摘要

实证研究表明，针对外语学习者在词汇教学的本质和方法上的探究还不够深入和全面。大量现有的研究已普遍认识到词汇习得的重要性，但对如何设计合适的词汇练习任务，如何输入词汇仍缺乏细致深入的探讨。

基于上述出发点，本文对词汇习得的特点及其在教材中的分布情况进行了分析，并且有目的的对词汇进行了仔细的规划研究，从教学中互动的双方（教师和学生）层面上寻找可行的指导性方法，以利于词汇的掌握。

对词汇的学习掌握有赖于词汇的重复出现。在反复出现的过程中词汇才能永久性地被储存在记忆中。相关研究表明，在可理解的语境中不断复现词汇，会使词汇的学习更为有效。通过研究，作者力求回答以下问题：教师怎样向学生呈现词汇才能实现有效的输入？教师怎样才能确保学生在输出中正确地使用词汇？

本文的研究主要采用了基于语料库的研究方法。全文分为六章：

第一章引言。对词汇研究的背景、现状及目前在词汇教学中所采用的一般方法进行了综述，阐明了改进现有教学方法的必要性。

第二章相关理论研究。运用词汇描述与习得的相关理论，对心理词汇的形成、输入与存储的关系、词频对词汇掌握的影响、以及词汇习得的特性和过程进行了详尽的分析。

第三章语料库语言学及语料库在词汇教学中运用的基本原理。主要说明语料库的检索功能、词频统计、关键词的生成、词的搭配与排列等在词汇教学实施过程中的重要作用。

第四章大学英语课程语料库。介绍由作者与其导师共同开发设计的以大学英语课程教学大纲要求为标准的课程语料库。该课程语料库由词汇大纲数据库、教材数据库（二十一世纪英语读写教程）和软件包三部分构成。同时，为进一步实证语料库语言学在词汇教学实施过程中的优势，利用大学英语课程语料库所生成的信息资料及数据，对该套教材的词汇分布及词汇的复现率进行了分析和有针对

性的评价。

第五章大学英语词汇规划。分析上述信息资料及数据，从而为词汇的合理规划提供依据。针对课堂教学时间的局限性，提出了词汇合理规划的方法。从词汇分布、复现率对记忆词汇有效性的角度进行分析，生成对学生学习词汇有帮助的各类词表和补充词表（为教材中没有涵盖到的大纲词汇）；将词表和短语表有计划、有指导性地提供给学生去记忆和运用，形成对词汇不同形式的多次循环，以提高词汇的复现率，弥补教材编写时考虑的不足。值得注意的是，该方法能使教师在有限的课堂教学时间里，通过有步骤、有目的地演示教学，鼓励学生积极利用计算机和语料库，实现词汇的课堂外自主学习。

第六章总结。概括该研究的主要成果和局限性，为进一步深入的研究提出有益的建议。

本文对词汇教学实践中的研究有一定的现实指导意义。目前，在我国英语教学活动中，针对词汇教学的研究虽已比较活跃，但以语料库为基础对词汇规划的系统研究尚不多见。因而，充分利用语料库所生成的各类词表，提高对语料进行自动检索的功能研究，能为师生提供相应词汇和短语的常见搭配。如果能借助计算机和语料库的帮助，同时结合教师自己的课堂示范讲解和辅导，就能使学生在教师的指导下，根据自身的特点、水平和学习时间，选择合适的学习内容，实现教学的个性化和学习的自主性。

以语料库为基础的词汇规划研究能有效地提高词汇学习者的掌握速度和深度，为教与学的双方提供清晰的科学指导途径，从而达到教与学的最佳效果。

关键词：词汇教学；基于语料库的方法；大学英语；词汇规划

Abstract

Case studies show that much less is known about the nature and the methods of teaching vocabulary to foreign language learners. Even though vocabulary acquisition is considered of vital importance, teachers are often confused on how to design proper vocabulary learning tasks and how to input vocabulary. These two aspects have a great impact on the result of the learners' vocabulary output, as well as on their interests in learning.

Based on the background above, this thesis will be an attempt to find some appropriate methods for teaching vocabulary through analyzing the features of vocabulary learning and the distribution of the vocabulary in the textbooks. The author aims to make careful plans in the vocabulary teaching and learning process. Effectiveness of learning words can be enhanced by repeated encounters until the words become permanently stored in the memory. Such encounters are more effective in comprehensible contexts. Through the research, the author will attempt to answer the following questions: In what way(s) should the teacher present words to the students to make the input effective? How can the teacher make sure that the students can use them correctly in the output?

To address this topic, the corpus-based method is employed for the research.

Chapter One presents an introduction of the research background, the current situation and methods in common practice for vocabulary teaching. The necessity for improvement of vocabulary teaching is also explained.

Chapter Two illustrates vocabulary acquisition theories: the vocabulary description, vocabulary acquisition based on the analyses of the formation of internal lexicon, the relation between input and storage, the influence of word frequency on lexicon access, and the features and process of learning vocabulary.

Chapter Three introduces Corpus Linguistics and its application in vocabulary teaching. With the aid of automatic concordance of corpus, words frequency, key words

in context (KWIC) and phraseology and collocation are analyzed and described. It makes the clearness of the advantage of the corpus-based method during the practice.

Chapter Four illustrates College English Course Corpus, complied both by Mr. Zhang Yang, the author's supervisor, and the author, with the instruction of the College English Curriculum Requirements The Course Corpus consists of three parts: syllabus database, textbook database (Reading and Writing Course in 21st Century College English), and software package. With the help of the materials and data generated from the corpus, the author analyzes and evaluates the distribution and reoccurrence of vocabulary in the textbooks for further concrete evidence of the advantage of the corpus-based teaching method during the practice.

Chapter Five proposes appropriate vocabulary planning with the aid of information and statistics generated from the corpus. Throughout various angles of analyses of distribution and reoccurrence, all kinds of wordlists and supplementary wordlists (which are the words uncovered in the textbooks, but required in the syllabus), will be helpful for students. The circulation in different ways of the vocabulary reoccurrence will be planned carefully for students' effective memory of the words by giving different wordlists. It can compensate the weakness of compiling the textbooks by increasing the reoccurrence of vocabulary. It is noted that students' autonomous input and exercises can be encouraged out class through in-class input with examples step by step in a planned way, and clear guidance can be provided for both teachers and students as well.

Chapter Six summarizes the implications and limitations of the research and gives some suggestions for further study.

This research has some practical significance in vocabulary teaching. Although the study on vocabulary teaching and learning is more active than ever before, the systematical vocabulary planning has been less discussed. It is beneficial for both teachers and students to make full use of the special function of computer and corpora in generating all kinds of wordlists, automatic concordance, and providing the collocations of words and phrases. While taking advantage of the teachers' lectures and assistance, students can be assisted by computer and corpus in choosing the appropriate content of learning according to their specific needs, proficiency and schedule under the

guidance of teachers. Then the best effects of vocabulary teaching and learning can be achieved.

Key words: vocabulary teaching   corpus-based approach   College English vocabulary planning

独 创 性 声 明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

签名：＿＿＿＿＿＿＿＿＿ 日期：2007年 6月 14日

关于论文使用授权的说明

本学位论文作者完全了解电子科技大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权电子科技大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后应遵守此规定）

签名：＿＿＿＿＿＿ 导师签名：＿＿＿＿＿＿

日期：2007年 6月 14日

Acknowledgements

The accomplishment of this thesis results from the help of many: Hereby I would like to express recognition and gratitude to those who made this thesis possible.

First of all, I am particularly indebted to my supervisor—Professor Zhang Yang—whose valuable and constructive suggestion, comments, advice and encouragements were given to me throughout the duration of writing this thesis. I am deeply grateful to him for his tolerant efforts and patience in providing guidance and inspiration to me. It is a great honor for me to study under his guidance, for not only have I benefited from his profound and erudite knowledge, but also from his reserved and modest personality, and his attitudes towards life and academic study.

My sincere appreciation goes further to Prof. Zhang Wenpeng, Prof. Feng Dou, Prof. Di Aiying, Prof. Wu Zhongjie, Prof. Zhou Guangya, Prof. Zeng Lu, Prof. Luo Ruilin, Prof. Zhang Jinfan, Prof. Feng Wenkun, and Prof. Chu Jun, for their scholarly advice, meticulous study style, and generous help in the first stages of this thesis.

I have also been fortunate to have support from the graduate's offices of this university. I would like to thank the people there for their special assistance to me, especially for the kindness from Ms. Wei Zhongzhi, and Tang Junling during my stay here.

Last but not the least, I would like to extend sincere thanks to my beloved family—my husband, my daughter and my guest sister—Amanda Meyer, the volunteer of Peace Corp, for their unconditional love and support; to a number of friends who encouraged me during the past three years, including Wang Yi, Tang Yimei, Peng Lin, Luo Feng, and Yang Mingna.

It is just because of the contributions of these people, whether mentioned or not, that this thesis is able to be completed. To each and every one of them, I give my most sincere thanks.

Chapter One Introduction

Along with the increase of information and the frequency of international cultural and economical exchange, China has been giving an ever-increasing emphasis on English language teaching in its college education in the past two decades. As we enter the 21st century, the urgent acquisition of English is becoming more important in today society. Meanwhile, vocabulary acquisition has assumed a more important role, and as some would argue, the central role in learning a second language.(Lewis, 1993). But after learning English for almost ten years (7 years of primary school and 3 years of high school), most college students feel they still cannot grasp the true meaning of a word in a specific context and use properly while writing and speaking.

In China, we, both teachers and students of foreign language teaching and learning, are so accustomed to this model that we, more often than not, say or hear the statement that structure is the skeleton of language, and vocabulary is the organ and flesh (Harmer, 1991). Learning a language is synonymous with mastering the structures of the language and once you have mastered the system, you could generate any sentence you wish. After the dominance of grammar in English teaching and learning for years, linguists and educators have increasingly realized the importance of vocabulary teaching and learning. During the 1990s interest in vocabulary teaching and research increased. Every college student and experienced teacher of English knows just how important vocabulary is. After all, one word often is essential in understanding the meaning of an entire sentence, let alone the whole paragraph or passage. Students know they should master thousands of words that speakers and writers of English use. To bridge this gap, they spend much more time trying to remember words, but not in knowing the background or literary difference between Chinese and English. The result is that they still cannot understand the real meaning of the word in context. Is it really necessary to remember all the words that appear in the textbook? How much vocabulary does a student need? Is there any good solution for helping students memorize a new word when only using their textbook, besides some learned strategies?

For many years, word frequency and key words in context that were distributed in a course book gave little attention to proper planning for helping students learn vocabulary. With limited class hours and many tasks, vocabulary teaching tends to be taken on by the personal action of teachers. Under these circumstances, many students find that they make great efforts to memorize new words in a dictionary without any guidance. As a result, many of them still cannot call to memory and use words that they have "learned".

From her teaching experience, the author has found that the effectiveness of learning words would be enhanced by repeated encounters and that such encounters are more effective in comprehensible contexts. This study is a corpus-based research, with the aid of the College English Course Corpus compiled by the author's tutor Professor Zhang Yang, which gives information on the frequencies, reoccurrences and distribution of words and phrases in 21st Century College English. From the corpus we can find out whether words and phrases are repeated enough throughout the texts. Then, efforts will be made to focus on the distribution of words and their frequencies in the textbooks, to find a well-planned manner for presenting vocabulary items and some appropriate methods for teaching them through careful planning in the teaching and learning process. The author hopes that the results from this study will shed some light on English vocabulary teaching and learning for Chinese colleges and universities with the aid of corpus.

1.1 Background of the research

College English is a compulsory course in almost all colleges and universities in China. It is the single biggest university course in terms of duration—lasting for two years, totaling more than 200 classroom hours. To strengthen English language teaching, the Chinese government authorities have made several massive attempts in providing guidelines to college English teachers across the country. In 1985, *College English Syllabus (For Arts and Science)* was published. To the following year, *College English Syllabus (For Science and Engineering)* was released. In 1988, a common vocabulary list was issued for both syllabi. Then in 1999, these two syllabi were revised and combined into *College English Syllabus (Revised Edition)*. Finally, in 2004, *College*

English Curriculum Requirements (For Trial Implementation) replaced the combined syllabus. All these revisions were made in response to the nation's call for more progress in English language teaching to meet its needs for large numbers of scientists, engineers and managers who can communicate with their counterparts abroad.

According to the new *College English Curriculum Requirements*, the requirements of vocabulary as follows: 1) Words in Common Use (6,674 entries, including 2,354 active words marked "*", 1,055 words for higher requirements marked "▲" 1081 words for intermediate requirements marked "★", and 4,538 unmarked words for basic requirements); 2) phrases in Common Use (1,875 entries); compared to the requirements of vocabulary in the former syllabus, the number of words has increased by 1,000 entries to meet the needs of higher requirements (College English Curriculum Requirements,2004). The increase in word knowledge shows that the command of vocabulary has been paid much more attention to with the passing years, both in enlarged amounts and in the entire mastery of the words.

With the new requirements of vocabulary and a relatively longer period of English teaching, how to provide proper planning of vocabulary and how to input vocabulary would have a great impact on the result of the learners' vocabulary output as well as on their interests according to the objectives on vocabulary stipulated in *College English Curriculum Requirements*, which is a national course syllabus. In the worst cases, little progress has been made when the whole process of College English education is finished. Unfortunately, a large percentage of college student's proficiency in English, especially vocabulary, has been declining due to decreasing interest and the increase in new words. Students always complain that they have little spare time to review and reuse the words they learned in former units, for there are many other subjects that must also be learned.

1.2 Vocabulary teaching and learning in college English—the hardest nut

In College English teaching and learning, vocabulary acquisition is considered of vital importance. However, compared to grammar, phonology, and discourse studies,

much less is known about the nature and the methods of teaching vocabulary to foreign language learners. The distribution and planning of vocabulary required in the textbook seems less systematic. Pronunciation and grammar are emphasized in textbooks, but there is often little or no emphasis on vocabulary. In short, vocabulary has been neglected in programs for teachers and students during much of the twentieth century (Allen, 2002).

1.2.1 Reasons for neglecting vocabulary teaching and learning

Many linguists hold different neglecting attitudes towards vocabulary teaching and learning, which played a leading role in those years. There are several possible causes for this neglect:

As vocabulary is less amenable to generalization than closed systems like grammar or phonology, psycholinguists have reacted against vocabulary since it has been connected with associative learning rather than a learning process of hypothesis formation and testing, and an emphasis on the beginning stages of learning led to a focus on grammar (Laufer, 1985).

Next, learners often believe that all they need is a large number of words. They think they can master the language by just learning a certain number of English words, along with the meanings of those words in their own language. The term "vocabulary" refers to a list or set of words for a particular language or a list or set of words that individual speakers of a language might use. Since vocabulary is a list, learners would think that the only system involved is that of alphabetical order. (Hatch and Brown, 2001) Of course, this belief is one-sided. In addition to knowing English words and their meanings, one must also learn how words work together in English sentences, or why the emphasis is put upon grammar in teacher-preparation programs during the past decades.

There is a third reason why so little has been mentioned in theory and methodology courses about teaching words and their meanings. Based on Krashen's hypotheses (1981,1982), known as the "morpheme order studies", many investigations set out to determine whether there is a "natural sequence" in the acquisition of second language grammar and vocabulary (Nunan, 2001:43). Some specialists in methodology

seem to believe that the meanings of words cannot be adequately taught, so it is better not to try to teach them.(Allen, 2002). In the early 1970s Wilkins wrote: "Linguists have had remarkably little to say about vocabulary and one can find very few studies which could be of any practical interest for language teachers" (1972: 109), and almost a decade later Meara commented that vocabulary acquisition had 'received short shrift' from applied linguistics (1980: 221). Dubin and Olshrain (1986: 111-112) wrote that lay people believed that 'knowing a language' consisted of knowing words, while modern linguistic theories had placed little emphasis on vocabulary, focusing more on structures, functions, notions and communication strategies. Ellis and Sinclair (1989) expressed the view that the situation had not changed significantly.

The last reason is on the students and the teachers' parts—there seems to be a lack of clear goals in vocabulary learning, not only among the students but also among the teachers who barely have any systematic and holistic idea of what it is all about. The result of this blindness is the fact that many students and teachers fail to obtain self-confidence on vocabulary learning and teaching when they face so much vocabulary with in the textbooks.

These are just some of the reasons for the general neglect of vocabulary

1.2.2 Current situation of vocabulary teaching

Before students enter colleges and universities in China, all of them will have studied English for at least six years, having grasped approximately 2,000 words. But in no more than 2 years, they should acquire another 4,500 words (derivations excluded) and 700 phrases. To reach the newly requirement, students should now acquire 6,500 words and 1,700 phrases. Students complain that they are plunged into a giant whirlpool of vocabulary learning as soon as they begin their college English learning. The dramatic change in the quantity of vocabulary acquisition is a big headache to most college students. Vocabulary teaching and learning has, as a matter of fact, become a hard nut to crack.

The current situation has presented some problems for vocabulary teaching. First of all, words and phrases that occur in books may not have been well chosen. They are mostly decided by the topics and articles selected by the compilers. The distribution of

them may not be in good balance. As a result, vocabulary learning becomes a somewhat random process. Secondly, we cannot expect students to learn and remember words and phrases just through occasional encounters but through a repeated process. Thirdly, the word list provided in the syllabus is not presented as a whole picture both to the teachers as well as the students, then falling on the teachers' personal preference and judgment. Furthermore, little relevance exists between the students' majors and the vocabulary contained in most textbooks. This fact makes vocabulary teaching hard to meet the students' practical needs.

Considering the current situation of vocabulary teaching, it is time to take action for improving vocabulary teaching. Now, let's review the common attitudes in practice for vocabulary teaching.

1.3 Common attitudes in practice for vocabulary teaching

After decades of neglect, vocabulary is now recognized as central to any language acquisition process, whether native or non-native. What many language teachers might have intuitively known for a long time, is now being openly stated by some second language acquisition researchers (Laufer, 2002). The following quotations illustrate this:

No matter how well the student learns grammar, no matter how successfully the sounds of L2 are mastered, without words to express a wider range of meanings, communication in an L2 just cannot happen in any meaningful way. (McCarthy, 1990: viii)

Knowing words is the key to understanding and being understood. The bulk of learning a new language consists of learning new words. Grammatical knowledge does not make for great proficiency in a language. (Vermeer, 1992:147)

The lexicon may be the most important component for learners. (Gass and Selinker, 1994:270)

With the growing awareness of the importance of vocabulary teaching and learning, teachers are in favor of giving students a large number of lexical items. Often, vocabulary is presented via word lists, on which word form and meaning are usually the

only foci.(Schmitt, 2002) Teachers and students believe that the more vocabulary that can be learned, the better the contextual meaning can be understood. Certainly, this viewpoint is one-sided. Students should be taught more base words (very high frequency) rather than spend so much time on large quantity of unnecessary items appeared in the texts.

Furthermore, the teachers mostly follow the given arrangement of the textbooks, being led passively through the units. There is little guidance for teachers and students to teach and learn the required words that appeared in the text. Also, there is often no instruction for practice on newly acquired knowledge. Therefore how the students can learn new words partly depends on the compilation of the textbook, and partly depends on the teaching methods being employed.

1.4 Necessity for improvement

Seal (1991) divides vocabulary teaching into planned and unplanned activities. Unplanned vocabulary teaching happens when the student requests a meaning for a vocabulary item during a lesson, when the teacher realizes that a word that has just come up needs to be clarified, or when the teacher is not very clear with the requirement of new words. He noticed the importance of well-planned vocabulary in teaching and learning. With the huge amount of vocabulary appearing in the texts, well-planned vocabulary, both for students and the teacher, is hard work. Usually, the teacher must improvise on the spot. Certainly, we are not all equally good at improvising in a way that would promote learning. So, well-planned vocabulary teaching should be embodied.

On the other hand, a word with higher frequency helps to be comprehended, according to the study of word frequency by Foss (1969) in a phoneme monitoring study. A number of studies have shown that frequency influences response time in the lexical decision task, with higher-frequency words having shorter durations (Rubenstein Garfield, & Milliken, 1970; Whaley, 1978). Also, from real teaching experience, the author has found that the effectiveness of learning words would be enhanced by both repeated encounters and reoccurrence in various contexts. That is to say, repeating is

helpful for vocabulary learning.

The textbook, as a medium, conveys the essential information about English words for teaching and learning. How to compile the vocabulary planning of the textbook seems very vital for the learners and the teachers with the need of learning and teaching vocabulary effectively. On the other hand, the teachers need to be trained to understand the techniques in teaching vocabulary.

Therefore, with the advantages of the corpus-based approach with the use of computers, the goal, well-planned vocabulary concerning with vocabulary size, coverage, distribution, and reoccurrence in textbooks for learning and teaching, can be achieved successfully. New methods of vocabulary teaching will benefit both the students and teachers immensely.

Chapter Two   Vocabulary Acquisition Theories

In the previous chapter, the author made a brief introduction to the significance of the background of the study, current situation of vocabulary teaching and learning in College English, common attitudes, and the necessity for improvement. In this chapter, the related studies include the vocabulary description, vocabulary acquisition, which is based the analyses of the formation of internal lexicon, the relation between input and storage, the influence of word frequency on lexicon access, and the features and process of learning vocabulary.

2.1 Vocabulary description

When we know a word, we know its phonological, morphological, syntactic, and semantic attributes. Since the main vocabulary knowledge is learned through the conveying of the context of students' textbooks, the textbooks become the vital carriers of vocabulary knowledge for college students. In the present context of college English teaching in China, the authenticity issue seems to be out of question. Almost all College English textbooks use original texts by native English speakers. Although college students have learned a large number of English words in high school, they still feel puzzled to determine the specific meaning of the word they encounter during their reading. Just like Labov (1973) wrote, "Words have a habit of changing their meaning from one context to another." So knowing a word is not as simple as it might seem.

The importance of context in vocabulary learning is evident from two common-sense observations: what a word means on any given occasion is mediated by the many contexts in which it is used, and such contexts provide considerable input from which language users clearly pick up huge amounts of vocabulary knowledge, apart from any explicit vocabulary instruction they may receive (Nagy, 2002).

2.1.1 Meaning of vocabulary in context

A sample of main entries from *Webster's Third New International Dictionary,*

Unabridged (1964) shows that about 40 percent of words either has more than one meaning listed or belongs to a set of homographic main entries. There were an average of 2.3 meanings per entry, if all the subdivisions of meaning marked by numbers and letters were taken into account. Thus, the estimated 267,000 main entries in *Webster's Third* (Goulden, Nation and Read, op. cit.) represent a total of about 600,000 meanings. Webster's Third thus gives some indication of the sheer volume of contextual variation in meaning within the English language. However, compared to the distribution of words in running text, dictionaries contain a disproportionate number of low frequency words and technical terms which, unlike most common words, tend to have only a single meaning; and, dictionaries underestimate the contextual variability of meaning in language, in that they fail to record the full range of senses that are actually used. Green (1989) found that approximately 15 per cent of words in naturally occurring text were used in senses not included in existing dictionaries. He then concludes that 'it is possible to show that virtually every noun is polysemous (indeed virtually every verb, adjective, and preposition as well) and possible infinitely so'. The multiplicity of meanings found in *Webster's Third* might tell us more about traditions of lexicography, and the compulsive habits of lexicographers, than about what is actually stored in the heads of native speakers of the language.

Considering that, how can students access the real meaning from given context with the multiplicity of meanings of a word?

2.1.2 Collocation

It is contended that in order to explain the way in which meaning arises from language text, two different principles of interpretation have to be advanced. No single principle has been advanced which accounts for the evidence in a satisfactory way. The two principles are:

The open-choice principle is a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word, phrase, or clause), a large range of choice opens up and the only restraint is grammaticalness. This is probably the normal way of seeing and describing language. It is often called "slot-and-filler" model, envisaging texts as a series of slots which have to

be filled from a lexicon which satisfies local restraints. At each slot, virtually any word can occur. Since language is believed to operate simultaneously on several levels, there is a very complex pattern of choices in progress at any moment.

It is clear that words do not occur at random in a text, and that the open-choice principle does not provide for substantial enough restraints on consecutive choices. To some extent, the nature of the world around us is reflected in the organization of language and contributes to the unrandomness. Things which occur physically together have a stronger chance of being mentioned together. There is still far too much opportunity for choice in the model, and the principle of idiom is put forward to account for the restraints that are not captured by the open-choice model.

The idiom principle, proposed by Sinclair (1991), is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments. To some extent, this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation. The principle of idiom can be seen in the apparently simultaneous choice of two words, for example, of course. This phrase operates effectively as a single word, and the word space, which is structurally bogus, may disappear in time, as we seen in *maybe, anyway,* and *another*.

Some features of the idiom principle can be easily found: many phrases have an indeterminate extent; many phrases allow internal lexical variation; many phrases allow internal lexical syntactic variation; many phrases allow some variation in word order; many uses of words and phrases attract other words in strong collocation; many uses of words and phrases show a tendency to co-occur with certain grammatical choices; many uses of words and phrases show a tendency to occur in a certain semantic environment (Sinclair 1991). The overwhelming nature of this evidence leads us to elevate the principle of idiom from being a rather minor feature, compared with grammar, to being at least as important as grammar in the explanation of how meaning arises in text.

Collocation illustrates the idiom principle. On some occasions, words appear to be chosen in pairs or groups and these are not necessarily adjacent.

2.2 Vocabulary acquisition

Vocabulary is no longer a victim of discrimination in second language learning research, nor in language teaching. After decades of neglect, lexis is now recognized as central to any language acquisition process, native or non-native (Laufer, 2002).

Many factors appear to play a role in vocabulary development, though the exact nature of the role is not always clearly understood and the findings of some research studies seem to contradict the received wisdom of the ELT profession and what has been standard practice in the classroom. Some factors are to do with *input*, in other words the way in which vocabulary presents itself to learners, for example through teacher presentation, reading words in texts, learning words during peer exchange, or through self-access work of some kind. Other factors are to do with storing, organizing, and building vocabulary in the mental lexicon and being able to retrieve or recall it when needed.

2.2.1 Features of input

Frequency has been accorded a high level of significance in ELF for many years as a result of the use of word-frequency counting as a procedure informing syllabus and materials design. The rationale for this is quite simply that the most frequently occurring words in the English language will be those most useful to learners. The vocabulary list which form the basis of syllabus specifications for much material currently in use relate back to work done as early as the 1930s and 1940s, later revised and adjusted. Another way in which the concept of frequency has influenced ELT is in the repetition of words in texts (Hedge, 2002). A study by Kachru (1962), showed that most learners knew the words that appeared more than seven times, but they did not know half of the words that appeared only once or twice. That means the words that recur frequently in the text are automatically 'chunked' because our pattern recognition systems become preferentially tuned to perceive them in the future. The term "chunking" was coined by George Miller in his classical review of short-term memory (Miller, 1956). It is the development of permanent sets of associative connections in long-term memory and is the process that underlies the attainment of automaticity and fluency in language. Melton (1963) showed

that, when learning letter or digit sequences, the more stimuli are repeated in short-term memory (SLM), the greater the long-term memory (LTM) for these items, and in turn, the easier they are to repeat as sequences in STM. This process occurs for all kinds of materials (Schmitt and McCarthy, 2002). So repetition of words in materials can aid the process of lexical inference and has been used as a principle for many years. Certainly, if learners are to be exposed to a wide range of word meanings and associations then it will be important for them to encounter words in a variety of different situations through extensive reading and listening. According to Kirsner's (1994) study, lexical recognition processes and lexical production processes are equally governed by a power law relation between access time and number of exposures. Textbooks, therefore, should recycle words in a thorough way and, if they do not do so, then the syllabus should take up the slack. This is clearly a key point for textbook compilers to bear in mind on two levels: learners need to be enabled by being made aware of the learning advantages of such a method, but teaching program itself should also build in repetition at increasing intervals.

2.2.2 Internal lexicon

Psycholinguists refer to the representation of words in permanent memory as our internal lexicon (also named mental lexicon). When a given word in our lexicon has been found, the properties we associate with the word become available for us. These properties include the meaning of the word, its spelling and pronunciation, its relationship to other words, and related information. Much of this is the stuff of which dictionaries are made, but our internal lexicon also contains information that is not strictly linguistics. A part of our knowledge of elephants, for example, is that they are said to never forget things, but this is not part of the meaning of the word in dictionaries (Carroll, 2000). We activate meanings through other words, since all words conjure up related words to varying degrees. To use words effectively in our daily lives, we must utilize their stored knowledge of words, which includes phonological, syntactic, morphological, and semantic aspects. These aspects enable us to pronounce words, create new forms of words, and understand the meanings of words. So for a language learner, the size of internal lexicon will play great role on the vocabulary acquisition.

2.2.3 Input and storage

Cognitive psychologists (for examples Craik and Lockhart 1972; Craik and Tulvig 1975) have suggested that learners are more likely to remember a word if they have worked on its meaning actively; in other words, input becomes intake if there is a depth of processing. There appears to be growing evidence to support the experience of teachers and learners that active mental involvement aids retention. In other words, as Craik and Lockhart (1972) suggested, as a learner moves from surface-level activity based on the sound and visual image of a word to a deeper level of meaning, and at the same time relates the word to existing knowledge, then memory traces become more permanent.

The internal lexicon has been characterized variously as a storehouse, a library, an encyclopedia, and a computer. All of these are legitimate to a degree in describing its many capacities, but it is at present unclear exactly how learners store and organize words in the mental lexicon and what kinds of relationships are built among words as they are stored.

In recent years, contradictory advice to teachers has been emerging from studies into the use of semantic links or networks in classroom materials and activities. The following quotations can illustrate it:

Semantic links play an important role in production. This suggests the use of semantic field based presentation methods...(Channel, 1998)

... in light of research motivated by interference theory and, more recently, the distinctiveness hypothesis, the possibility arises that the practice of presenting L2 students with their new vocabulary grouped together in sets of syntactically and semantically similar new words might actually impede rather than facilitate the learning of the words. (Tinkham, 1993)

The distinctiveness hypothesis (Hunt and Mitchell, 1982) referred to here posits that items of information are more easily learned if they are not similar. Clearly, if further students were to support this hypothesis, the findings would call into question much of standard ELT practice in which textbooks and teachers tend to make use of semantic links.

2.2.4 Cultural factors in the building of meaning

As learners develop their vocabulary knowledge, they acquire not only new words but also new meanings associated with words they have already learned. These are acquired gradually as words are met in different contexts and eventually a word might have extensive and complex meaning associations. Anglin's (1985) scheme for "dog" illustrates Eco's (1979) comment that every word is potentially a text. This implies that, in order to interpret meaning correctly and to choose vocabulary appropriately, learners need to become aware of such nuances. Teachers need to be aware that there will be gaps in learners' understanding of nuance and find ways of helping them to fill the gaps with further meaning associations.

Another culturally affected factor influencing acquisition is the phenomenon of prototypes. A simple definition of prototype is that it is the foremost example of a particular conceptual category, the one that springs most easily to mind when a learner hears a word. Rosch (1975) claimed that Americans from quite widely differing backgrounds perceived a robin as a classic example of a bird, a canary as a poorer example, and a penguin as even poorer. A study by Aitchinson (1992) seemed to suggest that the cabbage was the prototype among vegetables for German speakers. Prototype studies have pointed out that we store certain kinds of knowledge about words. Some knowledge is basic and universal, other knowledge is more abstract and relates to personal or cultural experience: this is known as schematic knowledge which is mental representations associated with the word which are activated in the mind when it is encountered.

Certainly, as Olshtain suggests, the kinds of vocabulary a learner acquires or forgets also depend, in part, on age, the learner's interests, methods of teaching, the amount of reading the learner does, and the amount and type of social interaction experienced.

2.3 Current trends in teaching second/foreign language vocabulary

With the shift in emphasis on vocabulary, the classroom teacher is faced with the challenge of how best to help students store and retrieve words in the target language.

For many of us, our perspective on teaching vocabulary was greatly influenced by the top-down, naturalistic, and communicative approaches of the 1970s and 1980s. Emphasis was implicit, incidental learning of vocabulary. Textbook emphasized inferring word meaning from context as the primary vocabulary skill. But exposure to a word in context is not enough to understand the depth of the word's meaning, providing incidental encounters with words in list is only one method to facilitate vocabulary acquisition. The arguments for not focusing solely on implicit instruction to facilitate second language vocabulary acquisition come from a number of potential problems associated with inferring words from context.

First of all, acquiring vocabulary mainly through guessing words in context is likely to be very slow process. Considering that many students have a limited amount of time to learn a body of words, it is not perhaps the most efficient way to approach the task (Stenberg, 1987; Carter and McCarthy, 1988; Schefer, 1993).

Secondly, inferring the meaning of a word is an error-prone process. Recent studies have shown students seldom guess the correct meanings (Pressley, Levin, and McDaniel, 1987; Kelly, 1990). Students, especially those with low-level proficiency in the target language, are often frustrated with this approach and it is difficult to undo the possible damage done by incorrect guessing.

Third, even when students are trained to use flexible reading strategies to guess words in context, their comprehension may still be low due to insufficient vocabulary knowledge (Haynes and Baker, 1993).

Furthermore, putting too much emphasis on inference skills when teaching vocabulary belies the fact that individual students have different, yet successful, styles of acquiring unfamiliar vocabulary.

In Hulstifn's (1993) study, he found that those good at inferring meaning could acquire vocabulary more easily than those who could not infer well but that the opposite was not true. Students with large vocabularies were not necessarily good at inferring, but they had used other means, such as word lists, to acquire a high level of word knowledge. Finally, and most importantly, guessing from context does not necessarily result in long-term retention.

Current research, therefore, would suggest that it is worthwhile to add explicit vocabulary to the usual inferring activities in the language classroom (Haynes, op. cit; Coady, op. cit,; Stoller and Grabe, 1993; Wesche and Paribakht, op. cit.). Nevertheless, the question on how best to implement this kind of vocabulary instruction in the classroom remains. Throughout literature, these pedagogical themes emerge: build a large sight vocabulary with the aid of computer-assistance, integrate new words with the old, and provide a number of encounters with words both in text and activities.

Chapter Three   Corpus Linguistics and Application in Vocabulary Teaching

3.1 General introduction to Corpus Linguistics

3.1.1 Definitions of Corpus Linguistics

Over the last three decades the compilation and analysis of corpora stored in computerized databases has led to a new scholarly enterprise known as corpus linguistics. It is perhaps safe at present to see corpus linguistics as the computer-assisted study of language or language use by means of naturally occurring text corpus in machine-readable form. This definition of corpus linguistics is a revision of a variety of definitions given by individual linguists in this field. Among others, the following are worth mentioning:

Corpus linguistics is based on bodies of text as the domain of study and as the source of evidence for linguistic description and argumentation. The definition of a corpus as a collection of texts in an electronic database can beg many questions for there are many different kinds of corpora (Kennedy, G. 2000: 3-5).

Corpus linguistics can be described as the study of language on the basis of text corpora. (Aijmer & Altenberg, 1991:1)

Corpus linguistics is the study of large, computer-help bodies of text, or 'corpora'. (Renouf, 1993)

Corpus linguistics is perhaps best described for the moment in simple terms as the study of language based on examples of 'real life' language use. (McEnery & Wilson, 1996:1)

Svartivik (1996:3) also states, "Corpora are becoming mainstream." With the emergence of machine-readable corpora and rapid development of corpus-based research, corpus linguistics is establishing itself as a new approach in language study. It is, as Leech (1992) has rightly pointed out, "not just a newly emerging methodology for

studying language, but a new research enterprise, and in fact a new philosophical approach to the subject."(in Svartvik, 1996:12, emphasis added)

As a matter of fact, any attempt to define corpus linguistics at present is only tentative, for corpus linguistics, as a science, is still developing, and developing at a tremendous pace. Corpus linguistics is not a new theory of language, or a new branch of linguistics, but rather a new or renewed approach to language study. On the one hand, the use of corpus evidence introduces a new approach to linguistic study. On the other, the use of corpus as a source of evidence is not necessarily incompatible with any linguistic theory, and does not exclude the use of other sources of evidence. Corpus-based evidence allows us to develop teaching materials that are more helpful and accurate for learning vocabulary of all levels.

3.1.2 History of Corpus Linguistics

Corpus linguistics is a relatively modern term, and the history of modern corpus linguistics, according to Leech (1991:8), can be traced back to "the era of post-Bloomfieldian structural linguistics in the USA", before the late 1950s. It basically adopts an empirical approach, which uses induction as its way of reasoning. One serious deficiency of early corpus linguistics is due to the fact that the so-called corpora at that time were not machine-readable, and thus no analysis of the corpus data could be carried out by the computer. Faced with a large amount of data, most human researchers would feel discouraged and depressed by the laborious and arduous work involved in finding patterns in the corpora. Only those who were ambitious and courageous enough persisted and carried on their work in analyzing corpora manually. Naturally, what they could derive from corpus analysis was definitely not in proportion to the great amount of time they had invested. In the late 1950s, the adequacy and validity of the corpus approach was seriously questioned, and the 1960s and 1970s saw a period of what many recognized as "discontinuity" or "interregnum". During this period, corpus research was definitely a minority under the shadow of Chomskyan linguistics. However, the corpus-based approach was not abandoned by all and it was in this period that famous corpora such as the Survey of English Usage (SEU) Corpus, the Brown Corpus, the

LOB Corpus, the London-Lund Corpus (LLC) were constructed. It was not until the 1980s that the corpus approach began to prosper. Since then, it has been developing at a tremendous speed, and has come into the mainstream of linguistic research.

3.2 Corpus annotation for lexical analysis

As Leech (1993a) has noted in an analysis of the formal requirements for corpus annotation, it would be an advantage if there were as much standardization as possible or harmonization of terminology, annotation systems and means of analysis. The main focus will be on the processes and procedures available for academic research of the type normally carried out on personal computers.

3.2.1 Lemmatization

If we want a computer to recognize, count or sort the number of words in a text, they should be dealt with in some way that the computer can do. For there are some problems, such as *bank*, occurs twice in a text but in different word classes (noun and verb), hyphenated words such as *by-election* or *by-product* may be counted as two words in each case unless the *by-* occurs at the end of a line, in which case, they may be counted as one word, and transcriptions of spoken language texts by different transcribers can also lead to inconsistencies in graphic word counts such as *fullback, full-back* or *full back*. If words such as these occur with high frequency and are inconsistently transcribed, then the total word count can be significantly affected. Problems in standardizing word forms in transcriptions of spoken texts are discussed by Nelson (1995).

As a way of dealing with these and other potential problems which can affect the counting of linguistic items, analysts commonly treat word forms which are inflectionally or derivationally related to each other as if they are instances of a single word family, known as a lexeme or a lemma. Lemmatization is a process of classifying together all the identical or related forms of a word under a common headword, just as in dictionary-making many of the various morphological inflections or derivations of a word are listed under a single entry (Kennedy, 2000). As corpus is large, manual

analysis can be too formidable a task to contemplate. So automatic lemmatizing is urged. And reliable automatic lemmatizing depends on reliable grammatical tagging and, as we shall see in 3.2.2, most automatic tagging still requires considerable manual input to achieve total accuracy.

3.2.2 Tagging and parsing

The computer could count and sort word forms, but in doing so it buried or distorted some important facts about the language. Unless they could be lemmatized, variant inflected forms of nouns and verbs would be treated as entirely different word types. De Rose (1991) suggested that a large corpus may show even large proportions of categorical ambiguity. Because manual annotation of each word token with its word class in the corpus would be prohibitively expensive, the solution adopted for the *Brown Corpus* was to design a computer program to annotate automatically every word in the corpus with a 'tag' to show the word class to which it belonged in context. Greene and Rubin (1971) describe the computer program known as TAGGIT, which they developed for this purpose----the tagging was designed especially to mark syntactically significant features---- the major word classes and their inflectional variants, major function words, certain important individual lexemes such as *not*, and a small number of discourse markers.

As we have seen, grammatical tagging assigns a tag to each word in a text to label the word class to which it belongs in context. Parsing is a more demanding task involving not only annotation but also linguistic analysis, according to some particular grammatical theory, to identify and label the function of each word or group of words in a phrase or sentence. Corpora which have been analyzed in this way are often called treebank because they are collections of labeled constituent structures or phrase markers. It provides a labeled analysis for each sentence to show how the various words function. Information about the functions of words in context is of course vital for language processing in any attempt to use machines to simulate natural language processing. Word-class tagging and syntactic parsing are closely related, with tagging being used as the basis for parsing and the whole process being viewed as progressive disambiguation, of possible alternative readings using both morphological and

contextual information.

3.2.3 Syllabus word list labeling

The syllabus word list of *College English* is chosen on the basis of data taken from *The Colins Bank of English*, provided by Harper Collins Publishers. It contains 14,000 words in common use from the corpus, as reference for making the word lists. They are *Reference Word List of College English Curriculum Requirements, Reference Phrase List of College English Curriculum Requirements* and *Active Word List of College English Curriculum Requirements*, including: 1) *Words in Common Use* (6,674 entries, including 2,354 active words marked "*", 1,055 words for higher requirements marked "▲", 1081 words for intermediate requirements marked "★", 4,538 words for basic requirements marked "√" or "HS" if they are among those 1872 words taught in high schools); 2) *Phrases in Common Use* (1,875 entries).

3.3 Corpus-based lexical descriptions

From the angle of corpus-based study, lexical descriptions mainly include the word frequency, key words in context, phraseology and collocation. They will be illustrated as follows:

3.3.1 Word frequency

As Nation (1993a) stated, vocabulary knowledge enables language use, language use enable the increase of vocabulary knowledge, and knowledge of the world enables the increase of vocabulary knowledge and language use. With these cautions in mind, the vocabulary size and coverage play important roles for students' understanding. Research by Laufer (1988a) suggests that 95 percent coverage is sufficient to allow reasonable comprehension of a text. A larger vocabulary size is clearly better. Although a language makes use of a large number of words, not all of these words are equally useful. One measure of usefulness is word frequency, that is, how often the word occurs in normal use of the language. The availability of powerful computer and very large corpora make the development of such lists a much easier job than it was when

Thorndike and Lorge (1994) and their colleagues manually counted 18.000,000 running words(Nation & Waring: 2002)The recently published *College English Curriculum Requirements* is regarded as the standard requirement of English learning for Chinese colleges. The entries of its vocabulary lists are taken from *The Colins Bank of English*, Being all vocabulary with high frequency, however the focus of the word frequency in this study is the frequency of key words and active words, also the vocabulary required in the *College English Curriculum Requirements*, that occur in textbooks.

3.3.2 Key words in context (KWIC)

With concordance, the computer can display all the occurrences or tokens of a particular type in a corpus. The type is usually called a keyword. The format has the keyword in the center of the line as illustrated in Figure 1, but with more context on each side of the keyword.

2436	let us speculate a little	on	the maximum size of the python
2449	and there is nothing at all	on	the amethystine python
2468	data	on	the boa constrictor about match
2476	United States, could supply data	on	the maturing period for
2482	the following information	on	snakes varying greatly
2508	amount of agreement	on	some of the giants
2512	there are three levels	on	which to treat the subject
2514	proof and therefore may err	on	the conservative
2521	the third level leans	on	a belief that a lot of smoke
2536	but	on	the third level, and is chiefly
2544	detailed information	on	record lengths of the giants
2548	as far as possible, data	on	these aspects of growth

Figure 1 A Fragment of an unsorted concordance for *on* from the *Brown Corpus*

For many years now the KWIC format has been widely used in data-processing, it saves the researcher looking up each occurrence. The word-form under examination appears in the center of each line, with extra space on either side of it. And the context to the right of the keyword in each line has been the basis for the sorting of the whole line by a simple alphabetical ordering. Consequently, recurring word sequences show up

together. Both right- and left-sorted contexts are very important for the study of significant collocations, because they enable the researcher to identify, at a glance, the major recurring word sequences. Then based on the word frequency data, a keyword list can also be extracted from the same text.

However, it is becoming increasingly clear that language fluency is required. Not only is the mastery of the keywords of high frequency important, but the phraseology and collocation of basic active words is just as necessary.

3.3.3 Phraseology and collocation

Sinclair (1991), as a result of his experience directing the COBUILD project, the largest lexicographic analysis of the English language to date, proposed the principle of idiom mentioned in 2.1.2.

Rather than being a rather minor feature, compared with grammar, Sinclair suggests that for normal texts, the first mode of analysis to be applied is the idiom principle, since most of text is interpretable by this principle.

Lexical phrases, prefabricated routines and patterns (Hakuta, 1974), formulaic speech (Wong-Fillmore, 1976), memorized sentences and lexicalised stems (Pawley and Syder, 1983), lexical chunks (Lewis, 1993), or formulas (R, Ellis, 1994) are as basic to SLA as they are to L1 and so instruction relies as much on teaching useful stock phrases as it does on teaching vocabulary and grammar.

An important index of nativelike competence is that the learner uses idioms fluently. So language learning involves learning sequences of words (frequent collocations, phrases, and idioms) as well as sequences within words. Just as repetition aids the consolidation of vocabulary, so it does the long-term acquisition of phrases (Ellis and Sinclair, in press). So the high frequency of reoccurrence of the vocabulary and phrases should be considered in the textbooks. With the aid of computer, building a course corpus to study learner input, to evaluate the distribution of the vocabulary, to help the teacher remedy the defects of the textbook can be possible.

3.4 Application in vocabulary teaching

Corpus-based lexical description opened a new way of vocabulary teaching and learning for teacher and students. Today, advances in computer technology have given corpus-based vocabulary teaching several advantages over earlier work. First, computers have made possible the collection and storage of very large corpora from a variety of sources. With the ability to store more texts, we can also design corpora that more representative of the language overall, rather than being limited to a select group of authors.

In addition, computers facilitate analyses that are more complete and reliable. No occurrences are lost. Furthermore. Computers can analyze the patterns of word associations on a far more complex scale than is possible by hand.

Meanwhile, corpus-based method allows us to gain a variety of information such as frequency, key words in context, phraseology and collocation of the vocabulary in the textbook set, then real-time exercises of phraseology and collocation of the vocabulary can be designed, which are more helpful and clear for language students of all levels. A corpus-based listing can provide an exhaustive listing of all occurrences and contexts. Such tools describe the ways and frequency that vocabulary is actually used in natural context, identify the uses that students will most likely encounter, rather than relying on authors' impressions of meaning and use (Yang Huizhong, 2002).

As we all know, most frequent words play a very important role in English learning. Learning the most frequent words in English can be very productive. Analyzing one text for young native speakers and another for native speakers on the secondary level, Nation (1990) once found that 87 per cent of the words were on the high frequency list. Coady, et al. (1993) experimented with computer-assisted learning of the 2,000 most-frequent words in English and concluded that using computers to learn the list was an efficient use of time and that emphasizing the list was valuable because it resulted in higher reading proficiency.

Most concordancing programs can generate a frequency list of all the words in a corpus. With the aid of automatic concordance of corpus, various frequency lists can be easily generated from the corpus, which can help the teacher and students have a clear

picture towards the distribution and occurrence of the vocabulary in text, unit, book, even in the whole course book set. They are illustrated as follows:

Frequency list — first occurrence For knowing how often each different work-form occurs in a text, the simplest operation is to turn it into a list of the word-forms in the order of their first occurrence, noting the frequency of each. Each successive word-form is compared with each previous one. If it is a new word-form it is provided with a counter set at 1; if the word-form has occurred before, it is deleted from the text and 1 is added to the counter at the place of first occurrence of the word-form.

This kind of list is helpful as a quick guide to the way words are distributed in a text, and for a number of more specific purposes. If two word-forms have about the same frequency, but one occurs early in the text and the other does not occur until late, then this difference may alert the student to look for an interpretation. Again, if, in a technical text, there is very little technical vocabulary for some time and then a rush of it, that may be a clue to a high-level structural boundary in the text, perhaps the end of a general, layman's introduction to a technical subject. Such observations can help a lot in the selection of texts for vocabulary teaching.

Frequency list — alphabetical The information that is now available in the frequency list can be rendered in several ways. The most common are alphabetical order and frequency order. Both these lists can, of course, be placed in ascending or descending order.

The main use of alphabetical lists is for reference, but they are occasionally useful as objects of study. They are often helpful in formulating hypotheses to be tested, and checking assumptions that have been made. Mostly, alphabetical lists are available when there is a need to check the frequency of a particular word-form.

Frequency list — frequency order The same information can be sorted again so that the list begins with the most frequent item, and continues down to the single occurrences. The listing for a particular text can be compared to that of other texts or to large collections of language. The most frequent items tend to keep a stable distribution, and so any marked change in their order can be significant.

The automatic concordancing is another remarkable function of corpus, which has

been illustrated in 3.3.3.

It has special educational function to introduce corpus into the process of language teaching (He Anping, 2004). First, the observation and analyses with the aid of corpus can develop the learners' intelligence, because information processing requires the intelligence of human brain in three ways: attention, memory, and reasoning (Robinson, 2001). The corpus can provide a lot of real examples in authentic context in form of data and concordance to the students. In recent years, many foreign researchers as Sinclair(1991), Skehan (1993), and Aston (1997), once discussed the positive impact on students' schemata construction based on cognition of vocabulary acquisition. And the function of concordance specialized in corpus can present the same language phenomenon in different context repeatedly. It will be helpful for students' schemata construction of the language.

With the advantages of the function provided by corpus, building related two corpora for effective teaching and learning vocabulary becomes an urge ---corpus for teaching and corpus for researching. The former contains the materials from the students' textbooks and applied reading materials for providing students with the familiar materials; the latter gathers plenty of real natural language materials out of the in-class teaching for providing common word lists and examples for the teaching corpus, the materials for teachers to design learning tasks and for the students to extend their learning scale, which can be any given corpus related the teaching corpus, such as sample corpora, Learners' corpus.

Concerning with the aspects mentioned above, compiling a corpus for teaching vocabulary is becoming necessary. Then the College English Course Corpus, built by Prof. Zhang Yang, the author's supervisor, and the author, will be introduced in the following part.

Chapter Four   College English Course Corpus

4.1 Course corpora

College English Course Corpus is a pedagogical corpus, focusing on the learner input. The major concern is whether the language sampling represents the most prominent features of English as a whole or not, and how teachers can present these features to the students in a systematical way. With the indication of recently published College English Curriculum Requirements, the entries of its vocabulary lists are chosen on the basis of data taken from The Colins Bank of English, three lists of words and phrases were put into College English Course Corpus as important references to evaluate the textbooks. Here the textbooks are all the five volumes for Reading and Writing Course in 21st Century College English coursebook set jointly published by Higher Education Press and Fudan University Press. The major components of the course corpus are shown in Fig. 2.
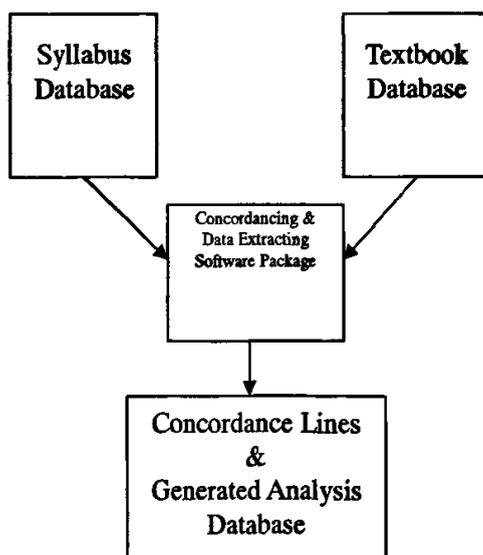
Fig. 2 College English Course Corpus

4.2 Syllabus database

The syllabus database includes the following tables taken from *College English Curriculum Requirements*: 1) *Words in Common Use* (6,674 entries, including 2,354 active words marked "*", 1,055 words for higher requirements marked "▲", 1081 words for intermediate requirements marked "★", 4,538 words for basic requirements marked " √ " or "HS" if they are among those 1872 words taught in high schools); 2) *Phrases in Common Use* (1,875 entries); 3) *Grammar* (including such categories as voices, moods, nonfinite verbs and types of clauses); 4) *Sentence Patterns* (the 5 basic pattern like SVCs, SV, SVO, SVOiOd and SVOdCo, the inversion, the emphatic and so on).

4.3 Textbook database

The textbook database constitutes the greatest part of the corpus. It contains all the 140 texts from the *Preliminary Book* and *Books 1* through *4* of *The 21st Century College English*, which is being used at Chengdu University. The texts are grouped into units of 3 in all books except the *Preliminary* in which each unit contains 2 texts. The texts have been tagged with part-of-speech markers and lemmatized by using the tagging tool of *Tree-tagger*. Further annotation has also been carried out in accordance with the *Grammar* and *Sentence Pattern* tables for data extraction on higher lexical levels.

4.4 Software package

The software package is incorporated into the corpus itself and is designed using *Microsoft FoxPro 7.0* object-oriented relational database management system. The package includes a number of useful tools. The *Concordance* produces concordance lines with the node being a word as a token or a type or a lemma, or being a phrase. The *Word List* generates a word list of a text, a unit, a book, or the whole coursebook series. The *Collocation* extracts phrases or collocations from a text, a unit, a book, or the whole coursebook series. The *ParaConc* aligns the original texts with their Chinese translations. There is also a *Test Engine*, which automatically produces different kinds

of homework assignments or test papers according to certain specifications. Finally, the *Evaluator* gives statistics about the teaching materials and lists all words and phrases required by the syllabus but not included in the coursebooks.

4.5 Generated database

A variety of lists and tables, as well as concordance lines, can be generated from the corpus. They provide useful information on the teaching materials. Information generated from the course corpus is dynamic in that it is organized and can be reorganized in different forms according to the user's special requirements. Traditionally, a printed textbook only contains a general vocabulary list attached to the end of the book and/or separate word list at end of a text or a unit. Usually, a word or phrase appears just once in one word list and will not occur anymore in other lists in the same book. Once the book is printed, everything becomes fixed. It is impossible to reorganize the information. However, vocabulary teaching and learning is a very complicated process with more diversified needs than fixed printed materials can satisfy.

The dynamic feature of the corpus-based approach can be illustrated in chapter five with all kinds word lists electronically generated from the same database.

4.6 Textbook evaluation with regard to vocabulary

With the aid of College English Course Corpus, the textbook evaluation, with regard to vocabulary, can be accomplished as follows:

1) The information about the distribution of words and phrases throughout the teaching materials are easily obtained from the corpus. According to data extracted from the corpus, only 28 out of the 6,648 syllabus words are not include in the whole coursebook set. As far as vocabulary is concerned, the coursebook set is very much in agreement with the syllabus requirements.

2) According to Kachru's study (1962), which has been mentioned in 2.1.1, most learners knew the words that appeared more than seven times, but they did not know half of the words that appeared only once or twice. With the frequency of reoccurrence of the keywords in the coursebook set, seven times or more than seven times of

reoccurrence as the standard, only 76 words of 869 words in Elementary Book(8.7%), 179 words of 1704 words in Book One (10.5%), 236 words of 1939 words in Book Two (12.2%), 223 words of 2294 words in Book Three (9.7%), and 276 words of 1840 words in Book Four (15%) of 21st Century College English (from unit one to unit ten) meet the needs (See Appendix II and III). The percentage above shows the frequency of reoccurrence is far more less than seven times. Furthermore, most of word or phrases appear just once or twice in one word list and will not occur anymore in other lists in the same book.

3) With the information showed in Appendix I, the average word length of each text, each unit, each textbook can be gained. As we all know, in general, the shorter the word length is, the more important it is, and easier it is to be remembered. From the statistics in Appendix I, the range of average word length of each unit is between 4.25 and 5.37 letters. The average word-length in one unit is 4.65 letters.

With the information gained from the corpus, global understanding of vocabulary will be clear. The strategies concerning this method give a certain idea, which combined vocabulary planning in general with in-class examples and demonstration as well as out-class expanding exercises.

Chapter Five   College English Vocabulary Planning

With the textbook evaluation in regards to vocabulary, the improvement for the defects can be designed. Therefore, with the help of the College English Course Corpus, the vocabulary planning can be reached. The tables in Appendix I and II contain proper evaluative statistics on all the 140 texts from the coursebook set, which can help the author analyze the distribution of the words and make a proper vocabulary planning in a large scale way, providing word lists for all kinds of forms for one unit can help students review the vocabulary in certain intervals as many as possible by rigorous planning. Then different word lists can be extracted from the corpus to meet the needs of sufficient high frequency of the words' reoccurrence in other ways.

5.1 Making use of course corpus statistics

5.1.1   Distribution

On a global view of the distribution of words, only 28 out of the 6,648 syllabus words are not included in the whole coursebook set. As far as vocabulary is concerned, the coursebook set is very much in agreement with the syllabus requirements of coverage. However, the distribution of most of active words such as *absolute* and *elect*, required in the syllabus fails to arrange in balance: The total frequency of *absolute* and *elect* in the coursebook set is 4, far from the requirement of high reoccurrence. And the total frequency of most of the active words is below 4 times. Furthermore the frequency of them in book, unit, and text is only once. While key words like *kite*, with high frequency in total is 33, but the frequency in the coursebook is still 33 and 3 times both in unit and text. The statistics shows the occurrence of *kite* appears with high intensity in one book and then it fails to be represented in other context.

So first, a total word list required by the syllabus and a supplementary word list of these 28 words would be provided to the students for a full understanding of the requirements that must be met. The supplementary word list of these 28 words would be

listed clearly.

Second, a list of words learned in high school can be extracted from each text to check the students' lexical competence, which should be the basic vocabulary level for college students when they enter into college. If some of them have difficulty with many of the words in the list, the teacher knows that these students may be very weak in vocabulary and that special measures must be taken to help them.

Then the vocabulary will be divided into many lists according to text, unit by means of word length, low frequency reoccurrence (below 3 times), and active wordlist to remember, while being practiced repeatedly with overlapped exercises in given period of time.

5.1.2 Reoccurrence

With the frequency of reoccurrence of the keywords in the coursebook set, seven times or more than seven times of reoccurrence as the standard, only 76 words of 869 words in Elementary Book (8.7%), 179 words of 1704 words in Book One (10.5%), 236 words of 1939 words in Book Two (12.2%), 223 words of 2294 words in Book Three (9.7%), and 276 words of 1840 words in Book Four (15%) of 21st Century College English (from unit one to unit ten) meet the needs (See Appendix II and III).

The percentage above shows the frequency of reoccurrence is far more less than seven times. Furthermore, most words or phrases appear just once or twice in one word list and will not occur anymore in other lists in the same book. However, language teaching and learning is a very complicated process with more diversified and repeated needs than fixed printed materials can be satisfied. So the practices and exercises concerning a word or phrase should be practiced at least seven times in order to reach the goal of vocabulary acquisition. Corpus-based method makes it possible. The supplementary word lists below will be useful for achieving the goal in well-balanced way.

5.1.3 Supplementary word lists

First, a list of words learned in high school can be extracted from each text to check the students' lexical competence. This is the necessity for the check of vocabulary acquisition for freshmen after they enter into college. The teacher will then know the students' lexical competence and special measures can be taken to help them.

A	cause	have	most	rise	think
about	close	heat	move	sail	this
absolute	come	here	my	say	to
across	compare	hold	necessary	scientist	too
After	continue	home	need	see	turn
again	cool	how	next	shine	two
Air	course	human	night	side	unless
Alive	dawn	I	no	small	up
All	day	ice	not	snow	us
amongst	degree	If	nothing	so	very
An	die	important	object	solid	voyage
And	difference	In	of	something	warm
another	do	into	off	space	water
appropriate	down	It	often	special	we
Ask	earth	Its	on	star	week
At	energy	just	once	start	what
atmosphere	every	keep	one	sun	when
atomic	fact	large	only	supply	which
auxiliary	far	less	onto	suppose	while
away	fine	life	other	surface	who
Back	follow	light	over	tell	wind
Be	foolish	love	perhaps	temperature	would
because	for	main	pour	than	year
before	freeze	make	power	that	you
below	fuel	mark	push	the	zero

34

Blow	get	mile	put	then
Boat	go	moon	rain	there
But	good	more	reason	these
By	ground	morning	remain	thing

Tab. 1   High school words from *Text A, Unit 1, Elementary Book, 21st century College English*

Secondly, a list of active words from different texts can be gained, which students should have the knowledge of them in college after their first two years. The active word list is based on some international sources such as *Nation's First 1000 Vocabulary, Nation's Second 1000 Vocabulary, Nation's Academic Vocabulary, Longman language Activator Key Word List, Longman Defining Vocabulary (College English Curriculum Requirements*: 182).

a	blow	freeze	make	remain	thing
about	but	fuel	more	rise	think
absolute	by	get	most	sail	to
across	cause	go	move	say	too
after	close	good	necessary	see	turn
again	come	ground	need	shine	unless
air	compare	have	next	side	up
alive	continue	hold	object	small	very
all	cool	how	of	so	visible
amongst	course	human	off	solid	voyage
an	degree	ice	often	space	warm
and	die	if	on	special	water
another	difference	important	once	start	week
appropriate	do	in	only	suck	what
ask	down	into	other	supply	when
at	elect	just	over	suppose	which
away	energy	keep	perhaps	surface	while
back	extreme	large	pour	tell	who
be	far	less	power	than	wind
because	fine	life	push	the	would

before	follow	love	put	then
below	for	main	reason	there

Tab. 2   Active words from *Text A, Unit 1, Elementary Book, 21st century College English*

Thirdly, based on the word frequency data, a keyword list can also be extracted from the same text. This keyword list well reflects the theme and content of the text and therefore is very helpful to the teachers and students in teaching and understanding the text.

LEMMA	BAND	ACTIVOC	LEMMA	BAND	ACTIVOC
absolute	HS	*	moon	HS	
air	HS	*	myriad	▲	
appropriate	HS	*	nitrogen	✓	
atomic	HS		object	HS	*
auxiliary	HS		oxygen	✓	
below	HS	*	pour	HS	*
blow	HS	*	powder	✓	
boat	HS		prehistoric		
compare	HS	*	radiation	✓	
constantly			reappear		
cool	HS	*	refute	★	
dawn	HS		reporter	✓	
day-time			sage		
elect	✓	*	sail	HS	*
engine	✓		sailboat		
era	✓		shine	HS	*
extreme	✓	*	snow	HS	
Fahrenheit			solid	HS	*
foolish	HS		suck	✓	*
freeze	HS	*	sun	HS	
furnace	✓		sunset	✓	
galaxy	★		surface	HS	*
gravitational			temperature	HS	

grip				unlikely		
ground	HS	*		utterly		
heat	HS			vacate		
hell	✓			visible	✓	*
historic	✓			voyage	HS	*
interview	✓			worship	✓	
limited	✓			worshipper		
liquid	✓			zero	HS	
maintenance	✓					

Tab. 3 Keyword list from *Text A, Unit 1, Elementary Book, 21st century College English*

Then, with the result of investigating the words required in the syllabus vocabulary, only 28 out of the 6648 syllabus words are not included in the whole coursebook set and the list as follows will help the teacher reorganize the words in other ways to create chance for the students' learning.

worditem	worditem2	band	activoc
appal	appall	▲	
authorize	authorise	★	
ax	axe	v	
centimeter	centimetre	v	
cozy	cosy	★	
dispatch	despatch	★	
email	e-mail	v	
fiber	fibre	v	
generalize	generalise	★	
jeopardize	jeopardise	★	
jewelry	jewellery	▲	
kilo	kilogram(me)	v	
kilometer	kilometre	v	
millimeter	millimetre	v	
mobilize	mobilise	★	
patronize	patronise	▲	

penalize	penalise	▲	
rumor	rumour	▼	
skeptical	skeptical	★	
skilful	skillful	▼	
stabilize	stabilise	★	•
subsidize	subsidise	▲	
sympathize	sympathise	▼	•
tumor	tumour	▲	
utilize	utilise	▼	•
vapor	vapour	▼	
wagon	waggon	▼	
woolen	woollen	▼	

Tab. 4   Syllabus words not included in *21st Century College English*

The information above will help teachers and students have a clear picture of words of each unit which are required in the syllabus, and remedy the lack of low frequency words required in the syllabus by providing reading materials, practices, and exercises generated by the corpus.

5.1.4   College English vocabulary teaching plans

According to the analyses and statistics, College English vocabulary teaching plans should be considered in these three aspects: distribution, reoccurrence, and supplementary word lists.

The vocabulary teaching plans can be made with aid of a series of detailed plans for different duration of learning vocabulary --- for weekly, monthly and academic circling of vocabulary required in the syllabus. Take weekly plan as example (for the first circling):

The brief introduction of basic information of the textbook one of *21st Century College English* should be listed as follows:

The book one consists of 10 units, each unit includes 3 texts. Text A is required to learn intensively for 4 periods of class, text B needs 2 periods of class for learning, and

text C is extensive reading out class.

With the help of the data and statistics generated from College English Course Corpus, the information of distribution, the frequency of reoccurrence of key words and active words, which are required in the syllabus of each text, each unit and the whole book, can be the guidance for planning vocabulary teaching. Before the beginning of Text A, 1) a list of words learned in high school can be extracted from the text to check the students' lexical competence. This is the necessity for the check of vocabulary acquisition for freshmen after they enter into college. The teacher will have clear picture of the students' lexical competence and special measures can be taken to help them. Then as a task for preview, 2) a list of key words from Text A can be provided to students for their first glance towards the topic, content, and the teacher will encourage them to find out the possible meanings of the words by consulting dictionary, guessing meaning in context, or discussing with other students and write down them. The students will confirm the specific meaning of the word during 4 periods class of learning Text A in details. Before Text A is put to an end, 3) a list of active words and a list of phrases from Text A will be given for students to review, especially the ones with low frequency in reoccurrence should be stressed and underlined in exercises and tests.

When Text A will be finished, 4) the first list of keywords of Text B will be handed to students for finding the same words appeared in the former lists and the new ones. And during 2 periods class of learning Text B, 5) the second list of active words of Text B will be presented for the students. Still the students are expected to find out the words learned before and the new ones.

Students are asked to compile a list for recording the words appeared in Text A or Text B while they read Text C. In addition, comparing different students' lists can help them get a sufficient one. Till then, the first circling for week has been accomplished.

When beginning with Unit Two, the lists of keywords, active words, low frequency words, even word-length can be refined for the former unit. All kinds of these lists can be the circulating funds for next circling (monthly plan).

With the guidance of vocabulary teaching plans, the implementation of it by the means of in-class input, autonomous input, and appropriate exercises and test will be

effectiveness for vocabulary teaching and learning.

5.2 Implementation of College English teaching plans

Combined with the teaching plans on the basis of full understanding of vocabulary distribution, reoccurrence and kinds of word lists, the teachers' guidance of representation and examples in class, encouragement and instruction for students' self access, designed tasks for exercises and tests with aid of corpus, can be more effective to reach the goal---knowing words and practicing words both in text and in real context.

5.2.1   In-class input

Classroom hours can hardly ever be enough for teachers to cover everything the students are supposed to learn. Moreover, sufficient systematic guidance should be provided for students' autonomous learning effectively. So the most important thing teachers should do in class is to illustrate the methods, explain the planning of vocabulary, and present vivid examples by using all kinds of wordlists and corpus-based methods for learning collocation in concordance lines. Sinclair (1998) once pointed out foreign language teaching based on vocabulary is to choose the most common entries from large natural corpus, and learn the basic patterns of these words for usage, then learn the most common collocation of these words. This idea of the most common words as "lure" to enlarge vocabulary and master grammar at the same time is embodied in a series of activities in vocabulary teaching based on corpus, designed by Willis (1998:44-46). The following vocabulary teaching activities with different contents for learners are based on her design.

Teachers can make an impromptu teaching material of concordance in-class with students. The steps as follows:

A. Copy concerning samples of the keyword

• ask students to discover the sentences, including the keyword, in the text.

• copy the sentences and put the keyword in the center.

B. Observe and analyze the samples

    • ask students to find the two borders of discourse segment, a complete sentence or grammatical unit.

    • ask students to place the sentences, including the keywords, under different categories according to the meaning, function or grammatical structure.

    • let students think over which sentences, including the keywords, were learned before and so under which categories can be put.

    C. Develop and strengthen the result of learning

    • strengthen the usage of the keyword by using grammatical books or dictionaries, especially the same usage as the one category.

    • take notes with the useful phrases and examples.

    • continue to observe the samples in concordance, try to find meaningful features and go back to the text to find concerning examples and write them down.

    Sample 1

    After reading a material for moderate learners of grade one, the teacher selected the word "as", a multi-functional word, as the learning subject. The first step is to lead students to list all the sentences with "as" as follows:

1	maintenance costs are decreased proportionately	*as*	productivity rises
2	complex activity which includes such actions	*as*	the search for food
3.	of blood constituents have been suggested	*as*	possible signals including
4	…which receives signals from the body	*as*	a result of consumption of food
5	some agent associated with energy storage acts	*as*	a signal for the long term
6	that signals are received directly from the crop	*as*	is explained later
7	a variety of aromatic substances such	*as*	dill, aniseed, coriander and …
8	food intake and energy requirement suggests that,	*as*	with energy, intake should vary
9	to environment temperature in the same way	*as*	monogastric animals, in that
10	this can be considered	*as*	an aspect of energy balance in
11	digestibility here is expressed	*as*	the coefficient for food energy
12	appears to be relatively unimportant in grazing	*as*	animals will graze in the dark

    The second step is to ask students to identify the comparatively complete

grammatical and syntactic unit. It can lead to heated discussions to identify the grammatical and syntactic unit because the easy-to-see phrases like *in the same way as..., as a result of...* are rarely seen in the samples. For instance, if they want to distinguish the specific collocation, students have to trace back the context and syntactic relationship in front of "as" to make sure whether *"as is explained later"* should associate with the front sentence in the sixth sample. And the process will inspire the students to make the concepts in their unconsciousness clear, which can activate consciousness raising of language awareness.

The third step is to ask students to conclude the different meanings and functions of "as", the result as follows:

• indicating time (example 1)

• leading out examples (examples 2 and 7, students has learned "such as ",now there is a new expression "such+ noun+ as")

• showing the similarity (examples 8 and 9)

• following with the verbs like *suggested, acts, considered, expressed* , present the same or compound

• indicating cause

• forming stable collocation with some words like *as a result*, etc.

With the basis of the steps above, teachers should then require students to recall as many examples learned before as possible, which are in correspondence with the usage in the above categories, such as "functioned as", "as you know", "as a child....". Then teachers should encourage students to compare the descriptions in the dictionary and grammar book. They will then gain more examples to prove these usages, or find out other usages which have not occurred in the concordance but mentioned in the dictionary. Such as "as if" or "as...as", the whole process can urge the students to investigate and refer to the usages in larger corpus. Through observation and discussion with others, students will know when and where the words "as" can be used.

The last step is to ask students to observe other meaningful language phenomenon with the help of the present concordance. The students can find out "noun+ noun" and

"adjective+ noun" compound phrases of noun, and the verbs and verbal phrases used in research papers, including, "have been suggested as..., acts as..., and can be considered as..."etc.

The concordance in-class can extend so many language points and activities for observing and learning. It could be very helpful for both teachers and students.

5.2.2 Autonomous input

The autonomous learning puts emphasis on activating the language awareness of students or the consciousness rising of language forms. Two functions of consciousness raising: 1) one is to help learners cross the consciousness gap, helping learners make sense of the concept in the unconsciousness distinguishable through mass language in-put; 2) another is to help learners process language forms.(Berry 1994)). The condition for reaching these two functions is: the in-put language information with high frequency of reoccurrence; and distinct features of language form, just as Schmit (1990) once pointed out that the higher frequency of reoccurrence of some language form appear, the easier learners' attention can be drawn, and it can help the learners more easily enter into interlanguage systems. The strong support of data-driven learning model can illustrate more effective learning. Actually, it is a discovery process of language, and concordance can provide the unprecedented condition for activating learning strategy of concluding.

The learning models of opening and criticism can coordinate closely with autonomous learning by providing examples. With performing the whole process of observation, generalization, and conclusion, self-found grammatical regulation, expression of meaning, and features of pragmatics can be reached. The clear difference between lecturer-centered and passive accepting models in the past lies in the different purposes, which aim to motivate curiosity, the spirit of criticism and reflection of students. Corpus-based methods introduced through in-class teaching can develop a series of activities for cultivating creative thinkers. The goal of interactive, open and analyzing learning can be gained.

Interactive learning can help students experience the whole process of learning

together by compiling the materials for vocabulary learning with the aid of concordance, observe and conclude the law of language phenomenon, by discussing and solving problems.

Open learning can permit students to try different ways to solve problems and to discover many possible answers of one problem.

According to the differences or similarities of context, reflective learning system requires students to dig the regulation of using language by classifying and concluding the samples from corpus in multi-level. This is one of the most effective ways for fostering the spirit of criticism with the aid of corpus.

5.2.3 Exercises and output

With the functions of concordance in corpus, the exercises can be generated for checking the mastery and output of students.

Sample:

For distinguishing *above* and *over.*

1	ow men but also with nature and	*above*	all, with those Higher Power which
2	ys, his dirty merino collar rose	*above*	his shirt, and he smelled the same
3	ad accidents have alcohol levels	*above*	the legal limit for driving.
4	ights as well. If the red lights	*above*	your lane flash, you must not go
5	for I saw her nonce as if from	*above,*	moving through a kind of square
6	ates agency business. We now have	*over*	160 offices under the Cornerstone
7	injured are either under 15 or	*over*	60. The young and elderly may no
8	unlikely proposition. I did it,	*over*	a period of time, by having at leas
9	this process should be repeated	*over*	a two week period. If there are any
10	veys divide teenagers into those	*over*	and those under sixteen, which is
11	ate 1950s—went on being handed	over	every Friday until his death, even
12	on of paying for larger purchases	*over*	several months. The minimum amount
13	child may do the same. Don't be	*over*-suspicious and try not to over react	
14	response. I got letters from all	*over*	the country, from young mothers of
15	as it were, throw their rubbish	*over*	the fence into the neighbor's garde
16	for pedestrians. Do not climb	*over*	the guard rails or walk outside them
17	of the children we played with	*over*	the road was given to the youngest
18	leave home, when they've got	*over*	the shock, most are mainly concern
19	I remember feeling was	*over*	the top at the time) and tie a piec
20	work for dozens of people all	*over*	the world. And every year, hosts an
21	to many parts of Britain and met	*over*	thirty girls, mainly in their home
22	that looks pretty good, so I went	*over*	to the ski jumps and just had look

Steps of learning:

1. Whether "above" in samples 1 to 5 can be replaced by "over"?

2. Mark the number of the samples if "over" indicates the meaning in the following contexts.

    a) at a higher level or covering _____

    b) more than _____

c) from one side to the other _____

d) during (from beginning to end of a period of time)_____

e) finished _____

3. Look at sample 13 and 25, what about the meaning when "over" as a prefix?

4. Write down the meanings of the verbal phrases below through the observation of the contexts.

Hand over (sample 11 and 23) _____

Get over (sample 18) _____

Talk over (sample 24) _____

All over (sample 14 and 20) _____

Over the top (sample 19) _____

(From Tribble 1990:44)

Through making cards of the new words and its collocation by means of corpus, new and deeper understanding of the word will be gained. Sharing the information gained from autonomous learning and discussion in group also encourage students do more and help them grasp the meaning and usage of vocabulary by using effective corpus-based approach.

Chapter Six Conclusion

6.1 Implications to teaching

Equipped with the supportive theories in relation to vocabulary acquisition, especially the positive effectiveness of schemata construction and the processes of dealing with information in human brain in three stages: attention, memory, and reasoning, this thesis is an attempt to explore a new field—the application of corpus-based methods in vocabulary teaching and learning. With the analysis, there is finally a more detailed understanding on the notion of corpus linguistics and its role in vocabulary acquisition. A few implications are yielded as follows, which might help to shed light on and understand this effective and easy-going approach.

The corpus-based approach, as an important breakthrough in pedagogy, has its special function in vocabulary acquisition to achieve special aims:

1) With the help of all types of information generated from the corpus on the textbooks being used, thorough evaluation of the distribution and frequency of reoccurrence of vocabulary can be gained, better teaching plans can be made on a long-term basis, and language input then becomes more systematic, better-balanced and more efficient.

2) With provided information about distribution, reoccurrence and frequency of lexical items, College English teachers can view vocabulary teaching as a systematic process rather than random way by planning vocabulary properly. Both teachers and students can understand syllabus requirements in a more clearly visualized manner and put themselves in a better position to implement these requirements through vocabulary teaching and learning.

3) Supplementary wordlists can be obtained with awareness of what materials should be found for the students, with aid of the list of words and phrases required by the syllabus but not included in the textbook.

4) Many of the corpus-generated materials can also be made accessible to the

students for autonomous language learning. Sufficient systematic guidance provided for students' autonomous learning would be effective, and improved output can be expected.

6.2 Limitations

Although a few functions on corpus-based approach in vocabulary acquisition have been presented, the present study has not reached the point of satisfaction in many aspects owing to the inadequacy of the author's personal understanding. It is impossible for the author to exhaust all parts of effectiveness of corpus linguistics applied in vocabulary teaching. In short, the study here is an attempt in this field. This tentative study is not a synthetic or systematic study of the features of corpus-based approach, but an interpretation of the application of corpus-based approach in vocabulary teaching. It is limited in terms of both breadth and depth, also cannot be free from errors and deficiencies. So it is maximally hoped that putting an end to this thesis is not equal to sealing up the research.

First, similar to a variety of linguistic studies, a major weakness concerning this research is found lacking strong empirical improvement for the analyses or interpretations of the selected data although they are not fabricated by the author herself and have come from the coursebook as sources. Some aspects still need to be tested through further experimentation, therefore some statements or conclusions might have been drawn with subjectivity.

Second, the present study might suffer from the problem of inexhaustiveness and oversimplification to some degree. The analysis in this thesis is not yet complete. Consequentially, there are untouched cases ahead, which might require further investigation.

The systematic study of vocabulary teaching based on corpus still needs further investigation in practical teaching, needs broad consideration of the factors influencing the vocabulary teaching and learning, and, most importantly, requires the support of practical teaching.

To conclude, the corpus-based approach can provide a new platform both for

teaching and learning. Along with the rapid development of scientific technology of telecommunication, corpus linguistics has become the mainstream in linguistics and applied linguistics. Corpus Linguistics has not only created a new way to study linguistics, but opened a new field of language research. It boomed with a variety of data and relative theories, instead of traditional teaching method with chalks and blackboard.

By illustrating College English Course Corpus, helpful materials and data can be generated from the corpus. The analysis and evaluation in relation to the distribution and reoccurrence of vocabulary in the textbooks is necessary in order to properly plan vocabulary.

The contribution of the thesis is to propose appropriate vocabulary planning on the whole, with the aid of information and statistics generated from the corpus. In angles with analyses of distribution and reoccurrence, all kinds of wordlists and supplementary wordlists, which is the words uncovered in the textbooks, but required in the syllabus, will be helpful for students. Recycling of reoccurrence of vocabulary will be planned carefully for students' effective memory of the words by giving different wordlists step by step in a planned way. In-class input with showing examples will be vivid for being imitated for limitation of time in class. Students' autonomous input and exercises can be encouraged and carried out, a clear guidance for vocabulary teaching and learning can be provided for both teachers and students.

This research has some practical significance in vocabulary teaching. Although the study on vocabulary teaching and learning is more active than ever before, the systematical vocabulary planning has been less discussed. It is beneficial for both teachers and students to make full use of the special function of computer and corpora in generating all kinds of wordlists, automatic concordance, and providing the collocations of words and phrases. While taking advantage of the teachers' lectures and assistance, students can be assisted by computer and corpus in choosing the appropriate content of learning according to their specific needs, proficiency and schedule under the guidance of teachers. Then the best effects of vocabulary teaching and learning can be achieved.

Bibliography

Aijmer, K. & Altenberg, B. 1991. *Introduction* in Aijmer, K. & Altenberg, B.(eds)

Allen,V.F. 2002. *Techniques in Teaching Vocabulary.* Shanghai: Shanghai Foreign Language Education Press

Aston, G. 1997. Enriching the learning environment: Corpora in EFL. In A. Wchmann, S. Tligelstone, T, McEnery et al. (eds.). *Teaching and Language Corpora.* New York: Longman

Berry, R. 1994. Using concordance printouts for language awareness training. In D.Li. D. Mahoney &J. Richards (eds.). *Exploring Second Language Teacher Development.* Hong Kong: City Polytechnic of Hong Kong

Coady, J., J. Magoto, P. Hubbard, J. Graney and K. Mkhtari. 1993. High frequency vocabulary and reading proficiency in ESL readers. In T. Huckin, M.Haynes and J.Coady (Eds.) *Second Language Reading and Vocabulary Learning:* 217-28. Norwood, N.J.: Ablex Publishing Corporation.

De Rose, S. J. 1991 *An Analysis of Probabilistic Grammatical Tagging Methods,* in Johansson & Stenstrom (1991)

Dubin, F. & E. Olshtain. 1986. *Course Design.111-12.* Cambridge: Cambridge University Press.

Ellis, G. B.&B. Sinclair. 1989. *Learning to Learn English 106* Cambridge: Cambridge University Press.

Ellis, R. 1994. *The Study of Second language Acquisition.* London: Blackwell.

Ellis,N.C. and S. Sinclair (in press). Working memory in the acquisition of vocabulary and syntax: putting language in good order. *Quarterly Journal of Experimental Psychology. Special Issue on Working Memor*

Foss, D.J. 1969. *Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times.* Journal of Verbal Learning and Verbal Behavior, **8**, 457—462.

Gass, S. M. & L. Selinker. 1994. *Second Language Acquisition: An Introductory Course.* Amsterdam: John Benjamins. 270.

Goulden, R., P. Nation & J. Read. 1990. How large can a receptive vocabulary be? *Applied Linguistics II (4): 341-63.*

Green, G. M. 1989. *Pragmatics and natural language understanding.* Hillsdale, NJ: Erbaum.

Greene, B.B. and Rubin, G. M. 1971. *Automated Grammatical Tagging of English Providence,* RI: Department of Computer Science, Brown University

Hakuta, K. 1974. Prefabricated patterns and the emergence of structure in second

language acquisition. Language Learning, 24: 287-98.

Hatch,E & Brown,C 2001. *Vocabulary, Semantics and Language Education.* Cambridge: Cambridge University Press

Kennedy, G. 2000. *An Introduction to Corpus Linguistics.* Foreign Language Teaching and Research Press.

Krashen, S., 1982. *Principles and Practice in Second Language* Acquisition. Oxford: Pergamon.

Labov, W. 1973. The boundaries of words and their meanings. In C. J. Bailey and R. shuy (Eds.) *New ways of analyzing variation in English : 340-73.* Washington, D.C.: Georgetown University Press.

Laufe, B. 1986. Possible changes in attitude towards vocabulary acquisition research. *International Review of Applied Linguistics. 24 (I): 69-75.*

Laufe, B. 1988a. What percentage of text-lexis is essential for comprehension? In C. Lauren and M. Nordmann (Eds.) *Special Language: From Humans Thinking to Thinking Machines.* Clevedon: Multilingual Matters.

Laufe, B. 2002. What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words. *Vocabulary: Description, Acquisition and Pedagogy.* 140 Shanghai: Shanghai Foreign Language Education Press.

Leech, G. 1992. *Corpora and Theories of Linguistic Performance,* in Svartivik, J. (ed.) Directions in Corpus Linguistics. Berlin: Mouton de Gruyter.

Leech, G. 1993a. *Corpus Annotation Schemes* Literary and Linguistic Computing 8,4: 275-281

Lewis,M. 1993. *The Lexical Approach.* Hove and London, England: Language Teaching Publications

McCarthy, M. J. 1990. *Vocabulary.* Oxford: Oxford University Press

McEnery T. and Wilson,A. 1996. *Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Meara, P., 1980. Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics: Abstracts: 221*

Nagy, W. 2002. On the role of context in first- and second-language vocabulary learning. *Vocabulary: Description, Acquisition and Pedagogy.64-68.* Shanghai: Shanghai Foreign Language Education Press.

Nation, P and R. Waring 2002. Vocabulary Size, Text Coverage and Word Lists. *Vocabulary: Description, Acquisition and Pedagogy..* Shanghai: Shanghai Foreign Language Education Press.

Nation, I. S.P 1990. *Teaching and Learning Vocabulary.* New York: Newbury House.

Nelson,G. 1995. *Standardizing Wordforms in the spoken ICE Corpora* Working paper of the ICE project, Department of English, University College London

Nunan, D., 2001. *Second Language Teaching and Learning.* Beijing: Foreign Language Teaching and Research Press.

Pawley,A. and F. Syder, 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. Richards and R. Schmidt (Eds.) Language and Communication. London and New York: Longman.

Renouf, A. 1993. *Sticking to the Text: A Corpus Linguist's View of Language,* Aslib Proceedings 4(5), 131-136.

Robinson, P. 2001. Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. Applied Linguistics. Vol. 22. No. 1

Rubenstein, H., Garfield, L.,& Milliken, J.A. 1970. *Homographic entries in the internal lexicon.* Journal of Verbal Learning and Verbal Behavior, 9,487—494.

Schmit, R. 1990. The role of consciousness in second language learning. *Applied Linguistics* Vol. 11: 17-46

Schmitt, N. 2002. Vocabulary learning strategies. *Vocabulary: Description, Acquisition and Pedagogy.220.* Shanghai: Shanghai Foreign Language Education Press.

Seal, B. 1991. *Vocabulary learning and teaching.* In M. Celce-Murcia(Ed.), Teaching English as a foreign or second language 2d ed.(pp. 296-312). New York: Newbury House.

Sinclair, J. & Renouf,A. 1998. A Lexical Syllabus for Language Learning. In Carter, R. & M. McCarthy. (eds.). *Vocabulary and Language Teaching.* London: Longman

Sinclair, J. 1991. Corpus, Concordance, Collocation. Oxford: Oxford University Press

Skehan, P. 1993. Second language acquisition strategies and task-based learning. In Scarpis, V.D., Innocenli, L. Marucci, F. & A. Pajalich. (eds.). *Intreccie contaminazioni.* Venice: Supernova

Svartvik, J. 1996. *Corpora Are Becoming Mainstream,* in Thomas, J & Short, M. (eds.).

Thaondike, E.L. and I.Lorge 1994. *The Teacher's Word Book of 30,000 Words.* New York: Teachers College, Columbia University.

Tribble, c.& G. Jones. 1990. *Concordances in the Classroom.* London: Longman

Vermeer, A. 1992. Exploring the second language learner lexicon. In L. Verhoeven and J.H.L.de Jong (Eds) *The Construct of Language Proficiency: 147-62.* Amsterdam: John Benjamins.

Whaley, C.P., 1978. *Word non-word classification time.* Journal of Verbal Learning and Verbal Behavior, 17,143—154.

Wilkins , D.A., 1972. *Notional Syllabuses 346.*

Willis,J. 1998. Concordances in the classroom without a computer. In B. Tomlinson (ed.). *Materials Development in Language Teaching.* United Kingdom: Cambridge University Press

Wong-Fillmore, L. 1976. *The Second time around.* Stanford University: unpublished doctoral dissertation.

何安平. 2004. 《语料库语言学与英语教学》北京：外语教学与研究出版社

杨惠中. 2002. 《语料库语言学导论》上海：上海外语教育出版社

《大学英语课程教学要求》.教育部高等教育司"大学英语教学基本要求"项目组. 清华大学出版社, 2004

Appendix I

Statistics on all texts from *21ˢᵗ Century College English*

textid	wpf	spf	wps	cpw	whpf	dif_lev
0011	509	28	18.20	4.25	17	-10.90
0012	527	43	12.30	4.56	8	-11.10
0021	557	48	11.60	4.49	19	-11.20
0022	547	39	14.00	3.98	17	-11.50
0031	547	45	12.20	4.26	15	-11.40
0032	651	58	11.20	4.04	26	-11.70
0041	837	68	12.30	4.09	20	-11.60
0042	512	27	19.00	4.60	30	-10.40
0051	635	54	11.80	4.63	11	-11.10
0052	556	41	13.60	3.92	21	-11.60
0061	636	52	12.20	4.10	19	-11.60
0062	686	49	14.00	4.06	27	-11.40
0071	623	46	13.50	4.29	24	-11.30
0072	506	27	18.70	4.80	27	-10.30
0081	765	77	9.94	3.96	24	-11.90
0082	678	46	14.70	4.22	13	-11.20
0091	644	44	14.60	4.35	14	-11.10
0092	463	24	19.30	4.43	9	-10.60
0101	665	39	17.10	4.55	19	-10.70
0102	495	42	11.80	4.20	21	-11.50
1011	854	57	15.00	4.62	21	-10.80
1012	620	38	16.30	4.36	18	-10.90
1013	727	72	10.10	3.75	17	-12.10
1021	815	59	13.80	4.40	19	-11.10
1022	675	57	11.80	5.34	10	-10.40
1023	771	78	9.88	4.30	25	-11.60
1031	833	63	13.20	4.43	25	-11.20
1032	746	56	13.30	4.63	19	-10.90
1033	712	63	11.30	4.10	10	-11.70
1041	901	68	13.30	4.08	16	-11.50
1042	676	47	14.40	4.30	17	-11.20
1043	678	57	11.90	4.17	19	-11.50
1051	795	75	10.60	4.16	39	-11.60
1052	738	56	13.20	4.43	22	-11.20
1053	526	65	8.09	4.36	6	-11.70
1061	645	29	22.20	4.88	20	-9.87
1062	749	48	15.60	4.62	16	-10.70
1063	687	52	13.20	4.18	26	-11.40
1071	670	47	14.30	4.29	30	-11.20
1072	675	45	15.00	4.30	15	-11.10
1073	626	41	15.30	4.25	14	-11.10
1081	868	51	17.00	4.85	25	-10.40
1082	688	47	14.60	4.85	19	-10.60
1083	571	36	15.90	4.89	9	-10.40
1091	637	34	18.70	4.97	19	-10.10
1092	663	34	19.50	5.05	4	-9.95
1093	594	35	17.00	5.37	12	-9.88
1101	959	72	13.30	4.79	23	-10.80
1102	875	59	14.80	4.77	14	-10.70
1103	495	45	11.00	4.76	8	-11.00
2011	753	43	17.50	4.64	19	-10.50
2012	804	48	16.80	4.60	24	-10.60
2013	627	28	22.40	5.07	11	-9.68
2021	975	57	17.10	4.97	26	-10.30
2022	593	40	14.80	5.02	12	-10.40
2023	789	36	21.90	4.87	30	-9.89
2031	772	40	19.30	5.00	23	-10.00
2032	722	38	19.00	4.88	18	-10.20

2033	661	34	19.40	4.85	29	-10.20
2041	1107	71	15.60	4.46	48	-10.90
2042	1035	76	13.60	4.23	26	-11.30
2043	1022	69	14.80	4.13	22	-11.30
2051	939	58	16.20	4.51	12	-10.80
2052	609	44	13.80	4.14	17	-11.40
2053	579	50	11.60	4.55	16	-11.20
2061	964	51	18.90	4.72	23	-10.30
2062	852	48	17.80	4.72	26	-10.40
2063	989	92	10.80	4.24	33	-11.60
2071	867	55	15.80	4.43	26	-10.90
2072	680	46	14.80	4.81	16	-10.60
2073	865	71	12.20	4.55	23	-11.10
2081	739	56	13.20	4.30	22	-11.30
2082	559	46	12.20	4.21	12	-11.50
2083	475	28	17.00	4.11	15	-11.10
2091	977	61	16.00	5.09	22	-10.20
2092	536	24	22.30	4.77	18	-9.97
2093	877	53	16.50	4.93	21	-10.30
2101	909	51	17.80	5.15	15	-10.00
2102	841	89	9.45	4.94	15	-11.00
2103	604	34	17.80	4.68	22	-10.50
3011	1030	85	12.10	4.35	20	-11.30
3012	823	51	16.10	4.32	22	-11.00
3013	787	63	12.50	4.90	15	-10.80
3021	750	60	12.50	4.79	19	-10.90
3022	676	47	14.40	4.14	28	-11.30
3023	668	46	14.50	4.31	16	-11.10
3031	900	35	25.70	4.44	29	-9.96
3032	868	48	18.10	4.43	29	-10.70
3033	713	45	15.80	4.13	26	-11.20
3041	1057	66	16.00	4.56	31	-10.80
3042	956	48	19.90	5.02	28	-9.92
3043	725	38	19.10	5.23	12	-9.81
3051	1022	93	11.00	4.78	33	-11.00
3052	985	54	18.20	5.10	49	-10.00
3053	559	54	10.40	5.15	8	-10.70
3061	768	71	10.80	4.38	36	-11.40
3062	910	57	16.00	5.03	20	-10.30
3063	838	48	17.50	4.42	30	-10.80
3071	1044	73	14.30	4.28	28	-11.20
3072	843	47	17.90	4.63	24	-10.50
3073	660	35	18.90	4.53	29	-10.50
3081	668	40	16.70	4.31	37	-10.90
3082	779	56	13.90	4.33	28	-11.20
3083	671	46	14.60	4.55	34	-10.90
3091	995	57	17.50	4.86	26	-10.30
3092	962	50	19.20	4.64	28	-10.40
3093	827	59	14.00	4.73	17	-10.80
3101	1292	96	13.50	4.69	32	-10.90
3102	1002	59	17.00	4.75	26	-10.50
3103	663	48	13.80	4.22	22	-11.30
4011	1208	79	15.30	4.80	37	-10.60
4012	907	54	16.80	4.43	27	-10.80
4013	751	38	19.80	5.10	21	-9.86
4021	1064	61	17.40	4.42	40	-10.70
4022	736	31	23.70	4.90	25	-9.69
4023	1066	66	16.20	4.49	29	-10.80
4031	1077	63	17.10	4.49	31	-10.70
4032	859	53	16.20	4.46	30	-10.80
4033	840	56	15.00	4.77	23	-10.60
4041	1113	67	16.60	4.56	32	-10.70
4042	1228	70	17.50	4.89	40	-10.30
4043	1167	107	10.90	4.14	24	-11.70
4051	1252	74	16.90	4.35	52	-10.80
4052	999	101	9.89	4.74	29	-11.20
4053	1138	60	19.00	4.16	35	-10.80
4061	1130	62	18.20	4.78	42	-10.30
4062	789	60	13.20	5.10	30	-10.50

4063	1045	57	18.30	5.07	31	-10.00
4071	1312	95	13.80	4.44	29	-11.10
4072	988	50	19.80	4.71	31	-10.30
4073	944	57	16.60	5.10	9	-10.20
4081	1111	78	14.20	4.03	31	-11.40
4082	1500	139	10.80	3.85	47	-11.90
4083	1399	123	11.40	4.12	33	-11.60
4091	1138	57	20.00	4.85	31	-10.10
4092	1042	39	26.70	4.79	26	-9.53
4093	803	30	26.80	4.95	29	-9.36
4101	1620	94	17.20	4.51	52	-10.70
4102	887	61	14.50	4.62	31	-10.80
4103	1357	57	23.80	4.40	43	-10.10

Appendix II

Keywords list from Book I, 21th Century English

课文	词项	总词频	册词频	单元词频	课词频	词长
0011	sun	23	9	9	9	3
0012	rid	2	1	1	1	3
0012	fly	20	7	6	6	3
0022	lid	2	2	2	2	3
0022	box	6	4	2	2	3
0031	cap	10	4	4	4	3
0032	mom	6	6	2	2	3
0032	lip	5	2	1	1	3
0041	pye	5	3	3	3	3
0041	hay	1	1	1	1	3
0041	sip	1	1	1	1	3
0041	wit	2	1	1	1	3
0041	log	2	1	1	1	3
0051	dog	7	5	3	3	3
0051	ice	22	8	7	7	3
0051	row	8	3	2	2	3
0051	hug	4	1	1	1	3
0052	No.	6	2	2	2	3
0061	mom	6	6	2	2	3
0062	bus	27	11	1	9	3
0062	cup	5	1	1	1	3
0071	pad	1	1	1	1	3
0081	dad	6	5	4	4	3
0081	map	5	1	1	1	3
0082	hue	1	1	1	1	3
0082	lip	5	2	1	1	3
0082	fat	1	1	1	1	3
0092	hat	20	20	2	2	3
0092	tea	5	1	1	1	3
0102	sky	13	4	4	3	3
0102	lay	5	1	1	1	3
1013	hug	4	1	1	1	3
1021	hit	28	11	6	6	3

1021	pin	3	1	1	1	3
1022	rap	1	1	1	1	3
1023	dig	4	2	1	1	3
1023	com	1	1	1	1	3
1023	con	1	1	1	1	3
1023	syn	1	1	1	1	3
1023	ant	3	1	1	1	3
1031	toy	10	4	4	4	3
1033	van	3	2	2	2	3
1041	pot	4	1	1	1	3
1041	tea	5	1	1	1	3
1041	lip	5	1	1	1	3
1041	ill	10	2	2	2	3
1042	sir	1	1	1	1	3
1043	rip	1	1	1	1	3
1043	cab	2	1	1	1	3
1051	alp	1	1	1	1	3
1052	flu	1	1	1	1	3
1053	No.	6	3	3	3	3
1053	bus	27	8	8	8	3
1053	hem	2	1	1	1	3
1053	lap	2	1	1	1	3
1053	rob	1	1	1	1	3
1053	hum	2	1	1	1	3
1062	fad	15	15	1	1	3
1062	rat	5	1	1	1	3
1071	pet	1	1	1	1	3
1071	fan	2	1	1	1	3
1071	dig	4	2	1	1	3
1071	gum	1	1	1	1	3
1072	pop	3	2	1	1	3
1073	bag	9	3	3	3	3
1081	kit	1	1	1	1	3
1081	map	5	2	2	1	3
1082	ray	2	2	1	1	3
1082	cup	5	1	1	1	3
1083	fax	3	1	1	1	3
1083	net	6	2	2	2	3
1083	map	5	2	2	1	3
1083	KPS	1	1	1	1	3

1091	cap	10	3	2	2	3
1091	sea	10	4	2	2	3
1091	ton	3	2	2	1	3
1091	dip	1	1	1	1	3
1092	ton	3	2	2	1	3
1092	oil	9	3	3	2	3
1092	ski	1	1	1	1	3
1093	ray	2	2	1	1	3
1093	CFC	1	1	1	1	3
1093	pit	2	1	1	1	3
1101	egg	5	3	3	3	3
1101	pig	1	1	1	1	3
1101	cow	1	1	1	1	3
1101	ban	1	1	1	1	3
1101	fox	2	1	1	1	3
1101	fur	2	1	1	1	3
1102	dye	4	3	3	3	3
1102	lab	3	1	1	1	3
1103	pop	3	2	1	1	3
1103	raw	2	1	1	1	3

Appendix III

Distribution of some syllabus phrases in the *Preliminary Book, 21st century College English*

phrase	c_0011	c_0012	c_0021	c_0022	c_0031	c_0032	c_0041	c_0042	c_0051	c_0052	c_0061	c_0062	c_0071	c_0072	c_0081	c_0082	c_0091	c_0092	c_0101	c_0102	total
a bit																					4
a few												1			2		1		1	1	57
a good deal/ a great deal																					7
a good many/ a great many							1														2
a little					1		1														36
a lot	1		1	1	1					1		2			1						48
a lot/ lot of	3		1	2	1					2		3			2						87
a number of																					6
a sea of																					4
a variety of	1																				3
above all																					1
accord to																2					19
account for															1						4
act on					1																3
add on																					1
add up																					1
add up to																					1
after all			2																		17

个人简历及发表的学术论文

个人简历

姓　　名：沈　岚
出生日期：1970 年 6 月 1 日

学习经历：

　　　1989 年 9 月----1993 年 7 月　　　　四川师范大学获英语专业文学学士学位

　　　2003 年 9 月----2007 年 6 月　　　　成都电子科技大学外国语学院攻读硕士学位

发表论文

沈　岚，2006，性别差异和交际失误，《电子科技大学研究生学报》 第 32 期，45—48 页。