Measuring Text Readability by Lexical Relations Retrieved from WordNet

Shu-yen Lin, Cheng-chao Su, Yu-da Lai, Li-chin Yang, Shu-kai Hsieh
English Department
National Taiwan Normal University
sylin@ntnu.edu.tw

Abstract

Current readability formulae have often been criticized for being unstable or not valid. They are mostly computed in regression analysis based on intuitively-chosen variables and graded readings. This study explores the relation between text readability and the conceptual categories proposed in Prototype Theory. These categories form a hierarchy: Basic level words like *guitar* represent the objects humans interact with most readily. They are acquired by children earlier than their superordinate words (or hypernyms) like *stringed instrument* and their subordinate words (or hyponyms) like *acoustic guitar*. Therefore, the readability of a text is presumably associated with the ratio of basic level words it contains. WordNet, a network of meaningfully related words, provides the best online open source database for studying such lexical relations. Our preliminary studies show that a basic level word can be identified by its frequency to form compounds (e.g. *chair* → *armchair*) and the length difference from its hyponyms in average. We compared selected high school English textbook readings in terms of their basic level word ratios and their values calculated in several readability formulae. Basic level word ratios turned out to be the only one positively correlated with the text levels.

Keywords: Readability, Ontology, Prototype Theory, WordNet, Basic Level Word

1. Introduction

Reading process is the core of language education. Teachers now have access to a vast amount of texts extractable from the Internet inter alia, but the materials thus found are rarely classified according to comprehension difficulty. It is not uncommon to see foreign language teachers using texts not compatible with the students' reading abilities.

Traditional methods of measuring text readability typically rely on the counting of sentences, words, syllables, or characters. However, these formulae have been criticized for being unstable and incapable of providing deeper information about the text. Recently, the focus of readability formula formation has shifted to the search for meaningful predictors and stronger association between the variables and the comprehension difficulty.

We start our research by assuming in line with Rosch et al.'s Prototype Theory [1] that words form conceptual hierarchies in that words at different hierarchical levels pose different processing difficulties. This processing difficulty is presumably correlated with the reading difficulty of the text containing the words. Putting the logic into templates, the measurement of text readability can be done by calculating the average hierarchical levels at which the words of a text fall.

Our study comprises two stages. In the preliminary experiments, we utilized WordNet [2], an online lexical database of English, to identify basic level words. In the subsequent experiment, we compared selected readings in terms of their basic level word ratios and their values calculated in several readability formulae. Basic level word ratios turned out to be the only one positively correlated with the text levels.

The remainder of this paper is organized as follows: Section 2 reviews the common indices the traditional readability formulae are based on and the criticism they have received. In Section 3, we first review an approach that centers on ontology structure, and then propose our own ontology-based approach. Section 4 is about methodology – how to identify basic level words, and how to assess the validity of our method against other readability formulae. Section 5 reports the results of the assessment and discusses the strength and weaknesses of our approach. In this section, we also suggest what can be done in further research.

2. Literature Review

In this section we first summarize the indices of the traditional readability formulae and then give an account of the criticism these formulae face.

2.1 Indices of Readability – Vocabulary, Syntactic, and Semantic Complexity

The earliest work on readability measurement goes back to Thorndike [3] where word frequency in corpus is considered an important index. This is based on the assumption that the more frequent a word is used, the easier it should be. Followers of this logic have compiled word lists that include either often-used or seldom-used words whose presence or absence is assumed to be able to determine vocabulary complexity, thus text complexity. Vocabulary complexity is otherwise measured in terms of word length, e.g., the Flesch formula [4] and FOG formula [5]. This is based on another assumption that the longer a word is, the more difficult it is to comprehend [6].

Many readability formulae presume the correlation between comprehension difficulty and syntactic complexity. For Dale and Chall [7], Flesch formula [4], and FOG index [5], syntactic complexity boils down to the average length of sentences in a text. Heilman, Collins-Thompson, Callan, and Eskenazi [8] also take morphological features as a readability index for morphosyntactically rich languages. Das & Roychoudhury's readability index [9] for Bangla has two variables: average sentence length and number of syllables per word.

Flesch [4] and Cohen [10] take semantic factors into account by counting the abstract words of a text. Kintsch [11] focuses on propositional density and inferences. Wiener, M., Rubano, M., and Shilkret, R. [12] propose a scale based on ten categories of semantic relations including, e.g., temporal ordering and causality. They show that the utterances of fourth-, sixth-, and eighth-grade children can be differentiated on their semantic density scale.

Since 1920, more than fifty readability formulae have been proposed in the hope of providing tools to measure readability more accurately and efficaciously [13]. Nonetheless, it is not surprising to see criticism over these formulae given that reading is a complex process.

2.2 Criticism of the Traditional Readability Formulae

One type of criticism questions the link between readability and word lists. Bailin and Grafstein [14] argue that the validity of such a link is based on the prerequisite that words in a language remain relatively stable. However, different socio-cultural groups have different core vocabularies and rapid cultural change makes many words out of fashion. The authors also question the validity of measuring vocabulary complexity by word length, showing that many mono- or bi-syllabic words are actually more unfamiliar than longer polysyllabic terms.

These authors also point out the flaw of a simple equation between syntactic complexity and sentence length by giving the sample sentences as follows:

(1) I couldn’t answer your e-mail. There was a power outage.

(2) I couldn’t answer your e-mail because there was a power outage.

(2) is longer than (1), thus computed as more difficult, but the subordinator “because” which explicitly links the author’s inability to e-mail to the power outage actually aids the comprehension. The longer passage is accordingly easier than the shorter one.

Hua and Wang [15] point out that researchers typically select, as the criterion passages, standard graded texts whose readability has been agreed upon. They then try to sort out the factors that may affect the readability of these texts. Regression analyses are used to determine the independent variables and the parameters of the variables. However, the researchers have no proof of the cause-effect relation between the selected independent variables and the dependent variable, i.e., readability.

Challenge to the formula formation is also directed at the selection of criterion passages. Schriver [16] argue that readability formulae are inherently unreliable because they depend on criterion passages too short to reflect cohesiveness, too varied to support between-formula comparisons, and too text-oriented to account for the effects of lists, enumerated sequences and tables on text comprehension.

The problems of the traditional readability formulae beg for re-examination of the correlation between the indices and the readability they are supposed to reflect.

3. Ontology-based Approach to Readability Measurement

3.1 An ontology-based method of retrieving information

Yan, X., Li, X., and Song, D. [17] propose a domain-ontology method to rank documents on the generality (or specificity) scale. A document is more specific if it has broader/deeper Document Scope (DS) and/or tighter Document Cohesion (DC). DS refers to a collection of terms that are matched with the query in a specific domain. If the concepts thus matched are associated with one another more closely, then DC is tighter. The authors in their subsequent study [18] apply DS and DC to compute text readability in domain specific documents and are able to perform better prediction than the traditional readability formulae.

In what follows we describe the approach we take in this study, which is similar in spirit to Yan et al.'s [18] method.

3.2 An Ontology-based Approach to the Study of Lexical Relations

In this small-scaled study, we focus on lexical complexity (or simplicity) of the words in a text and adopt Rosch et al.'s Prototype Theory [1].

3.2.1 Prototype Theory

According to Prototype Theory, our conceptual categorization exhibits a three-leveled hierarchy: basic levels, superordinate levels, and subordinate levels. Imagine an everyday conversation setting where a person says “Who owns this piano?”; the naming of an object with ‘piano’ will not strike us as noteworthy until the alternative “Who owns this string instrument?” is brought to our attention. Both terms are truth-conditionally adequate, but only the former is normally used. The word ‘piano’ conveys a basic level category, while ‘string instrument’ is a superordinate category. Suppose the piano in our example is of the large, expensive type, i.e., a grand piano, we expect a subordinate category word to be used in e.g. “Who owns this grand piano?” only when the differentiation between different types of pianos is necessary.

Basic level is the privileged level in the hierarchy of categorical conceptualization. Developmentally, they are acquired earlier by children than their superordinate and subordinate words. Conceptually, basic level category represents the concepts humans interact with most readily. A picture of an apple is easy to draw, while drawing a fruit would be difficult, and drawing a crab apple requires expertise knowledge. Informatively, basic level category contains a bundle of co-occurring features – an apple has reddish or greenish skin, white pulp, and a round shape, while it is hard to pinpoint the features of ‘fruit’, and for a layman, hardly any significant features can be added to ‘crab apple’.

Applying the hierarchical structure of conceptual categorization to lexical relations, we assume that a basic level word is easier for the reader than its superordinate and subordinate words, and one text should be easier than another if it contains more basic level words.

3.2.2 WordNet – An Ontology-Based Lexical Database of English

WordNet [2] is a large online lexical database of English. The words are interlinked by means of conceptual-semantic and lexical relations. It can be used as a lexical ontology in computational linguistics. Its underlying design principle has much in common with the hierarchical structure proposed in Prototype Theory illustrated in 3.2.1. In the vertical dimension, the hypernym/hyponym relationships among the nouns can be interpreted as hierarchical relations between conceptual categories. The direct hypernym of 'apple' is 'edible fruit'. One of the direct hyponyms of 'apple' is 'crab apple'. Note, however, hypernyms and hyponyms are relativized notions in WordNet. The word 'crab apple', for instance, is also a hypernym in relation to 'Siberian crab apple'. An ontological tree may well exceed three levels. No tags in WordNet tell us which nouns fall into the basic level category defined in Prototype Theory. In the next section we try to retrieve these nouns.

4. Methodology

4.1 Experiment 1

We examined twenty basic level words identified by Rosch et al. [1], checking the word length and lexical complexity of these basic level words and their direct hypernyms as well as direct hyponyms in WordNet [2]. A basic level word is assumed to have these features: (1) It is relatively short (containing less letters than their hypernyms/hyponyms in average); (2) Its direct hyponyms have more synsets[1] than its direct hypernyms; (3) It is morphologically simple. Notice that some entries in WordNet [2] contain more than one word. We assume that an item composed of two or more words is NOT a basic level word. A lexical entry composed of two or more words is defined as a COMPOUND in this study. The first word of a compound may or may not be a noun, and there may or may not be spaces or hyphens between the component words of a compound.

Table 1: Twenty basic level words in comparison with their direct hypernyms and hyponyms on (average) word length, number of synsets, and morphological complexity*

Item	Basic Level		Direct Hypernym			Direct Hyponym		
	W. Length	M. Complexity	W. Length	Synsets	M. Complexity	W. Length	Synsets	M. Complexity
guitar	6	A	18	1	B	10	6	A, B
piano	5	A	18	3	B	10	3	A, B
drum	4	A	20	1	B	7.4	8	A, B
apple	5	A	7.5	2	A, B	10.67	3	A, B
peach	5	A	9	1	B	9	0	N/A

[1] A synset is a set of synonyms. The direct hypernym of 'piano', for instance, is grouped into three synsets: (1) keyboard instrument, (2) stringed instrument, and (3) percussion instrument, percussive instrument.

grape	5	A	11	1	B	11.8	3	A, B
hammer	6	A	8	1	B	9.7	9	A, B
saw	2	A	8	1	B	8.7	7	A, B
screwdriver	11	A	8	1	B	19.8	3	B
pants	5	A	7	1	A	8.9	18	A, B
socks	4	A	7	1	A	7.2	5	A, B
shirt	5	A	7	1	A	7.56	9	A, B
table	5	A	5	1	A	13.8	6	A, B
lamp	4	A	20	1	B	9.88	17	A, B
chair	5	A	4	1	A	11.2	15	A, B
car	3	A	12	1	A, B	7	31	B
bus	3	A	15	1	A, B	8	3	B
truck	5	A	12	1	A, B	8	11	B
dog	3	A	10	2	A, B	7	18	B
cat	3	A	6	1	A, B	9	2	B

*A refers to “single word” and B refers to “compound”.

The results confirm our assumption. First, the average word length (number of letters) of both the hypernyms and the hyponyms is much longer than that of the basic level words. Second, the hyponyms have a lot more synsets than the hypernyms. Third, in contrast to the basic level words which are morphologically simple, their direct hypernyms and hyponyms are more complex. Many of the hypernyms are compounds. The hyponyms are even more complex. Every basic level word (except ‘peach’) has at least one compounded hyponym.

4.2 Experiment 2

In this experiment, we examined the distribution of the compounds formed by the basic level words and their hypernyms and hyponyms. We also randomly came up with five more words that seem to fall into the basic level category defined by Rosch et al. [1]. These basic level words (e.g. ‘guitar’) are boldfaced in each item set in Table 2 below. Above each basic level word is its (or one of its) direct hypernym(s) (e.g. ‘stringed instrument’), under the basic level word is the first-occurring direct hyponym (e.g. ‘acoustic guitar’). When the basic level word has more than one level of hyponym, the first word at the second hyponymous level was also examined (e.g. ‘movable barrier’, ‘door’, ‘car door’, ‘hatchback’). For words that have more than one sense, we focused only on the sense defined in Rosch et al. [1]. For example, the noun ‘table’ has six senses in WordNet; we only focused on the sense ‘a piece of furniture’.

For each target item, we clicked on its FULL HYPONYM in WordNet 3.0 [2] to find the compounds formed by the target item. The next step was to count the compounds formed by the target words. For example, among the twelve hyponyms of ‘guitar’, five are compounds formed by ‘guitar’ – ‘acoustic guitar’, ‘bass guitar’, ‘electric guitar’, ‘Hawaiian guitar’, and ‘steel guitar’. In contrast, only one hyponym of ‘stringed instrument’ is a compound containing ‘stringed instrument’. As for ‘acoustic guitar’, it has no hyponyms. We assume that basic level words are more apt to form compounds than their hypernyms as well as hyponyms, so their compound ratios are calculated: Number of compounds is divided by number of hyponyms. We also keep record of the level where a compound occurs.

Table 2: Compound ratios and distribution of compounds in hyponymous levels

Hypernym Basic Level Word Hyponym	Cpd # / Hyponym #	Cpd Ratio (%)	Number of Compounds at Hyponymous Levels					
			1st Level	*2nd Level*	*3rd Level*	*4th Level*	*5th Level*	*6th Level*
stringed instrument	1 / 86	1	1	0	0	0		
guitar	5 / 12	42	5					
acoustic guitar	0 / 0	0						
keyboard	0 / 35	0	0	0	0			
piano	8 / 16	50	4	4				
grand piano	3 / 8	38	3					
baby grand piano	0 / 0	0						
percussion	0 / 68	0	0	0	0			
drum	5 / 14	36	5					
bass drum	0 / 0	0						
edible fruit	0 / ...	0	0	0	0	0		
apple	5 / 29	17	5	0	0			
crab apple	2 / 8	25	2					
Siberian crab	0 / 0	0						
N/A	N/A	N/A						
peach	0 / 0	0						
N/A	N/A	N/A						
edible fruit	0 / ...	0	0	0	0	0		
grape	6 / 17	35	3	2	1			
muscadine	0 / 0	0						
hand tool	0 / ...	0	0	0	0	0		
hammer	7 / 16	44	7	0				
ball-peen hammer	0 / 0	0						
hand tool	0 / ...	0	0	0	0	0	0	
saw	25 / 30	83	13	12	0			
bill	0 / 0	0						
hand tool	0 / ...	0	0	0	0	0	0	
screwdriver	4 / 4	100	4					
flat tip screwdriver	0 / 0	0						
garment	4 / 448	0	3	1	0	0	0	
pants	9 / 49	18	8	1				
bellbottom trousers	0 / 0	0						
hosiery	0 / 29	0	0	0				

socks	5 / 13	38	5					
anklet	0 / 0	0						
garment	4 / 448	0	3	1	0	0	0	
shirt	8 / 17	47	8	0				
camise	0 / 0	0						
furniture	4 / ...	0	4	0	0	0	0	
table	39 / 79	49	32	7	0	0		
alter	0 / 0	0						
source of	0 / 108	0	0	0	0	0	0	
lamp	27 / 68	40	14	12	1	0		
Aladdin's lamp	0 / 0	0						
seat	6 /102	6	2	3	1	0		
chair	31 / 48	65	17	14	0			
armchair	0 / 10	0	0	0				
captain’s chair	0 / 0	0						
motor vehicle	0 / 153	0	0	0	0	0		
car	21 / 76	28	19	2				
amphibian	0 / 2	0	0					
public transport	0 / 38	0	0	0	0			
bus	3 / 5	60	3					
minibus	0 / 0	0						
motor vehicle	0 / 153	0	0	0	0	0		
truck	15 / 48	31	10	5	0			
dump truck	0 / 0	0						
canine	0 / ...	0	0	0	0	0	0	0
dog	51 / 279	18	13	20	16	2	0	
puppy	0 / 0	0						
feline	0 / ...	0	0	0	0			
cat	35 / 87	40	4	31				
domestic cat	0 / 33	0	0					
kitty	0 / 0	0						
publication	1 / 211	0	0	1	0	0	0	
book	39 / 145	27	21	14	4	0	0	
authority	0 / 7	0	0					
power of	0 / 0	0						
language unit	0 / ...	0	0	0	0	0	0	0
word	35 / 220	16	28	7	0	0	0	
anagram	0 / 1	0	0					
antigram	0 / 0	0						
material	16 / ...	0	14	2	0	0		
paper	59 / 210	28	40	18	1			

card	14 / 57	25	6	8					
playing card	0 / 48	0							
movable barrier	0 / 46	0	0	0	0				
door	18 / 23	78	13	5					
car door	0 / 1	0	0						
hatchback	0 / 0	0							
leaf	2 / 23	9	2	0	0				
page	5 / 20	25	5	0					
full page	0 / 0	0							

Note: The symbol "#" stands for "number". Cpd refers to "compound". The three dots indicate that the number of hyponyms is too many to count manually. The number is estimated to exceed one thousand.

The most significant finding is that basic level words have the highest compound ratios. In comparison with their hypernyms and hyponyms, they are much more frequently used to form compound words. Although some hyponyms like 'grand piano' and 'crab apple' also have high compound ratios, they should not be taken as basic level items because such compounds often contain the basic level words (e.g. 'Southern crab apple'), indicating that the ability to form compounds is actually inherited from the basic level words.

Our data pose a challenge to Prototype Theory in that a subordinate word of a basic level word may act as a basic level word itself. The word 'card', a hyponym of 'paper', is of this type. With its high compound ratio of 25%, 'card' may also be deemed to be a basic level word. This fact raises another question as to whether a superordinate word may also act as a basic level word itself.

Many of the basic level words in our list have three or more levels of hyponym. It seems that what is cognitively basic may not be low in the ontological tree. A closer look at the distribution of the compounds across the hyponymous levels reveals another interesting pattern. Basic level words have the ability to permeate through two to three levels of hyponyms in forming compounds. By contrast, words at the superordinate levels do not have such ability, and their compounds mostly occur at the direct hyponymous level.

4.3 Experiment 3

The goal of this experiment is to show that whether a word belongs to the basic level affects its readability. This in turn affects the readability of a text and should be considered a criterion in measuring text readability. An easy text presumably contains more basic level words than a difficult one. Put in fractional terms, the proportion of basic level words in a text is supposed to be higher than that of a more difficult text.

To achieve this goal, we need independent readability samples to be compared with our prediction. As readability is subjective judgment that may vary from one person to another, such independent samples are extremely difficult, if ever possible, to obtain. In this study, we resorted to a pragmatic practice by selecting the readings of English textbooks for senior high school students in Taiwan. Three textbooks from Sanmin Publishing Co., each used in the first semester of a different school year, were selected. We tried to choose the same type of text, so that text type will not act as a noise. Furthermore, since we do not have facility to run large-scale experiment yet, we limited the scope to two-hundred-word text at each level. Accordingly, the first two hundred words of the first reading subjectively judged as narrative were extracted from the textbooks (Appendix 1). All the nouns occurring in these texts, except proper names and pronouns, were searched for in WordNet [2]. Considering the fact that for a word with more than one sense, the distribution of hyponyms differs from one sense to another, we searched for the hyponyms of the word in the particular sense occurring in the selected readings. We know that this practice, if used in a large-scale study, is applicable only if sense tagging is available, and we hope that it will be available in the near future.

Based on the results of the two preliminary experiments, we assume that basic level words have at least the following two characteristics: (1) They have great ability to form compounded hyponyms; (2) Their word length is shorter than the average word length of their direct hyponyms. These characteristics can be further simplified as the **Filter Condition** to pick out basic level words:

(1) Compound ratio of full hyponym ≧ 25%;

(2) Average word length of direct hyponym minus target word length ≧ 4.

Note in passing that the second criterion differs fundamentally from the commonly used criterion of word length. Ours compares the target word with its full hyponyms. Word length is measured in relative terms: What is counted is the word length difference, not the word length itself. Based on the two assumed characteristics of our filter condition, the information for each noun we need includes: (1) Length of the target word, i.e. how many letters the word contains; (2) Compound ratio of the target word, i.e. how many hyponyms of the word are compounds formed by the word. Note that here the hyponyms refer to the full hyponyms, so all the words in every hyponymous synset were counted; (3) Average word length of the direct hyponyms. The next section reports the computed information via WordNet [2].

5. Results and Discussion

The three selected readings contain sixty nouns in total, of which twenty-one conform to the proposed Filter Condition of basic level words. They are given in Table 3 below. A comprehensive list of all the sixty nouns are given in Appendix 2 at the end of this paper. Note in passing that the level numbers refer to the presumed difficulty levels of the selected readings. Level 1 is presumably the easiest; Level 3, the hardest. These numbers should not be taken as ratio measurement. Level 3, for example, is not assumed to be three times harder than Level 1. We intend these numbers to stand for ordinal relations.

Table 3: Basic Level Words from the 200-word Texts at Three Levels

Target Word	Level	Compound Ratio (%)	Length of Target Word	Average Length of Direct Hyponyms
food	1	53	4	8
apple	1	56.6	5	10
vinegar	1	60	7	11
potato	1	62.5	6	11
cold	1	66.6	4	8
test	1	72.7	4	9
orange	1	88.8	6	11
soap	1	93	4	9
language	2	35.12	8	12
job	2	37.5	3	8
heart	2	40	5	15
technology	2	47.22	10	19
factor	2	63.64	6	12
culture	2	85.19	7	19
physics	3	32.84	7	12.6
question	3	35.71	7	15
barometer	3	40	9	13.25
system	3	60.95	6	12.93
time	3	62.22	4	10
office	3	72.22	6	11.5
call	3	93.33	4	11

In order to measure the text difficulty, basic level word ratios of the selected texts were computed. Table 4 shows the statistics. Diagrammatically, it is clear in Figure 1 that the basic level word ratios are decreasing as the difficulty levels of the selected readings increase. The text from Level-1 has the highest basic level word ratio; the text from Level-3 has the lowest basic level word ratio. This finding conforms to the levels of these textbooks, and proves the usefulness of the basic level word concept in the measurement of readability.

Table 4: Basic level word ratio at different levels

	Number of nouns	Number of Basic Level Words	Ratio of Basic Level Words
Level-1	17	8	47.1
Level-2	15	6	40.0
Level-3	28	7	25.0

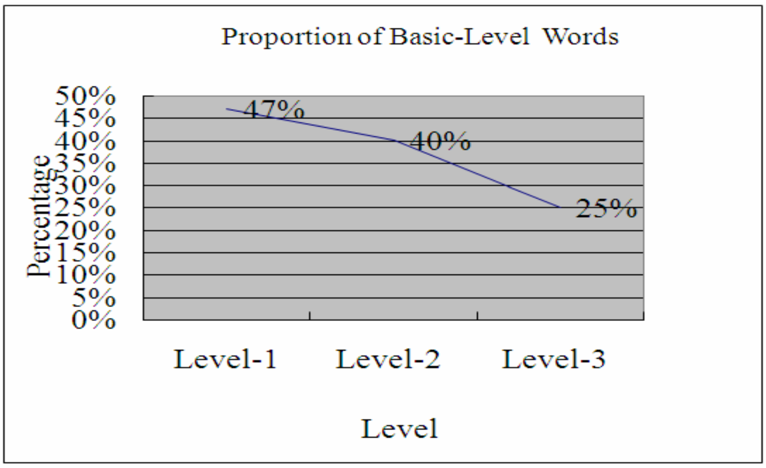

Figure 1: Basic Level Word Ratio of Selected Texts

Table 5 shows the readability scores of the selected readings measured by several readability formulae. Figure 2 displays the overall tendency computed by these formulae: Level-1 is the easiest, while Level-2 and Level-3 are at about the same difficulty level. The readability formulae seem not to be able to decipher the difference between the texts of Level-2 and Level-3 while our basic level word ratio can easily show their different difficulty levels.

Table 5: Readability of the 200-word Texts Computed by Several Readability Formulae

	Dale-Chall	Flesch Grade Level	FOG	Powers	SMOG	FORCAST	Spache
Level-1	4.6	2.1	7.8	4	6.4	7.7	2.4
Level-2	7.4	8.3	18.9	6.2	10.2	11.8	3.9
Level-3	6.3	9.5	16.4	5.9	10.5	9.1	4.8

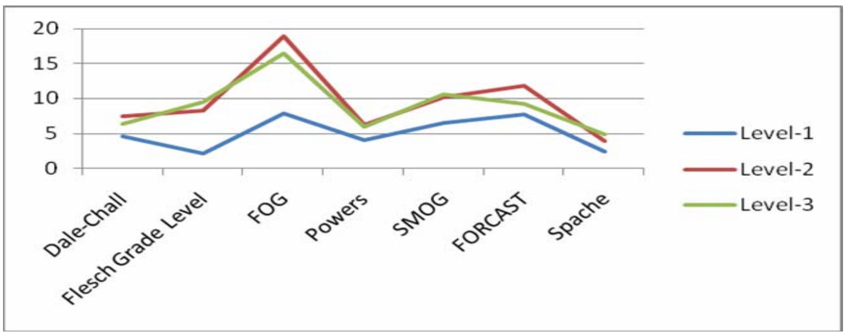

Figure 2: Readability of the 200-word Texts Computed by Several Formulae

This paper is just the first step to measure readability by lexical relations retrieved from WordNet [2]. Twenty-five percent of the twenty basic level words defined by Rosch et al. [1]

are NOT identified by our Filter Condition (e.g. ‘truck’, ‘shirt’, socks’). Among the identified basic level words in the three selected texts, some look rather dubious to us (e.g. ‘barometer’, ‘technology’). The filter condition proposed in this study certainly leaves room to be fine-tuned and improved in at least two respects. First, the two criteria of compound ratios and word length difference have been used as sufficient conditions. We will postulate the possibility of weighting these criteria in our subsequent research. Second, in addition to the lexical relations proposed in this study, there are presumably other lexical relations between basic level words and their hypernyms/hyponyms that are retrievable via WordNet [2]. Doubts can also be raised as to whether all basic level words are equally readable or easy. Can it be that some basic level words are in fact more difficult than others and some hypernyms/ hyponyms of certain basic level words are actually easier than certain basic level words?

We thank our reviewers for raising the following questions, and will put them in the agenda of our subsequent study: (1) The examined words in this study are all nouns. Can we find relationships between verbs, adjectives, and even adverbs like the hypernym/hyponym relationships with the basic level “nouns”? The tentative answer is yes and no. Take the example of the verb ‘run’. It has hypernyms in WordNet (‘speed’, ‘travel rapidly’, etc.). It also has subordinate lexical relation called ‘troponym’, which is similar to hyponym of nouns. Admittedly, English verbs do not constitute compounds so often as English nouns, but other lexical relations may exist between the verbs, and the relations are likely to be retrievable. (2) Although the small scale size of our experiments makes the validity of the results challengeable, the exciting findings of this study have provided the outlook of a large-scale project in the future. (3) Are basic level words frequent words in general? Can we use frequency to substitute for ‘basichood’ if the two criteria have approximately the same indexing power? We like to extend this question and ask whether the ontological relations between the lexical units in WordNet are correlated with word frequency. We hope we will be able to answer this question in a study of larger scale.

Laying out the groundwork for further research, we aim to tackle the following issues too. All traditional readability formulae implicitly suppose an isomorphic relation between form and meaning as if each word has the same meaning no mater where it occurs. We acknowledge that one of the biggest challenges and the most badly needed techniques of measuring readability is to disambiguate the various senses of a word in text since the same word may have highly divergent readability in different senses. Another tacit assumption made by the traditional readability formulae is that the units of all lexical items are single words. This assumption overlooks many compounds and fixed expressions and affects the validity of these formulae.

Although our research has provided the study of readability a brand new perspective and has offered exciting prospects, our challenges are still many and the road is still long.

References

[1] Rosch, Eleanor, Mervis, Carolyn, Gray, Wayne, Johnson, David, & Boyes-Braem, Penny, "Basic objects in natural categories," Cognitive Psychology 8: 382-439, 1976.

[2] WordNet, version 3.0. Princeton, N.J.: Princeton University. Retrieved from World Wide Web: http://wordnet.princeton.edu/perl/webwn?s=word-you-want, 2006.

[3] Thorndike, E.L., The Teacher's Word Book. New York: Teacher's College, Columbia University, 1921.

[4] Flesch, R., “A new readability yardstick”, Journal of Applied Psychology 32: 221-233, 1948.

[5] McCallum, D. R., & Peterson, J. L., "Computer-based readability indices," *Proceedings of the ACM ’82 Conference*, 1982.

[6] Chall, J., & Dale, E., *Readability revisited: The new Dale-Chall readability formula*. Cambridge, Massachusetts: Brookline Books, 1995.

[7] Dale, E., Chall, J., "Formula for predicting readability," *Educational Research Bulletin* 27 (1–20), 37–54, 1948.

[8] Heilman, Collins-Thompson, Callan & Eskenazi, "Combining Lexical and Grammatical Features to Improve Readability Measures for First and Second Language Texts," *Proceedings of the HLT/NAACL Annual Conference*, 2007.

[9] Das & Roychoudhury, "Readability Modelling and Comparison of One and Two Parametric Fit: A Case Study in Bangla," *Journal of Quantitative Linguistics, 13*, 17-34, 2006.

[10] Cohen, J.H., "The effects of content are material on cloze test performance," *Journal of Reading*, 19/3: 247-50, 1975.

[11] Kintsch, W., *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum, 1974.

[12] Wiener, M., Rubano, M., and Shilkret, R., "A measure of semantic complexity among predications," *Journal of Psycholinguistic Research*, Vol. 19, No. 2: 103-123, 1990.

[13] Crossley, S.A, Dufty, D.F., McCarthy, P.M., & McNamara, D.S., "Toward a new readability: A mixed model approach," *Proceedings of the 29th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society, 2007.

[14] Bailin, A. & Grafstein, Ann, "The Linguistic Assumptions Underlying Readability Formulae: A Critique," *Language and Communication* 21(3): 285-301, 2001.

[15] Hua, N. & Wang, G.., "Lun chuantong keduxing gongshi de bu-kexuexing," [On the non-scientific aspects of traditional readability formulae]. *KaoShiZhouKan* 18: 119-120, 2007.

[16] Schriver, K. A., "Readability formulas in the new millennium: What’s the use?" *ACM Journal of Computer Documentation, 24.3*: 138-140, 2000.

[17] Yan, X., Li, X., and Song. D., "Document generality: its computation for ranking," In: G. Dobbie and J. Bailey *Seventeenth Australiasian Database Conference (ADC2006)*, Hobart, Australia, 16-19 January, 2006a.

[18] Yan, X., Li, X., and Song. D., "Concept-based Document Readability in Domain Specific Information Retrieval," CIKM 2006: 540-549, 2006b.

Appendix 1: Three Pieces of 200-word-text from a Senior High School Textbook

Level 1: Book 1 Lesson 2

Scientist say that your tongue can recognize only four tastes. It can tell if something is sour (like vinegar) or bitter (like soap). But that's all. To tell different foods apart, we also have to use our noses.

Can you remember a time when you had a bad cold? Your food tasted very plain then. It seemed to have little taste at all. That wasn't because your tongue wasn't working. It was because your nose was stopped up. You couldn't smell the food, and that made it seem tasteless. You can prove this to yourself. Try eating something while you pinch your nose shut. It won't seem to have much taste.

Here's another test. It shows how important the nose is in tasting. First you blindfold a person. Then you put a piece of potato in his mouth. You tell him to chew it. At the same time, you hold a piece of apple under his nose. Then ask what food is in his mouth. Most people will say, " An apple." The smell of the apple fools them. The test works best when two foods feel the same in the mouth. It won't work well with apple and orange slices.

Level 2: Book 3 Lesson 2

When people from different cultures live and work together much more than before, change takes place. The languages of the world's dominant cultures are replacing the languages of the smaller cultures. You're learning English right now. Could this be the beginning of the end for the Chinese language? Of course not. *Mandarin* remains the healthy, growing language at the heart of Chinese culture. Mandarin steadily continues to spread among Chinese people worldwide. Elsewhere, *Swahili* grows in Africa. Spanish continues to thrive in *South America*. *Hindi* rules India. And of course almost everyone these days wants to learn English. However, many less common regional languages haven't been so lucky, because most young people have stopped learning them.

When less common languages disappear, two factors are to blame: trade and technology. Most international trade takes place in major world languages such as English or Mandarin. Cultures that isolate themselves from international business and major world languages have difficulty prospering.

Most children respect their own culture and traditions. But when it comes to getting a job, knowing a major world language if often essential. It may mean the difference between success and failure. For many, using a less common reginal language simply isn't

Level 3: Book 5 Lesson 2

Some time ago, I received a call from a colleague who asked if I would be the referee on the grading of an examination question. He was about to give a student a zero for his answer to a physics question, while the student claimed he should receive a perfect score and would if the system were not set up against the student. The instructor and the student agreed to submit this to an impartial judge, and I was selected.

I went to my colleagues' office and read the examination question: "Show how it is possible to determine the height of a tall building with the aid of a barometer." The student had answered: "Take the barometer to the top of the building, attach a long rope to it and lower the barometer to the street. Then bring it up and measure the length of the rope. The length of the rope is the height of the building."

I pointed out that the student really had a strong case for full credit, since he had answered the question completely and correctly. On the other hand, if full credit were given, it could well contribute to a high grade for the

Appendix 2: Nouns Extracted from the Three Pieces of 200-word-text.

Item	Target Item				Direct Hyponyms	
	Level	Cpd # / Hyponym #	Cpd Ratio (%)	Length	Avg. Length	Number
scientist	1	37/174	21	9	13	20
tongue	1	0/4	0	6	0	0
taste	1	4/34	11.7	5	6	9
vinegar	1	3/5	60	7	11	3
soap	1	14/15	93	4	9	8
food	1	1234/2310	53	4	8	15
nose	1	4/22	18	4	6	8
time	1	1/0	0	4	0	0
cold	1	2/3	66.6	4	8	1
test	1	8/11	72.7	4	9	5
person	1	3152/13235	23.8	6	8	401
potato	1	10/16	62.5	6	11	5
mouth	1	3/10	30	5	4	6
apple	1	17/30	56.6	5	10	3
smell	1	2/23	8.6	5	6	4
orange	1	8/9	88.8	6	11	3
slice	1	2/10	20	5	6	2
culture	2	23/27	85.19	7	19	7
language	2	425/1210	35.12	8	12	16
world	2	2/9	22.22	5	11	3
end	2	23/54	42.59	3	6	14
heart	2	2/5	40	5	15	2
factor	2	14/22	63.64	6	12	6

trade	2	16/66	24.24	5	10	3
technology	2	17/36	47.22	10	19	7
business	2	54/163	33.13	8	8	12
child	2	34/55	61.82	5	7	21
tradition	2	0/7	0	9	6	4
job	2	3/8	37.5	3	8	15
difference	2	0/11	0	10	10	9
success	2	24/58	41.38	7	7	5
failure	2	7/50	14	7	7	8
time	3	28/45	62.22	4	10	16
call	3	14/15	93.33	4	11	8
colleague	3	0/0	0	9	N/A	N/A
referee	3	0/0	0	7	N/A	N/A
grading	3	0/0	0	7	N/A	N/A
examination	3	20/32	62.5	11	9	24
question	3	10/28	35.71	7	15	3
student	3	16/48	33.33	7	9.25	20
zero	3	0/0	0	4	N/A	N/A
answer	3	0/2	0	6	8	2
physics	3	22/67	32.84	7	12.6	18
score	3	1/5	20	5	8.5	4
system	3	103/169	60.95	6	12.93	28
instructor	3	30/55	54.55	10	10.86	21
judge	3	7/33	21.21	5	7.33	3
office	3	13/18	72.22	6	11.5	8
height	3	0/7	0	6	7.5	2
building	3	212/485	43.71	8	9.76	54
aid	3	0/1	0	3	8	1
barometer	3	2/5	40	9	13.25	4
top	3	0/9	0	3	5.8	5
rope	3	15/37	40.54	4	7.21	19
street	3	22/32	68.75	6	8.95	21
length	3	1/19	5.26	6	8.8	5
case	3	0/2	0	4	8	1
credit	3	1/9	11.11	6	7	3
hand	3	0/1	0	4	4	1
grade	3	1/5	20	5	8.5	4

Note 1: Level ranges from 1 to 3, which respectively represents the English textbooks of Book I for the first-year senior high school students, Book III for the second-year, and Book V for the third-year senior high school students in Taiwan.

Note 2: Cpd ratio refers to the ratio of compounds formed by the target item to the total number of the target item's full hyponyms.

Note 3: Direct hyponyms refer to the lexical items at the level immediate below the target item.