ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0007

# A Review of Lexical Richness in China and Abroad

Meiying Long[1, a]

[1]School of language and literature, University of South China, Hengyang, China

[a]20192003110114@stu.usc.edu.cn

## Abstract

**This study is to fill the blanks of comprehensive lexical richness review so as to offer references both to scholars who are dedicated themselves to this area and to learners and practitioners who are still struggling with vocabulary. In order to do that, definitions, measurements, classifications and previous studies and trends constitute this research. We found that controversy exist in using the 5 categories (lexical density, lexical diversity, lexical sophistication, lexical originality and errors) to calculate lexical richness. Besides, we encourage the areas of lexical richness to goes beyond, such as the relation to text readability, which is seldomly examined and promising as well.**

## Keywords

**Lexical richness; Measurement; Disputes.**

## 1. Introduction

Vocabulary is an efficient indicator of one's second language proficiency (Linnarud, 1986). The period of 1980s witnesses the uprising studies on vocabulary with rapid development of measurement tools. In order to measure one's vocabulary level, lexical richness, as one of the measurement tools, was put forward.

Broadly speaking, lexical richness serves as a uniform concept. It consists of more than one single concepts, such as lexical diversity, lexical density, and lexical frequency profile. For different researchers, its component varies to meet different needs. Narrowly speaking, it is another term for lexical diversity or lexical complexity, a vital tool used for measuring vocabulary in writing (Daller et al. 2004; Read 2000). It involves both width and depth of lexical knowledge. Despite of a few supporters of narrow definition of lexical richness, definition of lexical richness in broad sense has been widely accepted.

Lexical richness was and continues to be the center of second language learning. Therefore, a comprehensive review of lexical richness is in great necessity. This research is dedicated to depict a full picture of recent studies on lexical richness through three dimensions, classification, measurement and previous studies at home and abroad.

## 2. Previous Studies on Lexical Richness Abroad

### 2.1. Three Classifications of Lexical Richness

As aforementioned, scholars have come up with different framework of lexical richness.

For Linnarud (1986), lexical richness can be classified into lexical individuality, lexical variation and lexical density.As far as Laufer & Nation (1995) concerned, four categories, namely, lexical variation, lexical density, lexical sophistication and lexical originality constitute the framework of lexical richness. Laufer & Nation's classification received some criticisms. For example, Read argue that lexical originality fails to measure learners' lexical ability. On this very note, Read come up with a new classification. According to his classification, lexical richness is categorized into lexical density, lexical diversity and lexical sophistication as well as lexical error. Although

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0007

this classification seems fixed, it was adjusted to meet scholars' needs. For instance, lexical error was excluded when learners are of high language proficiency together with repeated revisions.

From the above-mentioned classification, it can be detected that in Read's classification, lexical diversity, in some cases known as lexical variation, lexical density as well as lexical sophistication remain to be defining factors in most cases. And Read's classification has been proven to be valid in various research.

## 2.2. Measurement of 5 Categories

As for the measurement of lexical richness, it is closely related to various classifications of different scholars. In other words, one category is likely to be measured by its corresponding index. For example, lexical diversity is measured by STTR (standard-type-token-ratio). Generally speaking, despite the fact that some scholars incline to employ one single index. For instance, lexical quality formula is applied in Arnaud's research. Increasing number of scholars believe using multiple indexes is of higher validity. The following are detailed measurement descriptions of lexical density, lexical diversity, lexical sophistication, lexical originality as well as lexical errors.

### 2.2.1. Measurement of Lexical Density

Lexical density can be calculated through the ratio of lexical words in the texts according to Ure's definition. Based on Ure's formula proposed in 1971, some adjustments are made by Halliday (1985), Laufer (1991), and Laufer & Nation (1995). These formulas have been employed in research their validity has been proven. For instance, Lu (2012) applies Ure's formula to examine second language learners' performance in oral English tests. Besides, Zhu and Wang (2013) used Halliday's formula to investigate developmental traits of 30 English majors from the first year to the last year. These formulas are illustrated as follows.

LD=the total of lexical words/ the total of words (Ure, 1971)

LD=the total items of words*100%/ the total of sentences (Hallliday, 1985)

LD=the total of lexical words*100%/ the total of lexemes (Laufer, 1991)

LD=the total of word types*100/ the total of word types (Laufer & Nation, 1995)

These Throughout the years, it has received several doubts from scholars such as Zhu &Wang (2013), Wang (2017) in view of its validity in predicting writing quality. Specifically, Zhu & Wang argue that the ratio calculation does not consider the situation when one single word was used for multiple times. Although higher lexical density in general means large vocabulary and thus excellent writing ability. In this scenario, it not only fails to illustrate large text information despite lexical density being high, but might mislead the researchers' assessment on learners' vocabulary and writing skills. Likewise, in Laufer & Nation's (1995) opinion, few functional words in a text may indicate low level of grammar, such as little use of subordinate clause, participle clause and ellipsis. Though they are not lexical words, but structural features can be reflected which are also part of one's language proficiency.

### 2.2.2. Measurement of Lexical Diversity

Lexical diversity can be measured in various ways. Historically, Read (2000) applied TTR (Type-token Ratio), which consists of its types and tokens. Higher ratio indicates larger vocabulary of the learner and thus higher lexical richness. However, it is highly sensitive to length of the data sample. That is to say, shorter texts usually have disproportionally higher TTR compared to longer texts, thus the value of the indices is less reliable. With the aim to overcome its deficiency, scholars have come up with various solutions including Root TTR (Guiraud 1960), T method (Tuldava 1993), D value method (Malvern & Richard 2002), MTLD method (McCarthy 2005), MATTR (Moving-Average-Type-Token-Ratio) (Covington 2007; Covington & McFall 2010), Hypergeometric Distribution (McCarthy& Jarvis, 2010) as well as

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0007

STTR (Standard Type Token Ratio). As one of the latest methods, STTR, proposed by Scott in 2008, is not only consistent irrespective of text length but can be calculated conveniently. It was used for lexical diversity measurement, such as in Zhu & Wang (2013)'s research. However, it does not indicate ineffectiveness of other methods mentioned above. Different methods are employed to satisfy research needs. Generally speaking, adjusted methods have all successfully overcome the shortcoming of TTR through text length control. For instance, MTLD method proposed by McCarthy in 2005 employed strict control of text length. To be specific, formula L/n (L means text length and n refers to the total number of parsed texts) is applied only when the value of each sample text reaches 0.72.

In a word, despite slight differences in calculation, these methods are equally effective for measuring lexical diversity.

### 2.2.3. Measurement of Lexical Sophistication

Lexical sophistication refers to the frequency of unusual word or advanced words in a text. It is measured either by formulas or software. When it comes to formulas, Linnarud (1986) and Hyltenstam (1988) first calculated lexical sophistication through the ratio of sophisticated words in texts. Subsequently, it is measured by Lexical Frequency Profile, the most frequently used method put forward by Laufer & Nation in 1995. This measurement highlights the distribution of words varying from frequency levels. Concerning the details, it includes three wordlists: first 1000 most frequent words, second 1000 most frequent words and the 570 most frequent academic words. Quite a few studies have testified the its reliability. For instance, Chen (2011) applied Lexical Frequency Profile in research on different genre writing. He came to conclusion that on the one hand, compared to narrative writing, learners employ more academic words in expository writing. On the other hand, a variety of word types and comparatively complex words were used. Furthermore, Higgibotham & Reid (2019) revealed that thesis with high grades tend to use more low-frequency academic words while their counterparts rely on high-frequency ones. Besides, lexical sophistication is also measured by its squared version (Chaudron & Parker 1990) and corrected type-token ratio (Wolfe-Quintero et al. 1998). What is more, calculating lexical sophistication through software is also accepted. Judit Kormos (2011), for example, uses Coh-Metrix 2.0. Although there are two kinds of ways to measure lexical sophistication, the core, calculating through word lists, remains the same. However, these measurements have received some criticism. Wang (2017) questioned the definition of lexical sophistication by asking "to what extent does a word can be defined as a complicated word for learners?" Indeed, using word lists that mainly contains rare words cannot accurately determine lexical sophistication since sophisticated words varies for different individuals. For example, if we are to examine both kindergarten children and senior high school students, should we use the same word lists? Apparently, age and knowledge background will affect lexical sophistication. As Laufer & Nation (1995) pointed out: "Clear definition of sophisticated words determines lexical sophistication and different results analysis will be achieved under different definitions. If the research subjects are selected under different education system, the term advanced or complex words should be clearly defined."

To sum up, the measurement of lexical sophistication is currently stuck using word lists either are embedded in software or formulas counting sophisticated words in total texts. Besides, what deserves our attention is the concept of being sophisticated is rather subjective.

### 2.2.4. Measurement of Lexical Originality

Concerning lexical originality (LO), known as lexical individuality (Linnarud, 1986) or lexical specificity (Birgit Harley & Mary Luo King 1989). It means uniqueness which can be reflected by counting unique words (Laufer & Nation, 1995). As we are able to detect from this definition, uniqueness is a comparative concept. Generally, distinctive words employed by learners

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0007

demonstrate his or her depth and width of vocabulary, thus, with higher language proficiency level (Linnarud, 1986; Laufer, 1991). The formula follows in below.

LO= the total of lexeme of one author* 100%/ the total of lexemes (Rod Ellis, 1984)

However, this statement was criticized primarily for three reasons. To begin with, the limitations on data. Owing the limited subjects, research results have very little applicability. In other words, if any factors are to be changed, for instance, the number of the subjects, the ultimate results vary since lexical originality is not entirely determined by one but group members. Similarly, Read (2000) also supports that this method is not reliable since it varies with the size and language proficiency of research subjects. Furthermore, limited applications. To be specific, it applies solely to small samples, such as a class quiz test. When the number of subjects grows, large vocabulary is likely to be covered. Under this circumstance, there is slight possibility that learners use unique words. Even if one does employ several unique words, it is possible not to be detected out of limited numbers and restricted time for the assessors. Besides, another reason is the ambiguous definition of "originality". To put it in another way, in what sense a word is justified to be an original or unique word? Without fixed standards, different evaluators incline to have different judgement. Due to these three reasons, quite a few scholars have reach common sense that it remains unsolved whether it should be applied when measuring one's lexical richness.

In a nutshell, lexical originality is measured by the number of unique words occurring in the text. Even if it can reflect one's lexical ability in small samples, it bears three disadvantages, including limited data and practicability as well as vague definitions of "originality".

### 2.2.5. Measurement of Lexical Errors

Lexical errors, a term with a long history, were derived from error analysis in the 1960s. It is also a term scholars and practitioners find familiar with. It is evitable in the process of second language learning and its analysis greatly contribute to second language teaching. And we are aware that lexical errors are one of the factors affecting writing since writing requires not only grammar, cultural knowledge but vocabulary. Lexical errors are usually calculated through corpus, though its measurement varies with different classification. There are three classifications in total. Granger (1994) stands that lexical errors are composed of lexico-grammatical errors and pure lexical errors. For James (2001), lexical errors can be classified into formal errors and semantic errors. And formal errors are further divided into distortion, formal mis-selection and mis-formation. Besides, semantic errors to be cut into collocation errors and meaning confusion. Recently, Gui and Yang (2003) put forward a novel classification which consists of forms, verbs, nouns, pronouns, adjectives, adverbs, prepositions, conjunctions, words, collocations and syntax. These classifications determined the way how lexical errors are calculated.

## 2.3. Previous Studies on Lexical Richness in China

Studies on lexical richness abroad can be traced back to 1980s, which can be divided approximately into three perspectives: (1) The correlational studies on lexical richness and second language (SL)writing quality. Engber (1995) analyzed 66 compositions written by foreign students within limited time. The results illustrate that lexical richness plays a vital role in passage construction. (2) The contrastive studies on lexical richness between SL learners and native speakers. Linnarud (1986), for example, writings by Swedish students significantly differ from English native speakers in the same grade in lexical errors, lexical variation, lexical collocations, lexical originality and so on. Similarly, Lei (2020) conducted research on the lexical richness difference in thesis by English native speakers and Chinese advanced language learners. Regardless of the difference between them, the study acknowledged the significance of academic expertise. (3) The developmental characteristics of lexical richness among SL learners. Siegler (2006) believes that learners' lexical richness is neither one-way forward nor

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0007

linear. That is to say, it is normal if lexical richness goes backward. Capsi & Lowie's study in 2010 also supports this finding by stating that various features can be observed in different stages and we shall attach great importance to those traits.

Not until early 21st century, lexical richness in SL writing attracts Chinese scholars' attention. The past 20 years has witnessed its flourishing fruits, which can be categorized primarily into three aspects: (1) The correlational studies on lexical richness and SL writing quality. Due to the significance of lexical richness as a pivotal indicator for one's vocabulary, a crucial component of one's writing performance, the relationship between lexical richness and writing performance has been carefully researched. For instance, Qin &Wen (2007) found out that lexical diversity and lexical sophistication is directly proportional to one's writing quality after analyzing 240 compositions by English majors varying from the first grade to the fourth grade. (2) The studies on pedagogical implications. Some scholars saw great value in how teaching methods and teaching activities improve students' lexical richness, thus offering guidance for vocabulary teaching. Dai (2019), for instance, shed some light on the effect of continuation task types and second language proficiency on lexical richness. The results suggest that continuation task is conductive to the improvement of learners' improvement. Concerning language proficiency, advanced learners perform better than beginners, which testify lexical richness can indeed predict learners' vocabulary to some extent. Besides, low-level learners benefit from continuation task in lexical richness, lexical density and lexical diversity in particular. Other than continuation task, scholars have also tapped into other teaching activities. Xiao (2018) focused on how two writing tasks, namely, story writing according to pictures and abridged writing after reading affect inter-mediate learners' learning. It was demonstrated that writing tasks greatly impact lexical diversity and lexical originality and the same writing tasks influence language learners' lexical richness but spare the native speakers. Research was conducted from this perspective to offer practical teaching implications. (3) The longitudinal study on lexical richness. Wang (2012) detected the developmental features of lexical richness of 30 non-English majors. That is the index for lexical diversity, lexical sophistication is ascending as learners' language proficiency improves. Similarly, Wan (2010) selected 200 pieces of compositions for TEM-4 and TEM-8 from the same group of students as research samples. Her study back up Wang's finding that language proficiency is an important indicator for lexical sophistication and lexical diversity. It also reveals that the longer the article, the more lexical errors, especially errors in articles. Likewise, after intensive study on lexical richness of non-English majors in the first and second grade and junior and senior English majors, Bao (2008) came to conclusion that lexical originality, lexical diversity, lexical density illustrates no signs for linear development. Lexical sophistication is an exception though.

Compared to studies abroad, we seek to learn from comparative studies on lexical richness by international scholars while continuing relevant research with intangible pedagogical implications. In a nutshell, there are few studies on the effect of vocabulary on text readability, but as an effective means to assess learners' vocabulary knowledge, lexical richness is worth investigating.

## 3. Specific Trends of Lexical Richness

To be specific, considering lexical density alone, it has mainly been to two areas. To start with, it was used to predict writing quality. Some researchers hold the position that lexical words are used express meanings (Engber, 1995). Therefore, higher lexical density entails higher writing quality. In addition, it also applies to differentiate genres. For instance, Ure (1971), Halliday (1985) and Stubbs (1986) found out that lexical density is higher in written language than in spoken language. As stated earlier, it was supported by Zhu & Wang's (2013) research in which lexical density of college students majoring in English from the first to the last year climbs as a

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0007

whole. However, some scholars stand in the opposite position. They believe whether lexical density should be excluded from the framework remains unresolved. The reasons for that are due to lexical density fails to reflects learners' language proficiency if same words are used repeatedly or simple sentences fulfill the whole text. Briefly, lexical density studies are either on writing quality prediction or genre differentiation. And its validity was questioned due to its own limitations.

Concerning lexical diversity, it is primarily employed to investigate the relationship with writing quality as well as discrepancy between first and second productive vocabulary. For the former, quite a few scholars proved the reliability of lexical diversity. For example, Engber (1995) observed advanced English learners in American university and found out that lexical diversity is closely related to their writing quality. What is more, Durán et al (2004) revealed that enfants ranging from 18 to 42 months develop higher diversity with age. However, some argue lexical diversity fails to depict the differences in spontaneous spoken language (Vermeer 2000). In addition, studies on lexical diversity discrepancy in first and second language speakers are also conducted. To name a few, Linnarud (1986) compares Sweden's English speakers (age 17) with 9 years learning experiences to their native counterparts. Although its validity seems to be confirmed, some (Foste & Tavakoli, 2009) found no lexical diversity differences in first and second language learners. From the above, we can summarize that despite affluent research the validity of lexical diversity is still at odds.

When it comes to lexical sophistication, we still focus on whether lexical sophistication can be used to predict learners' language proficiency. As Bao (2008) pointed out, it distinguishes learners in different levels. Kormos (2011) also discovered that native speakers' lexical sophistication is far higher than second language learners. However, as we said, its issue lies in the concept of "sophisticated" as it is hard to be clearly defined. A few wordlists seem lack validity to reflect lexical sophistication, especially for learners with different language proficiency. To deal with it, wordlist should at least be adjusted.

With regard to lexical originality, theoretically, it is believed to be valid. Generally speaking, unique words usage indicates large vocabulary and thus higher language proficiency in comparative sense. However, in Read's research, he found that comparing to other dimensions of lexical richness, lexical originality is not stable since it varies when the number of groups members grows. Because of these severe drawbacks, it is rarely used in studies abroad and at home.

In respect to lexical errors, previous studies abroad are prevalent in 21st century. For example, Grauberg proposed lexical errors dominance by investigating advanced German learners. In his study, 102 lexical errors were found in 193 errors recorded. From his study, we see lexical errors paly a dominant role in errors which further proves the significance of vocabulary in learning a second language. In view of domestic studies, with burgeoning corpus, such as Chinese Learners English Corpus (CLEC) and Spoken and Written English Corpus of Chinese Learners (SWECCL). Chinese scholars have conducted inspiring research.

## 4. Conclusion

Through a complete and thorough presentation and discussion, two implications can be retrieved.

(1) Disputes on lexical richness measurement. As we infer, all categories have received more or less criticisms, either on vague definitions or measurement tools. But it should be kept in mind that plentiful results have achieved as well.

(2) Limited research areas. From previous studies, we can see that two topics, writing quality prediction and relationship with learners' language proficiency. Lexical richness certainly can go beyond. For example, as we aware, written language are shall to be read and spread in most

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0007

cases. However, studies on the relationship between one's lexical richness and text readability has rarely been studied.

## Acknowledgments

This work was supported by grants from Postgraduate Scientific Research Innovation Project of Human Province "Corpus-based Study on Lexical Richness and Text Readability of Chinese and International Medical Journal Articles" (No. CX20210928).

## References

[1] Birgit H. & Mary L. Verb Lexis in the Written Compositions of Young L2 Learners[J]. Studies in Second Language Acquisition, Vol.11(1989) No. 4.

[2] Capsi T. & Lowie W. A Dynamic Perspective on L2 Lexical Development in Academic English[A]. In R. Chacon-Beltran, C. Abello-Contesse & Torrebalance-Lopez (Eds.), Insights into Non-native Vocabulary Teaching and Learning [M]. Bristol: Multilingual Matters: 2010.

[3] Daller H, Van R. & Chipere N. Development Rends in Lexical Diversity [J]. Applied Linguistics, Vol.2(2004), p.220-242.

[4] Durán, P, David M., Brian R. & Ngoni C. Developmental Trends in Lexical Diversity. Applied Linguistics [J]. Vol.2(2004), p.25.

[5] Engber A. The Relationship of Lexical Proficiency to the Quality of ESL Compositions [J]. Journal of Second Language Writing, Vol.2(1995), p.139-155.

[6] Foster P. & Parvaneh T. Native Speakers and Task Performance: Comparing Effects on Complexity, Fluency, and Lexical Diversity[J]. Language Learning, Vol.4(2009), p.59.

[7] Granger S. New Insights into the Learner Lexicon: A Preliminary Report from the International Corpus of Learner English[A] .In Flowerdew L & Tong K(eds.) .Proceedings of the Joint Seminar on Computers and Lexicology[C] .Hong Kong: Hong Kong University of Science and Technology, 1994.

[8] Haliday M. An Introduction to Functional Grammar. London[M]: Edward Arnold, 1985.

[9] Higginbotham G, & Reid, J. The Lexical Sophistication of Second Language Learners' Academic Essays. [J]. Journal of English for Academic Purposes, Vol.37(2019), p.127-140.

[10] Hyltenstam K. Lexical Characteristics of Near-native Second Language Learners of Swedish[J]. Journal of Multilingual &Multicultural Development, Vol.9(1988), p.1-2.

[11] James C. Errors in Language Learning and Use: Exploring Error Analysis[M]. Beijing: Foreign Language Teaching and Research Press, 2001.

[12] Judit K. Task Complexity and Linguistic and Discourse Features of Narrative Writing Performance[J]. Journal of Second Language Writing, Vol.2(2011), p.20.

[13] Laufer B. & Nation P. Vocabulary Size: Lexical Richness in L2 Written Production[J]. Applied Linguistics, Vol.3(1995), p.307-322.

[14] Laufer B. The Development of L2 Lexis in the Expression of the Advanced Leaner [J]. The Modern Language Journal, Vol.4(1991), p.440-448.

[15] Lei, S. Yang, R. Lexical Richness in Research Articles: Corpus-based Comparative Study Among Advanced Chinese Learners of English, English Native Beginner Students and Experts[J]. Journal of English for Academic Purposes, Vol.47(2020), p.1-9.

[16] Linnarud M. Lexis in Composition: A Performance Analysis of Swedish Learners' Written English[M]. Sweden: CWK Greenup, 1986.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0007

[17] Lu X. The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives [J]. The Modern Language Journal, Vol. 2(2012), p.190-208.

[18] Read J. Assessing vocabulary in language teaching [M]. Cambridge: Cambridge University Press, 2000.

[19] Siegler, R. Micro-genetic Analyses of Learning[A]. In Damon W. & Lerner R.& Kuhn D.& R. Siegler S., Handbook of Child Psychology, Vol.2: Cognition, Perception, and Language: Wiley & Sons. 2006.

[20] Ure J. Lexical Density and Register Differentiation. In Perren G. & Trim J. (eds.) Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics p. 443-452. Cambridge: Cambridge University Press, 1971.

[21] Vermeer A. Coming to Grips with Lexical Richness in Spontaneous Speech Data[J]. Language Testing, Vol.17(2000), No.1.

[22] Bao G. A Multidimensional Study on the Development of Lexical Richness in Second Language Learners [J]. Audio-visual Teaching of Foreign Languages, Vol.5(2008), p. 38-44.

[23] Jian C. A Study on The Lexical Richness of Two Genres of College English Majors based on CEW Corpus [J]. Journal of Tianjin Foreign Studies University, Vol.18, (2011)No.4, p.55-61.

[24] Dai T. A Study on the Effects of Writing Task types and Second Language Proficiency on English Writing Vocabulary Richness [D]. Shandong University, 2019.

[25] Gui S. & Yang H. A Corpus of Chinese Learners' English [M]. Shanghai: Shanghai Foreign Language Education Press, 2003.

[26] Qin X. & Wen Q. Research on the Development Rules and Characteristics of Chinese College Students' Written Language Writing Ability [M]. Beijing: China Social Sciences Press, 2007.

[27] Wan Lifang. A Study on Vocabulary Richness in Second Language Writing of Chinese English Majors [J]. Foreign Language Community, Vol.1(2010), p.40-46.

[28] Wang Y. The Correlation Between the Vocabulary Richness and Writing Performance of Chinese Second Language Learners -- On multiple Linear Regression Model and Equation for Measuring Writing Quality [J]. Language and Character Applications, Vol. 2(2017), p.93-101.

[29] Xiao L. The Effect of Task Type on Writing Vocabulary Richness of Intermediate and Advanced Chinese Second Language Learners [J]. Language Teaching and Research, Vol. 6(2018), p.36-47.

[30] Hui Z. & Jun W. The Development of Lexical Richness in English Writing: A Longitudinal Study based on Self-built Corpus [J]. Foreign Languages, Vol. 6(2013), p. 77-86.