings for feedback on grammar and vocabulary.

It is also suggested that teachers should raise students' awareness of the importance of grammatical structure and vocabulary in their writings. Moreover, teachers need to organize writing activities focusing on these aspects in class on a regular basis. It is also a good idea to provide students with sources of both printed and online self-study materials which contain good samples of written products so that they can adapt into their own ones.

Limitations and recommendations for the further study

Lexical density and readability are wisely applied to evaluate the texture of writing. However, we can only assess the quality based on these two parameters, but cannot say much about the quality because an article have to be evaluated on many criteria such as structure, organization and the ability to transmit information to the reader.

In addition, in order to miscalculate, the incorrectly spelled words along with grammar mistakes are not concerned, which means those figures are theoretical and fairly higher than the reality.

For the limitation of time, this study could only cover a minor number of participants and data, which lead to the fewer databases to be analyzed and interpreted.

This study leaves many open questions for further investigations due to the fact that it only examined the lexical density and readability of non-English majored students' writings. It is suggested that we can compare two parameters of written tasks and spoken tasks, or compare with written tasks of English-major students at the same age, and on a bigger scale.

References

- Johansson, Victoria. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. Lund University, Dept. of Linguistic sand Phonetics.
- John J. Pikulski, Ph.D.(2002). Readability. University of Delaware.
- Halliday, M. A. K. (1985). Spoken and written language. Geelong Vict. Deakin University
- M. Linnarund. 1973. Lexical density and lexical variation – An analysis of the lexical texture of Swishes students' written work. University of Lund.
- Šárka Timarová. (2005). Corpus linguistic methods in interpreting research: A case study. Charles University, Prague.
- Vinh To, Si Fan, & D.Thomas. (2013). Lexical Density and Readability: A Case Study of English Textbooks. Internet Journal of Language, Culture and Society.
- Ure, J. (1971). Lexical density and register differentiation. Applications of linguistics, 443-452.
- William H.Dubay (2004). The Principle of Readability.