

## Effect of the Part of Speech Elements on Readability of Literary Texts in Secondary Vietnamese Textbooks

Như-Điệp Nguyễn<sup>1</sup>, An-Vinh Lương<sup>2</sup>, Diên Đình<sup>2</sup>

(1. Department of Linguistics, University of Social Sciences and Humanities, Vietnam National University Ho Chi Minh City, Vietnam; 2. University of Science, Vietnam National University Ho Chi Minh City, Vietnam)

**Abstract:** Readability has been the theme which is crucial and interesting in the world since the nineteenth century up to the present time. These research results are highly applied in a lot of the social fields, such as the Readability formula in Microsoft Office or The Common European Framework of Reference for Languages. In readability research, word is one of the unit in which its elements has been carefully investigated. Based on the comparison and contrast the elements effecting Readability on the “word” unit in English, we determine and select the part of speech in Vietnamese, having influences on readability; namely examining and evaluating readability of the part of speech in the proeses in Vietnamese textbooks for secondary students from the results of the statistical data. Then, the survey results are investigated to find out the frequency as well as correlation of the part of speech in Vietnamese. Initially, the findings can assist writers, editors; especially teachers in editing documents, reforming textbooks, and building exam question banks of Vietnamese language for native Vietnamese in general and foreigners in particular.

**Key words:** readability, part of speech, Vietnamese textbooks, secondary students

### 1. Introduction

The studies of readability have been done since the early nineteenth century. Among these achievements are the formulas for measuring readability, which are used as the tools for determining the complexity of the text. Therefore, they can help users select the appropriate text with different reading levels for the readers in efficiency, saving time and labor. The results of the research have applied in various areas of society, such as the integrated measuring the Flesch formula in Microsoft Office software or the same with the formulae: Flesch-Kincaid Cohmetrix, Idicies, Lexile Measures, etc. in the Common European Framework of Reference.

In forming a formula or a tool to measure text readability, linguistic elements or the linguistic components in a particular text, play a very important role, as shown in a lot of readability research. These linguistic elements are analyzed on the shallow/surface features, such as the average length of words by the number of syllables, the

---

Nhu Diep Nguyen, Lecturer, Ph.D. Student, Saigon Technology University; research areas/interests: linguistics, language proficiency, translation, and corpus linguistics. E-mail: [nhudiep2004@gmail.com](mailto:nhudiep2004@gmail.com).

An Vinh Luong, Lecturer, Ph.D. Student, University of Science, VNU-HCMC Vietnam; research areas/interests: informatics. E-mail: [anvinhluong@gmail.com](mailto:anvinhluong@gmail.com).

Dien Dinh, Assistant Professor, University of Science, VNU-HCMC Vietnam; research areas/interests: informatics and linguistics. E-mail: [ddien@fit.hcmus.edu.vn](mailto:ddien@fit.hcmus.edu.vn).

average numbers in a sentence, or the frequency of words; and the deep features of the language, such as the parsed syntactic features, the language modeling features or the part of speech- based features.

With the scope of this article, firstly, based on the contrast of the linguistic elements having influences on text readability in English, the work of the two authors William S. Gray and Bernice E. Leary, we define the part of speech (POS) as the linguistic elements, having effect on the text readability in Vietnamese. Then, we survey and evaluate the readability influences of the POS elements of the literary texts in Vietnamese subject textbooks for elementary school-aged children according to the statistics. The findings will help writers, editors, and especially teachers compile and select lessons as well build of the bank of questions at the basic grade level for Vietnamese learners in general and foreigners who learn Vietnamese in particular.

## 2. Methodology and corpus

Our corpus is taken from the prose texts in the Vietnamese textbooks for secondary students (from grades 6 to 9), Education Publisher 2016. In the preprocessing, finding some texts are in the forms of questions, puzzles or drawing annotations, we had to re-select them; and the left texts are 79, providing children with general knowledge and helping children practice reading skills. The linguistic elements with surface features are generally described in the following table<sup>1</sup>:

Grade	Number of Texts	Average number of words	Average number of Sentences
6	28	680.89	54.25
7	13	678.31	45.31
8	17	971.06	63.82
9	21	1451.52	105.62

Next, we used the “CLC-Vietnamese-Toolkit”<sup>2</sup>, generated by Computational Linguistics Center, University of Science, HCMC to handle the POS in each text and calculate their frequency. Besides, the Corpus is also analyzed the correlation to find the relationship between the POS with the text readability.

## 3. Effect of Linguistic Elements on Text Readability

### 3.1 Linguistic Elements Having Effect on Text Readability in English

In the work “What Makes a Book Readable”, published in 1935, S. Gray and E. Leary identified 288 elements effecting on English text readability. These elements were classified into four main categories: (I) Format or Mechanical Features, (II) General Features of Organization, (III) Style of Expression and Presentation, and (IV) Content.

Within the scope of their study, only investigating the linguistic elements of the “Style of Expression and Presentation”, they identified 82 language elements which function as the “potential elements” effecting text readability. These elements are classified into three units, “Word”, “Sentence” and “Paragraph/Passage”. Among them, the elements having influences on text readability in the word unit are 41 elements (1) (see Glossary). With

<sup>1</sup> We already presented these surface linguistic elements in the paper “Investigating Some Elements Effecting on the Readability of Vietnamese Text-Easy Category (Contrast With Those In English)” at *The 3rd International Conference on Vietnamese Studies* (2017); and “Examining the Text-length Factor in Evaluating the Readability of Literary Texts in Vietnamese Textbooks” at *The 9th International Conference on Knowledge and Systems Engineering* (KSE2017).

<sup>2</sup> <http://www.clc.hcmus.edu.vn/>.

the aim to conduct an experimental research and use the quantitative enumeration, 14 linguistic elements (without asterisk \*) are not selected for their further analysis research, due to the following reasons: (i) the linguistic elements do not meet the experimental process; (ii) they have not been formed by any clear definitions yet, and (iii) these linguistic elements cannot be measured or counted objectively in largely analyzed cases from their corpus.

However, there are many differences between English and Vietnamese, causing from its morphological typology, culture, such as having differences in phonetics, morphemes, word boundaries, the word forms, part of speech, word order (adjective and noun), sentence structure (theme-rheme). We cannot simply copy the same processing model in English and apply it in Vietnamese, and the adjustment based on the results of comparisons and contrasts between these two languages is crucial in this case<sup>3</sup>.... Hence, by comparing the similarities and differences of the linguistic elements between Vietnamese and English in the word unit, this article selects and surveys the POS elements at the word unit from the former-mentioned corpus.

### **3.2 Linguistic Elements Effecting on Text Readability in Vietnamese**

#### **3.2.1 Lexico — Grammatical Category**

Vocabulary in any languages is often very large. Therefore, words should definitely be classified to make the rules of forming and using them more comprehensible. There are several ways to classify them. For example, in terms of meanings, some words convey one meaning and the others with poly-semantics. In terms of original, there are cognates and borrowed words. Concerning word usage, there are popular words and other types such as slang, dialect lexicon, technical terminologies and so on. Based on word form, there are monosyllabic and multisyllabic words. In other words, these words can be classified according to their forming, such as single words and compound ones. Also, words can be classified according to its first letter, as in the dictionary. But in Vietnamese, there is a crucial way of word classification, which based on its lexical meaning and grammatical function, called lexico- grammatical category<sup>4</sup>.

Each grammatical category includes a set of different forms of a word, while each lexico-grammatical category includes a set of words. Determining grammatical category begins with the considering the forms of the same word to determine number; for example, in English, book (singular) with books (plural). In contrast, the establishment of lexical- grammatical category begins with the unifying the forms a single word; then based on its general meaning and grammatical characteristics to classify it. It can be seen that the part of speech in English can be basically analyzed in its grammar and then category them into content words (nouns, verbs, adjectives, adverbs...) and function words (articles, prepositions, conjunctions...). But in Vietnamese, classifying nouns, verbs, adjectives etc., based on both lexical meaning and grammatical activity. Therefore, there are two main kinds of words called “lexical words” instead of content words and “form words” instead of function words and in this case the term lexico and grammatical category are called.

In addition, from the different perspectives, Vietnamese is categorized into different criteria, and sometimes there are different names for the same subject. Therefore, to be consistent in this study, we classify the mentioned previously corpus based on semantic and grammatical characteristics and the POS classified of Committee of

---

<sup>3</sup> Original text: “... tiếng Anh và tiếng Việt có nhiều điểm khác biệt (do loại hình ngôn ngữ, do nền văn hóa,...). Chẳng hạn khác biệt về ngữ âm học, hình vị, ranh giới từ, sự từ vựng hóa (như : ox – bò đực, anh – elder brother,...); từ loại; trật tự từ (tính từ và danh từ), kết cấu câu (đề thuyết và cụm chủ vị),... Vì vậy chúng ta không thể áp dụng y nguyên các mô hình xử lý ngôn ngữ của tiếng Anh sang cho tiếng Việt được, mà phải có sự điều chỉnh nhất định dựa trên các kết quả so sánh đối chiếu giữa tiếng Anh và tiếng Việt.” Đinh Điền (2006), *Xử lý Ngôn ngữ tự nhiên*, Tp. Hồ Chí Minh, Nxb Đại học Quốc gia Tp. HCM, pp. 13–14.

<sup>4</sup> Nguyễn Thiện Giáp (Chủ biên)- Đoàn Thiện Thuật - Nguyễn Minh Thuyết, 2008, *Dẫn luận Ngôn ngữ học*, Nxb Giáo dục, p. 242.

Social Science (1993). In other words, we determine the POS elements on lexical-grammatical category.

According to Committee of Social Science, “part of speech include words with the same general meaning and grammatical characteristics...The general meaning of Vietnamese words are reflected in their grammatical characteristics. However, their characteristics, in such an isolating language like Vietnamese, are not shown in the phonology but their collocations with other words”<sup>5</sup>. For a long time, there has been a distinction between lexical words and form words in Vietnamese.

In order to make the classification more effective and useful in forming sentences, lexical words and form words are divided into two smaller groups. Lexical words are categorized into “nouns”, “verbs” and “adjectives”; whereas form ones are classified into “adjuncts” and “conjunctions”. In addition to these categories, there are also “pronouns”; and there are two another word classes namely “modifiers” and “interjections”, which do not belong to either lexical words or form words, and are also different from pronouns. In brief, POS in Vietnamese can be categorized into eight: nouns, verbs, adjectives, pronouns, adjuncts, conjunctions, modifiers and interjections, in which, six former groups are divided into sub-categories, for example nouns are classified into proper nouns, common nouns; or smaller categories including simple noun, general nouns, “cái, con” nouns, countable nouns, abstract nouns, concrete nouns, numerals and place nouns<sup>6</sup>.

### 3.2.2 POS Elements Effecting Text Readability in Vietnamese — The Word Unit — Secondary Textbooks

Word class is a hierarchical system, or a category which includes smaller categories. Besides, the same lexical meaning for the same POS, there are the same one for the same sub-categories. This significance covers a narrower scope of a word, but the meaning remains the general syntactic meaning<sup>7</sup>. Lijun F., Martin J., Matt H. and Noemie E. (2010, re-extracted from Heliman et al., 2007; Leory et al., 2008) shows that the characteristics of the part of speech in a text prove very useful in determining text readability.

Based on five common part of speech, consisting of nouns, verbs, adjectives, adverbs and prepositions in English as well as two categories of words in Vietnamese, namely lexical words and form words, according to the lexical meaning and grammatical characteristics, we have examined the POS of the corpus, which is Vietnamese, compatibly.

To determine the influences of the POS on readability in Vietnamese texts, we used the automatic supporting tool called CLC-Vietnamese-Toolkit, developed Computational Linguistics Center, University of Science, HCMC (2) (see Glossary) to identify 24 common POS in Vietnamese, to investigate the relationship of them in text readability with the different grade levels ( from grade 6 to grade 9).

The Vietnamese POS elements are first classified into basic categories, then divided into the smaller common groups of words. For example, nouns will specifically be classified into “nouns of quantity”, “proper nouns”, “nouns of time” and “concrete nouns”. The CLC-VN-POS Toolkit is used to label different types of words in Vietnamese. In case it is not impossible to determine exactly the POS, words will be labeled X (unidentified POS - Unknown).

The corpus is analyzed to determine the frequency of each POS, and listed all categories from the smallest to

---

<sup>5</sup> Original text: “từ loại bao gồm những từ có nghĩa khái quát giống nhau và đặc điểm ngữ pháp giống nhau... Nghĩa khái quát của từ tiếng Việt được thể hiện qua đặc điểm ngữ pháp của nó, nhưng đặc điểm ngữ pháp ấy, trong một ngôn ngữ đơn lập như tiếng Việt, không thể hiện ở hình thức ngữ âm của từ mà chủ yếu ở khả năng kết hợp của nó với từ khác”.

<sup>6</sup> Categorized according to Vietnamese Grammar, Vietnam Committee of Social Science (1993), Social Science Publisher, Hanoi, pp. 67– 95.

<sup>7</sup> Mai Ngọc Chừ (Chủ biên)- Nguyễn Thị Ngân Hoa - Đỗ Việt Hùng - Bùi Minh Toán, 2007, Nhập môn Ngôn ngữ học, Nxb Giáo dục, p. 270.

the largest in each grade level. We can see details in the following table.

**Table 1 The Frequency of the POS Elements Effecting Text Readability on Vietnamese — Secondary Literary Textbooks**

No.	Part of speech	Grade 6	Grade 7	Grade 8	Grade 9	Total	Note
1	Proper Nouns	0–68	2–51	3–42	0–143	0–143	Nr
2	Countable Nouns	3–105	3–35	3–90	0–149	0–149	Nc
3	Concrete Nouns	0–14	0–2	0–17	0–40	0–40	Nu
4	Temporal Nouns	2–52	3–60	2–65	3–92	2–92	Nt
5	Numerals	2–82	13–51	7–134	1–175	1–175	Nq
6	Nouns	16–404	78–340	56–326	59–868	16–868	Nn
7	Directional Verbs	0–21	0–14	0–18	0–36	0–36	Vd
8	State Verbs	1–24	3–15	0–52	3–52	0–52	Ve
9	Comparative Verbs	0–17	2–12	1–22	1–47	0–47	Vc
10	Volitive Verbs	27–347	43–399	44–680	47–693	27–693	Vv
11	Directions	1–29	0–39	0–24	0–63	0–63	D
12	Quantity Adjectives	0–3	0–5	0–5	0–3	0–5	An
13	Quality Adjectives	5–14	52–117	22–237	13–306	2–306	Aa
14	Demonstrative Pronouns	0–47	3–29	2–63	6–90	0–90	Pd
15	Personal Pronouns	4–204	13–184	4–371	11–315	4–315	Pp
16	Adverbs	7–182	23–156	17–392	24–410	7–410	R
17	Prepositions	1–82	14–60	12–104	10–144	1–144	Cm
18	Parallel Conjunctions	1–114	23–80	7–253	34–183	1–183	Cp
19	Subordinating Conjunctions	0–7	0–8	0–17	0–26	0–26	Cs
20	Modifiers	0–30	0–18	0–78	2–85	0–85	M
21	Emotion Words	0–12	0–11	0–14	0–12	0–12	E
22	Foreign Words	0–10	1–8	0–15	0–29	0–29	FW
23	Onomatopoeia	0	0	0	0–2	0–2	ON
24	Idioms	0–3	0–3	0–7	0–3	0–7	ID

According to the corpus analysis outlined above, in the term of quantity, we can identify 24 POS elements effecting on the text readability, the average level. In one side, the frequency of each element for each grade and the total (from grade 6 to grade 9) are identified, for example, the frequency of nouns in a text of average level is from 0 to 143; in details, in grade 6, the frequency is from 0 to 68, 2 to 51 for grade 7, 3 to 42 for grade 8, and 0 to 143 for grade 9, and so on. In other side, we can have a general view with three groups by the POS frequency: 9 elements with low frequency (from 0 to 50 times), including Concrete Nouns, Directional Verbs, Comparative Verbs, Quantity Adjectives, Subordinating Conjunctions, Emotion Words, Onomatopoeia, Idioms, Foreign Words; 2 elements with average frequency (from 0 to 80 times): State Verbs Directions; and 13 elements with high frequency (from 0 to 81 and over): Proper Nouns, Countable Nouns, Temporal Nouns, Numerals, Nouns, Volitive Verbs, Quantity Adjectives, Quality Adjectives, Personal Pronouns, Adverbs, Prepositions, Parallel Conjunctions, and Modifiers as in Table 2:

**Table 2 The POS Elements Effecting Text Readability on Vietnamese — Secondary Literary Textbooks**

Elements with low frequency	Elements with average frequency	Elements with high frequency
<ul style="list-style-type: none"> <li>– Concrete Nouns</li> <li>– Directional Verbs</li> <li>– Comparative Verbs</li> <li>– Quantity Adjectives</li> <li>– Subordinating Conjunctions</li> <li>– Emotion Words</li> <li>– Onomatopoeia</li> <li>– Idioms</li> <li>– Foreign Words</li> </ul>	<ul style="list-style-type: none"> <li>– State Verbs</li> <li>– Directions</li> </ul>	<ul style="list-style-type: none"> <li>– Proper Nouns</li> <li>– Countable Nouns</li> <li>– Temporal Nouns</li> <li>– Numerals</li> <li>– Nouns</li> <li>– Volitive Verbs</li> <li>– Quantity Adjectives</li> <li>– Quality Adjectives</li> <li>– Personal Pronouns</li> <li>– Adverbs</li> <li>– Prepositions</li> <li>– Parallel Conjunctions</li> <li>– Modifiers</li> </ul>

In this article, we investigate the deep linguistic elements on not only their frequency but also their correlations so that we can identify the “potential” linguistic elements effecting on text readability in Vietnamese. The correlation analysis results are presented in Table 3 as follows:

**Table 3 The Correlation between POS Elements and Text Readability**

Part of speech	The correlation coefficient	Parts of Speech	The correlation coefficient
Demonstrative Pronouns	0.420428156	Countable Nouns	0.305154497
Concrete Nouns	0.161392278	Common Nouns	0.369234368
Quantity Adjectives	0.290212626	Quality Adjectives	0.436047868
Idioms	0.038782703	<i><b>Numerals</b></i>	<i><b>0.49700568</b></i>
Proper Nouns	0.181840006	Personal Pronouns	0.293475393
Foreign Words	0.397330916	Adverbs	0.398622477
Directional Verbs	0.1546354	Onomatopoeia	0.213045569
Volitive Verbs	0.358210755	Modifiers	0.349916801
Comparative Verbs	0.42931023	Parallel Conjunction	0.448005754
Directions	0.23919875	State Verbs	0.322522351
Temporal Nouns	0.364036449	Subordinating Conjunctions	0.389009482
Emotion Words	0.145261501	<i><b>Major/Minor Conjunction</b></i>	<i><b>0.514357169</b></i>

From the correlation analysis results, it can be seen that the frequency of numerals and prepositions have a strong connection with the text readability (0.49 and 0.511). The 0.49 correlation coefficient of numerals means that nearly 30 ( $0.49 \times 0.49$ ) of the change of text readability has a link to the change in frequency of numerals in the texts with all the elements being analyzed. Similarly, the 0.51 correlation coefficient of prepositions; means, with all the elements being analyzed, about 30% ( $0.51 \times 0.51$ ) of the change of text readability is related to the change of the frequency of prepositions.

#### 4. Conclusion

The survey of 24 elements of POS effecting on text readability of the literary texts in Vietnamese textbooks for primary pupils can firstly help teachers, editors, and also learners determine the level of difficulty of texts qualitatively at basic level (easy), based on the POS elements in the following categories: (1) proper noun; (2)

concrete noun; (3) countable noun; (4) temporal noun; (5) numeral; (6) noun; (7) directional verb; (8) state verb; (9) comparative verb; (10) volitive verb; (11) directions; (12) quantity adjective; (13) quality adjective; (14) demonstrative pronoun; (15) personal pronoun; (16) adverbs; (17) preposition; (18) parallel conjunction; (19) subordinating conjunction; (20) modifier; (21) emotion word; (22) foreign word; (23) onomatopoeia; and (24) idiom. As a result, the users irrespective of specialists or non-specialists can select or compile the Vietnamese lessons appropriately for each grade with not only the qualitative approach but also the quantitative one.

The most important thing in evaluating POS elements effecting on text readability is that the corpus must be classified in different levels. However, at present, there is no tool or formula reliable or effective enough to measure the text readability for Vietnamese texts. Therefore, choosing a corpus collected from the textbooks which were already classified into different grade levels for secondary students is ideal for this study. Besides, there are still some issues in the corpus itself. Although the texts hierarchically divided in increasing levels from Grade 6 to Grade 9, there is no clear distinction. For example, the frequency of idioms in grade 6, and 9 is equal (0–3), while grade 8 has the highest frequency (0–7). Therefore, further studies with a larger corpus for this level as well as making a comparison between basis level and average or advanced one are necessary.

The noticeable thing, from the statistics of frequency, shows that the unidentified POS is used only 2 times for the total number of 79 texts at the average level. Therefore, the unidentified POS is not listed in the table. It is assumed that this linguistic element must be considered, evaluated and determined separately before giving suggestions for using in text readability at basic or average level as well as at advanced one. Besides, the correlation between POS elements and text readability also shows that numerals and prepositions are the “potential” linguistic elements having strong influence on Vietnamese text readability, these elements should be interpreted in other aspects of language, such as prepositions or numerals in ambiguity; on that basis the analysis of the corpus can show robust evidences about the effect on linguistic elements in Vietnamese text readability.

## References

- Bùi Mạnh Hùng (2008). *Ngôn ngữ học Đối chiếu (Contrastive Linguistics)*, TP. Hồ Chí Minh, Nxb Giáo dục (HCM City, Education Publisher).
- Cao Xuân Hạo and Hoàng Dũng (2005). *Từ Điển Thuật ngữ Ngôn ngữ học Đối chiếu Anh – Việt; Việt – Anh (English– Vietnamese; Vietnamese– English Contrastive Linguistics Dictionary)*, TP. Hồ Chí Minh, Nxb Khoa học Xã hội (HCM City, Social Science Publisher).
- Council of Europe (2010). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge Press, accessed on June 10, 2017, available online at: [http://www.coe.int/en/t/dg4/linguistic/source/framework\\_en.pdf](http://www.coe.int/en/t/dg4/linguistic/source/framework_en.pdf).
- Dale E. and Chall J. S. (1949). “The concept of readability”, *Elementary English*, Vol. 26, p. 23.
- Dubay H. W. (2004). *The Principles of Readability: Impact Information*, Costa Mesa, California.
- Đinh Điền (2006). *Xử lý Ngôn ngữ tự nhiên (Processing Natural Language)*, TP. Hồ Chí Minh, Nxb Đại học Quốc gia Tp. HCM (HCM City, Vietnam National University Publisher).
- Flesch R. (1943). *Marks of a Readable Style, Columbia University Contributions to Education*, No. 897, New York: Bureau of Publications, Teachers College, Columbia University.
- Kincaid J. P., Fishburne R. P., Rogers R. L. and B. S. Chissom (1975). “Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for navy enlisted personnel”, *CNTECHTRA Research Branch Report*, pp. 8–75.
- Harry Mc Laughlin (1969). “SMOG Grading — A new readability formula”, *Journal of Reading*, Vol.12, No. 8, pp. 639–646, doi: <http://dx.doi.org/10.2307/40011226>.
- William S. Gray and Bernice E. Leary. 1935. *What Makes a Book Readable*, The University of Chicago Press, Chicago, Illinois.

- Jamie D. (2014). "Investigating the relationship between empirical task difficulty, textual features and CEFR levels", *EALTA 2014*, 29 May – 1 June, University of Warwick.
- Klare G. R. (1973). *The Measurement of Readability*, Ames, Iowa: Iowa State University Press.
- Lijun F., Martin J., Matt H., Noémie E. (2010). "A comparison of features for automatic readability assessment", Beijing August 2010, Poster Volume, pp. 276–284.
- Mai Ngọc Chừ (Editor), Nguyễn Thị Ngân Hoa, Đỗ Việt Hùng and Bùi Minh Toán (2007). *Nhập môn Ngôn ngữ học [An Introduction to Linguistics]*, Nxb Giáo dục (Education Publisher).
- Mostafa Z. and Pooneh H. (2012). "Readability of texts: State of the art", *Theory and Practice in Language Studies*, Vol. 2, pp. 1, pp. 43–53, available online at: <http://doi.org/10.4304/tpls.2.1.43-53>.
- Nguyễn Thiện Giáp (Chủ biên), Đoàn Thiện Thuật and Nguyễn Minh Thuyết (2008). *Dẫn luận Ngôn ngữ học (General Linguistics)*, Nxb Giáo dục (Education Publisher).
- Ủy ban Khoa học Xã hội học Việt nam (1993). *Ngữ pháp tiếng Việt*, NXB Khoa học Xã hội–Hà Nội.
- Vũ Thị Phương Anh (2006). "Khung trình độ chung Châu Âu và việc nâng cao hiệu quả đào tạo tiếng Anh tại ĐHQG – HCM (Common European Framework and Improving the Efficiency in English Education at National University — HCM)", *Developing Science and Technology Journal*, Vol. 9, No. 10.
- Vu Thị Phương Anh (2011). "Text readability and testing languages", October 30, 2011, available online at: <http://ncgdnv.blogspot.com/2011/10/o-kho-cua-van-ban-va-viec-kiem-tra-ngon.html>.
- William S.Gray and Bernice E. Leary (1935). *What Makes a Book Readable*, The University of Chicago Press, Chicago, Illinois.



## Glossary

## 1. Linguistic Elements Effecting on Text Readability in English–Word Unit

No.	Các yếu tố (Elements)	
	Thuộc tính của Từ (properties of Words)	
*1	Số lượng từ dễ	Number of easy words
*2	Phần trăm số từ dễ	Percentage of easy words
*3	Số lượng từ khó	Number of different hard words
*4	Số lượng từ 90% học sinh lớp 6 không biết	Number of words not known to 90 percent of sixth-grade children
*5	Phần trăm số lượng từ 90% học sinh lớp 6 không biết	Percentage of words not known to 90 percent of sixth-grade children
*6	Số lượng những từ khác	Number of different words
*7	Phần trăm số lượng từ khác	Percentage of different words
*8	Phần trăm từ đơn tiết	Percentage of monosyllables
*9	Phần trăm từ có hai âm tiết	Percentage of bisyllables
*10	Phần trăm từ đa tiết	Percentage of polysyllables
*11	Đại từ ngôi thứ nhất	First-person pronouns
*12	Đại từ ngôi thứ hai	Second-person pronouns
*13	Đại từ ngôi thứ ba	Third-person pronouns
*14	Đại từ ngôi thứ nhất, thứ hai và thứ ba	First-, second-, and third-person pronouns
15	Danh từ (tên riêng, trừu tượng, cụ thể)	Nouns (proper, abstract, and concrete)
16	Động từ	Verbs
17	Tính từ (mô tả, hạn định)	Adjectives (descriptive, limiting)
18	Mạo từ (xác định, không xác định)	Articles (definite, indefinite)
19	Trạng từ (thời gian, nơi chốn, thể cách)	Adverbs (time, place, manner)
20	Kết từ (đẳng lập, chính phụ)	Conjunctions (co-ordinate, subordinate)
21	Thán từ	Interjections
22	Giới từ	Prepositions
*23	Từ chỉ nội dung/ thực từ	Content words
*24	Từ thuộc cấu trúc	Structural words
25	Từ mang nghĩa tu từ	Image-bearing words
26	Từ không mang nghĩa tu từ	Non-image-bearing words
*27	Từ bắt đầu bằng chữ cái w	Words beginning with w
*28	Từ bắt đầu bằng chữ cái h	Words beginning with h
*29	Từ bắt đầu bằng chữ cái b	Words beginning with b
*30	Từ bắt đầu bằng chữ cái i	Words beginning with i
*31	Từ bắt đầu bằng chữ cái e	Words beginning with e
32	Từ liên gắn liền với cuộc sống của người lớn	Words associated with adult living
33	Từ mô tả những khái niệm trừu tượng	Words expressing abstractions
*34	Từ chú thích, đồng vị, từ trong ngoặc đơn	Asides, appositives, parenthetical expressions
*35	Từ địa phương	Dialect words
*36	Từ cổ, từ hiếm	Archaic words, rare words
*37	Từ địa phương và từ tự tạo	Local expressions and coined words
38	Từ mang tính văn chương và thi pháp	Poetic and highly literary words
*39	Từ có tính chất thành ngữ	Idiomatic expressions
*40	Danh từ riêng mang tính thần thoại và lịch sử	Proper nouns of mythology and history
41	Từ kỹ thuật	Technical words

(Extracted from XXVI table, William S. Gray &amp; Bernice E. Leary “What makes a Book readable”, 1935, pp. 98, 99)

2. CLC–Vietnamese – Tool kit, Computing Linguistics–CLC–University of Science Ho Chi Minh City

Label	Từ loại tiếng Việt –Vietnamese Parts of Speech	Ví dụ Example	Từ loại tiếng Anh English equivalents
<b>Aa</b>	tính từ hàm chất	lộ thiên, đầy, mắc	quality adjective
<b>An</b>	tính từ hàm lượng	đầu tiên	quantity adjective
<b>Cm</b>	giới từ	giữa, của, trong, tại	major/minor conjunction
<b>Cp</b>	kết từ đẳng lập	cùng, với, và	Parallel Conjunction
<b>Cs</b>	kết từ chính phụ	nếu, thì, vừa, là	subordinating conjunction
<b>D</b>	phó động từ chỉ hướng	ra, vô, lên, xuống	direction
<b>E</b>	cảm từ	thưa, làm gì	emotion word
<b>FW</b>	từ nước ngoài	Miss, pH, super	foreign words
<b>ID</b>	thành ngữ	công ăn việc làm	Idiom
<b>M</b>	trợ từ	đến, riêng, được, có, đó	modifier
<b>Nc</b>	danh từ đơn thể	bộ, ngôi, bản, con, bài	countable noun
<b>Nn</b>	danh từ	nước, người, chuyện, ông	common noun
<b>Nq</b>	danh từ số lượng	một vài, phần lớn, mấy	numeral
<b>Nr</b>	danh từ riêng	Tuấn, Hồng, Thành, Hà Nội	proper noun
<b>Nt</b>	danh từ chỉ thời gian	sáng, tối, năm, khi	temporal noun
<b>Nu</b>	danh từ chỉ đơn vị	TP., tỉnh, khu phố	concrete noun
<b>ON</b>	từ tượng thanh	tách, bùm bụp, hì hì	onomatopoeia
<b>Pd</b>	đại từ không gian, thời gian	nào, này, đó, bao giờ	demonstrative pronoun
<b>Pp</b>	đại từ xưng hô	tui, con, anh, chị, ông	personal pronoun
<b>PU</b>	dấu câu	Dấu phẩy, dấu chấm	punctuation
<b>R</b>	trạng từ	được, đều, chưa, nào	adveRb
<b>Vc</b>	động từ so sánh	Là	comparative verb
<b>Vd</b>	động từ chỉ hướng	đến, ra, xuống	directional verb
<b>Ve</b>	động từ tồn tại	có, hết	state verb
<b>Vv</b>	động từ ý chí	viết, muốn, được, thay, ăn	volotive verb
<b>X</b>	không xác định	v.v	Unidentified POS