



THE IMPACT OF LEARNING MODES ON THE DEVELOPMENT  
OF COLLOCATIONAL COMPETENCE OF CHINESE EFL LEARNERS



ZHILIANG YUE

Graduate School Srinakharinwirot University

2023

ผลจากวิธีการเรียนรู้ที่มีต่อการพัฒนาความสามารถในการใช้กลุ่มคำ  
ที่ต้องปรากฏร่วมกันของผู้เรียนภาษาอังกฤษชาวจีน



ปริญญาานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปรัชญาดุษฎีบัณฑิต สาขาวิชาภาษาและการสื่อสารสากล  
คณะมนุษยศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ  
ปีการศึกษา 2566  
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

THE IMPACT OF LEARNING MODES ON THE DEVELOPMENT  
OF COLLOCATIONAL COMPETENCE OF CHINESE EFL LEARNERS



A Dissertation Submitted in Partial Fulfillment of the Requirements  
for the Degree of DOCTOR OF PHILOSOPHY  
(Ph.D. (Language and Global Communication))  
Faculty of Humanities, Srinakharinwirot University

2023

Copyright of Srinakharinwirot University

THE DISSERTATION TITLED  
THE IMPACT OF LEARNING MODES ON THE DEVELOPMENT  
OF COLLOCATIONAL COMPETENCE OF CHINESE EFL LEARNERS

BY  
ZHILIANG YUE

HAS BEEN APPROVED BY THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DOCTOR OF PHILOSOPHY  
IN PH.D. (LANGUAGE AND GLOBAL COMMUNICATION) AT SRINAKHARINWIROT  
UNIVERSITY

-----  
(Assoc. Prof. Dr. Chatchai Ekpanyaskul, MD.)  
Dean of Graduate School  
-----

ORAL DEFENSE COMMITTEE

..... Major-advisor	..... Chair
(Assoc. Prof. Dr.Sugunya Ruangjaroon)	(Assoc. Prof. Dr.Supakorn Phoocharoensil)
	..... Committee
	(Dr.Justin james Bartlett)
	..... Committee
	(Assoc. Prof. Dr.Nuntana Wongthai)
	..... Committee
	(Asst. Prof. Dr.Sakulrat Worathumrong)

Title	THE IMPACT OF LEARNING MODES ON THE DEVELOPMENT OF COLLOCATIONAL COMPETENCE OF CHINESE EFL LEARNERS
Author	ZHILIANG YUE
Degree	DOCTOR OF PHILOSOPHY
Academic Year	2023
Thesis Advisor	Associate Professor Dr. Sugunya Ruangjaroon

This study takes collocation, one of the most prevalent types of formulaic expression, as the focus, and aims to answer two questions: 1) what is the collocational competence profile of intermediate Chinese EFL learners? 2) how do specific learning modes affect the development of such competence? In essence, this study strives to provide reference value with regard to collocation competence profiling as well as its development. Two kinds of pedagogical treatments, namely, extensive reading and classroom training, serve as the independent variables. Four groups of Chinese intermediate learners were recruited and received different combinations of such treatments. During pretest and posttest, both spontaneous speaking and untimed writing data were collected and analyzed to probe into their implicit as well as explicit collocational knowledge. A random yet balanced piece of data from Corpus of Contemporary American English serves as a proxy of native speaker reference. Based on the analysis of six aspects (i.e. general lexical proficiency, collocational lexical proficiency, collocation density, collocation diversity, collocation accuracy rate, and collocation strength), it is concluded that one's collocational competence is related to and harder to develop than his/her general linguistic competence, and learners tend to produce less diversified and weaker collocations compared to native speakers. In terms of pedagogy, the matching between treatment type (i.e. implicit/explicit) and resultant knowledge type (i.e. implicit/explicit) is largely confirmed, and the combination of both treatments seem to provide the biggest gain, except for the aspect of collocation strength.

Keyword : collocational competence profiling, collocational competence development, learning mode

## ACKNOWLEDGEMENTS

If life is truly like a box of chocolates, then I am one of those who have been pleasantly surprised opening it. This journey of pursuing a Ph.D. degree was unforeseen in the first place, and unforgettable in the end. There are many I wish to thank and give credit to in this short note.

First of all, my deepest respect and gratitude goes to my advisor, Dr. Sugunya Ruangjaroon. Without her unflinching support through the years, this dissertation simply would not have taken shape by this day. I would like to thank her for all the guidance, love, care, encouragement, friendship, and laughter. She not only set a professional model for me, but also inspired me of the true meaning of life, that is, to help others grow and progress. Although my Ph.D. pursuit largely dovetailed with the pandemic and cultural challenges for a sojourner in Bangkok remained great, she managed everything to make my experience a smooth and rewarding one.

My gratitude is also due to other cherished members of the SWU Humanities family. I want to thank J for his professionalism and friendship, including meeting me at the airport and lending me his working space when I was desperately in need of it. I am grateful to Ajarn Nan for her sisterly support and always showing confidence in me. I am also indebted to Ajarn Anchalee for her guidance and tutorship, as well as to Ajarn Sakulrat for her insightful feedback and kind care. Furthermore, I would like to thank Khun Yee, Khun Ploy, Khun Tong, Khun Keng, Adeline, Amy, Jialun, Mingliang, Wanyi, Yuliang and Xueliang for their willingness to lend a helping hand at anytime.

Finally, I would like to thank my most precious family. My wife Xiaohong is everything that I could ask for from my Lord. Day in and day out, she set my mind at peace, my heart in love, and our household in order. My mom and dad are the source of my strength. I am truly blessed to have such parents and love them very much. If there is any accomplishment in this scholarly journey, I proudly acknowledge that it is ours rather than mine.

ZHILIANG YUE

## TABLE OF CONTENTS

	Page
ABSTRACT .....	D
ACKNOWLEDGEMENTS.....	E
TABLE OF CONTENTS.....	F
LIST OF TABLES.....	H
LIST OF FIGURES.....	J
CHAPTER 1 INTRODUCTION.....	1
1.1 Rationale .....	1
1.2 Objectives of the Study.....	10
1.3 Research Questions.....	10
1.4 Research Hypotheses.....	10
1.5 Scope of the Study.....	11
1.6 Research Framework.....	12
1.7 Definitions of Specific Terms .....	14
1.8 Significance of the Study .....	15
CHAPTER 2 LITERATURE REVIEW.....	17
2.1 An Overview of the Field .....	17
2.2 Delimiting the Term of Collocation.....	19
2.3 Mystery #1: Contrasting Levels of Difficulty for NS and NNS.....	26
2.4 Mystery #2: How to Account for NNS' Collocational Competence .....	30
2.5 Mystery #3: How to Effectively Develop NNS' Collocational Competence .....	35
2.6 Summary .....	39

CHAPTER 3 RESEARCH METHODOLOGY .....	40
3.1 Population .....	40
3.2 Participants .....	40
3.3 Research Design.....	40
3.4 Research Instruments .....	43
3.5 Data Collection.....	54
3.6 Procedure of the Experiment .....	56
3.7 Data Analysis .....	62
3.8 Ethics.....	62
CHAPTER 4 RESEARCH FINDING .....	63
4.1 Collocational Competence Profiling .....	67
4.2 Collocational Competence Development.....	84
CHAPTER 5 CONCLUSION AND IMPLICATIONS.....	99
5.1 Conclusion .....	99
5.2 Implications .....	104
5.3 Final Thought.....	105
REFERENCES.....	106
APPENDIX .....	117
VITA .....	171



## LIST OF TABLES

	Page
Table 1 Meta-analysis of Collocational Competence Profiling Studies .....	7
Table 2 Meta-analysis of Collocation Learning Mode Studies .....	9
Table 3 CEFR Levels (adapted from CEFR - Companion Volume 2020) .....	11
Table 4 Subcategories of V-N Combinations (revised from Nesselhauf 2003) .....	24
Table 5 Participants' Demographics .....	40
Table 6 Groups and Treatments .....	41
Table 7 In-Class Collocation Training Plan .....	51
Table 8 Teaching Plan for Group (e+i) .....	58
Table 9 Teaching Plan for Group (e) .....	59
Table 10 Teaching Plan for Group (i) .....	60
Table 11 Teaching Plan for Group (c) .....	61
Table 12 General Lexical Level Profiling .....	67
Table 13 Collocation Lexical Level Profiling .....	70
Table 14 Collocation Density Profiling .....	72
Table 15 Collocation Diversity Profiling .....	74
Table 16 Collocation Accuracy Rate Profiling .....	76
Table 17 Collocation Strength Profiling .....	78
Table 18 Covert Collocation Errors .....	82
Table 19 Collocation Density Development (token) .....	85
Table 20 Collocation Density Development (lemma) .....	88
Table 21 Collocation Diversity Development .....	90

Table 22 Collocation Accuracy Rate Development (lemma)..... 92

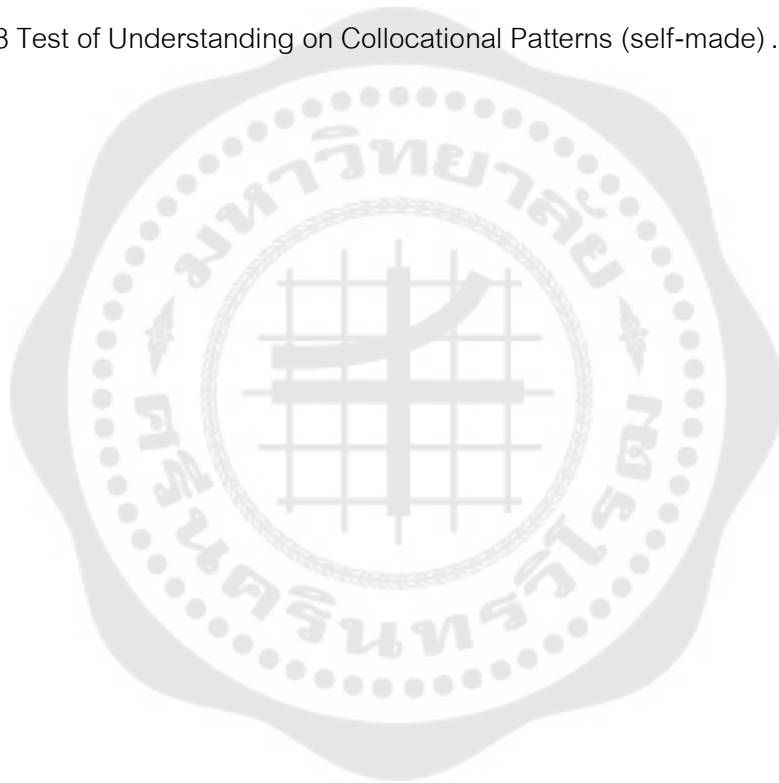
Table 23 Collocation Strength Development (lemma) ..... 94



## LIST OF FIGURES

	Page
Figure 1 Note on “idiomatic”, adapted from Merriam-Webster online dictionary .....	2
Figure 2 Top-down Mode of Language Acquisition (adapted from Wray 2002) .....	29
Figure 3 Bottom-up Mode of Language Acquisition (adapted from Wray 2002) .....	30
Figure 4 Learner’s Comprehension Process (adapted from Ellis 1996) .....	32
Figure 5 Overall Research Design (self-made) (self-made) .....	42
Figure 6 Cambridge Placement Test Interface (self-made) .....	44
Figure 7 Book Selection and Try-Out Form (self-made) .....	47
Figure 8 After-class Reading Log (self-made) .....	49
Figure 9 Data Processing Workflow (self-made) .....	66
Figure 10 General Lexical Level Profiling (self-made) .....	68
Figure 11 Collocation Lexical Level Profiling (self-made) .....	70
Figure 12 Collocation Density Profiling (self-made) .....	72
Figure 13 Collocation Diversity Profiling (self-made) .....	75
Figure 14 Collocation Accuracy Rate Profiling (self-made) .....	76
Figure 15 Collocation Strength Profiling (self-made) .....	79
Figure 16 Collocation Density Development (token) (self-made) .....	86
Figure 17 Collocation Density Development (lemma) (self-made) .....	89
Figure 18 Collocation Diversity Development (self-made) .....	91
Figure 19 Collocation Accuracy Rate Development (self-made) .....	93
Figure 20 Collocation Strength Development (self-made) .....	95
Figure 21 Recommended Resources – Dictionaries (self-made) .....	120

Figure 22 Recommended Resources – ER Website (self-made).....	121
Figure 23 Recommended Resources – Online Collocation Dictionary .....	124
Figure 24 Recommended Resources – COCA (self-made) .....	125
Figure 25 Collocate Matching Practice (self-made).....	133
Figure 26 A List of Intensifying Adverbial Synonyms (self-made).....	136
Figure 27 Collocation Pattern Identification Practice (self-made).....	137
Figure 28 Test of Understanding on Collocational Patterns (self-made) .....	137



## CHAPTER 1

### INTRODUCTION

#### 1.1 Rationale

“I wouldn’t say it that way. But it is surely understandable.”

Ashley, one of my friends from the States, grins when hearing my plan of “*wash the clothes*”. She just quizzed about how I was going to spend the afternoon.

“How would you say it then?”


“*Do the laundry.*”

At that moment, a mixed feeling of embarrassment and delight arises in me. The good thing is that I am having another opportunity to calibrate my English. The bad thing is that I don’t know whether this calibration process will ever come to an end. Having studied the language for a minimum of twelve years, got 114 on TOEFL, and immersed myself in the U.S. for almost three years, I still run into this kind of “calibration” moment from time to time. There seems to be a persistent linguistic gap between a fluent immigrant and native speakers. Compliments like “surely understandable” do not help bridge that gap a bit.

Years later, as I became a college English teacher back in China, it became clear to me that such problems are far from being my own business. Virtually every Chinese undergraduate, including those majoring in English, fails to attain native-like quality. It seems that thousands of hours of word memorization and grammar learning, as we have long been trained to do, is not sufficient for success. Again and again, aspiring students complain to me about their dampened faith in language learning:

- “*When watching that episode, I know almost each word individually, but I just can’t figure out how the actors are able to put words together to sound native.*”
- “*Why must I say ‘smart TV’ but not ‘clever TV’? Aren’t they synonyms?*”
- “*How come my essay didn’t receive a high mark? There is nothing wrong in my spelling and grammar!*”

Indeed, this whole matter is not about right and wrong. It is about whether your language is *idiomatic* or not. In fact, if one does a search of the word “idiomatic” in Merriam-Webster online dictionary, a little note will show up at the bottom of the page as below:

 **Did you know?**

The speech and writing of a native-born English-speaker may seem crude, uneducated, and illiterate, but will almost always be idiomatic—that is, a native speaker always sounds like a native speaker. For a language learner, speaking and writing *idiomatically* in another language is the greatest challenge. Even highly educated foreign learners—professors, scientists, doctors, etc.—rarely succeed in mastering the kind of idiomatic English spoken by an American 7th-grader.

Figure 1 Note on “idiomatic”, adapted from Merriam-Webster online dictionary

This explanation of the term is echoed in academia. For example, Lennon (1998) frequently equates “unidiomatic” with “non-nativelike”. Perhaps, what Merriam-Webster Dictionary terms as “the greatest challenge” will be a relief to many. After all, it would be a rare success to sound idiomatic among probably hundreds of millions of English learners worldwide, why should one care so much and be so hard on himself?

While such relief is helpful, we need to maintain a balanced view. In spite of the newer trends of World Englishes and translanguaging (for a review, see Xu, 2019), for the vast majority of L2 teachers and learners, the benchmark is still native speakers (e.g., Bestgen & Granger, 2014; Foster, 2001; Laufer & Waldman, 2011; Nesselhauf, 2003). Therefore, while the challenge remains great, there is no reason not to strive for a better result of all the efforts one makes. From such a perspective, that little note from Merriam-Webster is no less than an invitation to scholars in the field of Second Language Acquisition (SLA): Is there any way to improve the situation? Is there any hope that a foreign learner of the language, with proper training and practice, will eventually sound like a native? There are a lot of potential benefits hanging in the balance, not least those related to academics and career development (James, 1998).

This is how this study takes shape in the first place. As exemplified in one of the complaints earlier, to sound idiomatic, one needs to know how words combine to form utterances in specific contexts, which, in turn, goes beyond merely being semantically sensible and grammatically correct. Sinclair (1991) is one of the pioneers in bringing forth theories to lay bare this matter. Based on rigorous examination of patterns in his corpus, he claims that there are two principles at work, namely, the open-choice principle and the idiom principle. The former accounts for all kinds of “novel expressions”, the essence of which is to fill the slots with lexemes that meet the local semantic and grammatical requirements. This has traditionally received considerable attention from SLA researchers, teachers and students alike, and serves as the basis of almost all the EFL textbooks (Bestgen & Granger, 2014; El-Dakhs, 2015; Henriksen, 2013). The latter means that a large part of native speakers' utterances is prefabricated and readily available, rather than freshly composed every time. This idiom principle, as Sinclair argues, is the default mode for the sake of effort-saving for the speaker, and only switches to the other principle when necessary. It seems that most learners only focus on the first principle while ignoring the second, which leads to their unsatisfactory achievement (El-Dakhs, 2015; Fan, 2009). Sinclair's theory will be discussed in greater detail in Chapter 2.

If learners' inadequacy is indeed due to their ignorance of the idiom principle, then three subsequent matters call for explanation: why they ignore it, what specific linguistic components they are ignoring, and how they can make it right.

To address the first matter, a number of scholars have offered their thoughts and evidence (e.g., Wen, 1995; Pu, 2003, among others), but arguably the most widely accepted one is of Wray's (2002). Based on Sinclair's work and others, Wray postulates a dual model of language acquisition. The *top-down mode* accounts for how native speakers acquire their L1. While growing up as a toddler, we tend to absorb expressions around us as whole chunks, never breaking them down to smaller units unless there is a genuine need. This means that for a child, the earliest type of intake to her repertoire is a mixed bag: sentences, phrases and words are all present and

probably in a declining order of significance. The other model, called *bottom-up mode*, accounts for an ordinary non-native speaker's experience of acquiring an L2. As the learner is generally literate to some extent, he or she habitually pays attention to grammar rules and rote memorization of words. This is partly due to the way they are taught, and partly due to their own past learning experiences. As one reaps what he sows, the consequence is that learners' repertoire is filled up with individual words and a set of rules. When trying to compose larger units of expression, they inevitably suffer from too much freedom and all kinds of unidiomatic utterances come out accordingly (Darvishi, 2011; Laufer & Waldman, 2011; Nesselhauf, 2003). This distinct difference in schema between native speakers and non-native speakers could explain the persistent linguistic gap between the two groups.

Moving on to the second matter, the linguistic forms related to the idiom principle is often termed as "formulaic expressions" or "formulaic sequences" (Wray, 2002). This is an umbrella term, encompassing anything that is prefabricated rather than made on the spot. It covers several categories, including lexical bundles (Biber, 2009), idioms (Cowie, 2009), collocations (Nesselhauf, 2003) and so forth. While none of them is trivial, collocation seems particularly important and pervasive in English (Sun, 2004), and the mastery of it seems to be the strongest predictor of one's lexical proficiency (Vu & Peters, 2022). Collocation refers to the linguistic phenomenon that two or more words habitually show up together in expressions (Firth, 1957; Lewis, 1997). This is one of the areas in which learners have great difficulty: relying on semantics and syntactic rules alone, they are often blinded about which word goes with which word in actual use, again, as one of the aforementioned complaints of my students exemplifies. Fortunately, this matter has drawn increasing attention from the academics. As Harmer (2015) states, "word combinations (also known as collocations) have become the subject of intense interest in the recent past." This study joins the trend and takes collocation as its focus, aiming to profile the learners' competence and diagnose what their inadequacies are.

Then, turning to the matter of how learners can remedy the situation, we essentially touch the realm of pedagogy. Taking numerous testimonies and my personal



experiences into consideration, it seems obvious that the traditional way of teaching an L2 fails to bring about the desired result. Explanation of rules, requirement of rote memorization and endless drills, although not altogether meaningless, do not help a learner progress to idiomaticity (Bestgen & Granger, 2014; El-Dakhs, 2015). When it comes to collocation teaching, many either are simply not aware of the need or regard it as unteachable. For example, Swan (2006) likens teaching the formulaic language, such as collocations, to “trying to empty the sea with a teaspoon.” Henriksen (2013) also points out that individual words receive much greater attention than collocations in language teaching. On the other hand, not everyone shares such a pessimistic prospect (Dellar & Walkey, 2017; R. Ellis, 2014; Toomer & Elgort, 2019). Nevertheless, scholars seem to have reached a consensus that evidence-grounded pedagogical recommendations related to collocation development must be put forward.

This study partly aims to contribute to that endeavor. Besides, it also aims to advance the knowledge of learners’ collocational competence profile, which serves as the prerequisite of any pedagogical effect evaluation (El-Dakhs, 2015). Many researchers have already explored learners’ collocational competence from various angles, such as accuracy rate (Alangari, n.d.; Crossley et al., 2015; Nesselhauf, 2003; Peng et al., n.d.; Laufer & Waldman, 2011;), degree of use (Alangari, n.d.), and statistical measurements (Chen, 2021; Crossley et al., 2015; Durrant & Schmitt, 2009; Laufer & Waldman, 2011; Peng et al., n.d.; Siyanova-Chanturia, 2015). However, most of them focused on the advanced group, and explored mainly Verb-Noun type of collocation, as Table 1 shows. Furthermore, some of them evaluate collocational competence from a statistical point of view, and others from a categorical one, but hardly anyone from both angles. This is a gap that the current study aims to fill, through covering the intermediate group, three other common types of collocation (i.e., Adjective-Noun, Adverb-Verb, Adverb-Adjective), and encompassing both categorical and statistical measures.

As for pedagogical treatments, some researchers (e.g., Vu & Peters, 2022; Webb & Chang, 2020; Toomer & Elgort, 2019) have compared different modes of implicit learning, such as reading, reading while listening, reading with textual input enhancement, etc. Fewer researchers have touched on the matter of explicit teaching. For example, Chen (2016) examined the effect of dictionary use teaching on collocation development. But, as Table 2 demonstrates, there seems to be few, if any, scholars who have made a comparison between explicit treatment and implicit treatment, as well as their combination in terms of collocation gains. This is another gap that this study strives to bridge, through implementing three different learning modes (implicit treatment – extensive reading, or ER, explicit treatment – classroom instruction, and the combination of the two).

Therefore, the current study is essentially one of corpus-based language learning research (LLR). It first depicts the collocational competence of the target group of learners, which will be followed by a quasi-experiment to explore the effect of three different learning modes (i.e., ER only, explicit teaching only, and the combination of the two) on its development. This study contributes to the body of existing knowledge in filling three gaps: empirical data (in terms of proficiency level, production type, collocation coverage, and scope of profiling), methodological approaches (both implicit and explicit independent variables involved), and pedagogical materials (collocation lessons/syllabus devised).

Table 1 Meta-analysis of Collocational Competence Profiling Studies

Study	Subjects	Collocation Coverage	Scope of Profiling	Main Findings
Nesselhauf, 2003	Advanced German EFL learners	<ul style="list-style-type: none"> <li>Verb-Noun type</li> </ul>	overt and covert errors	<ul style="list-style-type: none"> <li>Learner's L1 seems to affect collocation accuracy the most.</li> <li>The degree of restriction has a minor impact on collocation accuracy as well.</li> </ul>
Alangari, (n.d.)	Advanced Saudi EFL learners	<ul style="list-style-type: none"> <li>Verb-Noun and Adjective-Noun types</li> </ul>	overuse, underuse, and misuse	<ul style="list-style-type: none"> <li>Learners produce more verb-noun tokens and a comparable number of verb-noun types compared to native speakers.</li> <li>No significant difference regarding Adjective-Noun collocation use between the two groups.</li> </ul>
Chen, 2021	Beginning, intermediate, and advanced Asian EFL learners	<ul style="list-style-type: none"> <li>All bigrams extracted</li> </ul>	exclusivity, directionality, dispersion, and unseen rates	<ul style="list-style-type: none"> <li>Learners develop their collocational competence as their proficiency grows.</li> <li>Learners show a decreasing use of unseen bigrams but increasing use of domain-specific bundles.</li> </ul>
Laufer & Waldman, 2011	Beginning, intermediate, and advanced Hebrew EFL learners	<ul style="list-style-type: none"> <li>Verb-Noun</li> </ul>	frequency and correctness	<ul style="list-style-type: none"> <li>Learners at all levels produced far fewer collocations than native speakers.</li> <li>Errors continued to persist even at advanced level of proficiency.</li> </ul>

Table 1 (Continue)

Peng et al., (n.d.)	Advanced learners of Chinese	<ul style="list-style-type: none"> <li>• Verb-Noun</li> </ul>	token, type, frequency, correctness	<ul style="list-style-type: none"> <li>• Learners produce significantly less Verb-Noun type collocations than native speakers, both in terms of token and of type.</li> <li>• Most incorrect collocations do not affect communication.</li> </ul>
Durrant & Schmitt, 2009	ESL/EFL learners with various L1 background	<ul style="list-style-type: none"> <li>• Modifier-Noun</li> </ul>	frequency, associative strength	<ul style="list-style-type: none"> <li>• Learners rely heavily on high-frequency collocations.</li> <li>• Learners underuse less frequent, strongly associated collocations.</li> </ul>
Syanova- Chanturia, 2015	Beginning Chinese learners of Italian	<ul style="list-style-type: none"> <li>• Noun-Adjective</li> </ul>	token, associative strength	<ul style="list-style-type: none"> <li>• Learners show an increase in strongly-associated collocations.</li> <li>• Learners show an increase in higher frequency collocations.</li> </ul>
Crossley et al., 2015	Beginning, intermediate, and advanced English learners with various L1 backgrounds	<ul style="list-style-type: none"> <li>• (not mentioned)</li> </ul>	collocation accuracy, lexical diversity, word frequency	<ul style="list-style-type: none"> <li>• The analytic judgements of the three parameters are highly predictive of holistic judgements of lexical proficiency.</li> <li>• Depth of knowledge related to collocation accuracy are the strongest predictors of lexical accuracy.</li> </ul>

Table 2 Meta-analysis of Collocation Learning Mode Studies

Study	Subjects	Learning Mode(s)	Measurements	Main Findings
Vu & Peters, 2022	pre-intermediate Vietnamese EFL learners	reading while listening, reading with textual input enhancement, and the combination of the two	form recall	<ul style="list-style-type: none"> <li>• Reading-while-listening plus textual input enhancement and reading-while-listening are more effective.</li> <li>• Learners' prior vocabulary knowledge and collocational congruency have an impact on learning.</li> </ul>
Webb & Chang, 2020	Taiwanese EFL learners	reading, listening, and reading while listening	word matching and meaning recall	<ul style="list-style-type: none"> <li>• Reading while listening proved to be the most effective.</li> <li>• Reading and listening contribute to similar gains.</li> </ul>
Chen, 2016	intermediate Chinese EFL learners	dictionary-assisted learning	fill-in elicitation technique	<ul style="list-style-type: none"> <li>• Dictionary significantly improved learners' productive collocation knowledge.</li> <li>• The retention rate is quite low after one week.</li> </ul>
Toomer & Elgort, 2019	advanced ESL learners	reading, reading with textual input enhancement, and reading with textual input enhancement plus glossing	explicit measures and implicit measures	<ul style="list-style-type: none"> <li>• All modes brought about gains in the explicit knowledge.</li> <li>• No mode brought about gains in the implicit knowledge.</li> </ul>

## 1.2 Objectives of the Study

This study has two objectives as below:

1.To profile collocational competence of intermediate Chinese learners of English using a corpus-based approach.

2.To explore the efficiency of three learning modes (ER only, explicit teaching only, and the combination of the two) on collocational competence development.

## 1.3 Research Questions

The research questions of this study are framed as follows:

1.What is the collocational competence profile of intermediate Chinese EFL learners?

2.How do different learning modes (ER only, explicit teaching only, and the combination of the two) affect the development of their collocational competence?

## 1.4 Research Hypotheses

Based on the existing body of research, this study makes the following hypotheses:

1.Intermediate L2 learners' collocational competence is markedly lower than that of native speakers (NS) in every aspect, such as diversity and accuracy (Nesselhauf, 2003). Additionally, their use of collocations should demonstrate a different statistical distribution (e.g. frequency, associative strength) from that of NS (Granger, 1998).

2.The combination of explicit instruction and implicit treatment (i.e., ER) will result in greater gain of collocations than explicit instruction only or implicit treatment only (Loewen, 2020:27).

3.Explicit instruction will mainly lead to an increase of explicit collocational knowledge while implicit treatment will mainly lead to an increase of implicit collocation knowledge. Besides, there may also be some minor crossing effects (Loewen, 2020:32-36).

## 1.5 Scope of the Study

This study focuses on a specific group of learners: those of intermediate level. This level, in the broader sense, according to the Common European Framework of Reference for Languages (CEFR), includes B1 and B2 (as shown in Table 3). Intermediate learners are largely under-researched when it comes to collocational competence, as the vast majority of existing studies (e.g. Alangari, 2019; Chang et al., 2008; Nesselhauf, 2003) in the literature are about advanced learners only. However, intermediate learners are no less important, since if any remedial intervention is necessary, it is better to be put in place earlier than later. On the other hand, those at beginner level possess a too-limited linguistic repertoire to produce considerable tokens of collocation for investigation, and thus are excluded in this study. Table 3 shows the scale of proficiency based on CEFR levels.

Table 3 CEFR Levels (adapted from CEFR - Companion Volume 2020)

<b>Proficient user</b>	<b>C2</b>	Can understand virtually all types of texts. Can summarise information from different oral and written sources, reconstructing arguments and accounts in a coherent presentation. Can express themselves spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	<b>C1</b>	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express themselves fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
<b>Independent user</b>	<b>B2</b>	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in their field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with users of the target language quite possible without imposing strain on either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	<b>B1</b>	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise while travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
<b>Basic user</b>	<b>A2</b>	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of their background, immediate environment and matters in areas of immediate need.
	<b>A1</b>	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce themselves and others and can ask and answer questions about personal details such as where someone lives, people they know and things they have. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Regarding actual data, only learners' production, both oral and written, were collected and analyzed. This is in line with most studies. Ideally, their comprehension should also be examined, but according to Corder (1974), any measure of comprehension ability has to be indirect and tricky in nature, and even if we are able to locate certain patterns, such as overt errors, it is very difficult to associate them with definite causes. This will be discussed in more detail in Chapter 2.

In terms of types of collocation, this study investigates three of the most common ones, namely, Adj-N, Adv-Adj, and Adv-V. Another pervasive type, V-N, has already been quite thoroughly studied by others (e.g. Alangari, 2019; Chen, 2017; Nesselhauf, 2003; Peng, 2016) and thus is skipped here. On the other hand, other minor types, such as V-Adj, do not warrant as much attention as these above ones, at least at this stage.

Lastly, this study explores both implicit knowledge and explicit knowledge of the learners in order to arrive at a fuller picture of their competence. As I side with those scholars (e.g. Bialystok, 1978) who take an interface position, both pieces are of significance to L2 acquisition and we cannot afford to miss either. The way to do this, as will be explained in greater detail in Chapter 3, is to have participants give free talks which are largely based on implicit knowledge only (Suzuki, 2017), and have them compose essays with sufficient time to prepare and revise so that they will rely on both implicit and explicit knowledge (Loewen, 2020). Such free responses are closest to natural use of the language (Ellis, 2008) and thus arguably best reflects learners' genuine competence.

## 1.6 Research Framework

On the psycholinguistic side, this study adopts the models proposed by Sinclair (1991) and later developed by Wray (2002). As mentioned earlier, Sinclair identifies two principles of text interpretation, the open-choice principle and the idiom principle. He argues, "one is not enough. No single principle has been advanced which accounts for the evidence in a satisfactory way." The open-choice principle, as is familiar to all, means that a language user fills the slots of text from a lexicon to satisfy local restraints,



semantically and syntactically. This process is energy-consuming in nature and could potentially generate all kinds of novel utterances. The idiom principle, on the other hand, means that a language user has available to him a stock of semi-prefabricated chunks that serve as single units. When relying on this principle, one's production becomes much more efficient and predictable, fitting the general patterns of a community. Sinclair points out that most learners unfortunately fail to establish their idiom mode of thinking in their L2, which is responsible for their unidiomatic way of expression.

Built on Sinclair's notion, Wray (2002) proposed a dual model of language acquisition. She contends that there is fundamental difference in one's L1 and L2 acquisition: the former is governed by the top-down mode while the latter by the bottom-up mode. The top-down mode means that a baby, when hearing an utterance from others (e.g. caregivers), tends to take it in as a whole for later imitation. As time passes by, he or she may gradually discover some rules, and break the previous intakes down to smaller units to compose novel utterance as context calls for it. In this way, a native speaker will always establish the idiom mode first, which largely shapes his or her language patterns, and only later become able to make use of open-choice mode. This explains why native speakers always sound native. L2 learners, on the other hand, often being literate already when studying the language, tend to master the rules as quickly as possible, and are keen to expand their vocabulary through rote memorization. Thus, they have nothing else to rely on except the open-choice mode, which in turn will result in all kinds of utterances that do not fit common patterns of native speakers.

On the pedagogical side, this study makes a distinction between implicit knowledge and explicit knowledge and mainly draws on Krashen's Monitor Theory (1982, 2003) to develop the former. As probably the first theory dedicated to SLA, it consists of five hypotheses which have lasting effect in this field. The researcher tried out a curriculum targeted at developing learner's collocational competence based Krashen's key notions. According to the comprehensible input hypothesis, acquisition is likely to take place when learner is provided with ample, comprehensible ( $i+1$ ) and engaging input. This was realized in after-class weekly reading assignments. According

to the affective filter hypothesis, learners “should not be put on the defensive”, meaning a low-stress and even enjoyable atmosphere should be established by the teacher in order to maximize the gain. On the other hand, as mine is an interface position (DeKeyser, 2015; Ellis, 2007), different from Krashen’s non-interface one, some modifications, such as trainings of awareness raising and patterns explanation, were made to facilitate the development of learners’ explicit knowledge as well. Also, as some studies illustrate (Trahey & White, 1993; Loewen et al., 2009), a mere input flood may not be sufficient to improve learners’ accuracy, and explicit teaching which offers negative evidence should be conducted to enhance their learning efficiency.

### 1.7 Definitions of Specific Terms

*Collocation:*

This study takes a mixed approach in defining collocation, considering both frequency and phraseological factors. Accordingly, a collocation is two or more words that 1) fit predetermined syntactic patterns (i.e. Adj-N, Adv-Adj, Adv-V), and 2) manifest themselves in collocation lists of a large reference corpus (i.e. Corpus of Contemporary American English).

*Collocational competence:*

This study focuses on actual ability rather than abstract knowledge of the learners. Therefore, collocational competence is defined as one’s actual ability to use his or her collocational knowledge to establish effective communication in specific contexts.

*Implicit knowledge:*

Implicit knowledge of a language refers to one’s intuitive and tacit system of knowledge that is readily available for online use (Ellis, 1996; Krashen, 1981). It is closely associated with Krashen’s acquisition-learning distinction hypothesis.

*Explicit knowledge:*

Explicit knowledge of a language refers to the rules and patterns one consciously learns through analysis (Ellis, 1996; Krashen, 1981). Making use of this kind

of knowledge usually requires time and effort, and thus it is not instantly available. It is also closely associated with Krashen's acquisition-learning distinction hypothesis.

*Extensive reading (ER):*

Extensive reading (ER) is an approach to develop a learner's linguistic proficiency through reading easy and enjoyable texts for an extended period of time (Takayuki, 2014). It is considered as an implicit way of learning, and contrasts with intensive reading. The ER program adopted in this study practices seven of the ten principles (Day & Bamford, 2002), and thus could be considered as modified ER on the ER continuum (Day, 2015).

*Chinese EFL learners:*

Chinese EFL learners are those who 1) are Chinese nationals and 2) possess Mandarin as their L1, and 3) learn English as a foreign language. The vast majority of English learners in China fits this definition.

### **1.8 Significance of the Study**

This study, along with many others, aims to explore how to reach the ultimate goal of L2 learning: being idiomatic. In this overall picture, collocation, an indispensable piece, is chosen as the very focus of the research. According to McCarthy and O'Dell (2017), learning this aspect of English well could offer at least three benefits: to speak in a more natural and accurate way, to make your expressions more versatile and colorful, and to improve your style in writing.

There are three gaps in the literature that this study aims to fill. The first is one of empirical data. As Table 1 previously shows, virtually all existing collocation studies investigate advanced learners only, and most of them choose either categorical (e.g., types of errors) or statistical measures (e.g., frequency and associative strength) in profiling the competence. In an effort to address this gap, I take the group of intermediate learners as my target population, and strive to draw a fuller picture of their collocational competence, including both categorical and statistical measures. This should deepen our understanding of the issue.

The second gap is of methodological approaches, namely, the choice of learning modes. As Table 2 previously shows, most, if not all, existing studies examine either the effect of implicit treatments only, such as ER (e.g. Vu & Peters, 2022), or the effect of explicit treatment only, such as dictionary use instruction (Chen, 2016). The current study takes both types of treatment into account, and investigates the effect of different modes on collocational competence development with regard to implicit and explicit knowledge.

The third gap is about pedagogical materials. Grounded in the existing body of research, I have designed a seven-week training program on collocations, with each week corresponding to a different topic such as awareness raising, characteristics of collocation, resource using, etc. Along with that, teaching content will also be devised, which should enrich the reference materials of collocation pedagogy, especially concerning learners with L1 Chinese background.

Overall, this study intends to make a preliminary yet necessary contribution to the field of SLA. Stemming from the researcher's own experiences and observations, inspired and informed by other scholars' ideas, theories and resources, I hope that the outcome of this work will in turn become a stepping stone for many in future.

## CHAPTER 2

### LITERATURE REVIEW

*"You shall know a word by the company it keeps."*

J. R. Firth: *Papers in linguistics*. 1957

#### 2.1 An Overview of the Field

This memorable quote by Firth brings into light the inextricable relationship between a word and its textual surroundings. Indeed, words do not come together by chance. Nor do they interact only according to the restrictions of grammar (Sinclair, 1991:110). As Firth points out, each word has some specific "company" that it habitually co-occurs with. It is this notion of co-occurrence that we will focus on in the present study.

Numerous researchers have devoted their attention to this interesting linguistic phenomenon (see Wray, 2002 for a review). The general umbrella term adopted in the literature is *formulaic language* (also known as *formulaic sequences* or *formulaic expressions*). According to Ellis (1996), it consists of expressions which are learnt as unanalyzable wholes and employed on particular occasions. Arguably, the most cited definition is from Wray (2002:9) as follows:

*"...a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar."*

Such a definition, as Wray (2002) contends, aims to be "as inclusive as possible" and covers a number of relevant subordinate terms such as *idioms* (Nesselhauf, 2003; Tabossi et al., 2009), *collocations* (Alangari, 2019; Chang et al., 2008; Chen, 2017; Ferraro et al., 2013), *lexical bundles* (Biber 2009; Schmitt et al.,

2004), *phrasal verbs* (Kim & Kim, 2012) and the like. It stresses the holistic nature of such prefabricated sequences in storage and processing, an important point to which we shall return later. Wray's argument has been largely borne out by studies which demonstrated processing advantage of formulaic sequences over nonformulaic ones (Underwood et al., 2004; Jiang & Nekrasova, 2007; Kim & Kim, 2012). However, not every scholar agrees on this point. For example, Siyanova-Chanturia (2015) argues from a psycholinguistic perspective that formulaic sequences are not necessarily fully-analyzed, although she does acknowledge their facilitative effect in language processing.

One of the most-researched subcategories of formulaic language is *collocation*, which, according to Cowie, is "the largest group of set phrases in English" (Cowie, 2009:49). In a simple sense, collocation refers to the way words are naturally associated with each other (McCarthy & O'Dell, 2017:4). For example, *fast food* and *quick question* should have a natural sounding to most native speakers' ears while *quick food* and *fast question* could easily get frowned upon, although *fast* and *quick* are normally considered as synonyms. Collocations such as these mentioned above are prevalent in any language (Biber & Conrad, 1999), and collocational knowledge serves as an indispensable constituent of one's overall linguistic competence. Nation even goes as far as to say that "language knowledge is collocational knowledge" (Nation, 2001:318). In general, researchers have reached the consensus that collocations, like other types of formulaic language, play a crucial role in our daily communication. Chen (2017) states that the use of right collocations could help lessen listeners' burden of decoding and thus reduce the risk of misunderstanding. Vilkaite and Schmitt (2017) extends the benefits to both reception and production, arguing that word combinations which conform to collocation norms are processed faster and/or more accurately than those which do not. McCarthy and O'Dell (2017), in their well-written Cambridge textbook "English Collocations in Use", inform EFL/ESL learners of another added value in learning collocations: a more versatile style of expression, thereby a better impression on the examiners.

Indeed, McCarthy and O'Dell represent the many researchers and practitioners in the field of Second Language Acquisition (SLA) who express compelling interest in studying collocations. In their eyes, collocations are both fascinating and problematic. On the one hand, the mastery of collocations is the recipe for success: they make learners sound more nativelike with increased fluency. On the other hand, collocations are notoriously difficult for learners to fully acquire. For example, based on a thorough corpus analysis, Sinclair draws the conclusion that many learners avoid using the common lexical items to make up idiomatic phrases, and as a result, "their language sound stilted and awkward" (Sinclair, 1991:79). Ferraro et al. (2014), in the same vein, contend that even advanced learners who have thoroughly mastered the grammar of L2 are prone to collocation mistakes. On the contrary, collocations seldom, if ever, cause any difficulty to native speakers. The fact that everyone could naturally and effortlessly develop his/her L1 collocational competence seems truly remarkable, provided that no established "rule" could yet fully explain the semantic choice of words. How can something so easy for native speakers turn out to be so difficult for L2 learners, including those who have progressed far in their proficiency? In addition, two other relevant questions follow: how can we measure learners' collocational competence in a reliable and comprehensive manner? And how can collocations be learned effectively and properly? It is the pursuit of the truth behind these questions (referred to as the three "mysteries" thereafter) that conceived the present study. In the remainder of this chapter, I will first discuss various approaches of defining collocation, identifying the one that best fits the purpose of this research. Then detailed exploration of these three mysteries follows in sequence. Finally, I will conclude with the proposal of my research questions.

## 2.2 Delimiting the Term of Collocation

The very first step of studying anything is to define it properly. Unfortunately, although its concept seems fairly easy to grasp, the definition of collocation is far from straightforward. As Chen (2016) rightly points out, scholars usually define the term according to their various research areas or practical purposes. As a result, it is not

surprising to find many corpus linguists, with their access to large-scale data, tend to adopt a frequency-based definition (Biber, 2009; McEnery et al., 2006; Sinclair, 1991). On the contrary, the researchers who mainly work in the field of phraseology and anyone who shares the same schema of language prefer a phraseological approach (Cowie, 2009; Howarth, 1998; Nesselhauf, 2003). A third group of scholars make an effort to reap the benefits of both by incorporating them into one mixed approach (Chen, 2019; Peng, 2016). I will first briefly discuss the two main approaches with their respective strengths and weaknesses, and then form my own working definition of collocation for the present research.

### 2.2.1 Frequency-based Approach

Studies which adopt this approach generally regard collocation as statistically-significant co-occurrence of words within a predetermined distance of each other (Walker, 2011). Sinclair is the right figure to cite here since his pioneering work in corpus linguistics greatly informed those after him. According to Sinclair and his associates, a combination of two words is considered as a collocation if its frequency is above what “their respective frequencies and the length of the text in which they appear would predict” (Sinclair et al., 1970:10). He offers a clear illustration of this by using the phrasal verb *set off* in his classic book *Corpus, Concordance and Collocations* (1991:69-70) as follows:

- *set* occurs about 250 times per million words, or 0.00025
- *off* occurs about 550 times per million words, or 0.00055
- Therefore, the possibility of *off* to occur immediately after *set* should be:  $0.00025 \times 0.00055 = 0.0000001375$
- That means in Sinclair’s reference corpus of 7.3 million words, such a phrase should manifest itself only once ( $0.0000001375 \times 7,300,000 = 1$ ).
- However, the actual situation is that *set off* has 70 occurrences in the reference corpus, which qualifies it as a rather strong collocation.



In another publication (Sinclair, 1966:415), Sinclair introduces three useful terms for collocation discussion which are well taken up by the mainstream of academic community:

*“We may use the term node to refer to an item whose collocations we are studying, and we may define a span as the number of lexical items on each side of a node that we consider relevant to that node. Items in the environment set by the span we will call collocates.”*

Sinclair and Jones (1974) further propose a span of four words on either side of the node, which also becomes established as a common practice among researchers. However, Walker (2009) points out that, ideally, the cut-off point should vary with the syntactic pattern of different collocations. For example, while it is quite unlikely for an adjective to be far apart from the noun it modifies (e.g., heavy rain, a big boy, etc.), a verb could be quite distant from its nominal objective due to syntactic transformation (e.g., watch TV → The TV which my family used to watch all the time got broken last year.) Regrettably, this sound advice has rarely been heeded in the literature.

Later, another factor of associative strength is introduced into the picture to overcome certain shortcoming of solely relying on frequency. Although there is no lack of statistical measures available (O'Donnell et al., 2013), most studies seem to favor the one called Mutual Information (MI). MI is commonly used in the field of information science in assessing the degree to which multiple lexical items in a phrase co-occur more frequently than chance would predict. According to Davies (2008-), MI is calculated as follows:

$$MI = \log ((Freq_{AB} \times \text{size of corpus}) / (Freq_A \times Freq_B \times \text{span})) / \log (2)$$

$Freq_{AB}$  = frequency of the collocation AB in the corpus

$Freq_A$  = frequency of the node A in the corpus

$Freq_B$  = frequency of the collocate B in the corpus

*size of corpus* = total number of words in the corpus

*span* = number of adjacent words considered (e.g., for a span of 4:4 the value would be 8)  $\text{Log}(2) = 0.30103$

For example, in a billion-word corpus such as the Corpus of Contemporary American English (COCA), the frequencies of the collocation “*concede defeat*”, and its constituents “*concede*” and “*defeat*” are 221, 4267, and 21950, respectively. Therefore, the MI of *concede defeat* is calculated as:

$$MI = \log \left( \frac{221 \times 1,000,000,000}{4267 \times 21950 \times 8} \right) / \log(2) = 8.20$$

As the formula infers, a higher MI score means a stronger association between the words. A note of caution is that MI tends to appear distorted for extremely low-frequency collocations (Evert, 2005), and therefore a threshold of minimum frequency needs to be put in place to avoid this problem.

Frequency-based approach has several apparent advantages. It is usually grounded in large-scale corpus analysis with minimum human predetermination involved. For example, Sinclair claims that “no account is taken of syntax, punctuation, change of speaker, or anything other than the word-forms themselves” (1991:117) – and therefore its findings are generally considered to be objective and statistically reliable. Besides, since the methodology is straightforward, the studies which adopt this approach are easy to replicate, provided that the same data set is publicly accessible. However, this approach is not without certain shortcomings. Most notably, if one *only* relies on raw frequency, then the list is likely to be topped by some uninteresting binary units, such as “*and a*” or “*but the*”, both of which have an occurrence of close to half a million in COCA. In fact, even some of the most ardent proponents of the frequency-based approach acknowledge the necessity of “a preliminary step ... where the linguist chooses ‘interesting’ target words” (Biber, 2009). Nevertheless, the key concepts and terms proposed by this approach are well established to the present day.

### 2.2.2 Phraseological Approach

The phraseological approach took shape decades before large text corpora became available. It was originated by some Russian linguists back in the 1940s and later taken up and developed in the western academic circle (Cowie, 1998; Howarth, 1998). Instead of dealing with statistical measurements, this approach puts its emphasis on syntactic structure and semantic properties, both of which, in turn, are rooted in one's intuitive judgement. For example, researchers who belong to this camp (Chen, 2017; Cowie, 2009; Nesselhauf, 2003) accept the traditional distinction between lexical words (also known as open class words) and function words (also known as closed class words), and on this ground propose the syntactic patterns of their interest, such as:

**verb + noun:** grab lunch, attend meeting, do homework

**adjective + noun:** grand hotel, brilliant idea, desperate need

**adverb + adjective:** blissfully ignorant, stunningly beautiful, clearly visible

**adverb + verb:** fully understand, carefully choose, abruptly end

As it shows, word pairs which fit such predetermined syntactic patterns are likely to make intuitive sense. Of all these patterns, “verb + noun” type seems to be the most researched one among SLA researchers (Alangari, 2019; Chen, 2017; Nesselhauf, 2003; Peng, 2016; Vilkaite & Schmitt, 2019), probably because it is where the most prevalent collocational errors seem to occur (Chang et al., 2008). Nevertheless, other types also have received some attention. Walker (2011) exemplifies an investigation into the “adjective + noun” type of collocation. Through examining some of the top adjectival collocates (e.g., *long*, *lengthy*, *slow*, *difficult*, *painful*, etc.) of a node (i.e., *process*) in the domain of Business English, Walker demonstrates that *process* generally has a negative semantic prosody, which in turn could be passed on to L2 learners as a valuable piece of information for their vocabulary building.

On the basis of such predetermined syntactic patterns, researchers further classify the word pairs by using two other semantic criteria: transparency and substitutability. Transparency refers to a) whether the components have a literal meaning, and/or b) whether the meaning of the entire unit can be deduced from the

knowledge of each component. Substitutability means the degree to which a component can be replaced by a synonym without affecting the idiomaticity of the whole. For example, Nesselhauf (2003) divides the “verb-noun” combinations into three subgroups as follows:

Table 4 Subcategories of V-N Combinations (revised from Nesselhauf 2003)

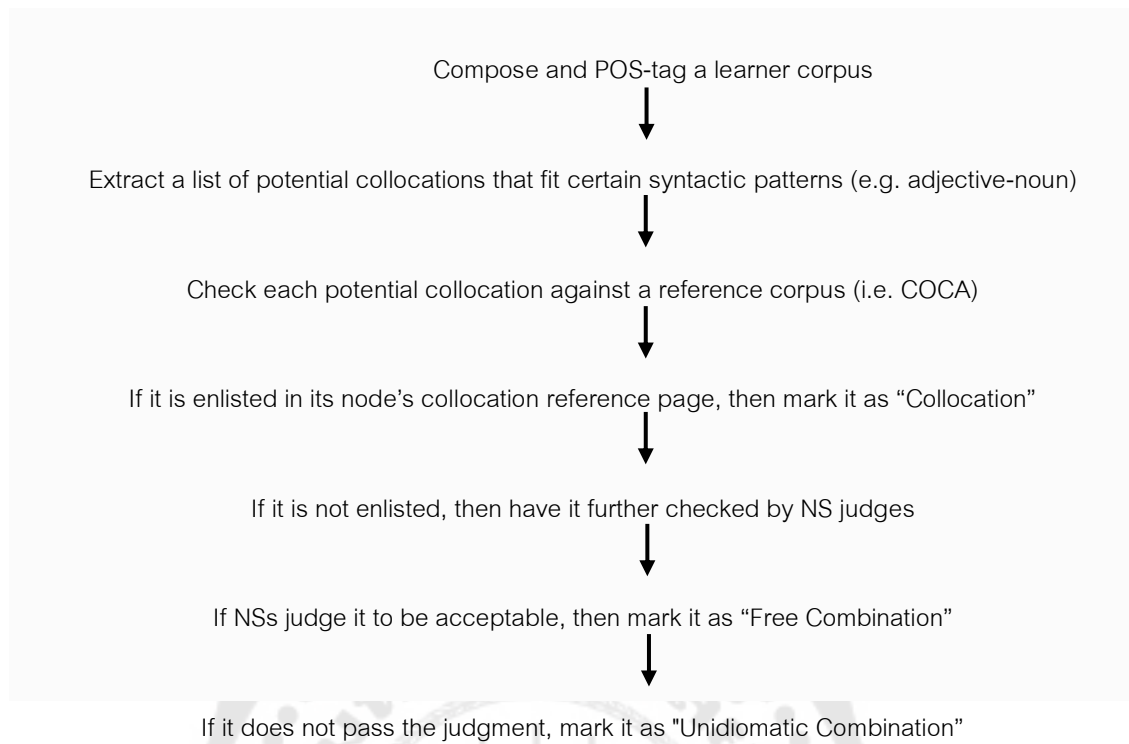
	Transparency	Substitutability	Examples
<i>free combinations</i>	Both components are used in their literal sense; the meaning of the whole can be deduced from knowledge of its parts.	Both components can be quite freely substituted	<i>want a car</i> <i>clean the room</i>
<i>collocations</i>	The noun has a literal meaning while the verb has a literal or non-literal meaning; the meaning of the whole may or may not be deduced from knowledge of its parts.	The verb can be freely substituted but the noun is restricted	<i>take a photo</i> <i>run the risk</i>
<i>idioms</i>	Both components are used in their non-literal sense; the meaning of the whole cannot be deduced from knowledge of its parts.	Neither the verb nor the noun can be freely substituted. The combination is lexically “fixed”.	<i>sweeten the pill</i> <i>pull one’s leg</i>

It should be noted that these criteria devised by Nesselhauf, admirable as they are, do not prove to be clear-cut in practice, thus rarely adopted in its exact form. In fact, some researchers do not even differentiate between the first two subgroups, and present them all together under the cover term *collocation* (McCarthy & O’Dell, 2017).

Phraseological approach has its face validity. After all, this is how language is traditionally understood: syntax governs how sentences are made up of constituents or phrases formed with rules, and semantics determines which words could habitually co-occur in accordance with their meaning relationships. As a result, a list of collocations rendered this way is likely to sound reasonable and relevant. However, as Stubbs (2002) points out, studies which only focus on “a small lexical sets” are subject to over-emphasizing idiosyncratic cases. In other words, if one only relies on his intuition and personal observation of a limited range of text, then there is a good chance that the collocations he compiles to be unrepresentative of the whole. Furthermore, while some cases are straightforward, others could be more ambiguous, thus arousing controversy among researchers. In essence, this approach eventually boils down to a matter of personal judgement.

### 2.2.3 A Mixed Approach

As shown above, frequency-based approach and phraseological approach are like two ends of the same stick. One emphasizes evidence while the other values intuition, and the optimal solution appears to be somewhere in between. As Stubbs (2002) suggests, “[t]he ideal would be to combine the best of both approaches, so as to make more precise quantitative generalizations about collocations across the whole of the vocabulary of a language.” Unsurprisingly, a growing number of researchers have already embarked on this mixed approach (Alangari, 2019; Peng, 2016; Vilkaite & Schmitt, 2019). The present study, in the same vein, adopts a mixed approach to define, identify, and evaluate collocations. As the details will be explained in the Methodology Section, here only a brief account is provided as follows.



Next, we shall turn our discussion to the three mysteries mentioned at the start.

### 2.3 Mystery #1: Contrasting Levels of Difficulty for NS and NNS

The first mystery, as briefly touched on in the opening section, is “*why are collocations so easy for native speakers yet so difficult for L2 learners?*” While numerous researchers have attested the challenge collocations pose to non-native speakers both in comprehension (Vilkaite & Schmitt, 2019) and production (Chang et al., 2008; Chen, 2019; Ferraro et al., 2014), the situation is drastically different for native speakers. As Wray (2002:1) puts it, native speakers “seem to find formulaic language an easy option.” In fact, prefabricated chunks are the first to manifest in a child’s utterances (Bolinger, 1975:100). Later, as one’s linguistic competence naturally develops to the fullest, a native speaker acquires a “feel” about what sounds right and what does not in her mother tongue.

In explaining this, one group of scholars belonging to the Chomskian school take a purely linguistic approach (also known as the *Internalized Approach*, or *I-Approach*). As Chomsky (1986) explains, the I-Approach makes use of one’s intuitions

about grammaticalness to probe how the grammar of any language is represented in a speaker's mind. He credits any human being with an innate *Language Acquisition Device* (LAD) which enables him/her to develop the linguistic competence to the fullest with limited stimuli of input. He also identifies a number of principles, some of which are believed to be parameterized, that comprise our *Universal Grammar* (UG). It is not my intention to discuss this theoretical stance in any detail, but only to point out that, despite its remarkable accomplishment in illuminating the underlying rules of language, Chomsky's theory does not seem to satisfactorily account for the subtle semantic differences between lexical items, which is the key to understanding the formation of prefabricated chunks, including collocations (Grace, 1995; Sinclair, 1991; Wray, 2002).

Another group of scholars adopt a usage-based stance. Their main arguments include: language is perceived to be a dynamic rather than static system; frequency plays a crucial role in language acquisition; one's linguistic representation is emergent in nature. Apparently, this approach puts great emphasis on the input one receives. For example, Ellis et al. (2008) draws on several psycholinguistic studies to show that the frequency of linguistic features is closely related to their acquisition by a learner. Langacker (2000) further explains that each occurrence of a form "leaves a trace in our brain", thereby facilitating its future re-occurrence. Some others make a distinction between word frequency and phrasal frequency, arguing that both contribute to the development of one's linguistic competence (Siyanova-Chanturia et al., 2011). According to Vilkaite and Schmitt (2019), usage-based theories are well capable of explaining the faster processing of formulaic sequences than nonformulaic ones. However, plausible as it sounds, the usage-based approach does not satisfactorily account for why L2 learners' interlanguage largely fails to match the probabilistic norm of the native speakers. Indeed, in the modern world where learners could easily access inexhaustible authentic English materials, such as texts, audio and video clips, one would assume that learners' idiosyncratic frequency profile would gradually get adjusted by the input and eventually conform to the norm. Nonetheless, as literature suggests, this rarely, if ever, happens.

A third explanation, and the most reasonable one in my eyes, rests on the notion of a *dual system*. Sinclair was the one who paved the ground by proposing his two principles of interpretation. As he states, “one is not enough. No single principle has been advanced which accounts for the evidence in a satisfactory way.” (Sinclair, 1991:109) Consequently, the open-choice principle sees texts as a series of slots to be filled from one’s lexicon, and the only restraint to govern the selection process is grammatical rules. This is in line with most traditional grammars. The idiom principle, opposing yet complementary to the first principle, acknowledges the existence of many semi-preconstructed phrases that constitute single choices to a user. Under this principle, these phrases are *not* decomposed into their lexical components, although they appear to be if viewed from the open-choice principle. Unlike the linguists of Chomskian school who treat such phrases as a trivial exception, Sinclair argues that they are at least as important as grammatical rules, and presents overwhelming evidence from corpus analysis. Later, Wray develops the idea a step further. She states in the concluding chapter of her book “Formulaic Language and the Lexicon” (2002:312) as follows:

*“Within a dual systems model, where language is processed both holistically and analytically, it is possible to attribute different types of linguistic knowledge independently to, on the one hand, how grammar works and, on the other, how language use determines patterns of distribution and frequency.”*

To elaborate on this, she puts forward two hypotheses from the point of first and second language acquisition respectively. First, native speakers unconsciously apply a *needs-only* approach in absorbing input. That means, by default, they take in word strings holistically and never break them down, unless there is a good reason to do so. This results in a lexicon consisting of items of various sizes, including morphemes, polymorphemic words, collocations, sentences, and even entire texts. The nature of such a lexicon is dynamic, and thus these items could have a change of status, from a fixed unit to a free, rule-based composition and vice versa, if certain conditions are met.



Second, non-native speakers, especially those who start learning the L2 after childhood, tend to adopt an analytic approach. Their inclination is bolstered both by the traditional fashion of tuition and by their own literate mindset. The result is that L2 learners often strive to memorize individual words while ignoring the truly valuable information, that is, what these words occur with. To sum up, native and non-native speakers approach language acquisition in fundamentally different ways: while the former follow a top-down path, taking entire strings as their starting point and never breaking them down more than necessary, the latter follow a bottom-up route, starting with the basic units and trying to build them up. No wonder they get lost on the way! As Foster (2001:90) observes, non-native speakers are often caught between a rock and a hard place by trying to “construct a great proportion of their language from rules rather than from lexicalized routines.” The following two figures adapted from Wray’s book show different mechanisms of L1 acquisition and L2 acquisition, respectively.

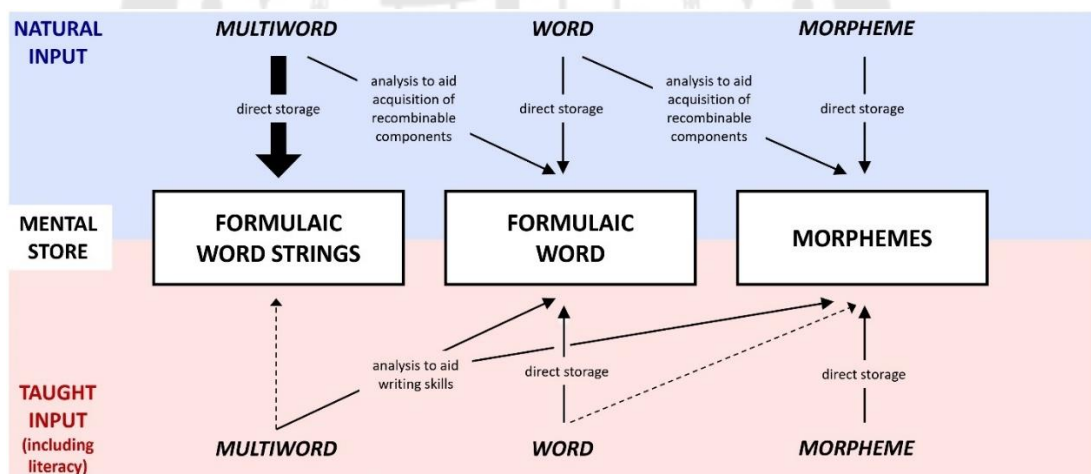


Figure 2 Top-down Mode of Language Acquisition (adapted from Wray 2002)

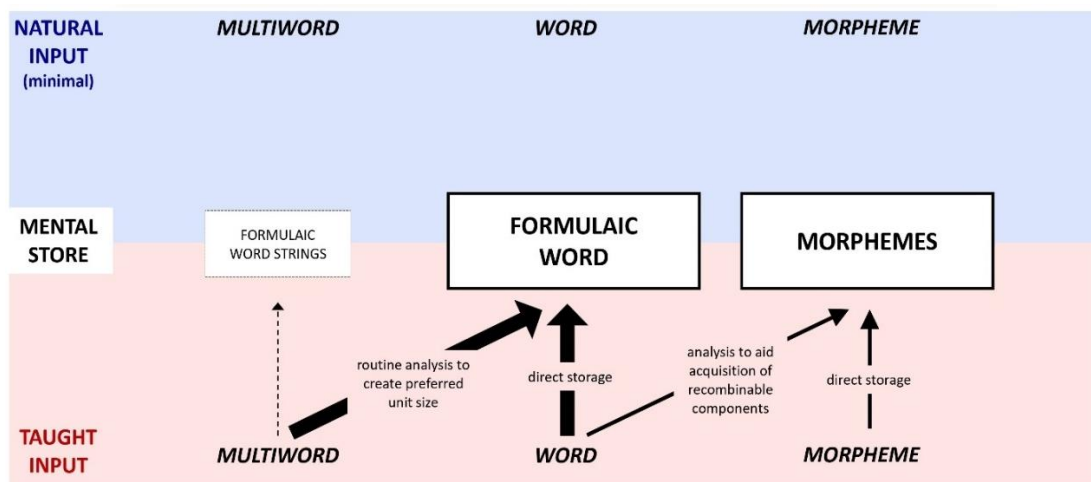


Figure 3 Bottom-up Mode of Language Acquisition (adapted from Wray 2002)

#### 2.4 Mystery #2: How to Account for NNS' Collocational Competence

While the first mystery is related to the schematic representations of collocation for native and non-native speakers, the second mystery, “*how to adequately account for learners' collocational competence*”, shifts our focus to *collocational competence* of non-native speakers only. Although some scholars explicitly mention this term in their work (Alangari, 2019; Ferraro et al., 2013; Peng, 2016), many more fail to make a direct reference to it (Biber, 2009; Chen, 2019; Vilkaite & Schmitt, 2019; Walker, 2011). Furthermore, hardly anyone bothers to provide a clear definition of it, despite the fact that their studies are centered on this concept. This is no trivial matter, since only when *collocational competence* is thoroughly understood, can we reliably measure it and effectively develop it. To start with, we shall examine the broader term, *competence*.

Ellis, in his encyclopedic book “The Study of Second Language Acquisition” (1996), presents an insightful discussion on this term. According to this account, there are two groups of scholars who hold different views. Chomsky (1965) considers competence as the mental representations of linguistic rules which constitute one's internal grammar. In this case, it is altogether abstract and implicit in nature. Others (Ellis, 1990; Hymes, 1971; Tarone, 1990; Taylor, 1988; Widdowson, 1983) see competence as one's *ability* to use the knowledge in specific contexts. In fact, scholars

who belong to this camp sometimes use other terms to make their points clear, such as *capability* (Tarone, 1990), *proficiency* (Ellis, 1990), and *capacity* (Widdowson, 1983). In this way, competence becomes more closely intertwined with one's actual performance, with an expanded scope of including explicit knowledge as well. As mine is a pedagogically-oriented study, I decided to side with the latter group to focus on the ability rather than the knowledge of L2 learners. On this basis, a tentative definition of *collocational competence* is proposed as follows:

*Collocational competence is one's actual ability to use his or her collocational knowledge to establish effective communication in specific contexts.*

With such a working definition, the next step is to explore how one's collocational competence can be reliably measured. Inherently, there are several decisions to be made.

#### 2.4.1 Comprehension vs. Production

Learners, like any users of language, apply their linguistic knowledge in two kinds of tasks, namely, receptive (i.e., comprehension) and productive (i.e., production). Therefore, measuring their collocational competence must entail evaluating their collocational performance in comprehension and/or production. However, only a handful researchers have devoted themselves to the former (Siyanova & Schmitt, 2008; Vilkaite & Schmitt, 2019) and their focus is largely narrowed down to the aspect of learner's processing speed. This is understandable on three grounds. First, as Nesselhauf (2003) points out, due to their transparent nature, comprehension of collocations is quite unproblematic for learners. This means that there is not much collocational to measure in comprehension: a learner either understands a collocation with sufficient word knowledge, or fails to understand it due to his lack of knowledge of its constituents. Second, even if we can test one's comprehension with regards to collocational competence, "it is very difficult to assign the cause of failures." (Corder, 1974:125) Third, some researchers (e.g., Read, 2000; Nation, 2007) posit that the best way to investigate the lexicon is through its use, especially when participants are not aware that their lexicon is being assessed. As a result, the outlook of such comprehension-based

studies remains gloomy with dampened enthusiasm from the scholars. Therefore, in accordance with the mainstream of literature, I will concentrate on evaluating learners' performance in production.

#### 2.4.2 Implicit vs. Explicit

Some SLA scholars make a distinction between two kinds of linguistic knowledge: *implicit* and *explicit*. This notion is closely associated with Krashen's work "Principles and Practice in Second Language Acquisition" (1982). In this book, he proposes his well-known *Monitor Theory* which consists of five hypotheses. In elaboration of the first one, the acquisition-learning distinction hypothesis, Krashen (1982:10) states that adult L2 learners have two contrasting ways (i.e. acquisition and learning) of developing their competence, each of which leads to a different kind of knowledge. He further argues in the Monitor hypothesis (the third of the five) that implicit knowledge gained from acquisition plays the central role in one's L2 development, while explicit knowledge from learning is peripheral with the sole function of monitoring one's output. However, Ellis (1996:349) argues that explicit knowledge could also aid one's comprehension in enabling one to notice and understand an input, as the following figure (adapted from Ellis, 1996:349) illustrates:

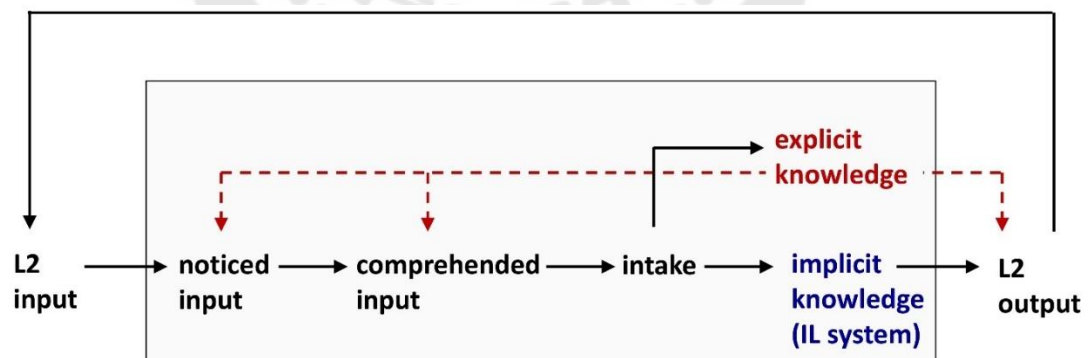


Figure 4 Learner's Comprehension Process (adapted from Ellis 1996)

The question then becomes: does measuring one's collocational competence entail assessing his/her ability to use the implicit knowledge, or explicit knowledge, or both? The author has taken "both" as the answer for this study. The rationale here is two-fold. First, according to Ellis et al. (2009), there is wide consensus from the scholars that one's L2 performance, which is closely related to his competence, utilizes a combination of both types of knowledge. While implicit knowledge undoubtedly makes up the backbone of one's competence, explicit knowledge also has its indispensable roles to play, such as monitoring production and facilitating comprehension (Ellis, 1996). Second, from the perspective of position on interface, unlike Krashen who takes a non-interface position in stating that implicit knowledge and explicit knowledge are totally separate, I incline to side with Bialystok (1978) that each type "leaks", and thereby explicit knowledge can become implicit knowledge and vice versa. This means what one consciously learns can eventually be integrated into the subconscious knowledge system and made available for online processing. In other words, although these two types of knowledge are distinctly different in nature, there exist "channels", the opening of which is based on certain conditions, to afford the possibility for one type to be converted into the other. In this case, explicit knowledge is also something worthy to assess.

#### **2.4.3 A Fuller Account of Learner Language**

As most SLA researchers choose to concentrate on production, learner language (i.e. the language a learner produces orally or in writing under various conditions) becomes the central piece of information and calls for careful handling. Unfortunately, most studies just stick to error analysis (EA) by exclusively focusing on collocational errors. That is, they strive to identify all the erroneous collocations in a learner corpus which fail to meet certain criteria (e.g. their frequency and associate strength being below a predetermined threshold in some reference corpus). For example, Liu (2002) examined the essays included in the English Taiwan Learner Corpus and identified 265 lexical miscollocations, the majority of which (88%) are verb-noun type. Nesselhauf (2003) analyzed 32 essays written by advanced learners whose

L1 is German, and found 65 miscollocations. Chen (2016) investigated the effects of dictionary use in learners' collocation production by their scores of a pretest and a posttest, which, in turn, depends on how many collocations they got "correct". All these collocation studies exemplify the general trend of taking errors as the sole focus.

However, as Schachter (1974) points out, EA falls short in providing a complete picture of learner language. To an SLA researcher, what a learner does right is as informative as what he does wrong. Furthermore, it is also worthwhile to investigate whether one's performance faithfully represents his scope of knowledge (i.e. whether strategies such as avoidance are at work). On this ground, this study adopts the analytical framework of Ellis (1996:302-306) in profiling learner language. This framework consists of four aspects, namely, errors, facilitation, underuse, and overuse. A brief illustration is provided as follows:

- **errors:** any collocation that fails to meet certain criteria, including overt errors (i.e. those that sound unidiomatic, such as *quick track* → *fast track*) and covert errors (i.e. those that are perfect in form but inappropriate in context, such as *After dropping her husband at the station, Susan got her way home.* → ... *made her way...*).
- **facilitation:** any collocation that 1) is judged to be correct based on the criteria, and 2) has an established equivalent form in learner's L1 through direct translation. Examples include 骑 (*ride*) 自行车 (*bike*), 缓解 (*alleviate*) 疼痛 (*pain*), 美丽的 (*beautiful*) 夕阳 (*sunset*), etc. This list could give learners a head start in learning collocations, and shed light on how languages shape each other as they interact in this modern age.
- **underuse:** learners are found to struggle with some parts of their knowledge which they are not confident about. When it comes to collocation, this can be underuse of certain syntactic types (e.g. adverb + verb), of some specific lexical items (e.g. delexicalized words), or of those that are figurative in nature (e.g. *a heated conversation*).

- **overuse:** on the other hand, learners could also make excessive use of certain parts of their collocational knowledge, and investigation into this will shed light on their competence from a different angle. Similar to underuse, overuse is also statistical in nature and takes the probabilistic norm of native speakers as the baseline for comparison.

To sum up, this study aims to give a fuller account of learner language by incorporating a number of other aspects in addition to errors. In this way, we should be able to assess learners' collocational competence in a more comprehensive, reliable, and enlightening manner.

### 2.5 Mystery #3: How to Effectively Develop NNS' Collocational Competence

The third mystery is pedagogical in nature: *How can L2 learners effectively develop their collocational competence?* On the one hand, the answer partly stems from the discussion on the first two mysteries. As already mentioned, Wray's (2002) dual model for first and second language acquisition seem to best account for an SLA practitioner's general observation. While native speakers process word strings holistically and break them down only when necessary, non-native speakers (i.e. L2 learners) routinely start from the most basic units and build them up according to whatever rules in their interlanguage. This inevitably grants them too much freedom, which often leads to unidiomatic expressions. Therefore, one way to promote the development of learners' collocational competence is to have their schema fundamentally modified. This could be achieved through awareness raising (Chen, 2016), focus shifting and so forth. Of course, any particular practice should be grounded in the comprehensive, fine-grained analysis of learner language (Nesselhauf 2003), as shown in the discussion to the second mystery. Furthermore, two particular treatments, namely, extensive reading (ER) and explicit instruction, will be discussed in detail here.

### 2.5.1 Extensive Reading

Extensive reading (ER) is a form of implicit learning to develop a learner's linguistic competence (Takayuki, 2014). According to Reber (1967), implicit learning is a process during which learners unconsciously acquire the knowledge present in the environment while focusing on something else. As they indulge themselves in reading materials which are easy, interesting and sufficient in length, learners are expected to naturally make progress in terms of fluency and comprehension. According to Day & Bamford (2002), there are ten principles for an ER program:

- 1) The reading material is easy.
- 2) A variety of reading material on a wide range of topics is available.
- 3) Learners choose what they want to read.
- 4) Learners read as much as possible.
- 5) The purpose of reading is usually pleasure, information and general understanding.
- 6) Reading is its own reward.
- 7) Reading speed is usually faster rather than slower.
- 8) Reading is individual and silent.
- 9) Teachers orient and guide their students.
- 10) The teacher is a role model of a reader.

The ER program adopted in this study follows the majority of these principles except the fourth, the sixth and the tenth, and thus could be considered as modified ER on the continuum (Day, 2015). The reasons of not following these three principles will be explained in section 3.4.2. ER's importance in language learning has been widely attested. For example, Vu and Peters (2022) point out that L2 learners, especially in EFL context, need extensive exposure to meaningful input for their collocational knowledge to effectively develop. Vilkaite-Lozdiene & Schmitt (2019) also calls for promotion of increased exposure through ER for collocation development. Ellis (2008) suggests a duty of the teachers to create opportunities such as ER programs for learners to obtain input outside the classroom, which is in line with Krashen's input



hypothesis. Accordingly, this study chooses ER as an independent variable to investigate its effect on learner's collocational competence development.

### 2.5.2 Explicit Instruction

Although developing implicit knowledge should be the main focus of any pedagogy, explicit knowledge is also worth the effort to teach (Ellis, 2008). The reasons include the value of the explicit knowledge in itself (Kormos, 1999) and its facilitative effect in developing implicit knowledge (DeKeyser, 1998; Ellis, 1993). As explicit knowledge is mostly from explicit learning/teaching, this study will incorporate the latter as another independent variable. Contrary to implicit learning, explicit learning takes place when analysis on the metalinguistic level is involved in the learning process (Zhang & Li, 2016). There are some researchers who have already confirmed the positive effect of explicit learning/teaching on the development of collocational competence (e.g., Hsu, 2002; Pirmoradian & Tabatabaei, 2012). El-Dakhs (2015) thus highly recommends training learners on the identification, memorization and retrieval of collocations. Therefore, this study explicitly trains the participants through a seven-week program on various aspects of collocation, such as awareness raising, characteristics, and learning strategies.

Finally, there is still the need to draw on some established SLA pedagogical theories to arrive at a systematic approach. We shall examine two of them here, with the conclusion about which one is the better fit for the purpose of this study. The first one is termed as "*focus on form*" by Long (1991). Unlike traditional instruction which tends to isolate linguistic features and teach them one at a time, focus on form involves "alternating in some principled way between a focus on meaning and a focus on form" (Long, 1991). This approach lies midway between the traditional "focus-on-forms" and the newer trend "focus-on-meaning", requiring teachers to adopt a task-based syllabus, and direct learners' attention to the target linguistic features sporadically during their classroom activities. Long's proposal is attractive in the way that it could effectively integrate fluency and accuracy. Many experiments on this method have been carried out with mixed results (Carroll et al., 1992; Doughty, 1991).

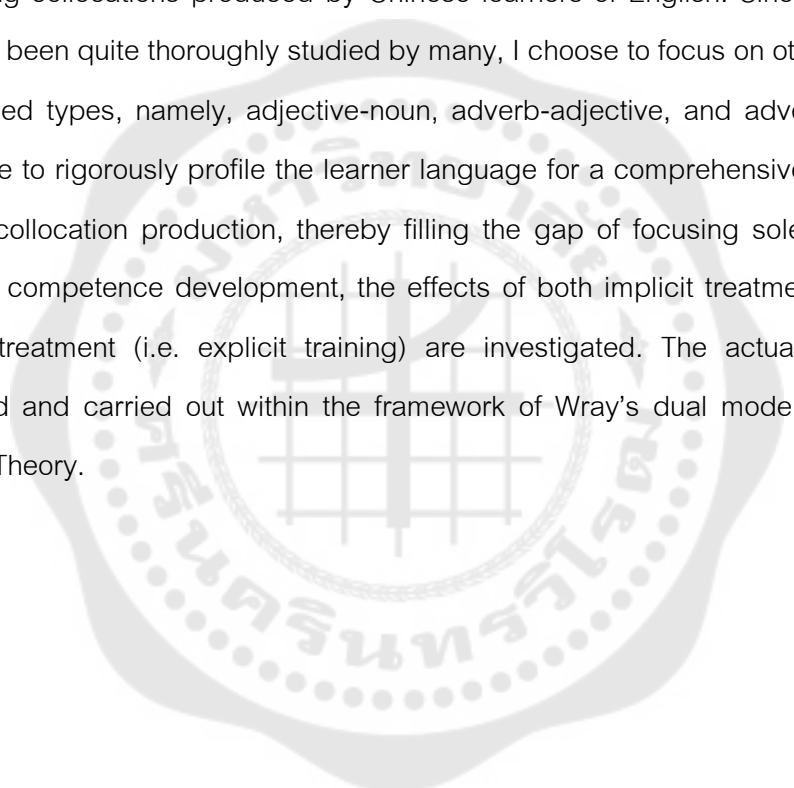
The second one is the aforementioned *Monitor Theory* by Krashen (1982). Of all its five hypotheses, two are particularly relevant to pedagogy: comprehensible input hypothesis and affective filter hypothesis. The former claims that a necessary but not sufficient condition for acquisition is input a little beyond the current level (i.e.  $i + 1$ ). On this basis, as Krashen (1982:161) puts it, the most important responsibility of a teacher is “to provide students with enough comprehensible input to bring their second language competence to the point where they can begin to ... read, and participate in conversations.” He further lists several criteria of “optimal” input: a) comprehensible; b) interesting; c) not grammatically sequenced; d) sufficient in quantity. The affective filter hypothesis states that affective variables play a mediating role in impeding or facilitating the delivery of input to one’s LAD. Krashen (1982:73) thus contends that teachers should avoid putting students “on the defensive” and make every effort to keep their affective filter as low as possible. In fact, he claims that any learning condition which satisfies these two requirements (i.e. optimal input combined with low affective filter) is sure to succeed.

As far as developing one’s collocational competence is concerned, *Monitor Theory* might be the more practical choice, although some modifications have to be made to it. Long’s focus-on-form stance, attractive as it sounds, requires teachers to design proper tasks featuring specific collocational patterns. This is no easy job, provided that such reference materials are scarce. It also demands teachers to adopt a principled way of shifting focus between meaning and form, which inevitably involves a lengthy trial-and-error phase by anyone novice to the concept. This simply seems unfeasible within the time frame of this study. *Monitor Theory*, on the other hand, requires teachers to collect sufficient input that is comprehensible to the students while creating a learning atmosphere favorable to them. These seem to be achievable for an experienced teacher. However, two additional important points are to be integrated into this approach: 1) Because this study aims to examine and develop learners’ implicit as well as explicit knowledge, whenever necessary, the teacher must intervene the activities and impart explicit knowledge by informing them of relevant patterns and rules;

2) As L2 learners are generally indulged with memorization of individual words but inexperienced to noticing collocations in input, the teacher needs to make an effort to raise their awareness of such prevalent multi-word structures.

## 2.6 Summary

To sum up, this study is centered on measuring and developing learners' collocational competence. Methodologically speaking, a mixed approach is adopted in identifying collocations produced by Chinese learners of English. Since the verb-noun type has been quite thoroughly studied by many, I choose to focus on other three under-researched types, namely, adjective-noun, adverb-adjective, and adverb-verb. Efforts are made to rigorously profile the learner language for a comprehensive understanding of their collocation production, thereby filling the gap of focusing solely on errors. In terms of competence development, the effects of both implicit treatment (i.e. ER) and explicit treatment (i.e. explicit training) are investigated. The actual experiment is designed and carried out within the framework of Wray's dual model and Krashen's Monitor Theory.



## CHAPTER 3

### RESEARCH METHODOLOGY

#### 3.1 Population

The target population for this research is Chinese intermediate learners of English, represented by undergraduates from a local college.

#### 3.2 Participants

The participants are four groups of English-major undergraduates who participated in a 7-week Extensive Reading course. The general background of the participants are follows:

Table 5 Participants' Demographics

Demographic Attribute	Value
Number of participants	84
Number of groups	4
Age Range	18-20
Major of Study	English
Proficiency Level	intermediate

#### 3.3 Research Design

Four classes were selected for this study. The qualified candidates (i.e. intermediate-level ones) of each class formed a corresponding experimental group, all together four groups (Table 6). The explicit treatment consisted of a 7-week series of in-class collocation training, and the implicit treatment was a weekly after-class extensive reading project. This study aims to find out how each type of treatment affected the development of learners' explicit knowledge and implicit knowledge, respectively. The overall design is shown in Figure 5 (next page).

Table 6 Groups and Treatments

	Experimental treatments		Normal class activities
	Explicit treatment	Implicit treatment	
Group (e+i)	+	+	+
Group (e)	+	-	+
Group (i)	-	+	+
Group (c)	-	-	+



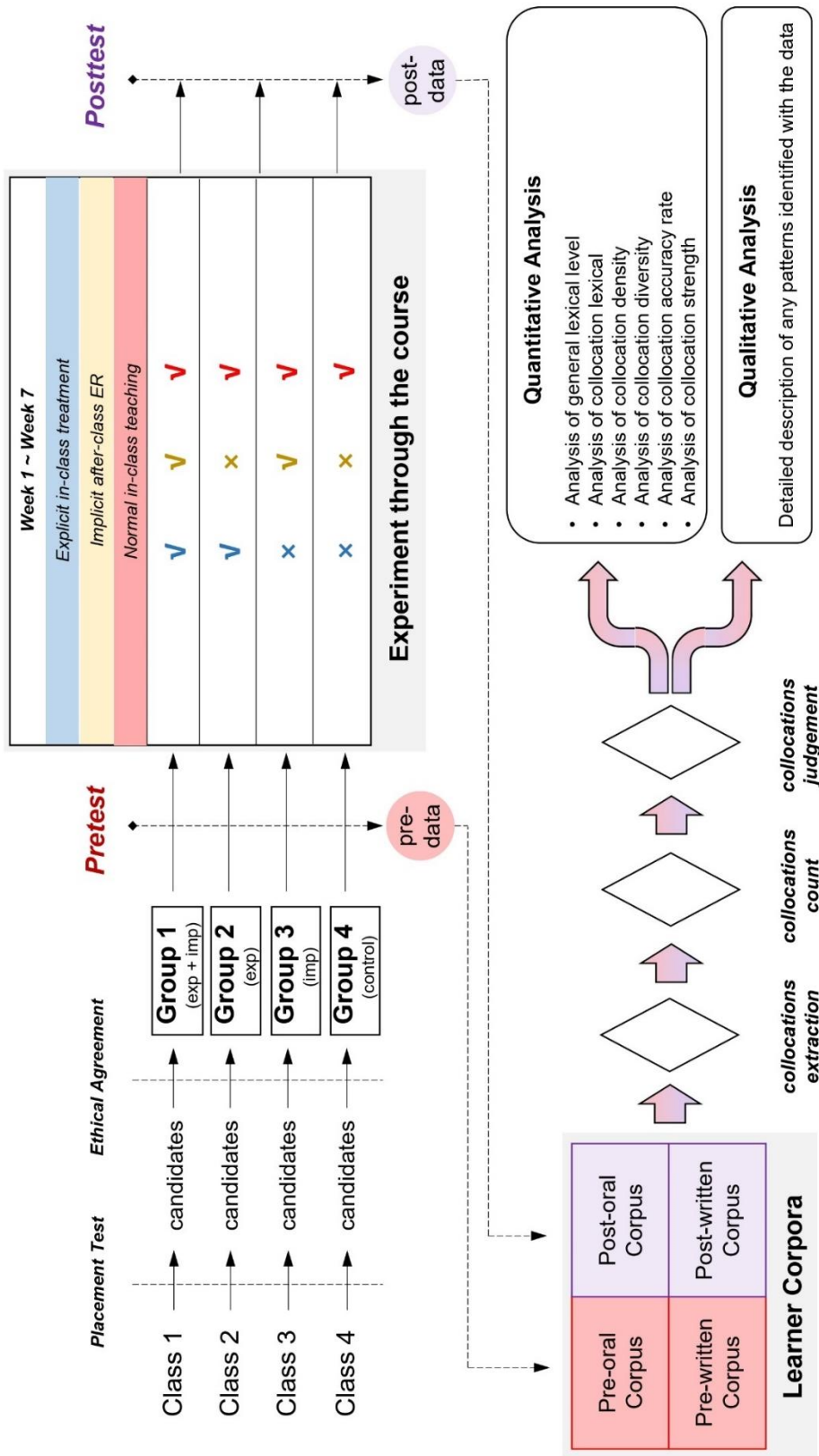


Figure 5 Overall Research Design (self-made)

Such a design, as balanced and reasonable as the researcher intends it to be, is unfortunately not bias-free. Since each group received a different treatment, the amount of exposure to English varied, with Group (e+i) having the most exposure and Group (c) the least. This factor alone, on top of the treatments themselves, could bring about some favorable effect on the experimental groups.

Data of spoken and written forms were collected before and after the experiment. The resultant four corpora, that is, pre-oral, pre-written, post-oral and post-written, went through a series of processing before becoming available for final qualitative and quantitative analysis. As spoken data are to be generated extemporaneously and written data are to be produced with sufficient time for preparation and revision, the former should reflect participants' implicit knowledge (Suzuki, 2017) while the latter their combination of implicit and explicit knowledge (Loewen, 2020). This gives us a comprehensive understanding of their collocational competence as well as how effective the treatments are regarding their competence development.

### 3.4 Research Instruments

#### 3.4.1 Placement Test

First, a placement test was carried out to pick out qualified candidates whose English proficiency is at intermediate level. According to CEFR, learners' proficiency can be divided into three levels: basic (A1/A2), intermediate (B1/B2), and advanced (C1/C2). In this study, the advanced group will not be examined because they have been quite thoroughly researched in the literature (e.g. Alangari, 2019; Chang et al., 2008; Nesselhauf, 2003). I also excluded the basic learners because their vocabulary and grammar are too limited to produce sufficient tokens of collocations. Therefore, the researcher only used a sample of intermediate learners for investigation.

The placement test of my choice is the online *Cambridge English Test – General English* (<https://www.cambridgeenglish.org/test-your-english/general-english/>). This test, designed by Cambridge University, is composed of 25 multiple-choice questions. As soon as a test taker finishes all the questions, the system will automatically inform him of

his proficiency level based on CEFR framework. The utilization of this standardized test has been deemed favorable due to its cost-effectiveness, reliability, and convenience for the participants. Additionally, the test encompasses the collocational aspect of one's linguistic competence, including *verb + noun* type ("order a pair of shoes", "unfasten one's seatbelt", "retrace one's steps") and *adverb + adjective* type ("highly reliable"). Thus, it is quite relevant to the scope of the present study. Logistically speaking, the researcher used the language labs on campus to have the groups finish the test within the time limit (15 minutes) under the monitoring of the researcher to ensure the reliability of the result. Below is a snapshot of the test interface:

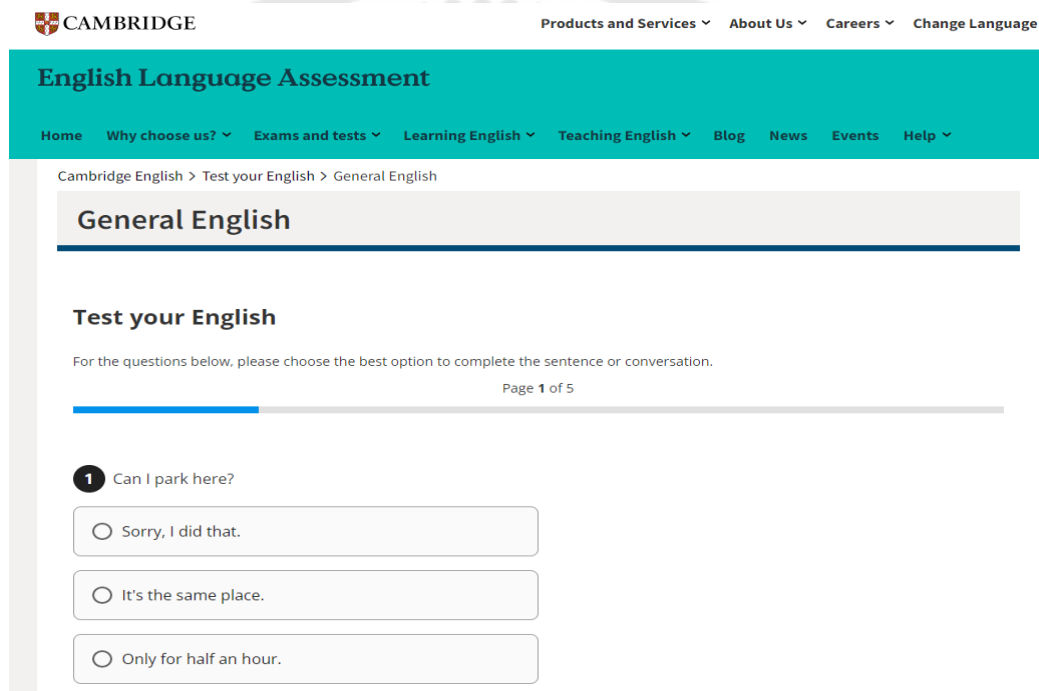


Figure 6 Cambridge Placement Test Interface (self-made)

After the test, all those who turn out to be at intermediate level were invited to participate in the experiment. They were informed of the following:

- Participation is totally voluntary;
- Participation (or not) does not affect their course grade in any way;
- Their data will be used in an anonymous fashion;
- They can withdraw at any point during the experiment;



- They will receive a certain amount of compensation upon full participation of the experiment.

All above are in accordance with Srinakharinwirot University's ethical requirements. The researcher recruited a minimum of 20 participants for each group (84 in total) in order to make the result statistically significant.

### 3.4.2 Extensive Reading Materials

According to Krashen's Monitor Theory, particularly the input hypothesis, the researcher requires the participants to finish a self-paced extensive reading task each week. Unlike intensive reading which is very focused and learning-oriented, extensive reading is less time-pressured and often done for pleasure (Harmer, 2015). In other words, extensive reading involves learners reading texts of their own choice and at their own pace, primarily for enjoyment rather than for language learning (Al-Homoud & Schmitt, 2009). According to Ender (2016), as no intention is paid to linguistic forms in the process, extensive reading belongs to incidental learning which seems to be especially important to enhance learners' implicit knowledge.

As previously mentioned in section 2.5.1, the ER program here upholds seven of the ten principles outlined by Day & Bamford (2002), including letting students choose their own books, referring them to a wide range of choices of books, the books being easy enough for their current level, etc. However, the other three principles were not strictly adhered to, namely, "reading as much as possible", "reading being its own reward", and "teacher being a role model". To begin with, the researcher would like to have some control over the amount of their reading time (i.e. 30-60 minutes per sitting, three sittings a week) to make this variable quantifiable to a certain extent. Furthermore, according to some scholars (Hill, 2013; Harmer, 2015), reading as the sole reward may not be sufficient for effective incidental learning. Therefore, other activities are deemed beneficial to supplement the reading itself. For example, Lyutaya (2011) integrated writing tasks by adopting a reading log for optimum outcomes. Lastly, the researcher serves a guide rather than a role model to the students through the program, in accordance with many previous studies (see Day, 2015 for a review).

As a result, before the course commences, each participant selected an English book of his or her own choice, which must be 1) interesting to the reader (in Krashen's word, the content needs to be "compelling"), 2) appropriate in difficulty (i.e.  $i+1$ ), and 3) sufficient in length ( $\geq 400$  pages). The last requirement is not only meant for providing ample input, but also for maximizing the opportunity for the learners to encounter certain collocations repeatedly so that they will be properly acquired (Loewen, 2020). Furthermore, participants were trained to form a habit of reading the book on a regular basis, with specific guidelines on the frequency, duration and number of pages to be read per week. This is to ensure that participants would receive ample comprehensible input for acquisition to take place, as Krashen claims.

In practice, as the participants are expected to have little prior English book selecting experiences, the researcher referred them to some accessible resources, including the on-campus library, e-books online and so on. Furthermore, they were given one week to try out whatever book they find to check if it is appropriate. That means they would need to read the first five pages of the book without the help of a dictionary to have a feel about whether it is sufficiently interesting and how much they can comprehend. They are required to turn in a try-out form to inform the researcher of the book they choose as well as the percentage of their word-recognition rate. According to Nation (2001), one should have a minimum of 95% lexical recognition in order to gain an adequate comprehension of a text. This means that for a reader to effectively guess out the meanings of unknown words from the context and reach a solid understanding of the text at large, such unknown words should account for no more than 5% of the words in total. However, in the process of this experiment, this percentage was given heed to but not taken literally due to several reasons. On the one hand, for these learners whose proficiency is at intermediate level, 95% of familiar word coverage effectively narrows the choices down to children's book, including graded reader series. Although such materials meet the requirement of being easy enough, they largely fail to arouse the students' interest, which is another important requirement. As some of the subjects suggested that such pre-K volumes sound irrelevant or even

insulting to a college student. On the other hand, if they could be given the freedom to choose from a little bit more challenging books (which constitute a much wider range of choices), there is a better chance for them to maintain the reading interest with encouragement and guidance from the researcher. As a result, I have decided to urge them to make every effort to find a book of which they know 90% of the words, and at the same time set the minimum percentage to 70% so that every student will have a reasonable chance to find a book of his or her choice of interest. Of course, if their lexical knowledge did not pass the threshold of 70%, they were required to change the book for a simpler one. The try-out form is as below:

### **The Book Selection and Try-Out Form**

Find your own book(s), which can be anything that is 1) in English; 2) interesting to you; 3) about 400 pages long; 4) largely understandable (>70%) without the help of a dictionary.

You can find it from the library, borrow it from a friend, purchase it from a bookstore, etc.

Please get it by (date X) and fill out the form below for the teacher to check:

**Student Name:**

**Student No.:**

Book title:			
Author:			
Publisher:			
Total pages:			
City of Publication:			
Year of publication:			
After reading the first 5 pages, I can understand ____ of the words in the text:			
<70%	70%-80%	80%-90%	>90%
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 7 Book Selection and Try-Out Form (self-made)

Throughout the 7-week period of the course, they read the book of their choice three sittings per week, 30-60 minutes per sitting. Besides the primary purpose of input provision, this practice is also expected to help them form a lifelong habit of extensive reading.

### 3.4.3 After-class Reading Logs

While interesting content is a vital factor in extensive reading, some studies have stated that learners need more incentive than mere pleasure to make the reading experience most fruitful (e.g. Hill, 2013). As Harmer (2015) suggests, it might be beneficial to ask the learners to keep a record of what they have read, just like a reading diary. Such tasks are crucial in fostering students' ability of autonomous learning, which is really the ultimate goal of any pedagogy. Consequently, along with the extensive reading task, the participants are required to fill up and submit a reading log every week. This is to ensure their genuine participation and cultivate their taste for the pleasure of reading. Additionally, participants would have a sense of ownership of their own development. In the log format, there is a box to note down what they enjoyed most in the current week's reading. The researcher guided them with the following requirements before the first week begins:

- Read regularly (3 sittings per week, 30-60 minutes per sitting)
- Enjoy the content without worrying about linguistic details
- Get rid of all distractions while reading
- Refrain from checking a dictionary frequently
- Maintain a steady pace while reading and minimize scanning back and forth

Below is the reading log format to be handed out to them weekly:

### Reading Log: Week \_\_\_\_\_

Book Title:

Author:

Total pages:

Finished pages:

Sittings	Date	Duration (min)	Place of reading	Pages read
1				
2				
3				
What I enjoyed most from reading this week (about 100 words):				

**Your Signature:****Class:**

Figure 8 After-class Reading Log (self-made)

#### 3.4.4 In-class Collocation Trainings

Many scholars (e.g. Ur, 2011; Long, 2017; Krashen, 2003) have reached the consensus that implicit knowledge is one's primary resource for spontaneous communication (i.e. *online* communication) while explicit knowledge is effortful to draw and thus not instantly available (Ellis, 2009). What many also agree is that explicit learning often results in explicit knowledge, and implicit learning (such as extensive reading) brings about implicit knowledge (Ellis, 2007; Krashen, 2003; Rebuschat, 2013). However, there is some evidence showing that explicit teaching will also enhance the students' implicit knowledge. For example, Spada and Tomita (2010) investigated the effects of explicit and implicit instruction on learners. They designed two kinds of tasks

(i.e. controlled task and free-constructed task) to measure learners' explicit knowledge and implicit knowledge, respectively. The result shows that explicit teaching seems to help in promoting both types of knowledge. Such a conclusion is borne out by other studies (e.g. Ellis et al., 2006). On such a basis, the researcher sides with Ellis (2007) that it may be optimal to support implicit learning (i.e. extensive reading) with explicit instruction (i.e. in-class collocation trainings).

This series of collocation trainings are incorporated into a reading course which consists of seven 90-minute weekly sessions. In each session, besides the normal course activities, the researcher took 30 to 40 minutes of the class time to conduct collocation-related trainings (only to Group (e+i) and Group (e) as explained earlier) to accomplish the following goals:

- to modify the participants' schema of L2 from bottom-up to top-down mode
- to raise their awareness of existence, importance, and general patterns of collocations, including:
  - arbitrary convention
  - a warning of L1 interfering effect
  - the trap of synonyms with regard to collocations
  - the rule of rhythm with regard to collocation formation (e.g. alliteration)
- to inform them of the procedure and resources of how to acquire collocations

The details of the training plan are shown in Table 7 (starting from next page).

Table 7 In-Class Collocation Training Plan

Wk	Theme	Activities	Time (min)	Notes	Theoretical Basis
1	Resource & Reading Strategy Introduction	<ul style="list-style-type: none"> <li>• Lead-in on the importance of collocation</li> <li>• Introduction of resources related to collocation learning: dictionaries, corpus, extensive reading websites, etc.</li> </ul>	40	This is to help them get on the path of self-learning and promote deeper processing of the content.	Learner autonomy being the ultimate goal for any language pedagogy (Harmer, 2015)
2	Schema Modification	<ul style="list-style-type: none"> <li>• “Chinglish” examples analysis</li> <li>• Three widespread myths about L2 learning</li> <li>• Two contrastive formulae for English learning</li> </ul>	40	Examples must be based on authentic data, preferably participants’ own utterances.	The Dual Model of Language Acquisition (Wray, 2002)
3	Collocation Awareness Raising – Pervasiveness	<ul style="list-style-type: none"> <li>• Lead-in with an excerpt from the textbook for collocation counting practice</li> <li>• Three benefits of collocation learning</li> <li>• Phraseological types of collocation</li> </ul>	40	Ask participants to take notice of collocation in their reading and record at least one example for each type in the log.	Noticing Hypothesis (Schmidt, 1990) Lexical dimension in language teaching (Timmis, 2008)

Table 7 (Continue)

Wk	Theme	Activities	Time (min)	Notes	Theoretical Basis
4	Collocation Awareness Raising – Arbitrary Convention	<ul style="list-style-type: none"> <li>Lead-in with some common examples to show the mismatch in collocation between Chinese and English</li> <li>Problems with the “mental translation” learning mode</li> <li>Emphasis of the importance of thinking in English</li> </ul>	40	Issue the participants a warning of potential L1 interference.	Collocation formation is based on the arbitrary convention of a language (Nesselhauf, 2003)
5	Collocation Awareness Raising – The Trap of Synonyms	<ul style="list-style-type: none"> <li>Myth of “free replacement” between synonyms</li> <li>Some examples for practice</li> <li>Explanation of certain patterns such as alliteration</li> </ul>	40	Inform participants of the true meaning of mastering a word: to know how to use it in contexts.	Synonyms are not freely interchangeable (Zaabalawi & Gould, 2017)



Table 7 (Continue)

Wk	Theme	Activities	Time (min)	Notes	Theoretical Basis
6	Collocation Awareness Raising – The Rule of Rhythm	<ul style="list-style-type: none"> <li>• Lead-in with some examples to introduce the concept of rhythm</li> <li>• Illustration of how words attract each other based on rhythm to form collocations</li> <li>• Practice of identifying rhythmic collocations in the textbook</li> </ul>	40	Explain how rhythm makes language sound beautiful.	Lexical Approach (Willis, 1990)  The mnemonic effect of alliteration (Boers & Lindstromberg, 2005)
7	Collocation Learning Demo	<ul style="list-style-type: none"> <li>• Lead-in with a short passage with excessive use of “very”</li> <li>• Presentation of 24 common intensifying adverbs in English</li> <li>• Demo of how to differentiate between <i>highly</i>, <i>deeply</i>, and <i>absolutely</i> through collocation study</li> </ul>	40	Participants' awareness should be further established concerning the uniqueness of each word in a language.	Involvement Load Hypothesis (Laufer & Hulstijn, 2001)

The researcher mainly consulted Cambridge's textbook "English Collocations in Use - Intermediate" (McCarthy & O'Dell, 2017) to get examples, ideas and teaching strategies. For example, in Chapter 10 and 11, examples of synonymous collocates were referred to when training participants about the "trap" of synonyms in collocation use. Besides, the Extensive Reading Foundation's website ([www.erfoundation.org](http://www.erfoundation.org)) is also consulted for ideas and principles.

The details of each training session will be further elaborated in the Appendix A. In general, through this series of explicit training, it is expected that participants' explicit knowledge of collocation and L2 learning in large would increase, which, as time passes by, should facilitate the increase of their implicit knowledge as well.

### 3.5 Data Collection

Two rounds of data were collected. The pretest was carried out before the experiment, and the posttest right after it. As stated earlier, their production rather than comprehension was examined.

Each test consists of two parts: speaking and writing. For speaking, each participant was interviewed individually and asked to express his/her thoughts on a common topic, such as:

- *Why do you choose English as your major?*
- *What is your most unforgettable experience this semester?*
- *Could you tell me something about your hometown?*
- *How do you like this city?*

The participants are advised to choose any topic about which they have plenty things to say. They are given about ten minutes to prepare and then asked to give a free talk on the topic for about 3 minutes. The interviewer (researcher) did not interrupt unless there occurs a long pause indicating that the participant has little else to say. In that case, another open question would be given to elicit more utterances. Such speaking data should shed light on their implicit collocational knowledge, since they are

giving an online performance with virtually no time to monitor or reconsider what they are uttering.

Each interview was conducted on a one-on-one basis to put the participants at ease, and their speaking was recorded by a voice recorder, which automatically transcribes the audio files into texts. Then the researcher manually checked the transcription to ensure accuracy. Later, all these txt files were combined to form a corpus for further analysis. Some specifics about the oral part are as follows:

- Format: 1-on-1 interview
- Topic: *Why do you choose English as your major?* (or other similar common topics)
- Duration: 3-4 minutes
- No. of participants: 84
- Total tokens (expected): 13,000 – 20,000

For the written part, their performance was offline in nature. Both topics are common and relevant to their English learning in general. The topic for the pretest is “*The Significance of Reading*” and the one for the posttest is “*How to Learn a Language*”. Participants are given 4 days to compose an essay of no less than 200 words on the specific topic. Their writing should help us probe into their implicit as well as explicit collocational knowledge, as they have ample time to conceive, compose and revise their works. Some specifics about the written part are as follows:

- Format: homework essay
- Topic: *the Significance of Reading* (pretest) / *How to learn a language* (posttest)
- Length: 200-250 words
- Time constraint: to be submitted within 4 days
- No. of participants: 84
- Total tokens (expected): 17,000 – 22,000 (from each test)

### 3.6 Procedure of the Experiment

As the earlier Figure 5 shows, the whole procedure can be divided into four stages. In stage 1, also known as the preliminary stage, the researcher randomly selected four classes of students and had them take the Cambridge Placement Test under monitoring. Based on the result, whoever falls into the category of intermediate level will be potential candidates. These candidates had a brief gathering, during which the researcher informed them of the nature of the experiment and relevant benefits, and anyone who understands and volunteers to participate in the experiment will sign an ethics form. Through this process, four groups were established: Group (e+i) would receive both explicit treatment in class and implicit treatment after class; Group (e) would receive explicit treatment only; Group (i) would receive implicit treatment only; Group (c), which is the control group, would receive no special treatment but only participate in the normal course activities.

In stage 2, also known as the experiment stage, all four groups first took their pretest. In the pretest, each participant was given four days to compose a 200-250 word essay with the title of "The Significance of Reading", and each of them took a one-on-one interview with the researcher, during which they are to give a 3-minute free talk on some common topics. These data composed the first two corpora: NNS\_Pre\_Oral corpus and NNS\_Pre\_Written corpus, each of which consists of four sub-corpora corresponding to the four groups.

During the week before the course commences (week 0), Group (e+i) and Group (i) were assigned a weekly self-reading task which is to be performed throughout the course. Furthermore, they are required to turn in a weekly reading log to the researcher for checking. Then as the 7-week extensive reading course starts, Group (e+i) would receive an explicit training in class each week on the topic of collocation learning, as explained in detail in 3.4.4. Besides, they will also finish the weekly self-reading task. Group (e) would receive the same explicit training but perform no self-reading task. Group (i) would do the weekly self-reading task but receive no explicit training. Group (c) (i.e. the control group) receives neither treatment. All four groups

underwent the normal course activities. For details, please see Lesson Plan Tables starting from next page for each of the groups.

After the course is completed, all four groups took the posttest, which is similar in form to the pretest, only with different topics. The data from the posttest constituted the other two corpora: NNS\_Oral corpus and NNS\_Written corpus.

In stage 3, also known as the data processing stage, all the corpora data were first manually checked and cleaned by the researcher, and then imported into AntConc software (Anthony, 2022) to extract collocation tokens. During this process, the information of collocation counts according to each category will also be obtained. After that comes the step of collocation judgement: each collocation was examined with reference to COCA and, if necessary, passed on to NS judges for a final call. Those tokens passing the check were marked as “idiomatic”, those which do not were marked “unidiomatic”.

In stage 4, also known as the data analysis stage, two kinds of analysis, qualitative and quantitative, were performed, both for collocational competence profiling and for collocational competence development. This will be explained in section 3.7.

Table 8 Teaching Plan for Group (e+i)

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
<b>Goals</b>	<ul style="list-style-type: none"> <li>Get to know the basics about ER</li> <li>Finish Unit 1 Primary Reading</li> <li>Understand the myths and formulae for English learning</li> <li>Get to know collocation learning resources</li> </ul>	<ul style="list-style-type: none"> <li>Finish Unit 1 Speed Reading</li> <li>Understand the myths and formulae for English learning</li> </ul>	<ul style="list-style-type: none"> <li>Get to know the three principles about ER</li> <li>Finish Unit 2 Primary Reading</li> <li>Get to know pervasiveness of collocations</li> </ul>	<ul style="list-style-type: none"> <li>Finish Unit 2 Speed Reading</li> <li>Understand the arbitrary convention for collocations</li> </ul>	<ul style="list-style-type: none"> <li>Get to know the skill of vocalization</li> <li>Finish Unit 3 Primary Reading</li> <li>Get to know the trap of synonyms</li> </ul>	<ul style="list-style-type: none"> <li>Finish Unit 3 Speed Reading</li> <li>Understand the rule of rhythm</li> </ul>	<ul style="list-style-type: none"> <li>Master the eye perception skill in ER</li> <li>Finish Unit 4 Primary Reading</li> <li>Learn how to acquire collocations through a demo</li> </ul>
<b>Materials</b>	<ul style="list-style-type: none"> <li>Textbook: Unit 1 Ancient Greece and Rome, Section 1 &amp; 2</li> <li>Collocation Training: Session 1</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 1 Ancient Greece and Rome, Section 3 &amp; 4</li> <li>Collocation Training: Session 2</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 2 Classical Methodology, Section 1 &amp; 2</li> <li>Collocation Training: Session 3</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 2 Classical Methodology, Section 3 &amp; 4</li> <li>Collocation Training: Session 4</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 3 History, Section 1 &amp; 2</li> <li>Collocation Training: Session 5</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 3 History, Section 3 &amp; 4</li> <li>Collocation Training: Session 6</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 4 Art, Section 1 &amp; 2</li> <li>Collocation Training: Session 7</li> </ul>
<b>Activities</b>	<ul style="list-style-type: none"> <li>Explanation of reading skills: basics about ER</li> <li>Primary Reading exercise</li> <li>Introduction of Collocation Learning Resource</li> </ul>	<ul style="list-style-type: none"> <li>Speed Reading exercise</li> <li>Illustration of "Chinglish" examples</li> <li>Discussion on the two contrastive formulae in English learning</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of reading skills: three principles about ER</li> <li>Primary Reading exercise</li> <li>Illustration of pervasiveness of collocations</li> </ul>	<ul style="list-style-type: none"> <li>Speed Reading exercise</li> <li>Illustration of mismatch between E and C collocations</li> <li>Discussion on the importance of thinking in English</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of reading skills: vocalization</li> <li>Primary Reading exercise</li> <li>Illustration of the trap of synonyms</li> </ul>	<ul style="list-style-type: none"> <li>Speed Reading exercise</li> <li>Illustration of rhythms</li> <li>Practice of identifying rhythmic collocations in texts</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of reading skills: eye perception</li> <li>Primary Reading exercise</li> <li>Demo of collocation acquisition</li> </ul>
<b>Assessment</b>	<ul style="list-style-type: none"> <li>Comprehension checks of P4-11 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P14-22 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P25-32 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P35-39 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P51-56 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P55-65 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P75-80 of the textbook</li> </ul>
<b>Homework</b>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>

Table 9 Teaching Plan for Group (e)

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
<b>Goals</b>	<ul style="list-style-type: none"> <li>Get to know the basics about ER</li> <li>Finish Unit 1 Primary Reading</li> <li>Get to know collocation learning resources</li> </ul>	<ul style="list-style-type: none"> <li>Finish Unit 1 Speed Reading</li> <li>Understand the myths and formulae for English learning</li> </ul>	<ul style="list-style-type: none"> <li>Get to know the three principles about ER</li> <li>Finish Unit 2 Primary Reading</li> <li>Get to know pervasiveness of collocations</li> </ul>	<ul style="list-style-type: none"> <li>Finish Unit 2 Speed Reading</li> <li>Understand the arbitrary convention for collocations</li> </ul>	<ul style="list-style-type: none"> <li>Get to know the skill of vocalization</li> <li>Finish Unit 3 Primary Reading</li> <li>Get to know the trap of synonyms</li> </ul>	<ul style="list-style-type: none"> <li>Finish Unit 3 Speed Reading</li> <li>Understand the rule of rhythm</li> </ul>	<ul style="list-style-type: none"> <li>Master the eye perception skill in ER</li> <li>Finish Unit 4 Primary Reading</li> <li>Learn how to acquire collocations through a demo</li> </ul>
<b>Materials</b>	<ul style="list-style-type: none"> <li>Textbook: Unit 1 Ancient Greece and Rome, Section 1 &amp; 2</li> <li>Collocation Training: Session 1</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 1 Ancient Greece and Rome, Section 3 &amp; 4</li> <li>Collocation Training: Session 2</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 2 Classical Methodology, Section 1 &amp; 2</li> <li>Collocation Training: Session 3</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 2 Classical Methodology, Section 3 &amp; 4</li> <li>Collocation Training: Session 4</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 3 History, Section 1 &amp; 2</li> <li>Collocation Training: Session 5</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 3 History, Section 3 &amp; 4</li> <li>Collocation Training: Session 6</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 4 Art, Section 1 &amp; 2</li> <li>Collocation Training: Session 7</li> </ul>
<b>Activities</b>	<ul style="list-style-type: none"> <li>Explanation of reading skills: basics about ER</li> <li>Primary Reading exercise</li> <li>Introduction of Collocation Learning Resource</li> </ul>	<ul style="list-style-type: none"> <li>Speed Reading exercise</li> <li>Illustration of "Chinglish" examples</li> <li>Discussion on the two contrastive formulae in English learning</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of reading skills: three principles about ER</li> <li>Primary Reading exercise</li> <li>Illustration of pervasiveness of collocations</li> </ul>	<ul style="list-style-type: none"> <li>Speed Reading exercise</li> <li>Illustration of mismatch between E and C collocations</li> <li>Discussion on the importance of thinking in English</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of reading skills: vocalization</li> <li>Primary Reading exercise</li> <li>Illustration of the trap of synonyms</li> </ul>	<ul style="list-style-type: none"> <li>Speed Reading exercise</li> <li>Illustration of rhythms</li> <li>Practice of identifying rhythmic collocations in texts</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of reading skills: eye perception</li> <li>Primary Reading exercise</li> <li>Demo of collocation acquisition</li> </ul>
<b>Assessment</b>	<ul style="list-style-type: none"> <li>Comprehension checks of P4-11 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P14-22 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P28-32 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P35-39 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P51-56 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P58-65 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P75-80 of the textbook</li> </ul>
<b>Homework</b>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>

Table 10 Teaching Plan for Group (i)

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
<b>Goals</b>	<ul style="list-style-type: none"> <li>• Get to know the basics about ER</li> <li>• Finish Unit 1 Primary Reading</li> </ul>	<ul style="list-style-type: none"> <li>• Finish Unit 1 Speed Reading</li> </ul>	<ul style="list-style-type: none"> <li>• Get to know the three principles about ER</li> <li>• Finish Unit 2 Primary Reading</li> </ul>	<ul style="list-style-type: none"> <li>• Finish Unit 2 Speed Reading</li> </ul>	<ul style="list-style-type: none"> <li>• Get to know the skill of vocalization</li> <li>• Finish Unit 3 Primary Reading</li> </ul>	<ul style="list-style-type: none"> <li>• Finish Unit 3 Speed Reading</li> </ul>	<ul style="list-style-type: none"> <li>• Master the eye perception skill in ER</li> <li>• Finish Unit 4 Primary Reading</li> </ul>
<b>Materials</b>	<ul style="list-style-type: none"> <li>• Textbook: Unit 1 Ancient Greece and Rome, Section 1 &amp; 2</li> </ul>	<ul style="list-style-type: none"> <li>• Textbook: Unit 1 Ancient Greece and Rome, Section 3 &amp; 4</li> </ul>	<ul style="list-style-type: none"> <li>• Textbook: Unit 2 Classical Methodology, Section 1 &amp; 2</li> </ul>	<ul style="list-style-type: none"> <li>• Textbook: Unit 2 Classical Methodology, Section 3 &amp; 4</li> </ul>	<ul style="list-style-type: none"> <li>• Textbook: Unit 3 History, Section 1 &amp; 2</li> </ul>	<ul style="list-style-type: none"> <li>• Textbook: Unit 3 History, Section 3 &amp; 4</li> </ul>	<ul style="list-style-type: none"> <li>• Textbook: Unit 4 Art, Section 1 &amp; 2</li> </ul>
<b>Activities</b>	<ul style="list-style-type: none"> <li>• Explanation of reading skills: basics about ER</li> <li>• Primary Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>• Speed Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>• Explanation of reading skills: three principles about ER</li> <li>• Primary Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>• Speed Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>• Explanation of reading skills: vocalization</li> <li>• Primary Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>• Speed Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>• Explanation of reading skills: eye perception</li> <li>• Primary Reading exercise</li> </ul>
<b>Assessment</b>	<ul style="list-style-type: none"> <li>• Comprehension checks of P4-11 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>• Comprehension checks of P14-22 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>• Comprehension checks of P28-32 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>• Comprehension checks of P35-39 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>• Comprehension checks of P51-56 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>• Comprehension checks of P58-65 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>• Comprehension checks of P75-80 of the textbook</li> </ul>
<b>Homework</b>	<ul style="list-style-type: none"> <li>• Vocabulary expansion</li> <li>• Weekly ER assignment</li> </ul>	<ul style="list-style-type: none"> <li>• Vocabulary expansion</li> <li>• Weekly ER assignment</li> </ul>	<ul style="list-style-type: none"> <li>• Vocabulary expansion</li> <li>• Weekly ER assignment</li> </ul>	<ul style="list-style-type: none"> <li>• Vocabulary expansion</li> <li>• Weekly ER assignment</li> </ul>	<ul style="list-style-type: none"> <li>• Vocabulary expansion</li> <li>• Weekly ER assignment</li> </ul>	<ul style="list-style-type: none"> <li>• Vocabulary expansion</li> <li>• Weekly ER assignment</li> </ul>	<ul style="list-style-type: none"> <li>• Vocabulary expansion</li> <li>• Weekly ER assignment</li> </ul>



Table 11 Teaching Plan for Group (c)

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
<b>Goals</b>	<ul style="list-style-type: none"> <li>Get to know the basics about ER</li> <li>Finish Unit 1 Primary Reading</li> </ul>	<ul style="list-style-type: none"> <li>Finish Unit 1 Speed Reading</li> </ul>	<ul style="list-style-type: none"> <li>Get to know the three principles about ER</li> <li>Finish Unit 2 Primary Reading</li> </ul>	<ul style="list-style-type: none"> <li>Finish Unit 2 Speed Reading</li> </ul>	<ul style="list-style-type: none"> <li>Get to know the skill of vocalization</li> <li>Finish Unit 3 Primary Reading</li> </ul>	<ul style="list-style-type: none"> <li>Finish Unit 3 Speed Reading</li> </ul>	<ul style="list-style-type: none"> <li>Master the eye perception skill in ER</li> <li>Finish Unit 4 Primary Reading</li> </ul>
<b>Materials</b>	<ul style="list-style-type: none"> <li>Textbook: Unit 1 Ancient Greece and Rome, Section 1 &amp; 2</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 1 Ancient Greece and Rome, Section 3 &amp; 4</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 2 Classical Methodology, Section 1 &amp; 2</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 2 Classical Methodology, Section 3 &amp; 4</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 3 History, Section 1 &amp; 2</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 3 History, Section 3 &amp; 4</li> </ul>	<ul style="list-style-type: none"> <li>Textbook: Unit 4 Art, Section 1 &amp; 2</li> </ul>
<b>Activities</b>	<ul style="list-style-type: none"> <li>Explanation of reading skills: basics about ER</li> <li>Primary Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>Speed Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of reading skills: three principles about ER</li> <li>Primary Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>Speed Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of reading skills: vocalization</li> <li>Primary Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>Speed Reading exercise</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of reading skills: eye perception</li> <li>Primary Reading exercise</li> </ul>
<b>Assessment</b>	<ul style="list-style-type: none"> <li>Comprehension checks of P4-11 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P14-22 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P28-32 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P35-39 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P51-56 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P59-65 of the textbook</li> </ul>	<ul style="list-style-type: none"> <li>Comprehension checks of P75-80 of the textbook</li> </ul>
<b>Homework</b>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>	<ul style="list-style-type: none"> <li>Vocabulary expansion</li> </ul>

### 3.7 Data Analysis

As mentioned above, this study will conduct both quantitative and qualitative analysis of the data, in order to arrive at a fuller picture of learner's collocational competence. The quantitative analysis means to take the figures to draw statistically sound conclusions. We will analyze six aspects, namely, general lexical level, collocation lexical level, collocation density, collocation diversity, collocation accuracy rate, and collocation strength.

The qualitative analysis is to follow the notable tradition of the field of SLA to render a description as robust as possible of the patterns identified in the data. Specifically, the researcher will be keen to learn whether their choice of words and phrases is different in speaking (which reflects their implicit knowledge) from writing (which reflects both their implicit and explicit knowledge), etc.

### 3.8 Ethics

I conducted my research in China, which exempted me from applying for research ethics approval in Thailand. However, in compliance with my university's regulations, I applied for the research ethics approval within my institution and adhered to procedures that conform to standard protocols. I informed all participants about the nature of the experiment and ensured their right to withdraw from the study at any point if they experienced discomfort. I undertook thorough precautions to prevent any harm to the participants, given the nature of my research. My study focused on teaching English collocations to the experimental groups using a specific teaching method. To guarantee the ethicality of the study, I obtained informed consent forms from all participants, thus ensuring their awareness and agreement to partake in the research.

## CHAPTER 4

### RESEARCH FINDING

This chapter is dedicated to data analysis and relevant findings. It consists of two parts, namely, collocational competence profiling and collocational competence development. This is in line with the two research questions we proposed in Chapter 1:

*RQ1. What is the collocational competence profile of intermediate Chinese EFL learners?*

*RQ2. How can different learning modes (ER, explicit teaching, and the combination of the two) affect the development of their collocational competence?*

In each part, we will first conduct quantitative analysis which is followed by qualitative analysis. For collocational competence profiling, we will make two kinds of comparisons. One is between NNS and NS, taking the latter as the benchmark, to shed light on the former's collocational inadequacies. The other is between NNS's oral and NNS's written data, to reveal any discrepancy between learners' implicit and explicit collocational knowledge. For collocational competence development, we will examine the difference between pretest data and posttest data for each of the four NNS experiment groups, so as to understand the effects of different treatments. Before presenting the details of the analyses, we would like to explain our data processing rules and evaluation criteria, to help our readers understand the data more clearly.

Firstly, regarding data processing, these are the principles we adhered to. In terms of counting, we did token and type for words, and token and lemma for collocations. Collocations are identified and counted this way: researchers manually scan the text with the help of AntConc POS-related commands, identify any word combinations that fit the three target syntactic patterns (i.e. Adj-N / Adv-Adj / Adv-V) regardless of the distance in between the two constituent words. We are flexible on the actual order of the constituents, except for Adj-N category (i.e. Adjective has to precede the Noun it modifies to be qualified as a Adj-N collocation). To make the count as

accurate as possible, when the pronoun *one(s)* is preceded by adjective(s), we make the count and revert such a pronoun to its antecedent before making an entry in the list. With this, we get the list of collocation tokens. To generate the list of collocation lemmas from it, we collapse any tokens which are made up of the same lemmas (e.g. *good friend* and *good friends* merge into *GOOD FRIEND*, *take seriously* and *took seriously* merge into *TAKE SERIOUSLY*, etc.). Also, collocations of different orders are collapsed (e.g. *greatly enhance* and *enhance greatly* merge into *GREATLY ENHANCE*). Furthermore, because we are interested in collocation rather than any free combination of words, any word pairs containing some extremely common words (e.g. *very* as an adverb) are not counted. Please see the following list of these extremely common words that we excluded:

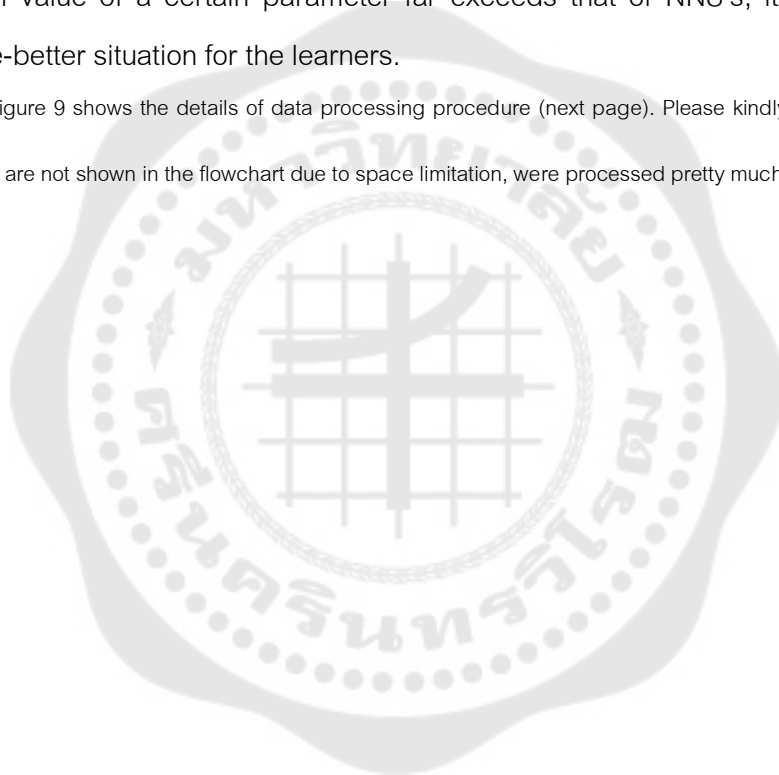
- Adverb: *very, still, so, mostly, really, usually, always, already, sometimes, never*
- Adjective: *many, much, little, only, same, different, other*

In terms of NS data collection, we used COCA to build the NS oral and written corpora. Of all the eight genres in COCA (i.e. SPOKEN, TV/Movies, FICTION, MAGAZINES, NEWSPAPERS, ACADEMIC, WEB-GENL, and WEB-BLOG), the first two are oral in nature while the remaining six are written. To build the NS oral corpus, we randomly picked ten 300-word-long pieces of text from the two oral genres (five from each genre), and by combining them together we have the NS oral corpus, the token count of which is about 3000. To build the NS written corpus, we randomly picked six 500-word-long pieces of text from the six written genres (one from each genre), and by combining them together we have the NS written corpus, the token count of which is also about 3000. This way, each of the NNS corpora has its corresponding reference NS corpus to compare with.

Secondly, regarding evaluation criteria, NS data serve as the sole benchmark for evaluating NNS data. This is what it means: in general, if genuine development took place, a learner should demonstrate improvement in a number of collocation-based statistical measures (e.g. collocation density). However, it is not a “the more, the better”

scenario. Why not? Because modifier-modified types of collocation (i.e. Adj-N / Adv-Adj / Adv-V) are just one way to make one's language more informative and flavorful. There are definitely other (and probably better) ways to achieve such desired result, such as using a relative clause or an appositional structure. On the other hand, an essay that is piled up with collocations in every sentence is unlikely to be of the highest quality. Therefore, there should be an optimum level of collocation usage, which we deem can only be revealed from study of large-scale NS production. Of course, when NS's statistical value of a certain parameter far exceeds that of NNS's, it becomes a the-more-the-better situation for the learners.

Figure 9 shows the details of data processing procedure (next page). Please kindly keep in mind that NS data, which are not shown in the flowchart due to space limitation, were processed pretty much the way as NNS data.



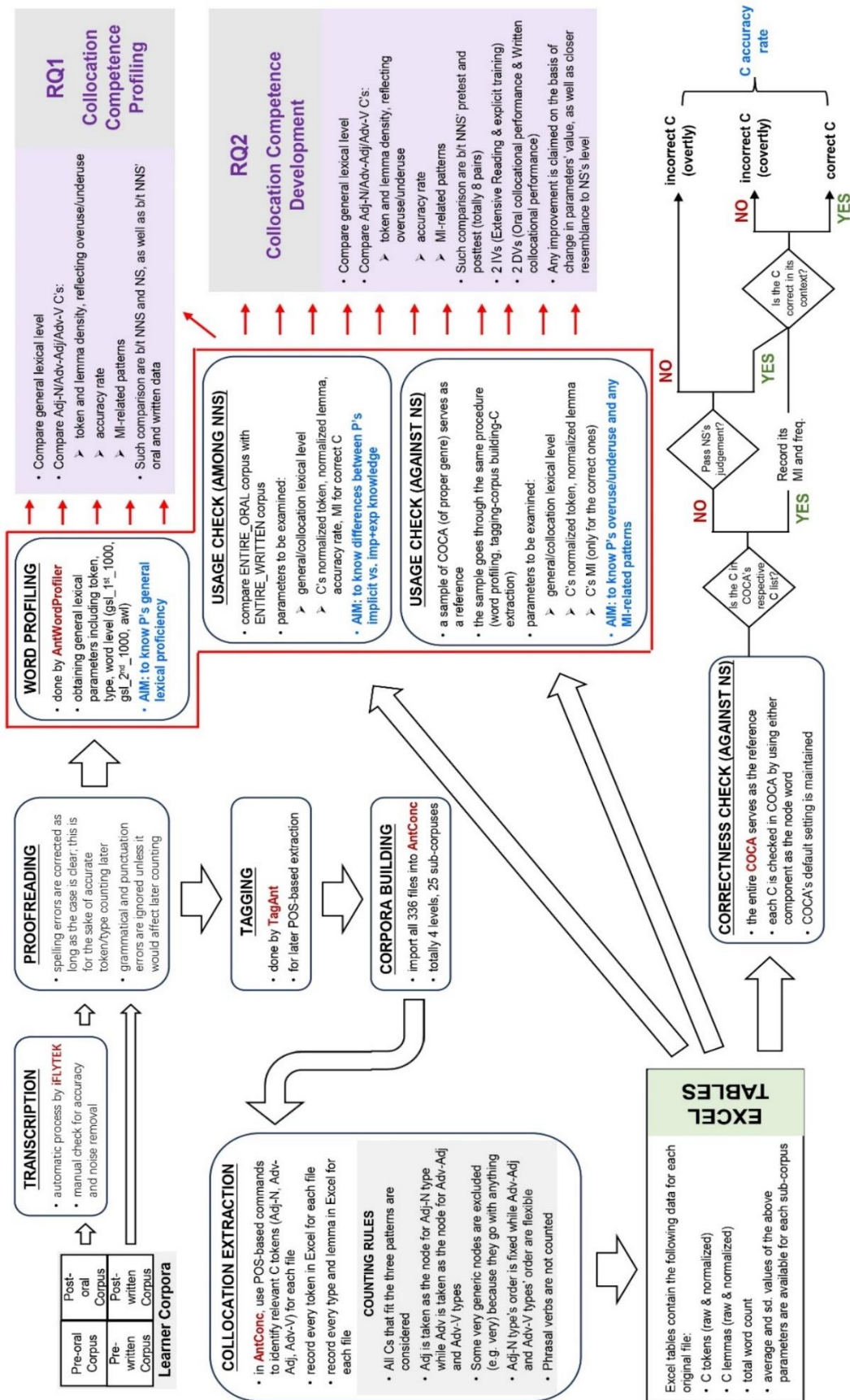


Figure 9 Data Processing Workflow (self-made)

Now we are ready to introduce the data and analyses, beginning with collocational competence profiling.

#### 4.1 Collocational Competence Profiling

This part aims to address the first research question: *how can learners' collocational competence be thoroughly profiled?* To answer it, we shall examine both their oral data and written data. The former reflects their implicit knowledge while the latter their overall knowledge (explicit plus implicit), as explained in section 2.4.2. Besides, native speakers' data is taken as the reference in every aspect. Below we shall first perform a quantitative analysis, and then a qualitative one, to draw the full picture of their collocational competence.

##### 4.1.1 Quantitative Analysis

In this section, we shall examine the characteristics of NNS's collocational competence in six aspects: general lexical level, collocation lexical level, collocation density, collocation diversity, collocation accuracy rate, and collocation strength. Comparisons will be made first between NNS and NS, and then between NNS's oral data and NNS's written data.

##### 1. General Lexical Level

Table 12 General Lexical Level Profiling

		Total tokens	Total types	gsl_1st_1000				gsl_2nd_1000				awl_570				not in lists			
				token	(%)	type	(%)	token	(%)	type	(%)	token	(%)	type	(%)	token	(%)	type	(%)
NNS	ORAL	36046	2696	32091	89.03%	1368	50.74%	1445	4.01%	463	17.17%	659	1.83%	237	8.79%	1851	5.14%	628	23.29%
	WRITTEN	40968	2849	36145	88.23%	1347	47.28%	1826	4.46%	423	14.85%	1563	3.82%	403	14.15%	1434	3.50%	649	22.78%
NS	ORAL	3160	904	2632	83.29%	536	59.29%	138	4.37%	96	10.62%	92	2.91%	62	6.86%	298	9.43%	210	23.23%
	WRITTEN	3031	1235	2203	72.68%	633	51.26%	202	6.66%	143	11.58%	163	5.38%	121	9.80%	463	15.28%	338	27.37%



Figure 10 General Lexical Level Profiling (self-made)

Before diving into any detail about the collocations learners produced, let us first take a look at their general lexical proficiency. Although each participant took the Cambridge placement test and is confirmed as being at the intermediate level, that test is not for lexical proficiency specifically, but overall linguistic proficiency. It is thus deemed beneficial to perform a deeper investigation into their lexical proficiency which should serve as a foundation to their collocation-related performance (see explanation in section 2.4.1).

The way to do this is to import any corpus of interest into a software called AntWordProfiler (Anthony, 2023) which will automatically calculate the percentage of each category based on General Service List (West 1953). These categories include “gsl\_1st\_1000” (i.e. the first thousand headwords), “gsl\_2nd\_1000” (i.e. the second thousand headwords), as well as “awl\_570” (i.e. the 570 headwords of the Academic



*Word List* by Coxhead, 1998) and “not\_in\_lists” (i.e. any word not included in the first three categories, generally of a more advanced vocabulary). To ensure accuracy of the result, we manually checked every file of the entire corpora to correct any spelling errors (which were actually minimal in the first place). Therefore, we have the confidence to say that the higher the percentage for “awl\_570” and “not\_in\_lists” is, the more advanced the vocabulary is proved to be.

In the comparison between NNS and NS, we can clearly see that NS possess a more advanced vocabulary than NNS. The gap is already quite visible in oral data, but even more pronounced in written data. For example, in writing, NS uses “not\_in\_lists” words four times more often than NNS (15.28% vs. 3.50%), and about one and half times more often with regards to “awl\_570” words (5.38% vs 3.82%). Many such advanced words of various parts of speech only manifest themselves in NS data, such as *dichotomous*, *convey*, *intellectual*, *curriculum*, *reconceptualize*, etc. This shows that NS’s implicit lexical level is clearly higher than NNS’s, and there is an even wider margin between them for explicit lexical level. This, in turn, will have an impact on their collocation performance because general lexical proficiency is a prerequisite for any collocation performance.

On the other hand, if we make an internal comparison between NNS’s oral data and their written data, there is not much difference in terms of token. However, in terms of type, their written vocabulary does seem a little more advanced (e.g. for “awl\_570” category, it is 14.15% vs. 8.79%). This means that learners’ explicit lexical knowledge could be slightly superior to their implicit knowledge, and in actual performance these advanced words are not used repetitively. Anyway, the difference is only marginal. So on the general lexical level, we are inclined to conclude that learners’ implicit and explicit knowledge are quite the same. This is important to keep in mind because, as we shall see, when examining NNS’s oral and written collocational performance, they are different in virtually every aspect. In other words, one’s general lexical proficiency and collocational proficiency are two related yet separate entities.

## 2.Collocation Lexical Level

Table 13 Collocation Lexical Level Profiling

		Total C tokens	Total C lemmas	gsl_1st_1000		gsl_2nd_1000		awl_570		not in lists								
				token	(%)	lemma	(%)	token	(%)	lemma	(%)	token	(%)	lemma				
NNS	ORAL	2340	1338	1816	77.61%	947	70.78%	196	8.38%	165	12.33%	144	6.15%	98	7.32%	195	8.33%	133
	WRITTEN	3854	2520	2665	69.15%	1603	63.61%	352	9.13%	265	10.52%	455	11.81%	364	14.44%	397	10.30%	308
NS	ORAL	178	162	99	55.62%	92	56.79%	12	6.74%	12	7.41%	30	16.85%	28	17.28%	37	20.79%	30
	WRITTEN	388	370	198	51.03%	185	50.00%	52	13.40%	48	12.97%	41	10.57%	41	11.08%	100	25.77%	96

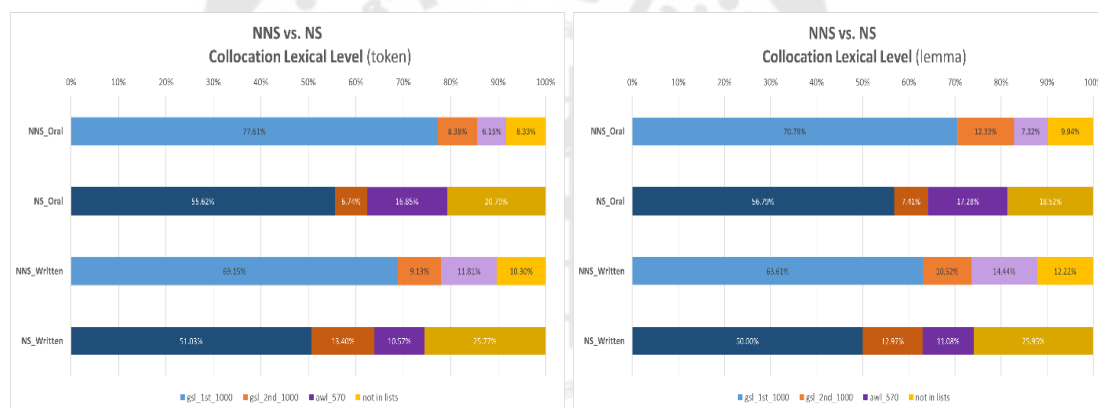


Figure 11 Collocation Lexical Level Profiling (self-made)

Now let us examine the aspect of collocation lexical level. What this means is that, instead of investigating NNS and NS production data in its entirety, we only focus on the collocations contained in it. Therefore, after extracting the collocations, we use them to build four collocation-based corpora (i.e. *NNS\_Oral\_collocations*, *NNS\_Written\_collocations*, *NS\_Oral\_collocations*, *NS\_Written\_collocations*). Then, we follow the same procedure of importing each corpus into *AntWordProfiler* to generate

the word level distribution, in order to check what kind of words users utilize to make up collocations.

As shown in the two charts above, the result is clear and consistent. Both in terms of token and lemma, NS use much more advanced vocabulary words (i.e. “awl\_570” and “not\_in\_lists”) to make up collocations than NNS do. Interestingly, the gap is larger for oral than for written data. If we combine “awl\_570” and “not\_in\_lists” together, the ratio of NS\_Oral to NNS\_Oral is 2.6:1 by token (2.1:1 by lemma), and the ratio of NS\_Written and NNS\_Written is 1.6:1 by token (1.4:1 by lemma). This means that, in oral production, which is based on their implicit collocational knowledge, NNS face a bigger challenge in using advanced words to make collocations. Many NS-specific collocations could illustrate this gap, such as *dichotomous terms*, *herbal antioxidant*, *Asian ginseng*, *commercial yeast*, etc. Therefore, in terms of general lexical level (as previously shown), the NNS-NS gap is wider for written data (based on combination of implicit and explicit knowledge), and in terms of collocation lexical level, the gap is wider for oral data (based on implicit knowledge only). This is quite interesting. One possible explanation, as suggested earlier, is that there is truly a gap between one's general lexical proficiency and collocation lexical proficiency, especially for the implicit part. Using a word in general is different from (and easier than) using a word in collocation.

Another observation from the charts is that, unlike general lexical proficiency for which NNS\_Oral and NNS\_Written are pretty much the same, for collocation lexical proficiency, NNS\_Oral is clearly inferior to NNS\_Written (14.48% vs. 22.11% by token, and 17.26% vs. 26.66% by lemma). For example, some advanced word combinations only appear in NNS writings, such as *interpersonal communication*, *idiomatic expressions*, *impeccable grammar*, etc. This means that, despite one's successful use of an advanced word in a general way, he or she may not be ready to use it in collocation making, especially when only implicit knowledge is relied upon. Again, this indicates that using a word in general and using a word in collocations are two related yet different kinds of proficiencies, especially concerning the implicit part.

## 3. Collocation Density

Table 14 Collocation Density Profiling

		Adj-N		Adv-Adj		Adv-V	
		Token	lemma	Token	lemma	Token	lemma
<b>NNS</b>	<b>Oral</b>	24.73	13.10	0.75	0.65	2.86	2.45
	<b>Written</b>	34.07	20.90	1.01	0.86	6.31	5.28
<b>NS</b>	<b>Oral</b>	20.66	18.83	0.78	0.78	1.83	1.57
	<b>Written</b>	47.05	44.48	2.28	2.28	5.99	5.99



Figure 12 Collocation Density Profiling (self-made)

Now we come to the third aspect, collocation density. The term here means normalized collocation token or lemma counts per 1000 words, which reflects the degree of “pervasiveness” of collocations produced. This, in turn, sheds light on whether learners are overusing or underusing some categories of collocations, and to what extent they are doing it.

First, we turn our attention to the comparison between NNS and NS. Again, as stated in the rationale at the beginning of this chapter, we use NS as the benchmark for any evaluation. (Unfortunately, due to the rather small size of NS data, for the least common Adv-Adj type, native speakers produced few tokens and thus the result may not be statistically sound.) For the oral part, interestingly, NNS produced more Adj-N collocation tokens and more Adv-V tokens than NS (24.73 vs. 20.66 and 2.86 vs. 1.83, respectively), while for Adv-Adj there is little difference (0.75 vs. 0.78). Does it mean that NNS are more proficient in producing collocations than their native counterparts? The answer is likely negative. Although some studies (e.g. Laufer & Waldman, 2011; Wang & Shaw, 2008) suggest NS produce more collocations than NNS with regards to certain type (i.e. V-N), our case here is not a “the more, the better” story, because truly proficient language users have other means at their disposal (e.g. relative clause) to achieve the same desired effect. Therefore, rather than regarding them as being incapable of making as many collocations, it is more plausible that NS are naturally striking an optimum balance between collocations and other competing linguistic means. On the other hand, lemmawise, we can see that the trend is reversed for the Adj-N category. NS produced more lemmas (per 1K words) than NNS do, which suggests a greater diversity of collocations. To put it another way, although NNS produced quite a lot of collocation tokens in speaking, they are simply repeating their limited repertoire of lemmas.

Moving on to written data, the picture is more straightforward. NS outperform NNS in almost every category (with the only exception of Adv-V token). Like the oral data, the gap between the two groups is more significant by lemma than by

token (for example, NS use twice as many Adj-N lemmas in writing as NNS). Together these could have two indications. First, NS exceed their non-native counterpart further in explicit collocational knowledge than in implicit knowledge. Second, NS have a much larger explicit collocation repertoire to draw from while NNS rely on repetitive use of a smaller repertoire.

Now, we shall make a comparison within NNS, that is, between their oral data and written data. The trend is quite consistent and clear: for each of the three target types of collocation (i.e. Adj-N, Adv-Adj, and Adv-V), they use more collocations in writing than in speaking, both by token and by lemma. This means that, when learners have access to both their implicit and explicit collocation knowledge, they will make more collocations than when they rely on their implicit collocation knowledge alone. In other words, the influence and value of explicit collocation knowledge is recognized.

To sum up, learners' explicit collocational knowledge plays a crucial role in their performance. Besides, in general, they produce fewer collocations than their NS counterpart, both by token and by lemma, probably due to their limited repertoire of collocations and underdeveloped collocational competence.

#### 4. Collocation Diversity

Table 15 Collocation Diversity Profiling

		Adj-N			Adv-Adj			Adv-V		
		Token	lemma	TLR	Token	lemma	TLR	Token	lemma	TLR
<b>NNS</b>	<b>Oral</b>	24.73	13.10	0.53	0.75	0.65	0.87	2.86	2.45	0.86
	<b>Written</b>	34.07	20.90	0.91	1.01	0.86	1.00	6.31	5.28	0.86
<b>NS</b>	<b>Oral</b>	20.66	18.83	0.61	0.78	0.78	0.85	1.83	1.57	0.84
	<b>Written</b>	47.05	44.48	0.95	2.28	2.28	1.00	5.99	5.99	1.00

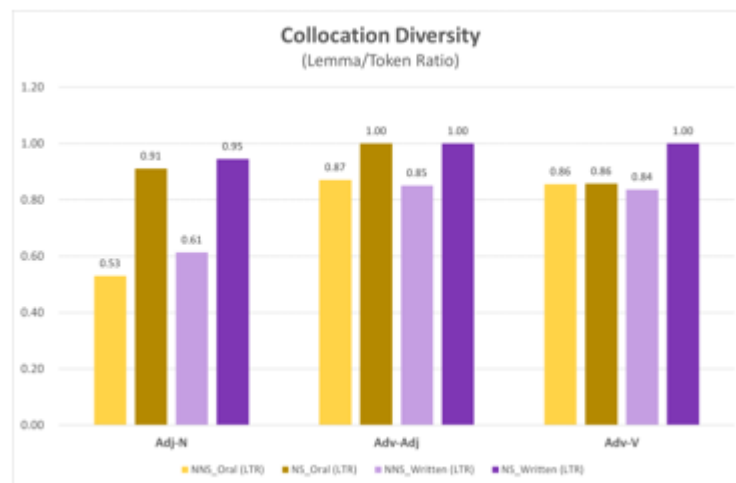


Figure 13 Collocation Diversity Profiling (self-made)

The fourth aspect we would like to examine is collocation diversity. This is measured by LTR, lemma token ratio. As we can see in the chart above, NS outperform NNS in all three categories, both oral and written. The widest gap lies in the Adj-N category (0.53 vs. 0.91 for oral, and 0.61 vs. 0.95 for written), and thus NNS seemingly need improvement in this category the most. To be more specific, while for NS no Adj-N collocation tokens were used more than 3 times, it is quite common to see NNS use some tokens for more than 6 times (e.g. *foreign languages*, *new language*, *human progress*, etc.). If we make a comparison between NNS\_Oral and NNS\_Written, it is hard to say which one is more diversified, as the former is slightly better in Adj-N while the latter is slightly better in Adv-Adj and Adv-V.

One last word of caution before we move on. Because the total collocation tokens and lemmas of NNS are much more than those of NS (see table 2), how comparable they are in terms of LTR is somewhat questionable. It is generally acknowledged that LTR generally drops as the number of token increases.

## 5. Collocation Accuracy Rate

Table 16 Collocation Accuracy Rate Profiling

		Total C Token	Total C Lemma	Collocation				Free Combination				Unidiomatic Combination			
				Token	(%)	Lemma	(%)	Token	(%)	Lemma	(%)	Token	(%)	Lemma	(%)
NNS	Oral	1169	668	537	45.94%	222	33.23%	541	46.28%	366	54.79%	91	7.78%	80	11.98%
	Written	1929	1260	845	43.81%	459	36.43%	818	42.41%	567	45.00%	266	13.79%	234	18.57%
NS	Oral	89	81	49	55.06%	46	56.79%	40	44.94%	35	43.21%	0	0.00%	0	0.00%
	Written	194	185	101	52.06%	95	51.35%	93	47.94%	90	48.65%	0	0.00%	0	0.00%



Figure 14 Collocation Accuracy Rate Profiling (self-made)

This is one of the most interest-arousing aspects of data analysis: collocation accuracy rate. After all, both teachers and students share the goal of avoiding errors in language learning. Before anything else, we will explain how tokens are categorized in the charts. First, all potential collocation tokens are identified from the NNS corpora. Then, each token is investigated in COCA with either constituent as the node. If the token is listed in the node's collocation page, then it is counted as a *collocation* (C). If not, it is passed on to two NS judges. If they agree that such a combination is acceptable, then it is counted as *free combination* (FC). Otherwise, it is counted as *unidiomatic combination* (UC). For the NS corpora, the procedure is a little simpler. If a token is not listed in COCA, it is automatically categorized as a free combination, with the assumption that NS do not make unidiomatic combinations.



First, we compare NNS and NS. The trend is crystal clear: NS outperform NNS in every case by producing higher percentage of collocations (55.06% vs. 45.94% for oral token, 52.06% vs. 43.81% for written token, 56.79% vs. 33.23% for oral lemma, 51.35% vs. 36.43% for written lemma). This indicates that, unsurprisingly, NS possess a better feel about which words should habitually go together, and NNS, being somewhat blind on this matter, produce more free combinations and unidiomatic ones (Lennon, 2998). This seemingly justifies Wray's position: NNS primarily rely on bottom-up mode which grants them too much freedom in composing utterances. Accordingly, a mode shift could be beneficial.

Shifting our focus to NNS\_Oral and NNS\_Written, we observe that the percentage of unidiomatic combinations is higher for written data than for oral data, and the difference is quite evident (13.79% vs 7.78% by token, 18.57% vs 11.98% by lemma). This suggests that in writing, when learners have access to their explicit collocational knowledge, they tend to make more mistakes. This calls into question the quality of their explicit collocational knowledge. Another thing to take notice of is that learners produced a higher percentage of collocations in oral than in written data by token, but not by lemma. This indicates that they make more repetitions of the "safe ones" in speaking (e.g., *great significance* was used five times in the control group's oral production), while in writing, they are more willing to take a risk to diversify their collocational performance (which could also partly explain why the error rate goes up in writing).

A last word before we move on. One should keep in mind that in reality, there is really no clear dividing lines between collocation and free combination, or even between free combination and unidiomatic combination. Statistically speaking, differences are continuous rather than categorical. Nevertheless, in practice, people widely adopt certain predetermined thresholds in order to make judgement on any token in question. For example, COCA takes a minimum MI of 2 for collocation establishment. One could rightly argue that a combination with an MI of 1.9 also sounds like a collocation. Similarly, our NS judges had a hard time making a call on some tokens, not

sure whether they should be categorized as free combinations or as unidiomatic ones. Indeed, this is a matter of continuous scale, not of definite categories. However, for the sake of presentation and discussion, we still have to adopt certain cut-off lines to categorize each token.

## 6. Collocation Strength

Table 17 Collocation Strength Profiling

		Adj-N		Adv-Adj		Adv-V	
		$\overline{MI}$ (token)	$\overline{MI}$ (lemma)	$\overline{MI}$ (token)	$\overline{MI}$ (lemma)	$\overline{MI}$ (token)	$\overline{MI}$ (lemma)
<b>NNS</b>	<b>Oral</b>	3.91	3.66	3.10	3.04	3.52	3.55
	<b>Written</b>	3.95	3.84	3.80	3.67	3.73	3.70
<b>NS</b>	<b>Oral</b>	4.05	4.11	5.53	5.53	5.36	5.54
	<b>Written</b>	4.49	4.43	4.79	4.79	4.68	4.68



Figure 15 Collocation Strength Profiling (self-made)

This is the last aspect for comparison: collocation strength. In all the previous aspects, every word combination that fit the three target syntactic patterns is included in analysis; but here, we make calculations only on those judged to be *collocations* by COCA (thus excluding free combinations and unidiomatic combinations). The parameter adopted to indicate collocation strength is mutual information (MI), the most widely used one in the literature. We will average the MI values up to learn how “strong” the collocations are, which, in turn, should indicate how good the user’s feel (which is part of her collocation competence) about collocations is.

First, let us compare NNS and NS. Clearly, NS outperform NNS in every single category, and the gap is considerable (e.g. 5.36 vs. 3.52 for the Adv-V category by token). This is in line with It shows that NS has a keener sense in perceiving and producing stronger collocations (e.g., *seemingly endless*, whose MI is 8.5), while NNS, probably due to lack of top-down mode and over relying on bottom-up mode, produce more free-combination-like collocations (e.g. *whole life*, whose MI is 2.98). Again, this bears out the position of Sinclair and Wray. Taking a closer look, we can see that the gap is larger for Adv-related collocation types (Adv-Adj and Adv-V), but smaller for Adj-N type. This indicates that while learners should strive to acquire strong collocation of all types, attention should be specifically paid to the ones involving adverbs.

Next, we turn our focus to the comparison between NNS\_Oral and NNS\_Written. Interestingly, their written data outperform their oral data in every collocation type, both in terms of token and lemma (although the difference is marginal in some cases, such as Adj-N by token). This means that, when having time to reflect, learners generally make a (slightly) better choice in composing collocations. Their explicit collocational knowledge’s contribution is again felt and appreciated.

#### 4.1.2 Qualitative Analysis

Besides quantitative analysis, we consider it necessary to carry out some qualitative analysis as well. This allows us to go beyond what mere numbers can reveal. We scanned through the subjects’ speeches and essays to understand a) what they

doing, b) how they are doing it, aiming to account for c) why they are doing it and d) how they can improve in future. Below are several things we found worth mentioning.

Inadequate word knowledge. Certain errors are clearly due to NNS's inadequate knowledge on the word level. Some are related to countability of nouns (e.g. *\*appropriate audios* → *appropriate audio*); some are related to POS of words (e.g. *\*calm word* → *calming word*, *\*extrovert classmate* → *extroverted classmate*); some are related to semantic range of words (e.g. *\*oral speaking*, in which oral and speaking are semantically redundant, *\*personal culture*, in which culture entails a mass of people, not an individual). It seems that in order to produce correct collocations, one must possess sufficient word knowledge.

Inadequate collocation knowledge. This should not be surprising, but here we will characterize their collocational performance in a more comprehensive way. Some learners are insensitive to word order when making collocations, especially for Adv-V type (e.g. *\*learn seriously* → *seriously learn*, *\*remember firmly* → *firmly remember*, *\*grasp patiently* → *patiently grasp*). Other learners demonstrate a tendency of using too much of the basic words in making collocation (e.g. the "very + Adj" structure occurs 13 times in one subject's 366-word speech!) Others do not make the best choice in making collocations. In other words, their tokens are arguably acceptable, but there is a better way to say it (e.g. *convictive way* → *convincing way*, *comprehensive application* → *broad application*, *small action* → *small-scale action*, *white sky* → *gray-white sky*). Others fail to use intensifiers properly (e.g. *\*awfully essential* → *absolutely essential*, *\*deeply fast* → *incredibly fast*, *\*deeply shy* → *extremely shy*, *\*exceedingly crucial* → *absolutely crucial*, *\*greatly useful* → *extremely useful*, *\*hugely good* → *very good*). All of these are areas where NNS need to improve.

Covert collocation errors. So far, we have been largely dealing with errors in form. However, a correct form does not warrant correct usage in context, and learners should be made aware of this kind of covert errors. Table 14 shows some typical covert collocation errors from the subjects.

Table 18 Covert Collocation Errors

Group	Subject	Corpus	Concordance	Correction	Note
G(c)	No. 2	Pre_Oral	Because I think teacher is a <b>good job ...</b>	good profession	What the subject means is really good profession while <i>good job</i> is a compliment on someone's performance.
G(e+i)	No. 15	Pre_Oral	I want to be an English teacher because it is a <b>good job.</b>	good profession	See previous note.
G(i)	No. 18	Pre_Oral	I think teacher ... it is a <b>good work.</b>	good profession	See previous note.
G(i)	No. 7	Pre_Oral	...my first choice is <b>Chinese literature.</b>	Chinese language	The subject is talking about her major, and what she meant is Chinese language.
G(e+i)	No. 13	Post_Oral	I <b>got well</b> with all my English teachers ...	got along well	She really meant having a good relationship with the teachers.
G(i)	No. 5	Post_Oral	Fortunately, I passed the <b>first exam</b> and met so many friends in it...	first interview	The subject was talking about her interview to join a local club.
G(i)	No. 17	Post_Oral	I did some <b>voluntary activities...</b>	volunteer activities	The subject participated some activities as a volunteer.
G(c)	No. 1	Post_Oral	I tried riding an <b>electric car</b> again...	electric bike	The subject was talking about learning to ride an e-bike on campus.
G(e+i)	No. 1	Pre_Writte n	The <b>old word</b> said, "it is when you are using what you have learned from books..."	old saying	She meant old saying.

Therefore, in collocation teaching and learning, we should not only give heed to forms, but also to actual usage in contexts.

L1 influence. Of all the above covert errors, some are seemingly due to L1 interference. For example, the common Chinese collocation “**好工作**” is literally translated as “good job”, but actually it means “good profession” in context. Another one

is the expression “**中文专业**” which should be translated as “Chinese (language) major”. But as the subject equals “**文**” with “**文化**” (culture) in her mind, mistake took place. Of course, L1-induced error is not limited to covert type alone. Certain overt errors seem to be due to L1 influence too, as the corresponding L1 collocation forms are acceptable. Examples include: \**deep communication* **深度交流**, \**extracurricular book* **课外书籍**, \**fast-paced age* **快节奏的时代**, \**fragmented reading* **碎片化阅读**, \**whole daytime* **整个白天**, \**whole person* **全人**, \**remember mechanically* **机械记忆**, \**officially learn* **正式学习**, \**learn seriously* **认真学习**, \**master firmly* **牢牢掌握**, \**face positively* **积极面对**, \**follow tightly* **紧紧跟随**. Therefore, students will do well to maintain a suspicious stance with regard to the degree of transferability of their L1 patterns. However, we should also keep in mind that L1 is a two-edged sword which sometimes interferes with L2 learning and sometimes facilitates it. Many cases can testify the latter, such as *ancient era* **古时**, *basic expression* **基本的表达**, *favorite lesson* **最喜欢的课**, etc. Therefore, learners should be advised to watch out for discrepancies between languages during collocation learning, but not discarding their mother tongue completely.

One last word before we end this section: as language is often intertwined with culture, some errors could be analyzed and understood from the perspective of culture too. For example, “voluntary activities”, one of the examples discussed earlier, could be due to the typical docile Chinese mindset that “we voluntarily accept arrangements in life”, regardless of whether they are actually imposed by others or not. Such a cultural factor might shed light on their linguistic performance as well.

Now that we are done with competence profiling, let us continue with the second half of the picture, competence development.

## 4.2 Collocational Competence Development

This part addresses the second research question: *how can one's collocational competence be effectively developed?* And again, we take both implicit knowledge and explicit knowledge into consideration.

As we explained in Chapter 3, there are four experiment groups, each of which went through different treatment(s): implicit treatment is a weekly extensive reading task, and explicit treatment is a series of in-class training sessions on various aspects of collocation learning. Group (e+i) received both treatments; Group (e) only received the explicit treatment; Group (i) only received the implicit treatment; Group (c), the control group, received neither treatment. We will basically compare the pretest data and posttest data for each of the four groups to identify any development. The native speakers' data again serve as the sole benchmark.

### 4.2.1 Quantitative Analysis

In this section, we will examine NNS's development in four aspects, namely, collocation density, collocation diversity, collocation accuracy rate, and collocation strength. The other two lexical-level aspects (i.e. general lexical level and collocation lexical level) are excluded in the analysis here, since the development of one's lexical level is a rather slow process and does not progress much in a seven-week period. This is especially true, considering the fact that this experiment is not specifically targeted for developing learners' lexical proficiency.

#### 1. Collocation Density Development



Table 19 Collocation Density Development (token)

		Adj-N		Adv-Adj		Adv-V		
		Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	
NNS	Group (e+i)	Oral	26.90	24.82	0.52	1.74	2.44	5.03
		Written	32.26	35.61	1.49	0.83	4.29	7.66
	Group (e)	Oral	19.04	25.91	0.40	1.19	1.20	2.97
		Written	36.10	39.81	1.50	0.70	7.15	10.60
	Group (i)	Oral	25.21	24.40	0.00	1.18	1.95	1.57
		Written	32.78	33.15	0.71	1.79	6.38	13.59
	Group (c)	Oral	21.44	29.81	0.22	0.61	2.38	5.07
		Written	32.90	29.89	0.81	0.35	5.54	13.36
	NS	Oral	20.66		0.78		1.83	
		Written	47.05		2.28		5.99	



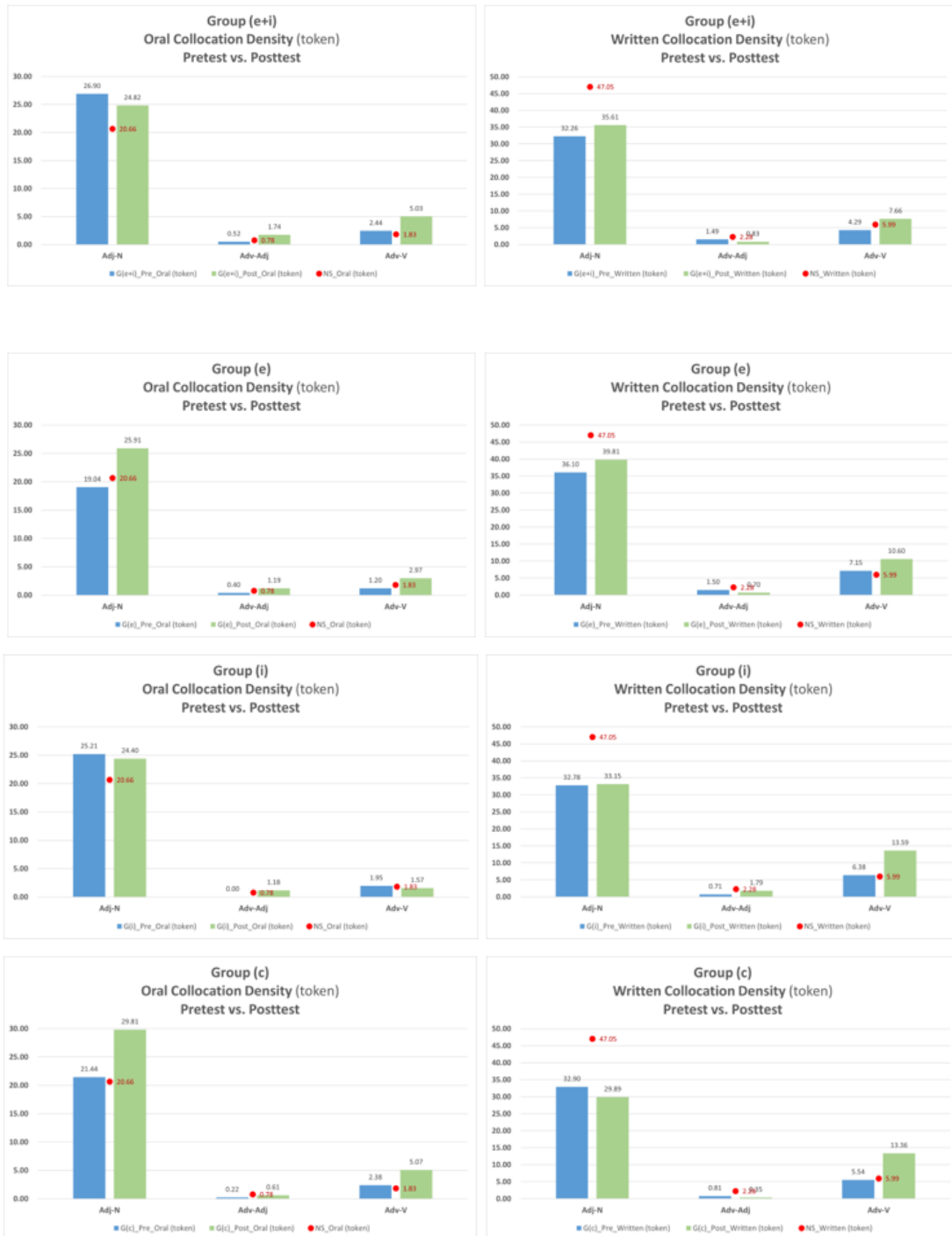


Figure 16 Collocation Density Development (token)

(self-made)

Before anything else, we would like to explain how the charts work. The blue columns denote pretest result while the green ones denote posttest result. And for each pair of columns, there is also a red dot denoting NS benchmark. It might be worthwhile to remind our readers that neither explicit nor implicit treatments actually encourage “more” production of collocation (please refer to section 3.4 for details), but simply strive to foster a “natural” use of collocation that is as close to NS counterpart as possible. Therefore, improvement is considered to have taken place if learners’ performance got closer to NS’s (but not necessarily making more tokens or lemmas per se) in the posttest.

Now, let us examine oral data first. Our expectation is that, since oral performance is based upon one’s implicit knowledge, the groups that receive implicit treatment (i.e. extensive reading) should demonstrate the biggest gain, while those receiving explicit treatment should also demonstrate moderate gain thanks to some crossing effect. Therefore, the order of improvement is expected to be Group (e+i) > Group (i) > Group (e) > Group (c). The results match our expectations quite well, especially for the most pervasive collocation type, Adj-N. For Group (e+i), the token dropped from 26.90 to 24.82, a step closer to NS’s 20.66. For Group (i), the token dropped from 25.21 to 24.40, also a (smaller) step closer to NS benchmark. For Group (e), the token rose from 19.04 to 25.91, thus having a status change from underuse to overuse compared to NS. For Group (c), the token rose from 21.44 to 29.81, showing an increase in the degree of overuse.

Next, we turn our focus to written data. Here we expect that, since written data is based on the combination of one’s implicit and explicit knowledge, the group that received both treatments should demonstrate the biggest gain, while the group that received explicit treatment should demonstrate the second biggest gain and the one that received implicit treatment the third (because explicit gain is supposed to manifest quicker than implicit gain). The control group should demonstrate the least gain, if any. The order should be Group (e+i) > Group (e) > Group (i) > Group (c). The result again is

quite in accordance with our expectation, especially for Adj-N type. For Group (e+i), the token rose from 32.26 to 35.61, a step closer to NS's 47.05. For Group (e), the token rose from 36.10 to 39.81, also a step closer to NS benchmark. For Group (i), the token rose from 32.78 to 33.15, a step closer to the benchmark, albeit a much smaller improvement. For Group (c), the token did not rise but fall, changing from 32.90 to 29.89, manifesting an increase in the degree of underuse! The trend for the other two collocation types is not as typical but still tells a similar story. For example, for the second most pervasive type (i.e. Adv-V), all groups produce excessive tokens, but the degree of overuse is lowest for Group (e+i) and second lowest for Group (e), meaning that these groups are doing better than the other two. It is encouraging to see that results based on token largely match our expectations. Next, we investigate the results based on lemma.

Table 20 Collocation Density Development (lemma)

		<b>Adj-N</b>		<b>Adv-Adj</b>		<b>Adv-V</b>		
		<b>Pretest</b>	<b>Posttest</b>	<b>Pretest</b>	<b>Posttest</b>	<b>Pretest</b>	<b>Posttest</b>	
<b>NNS</b>	<b>Group (e+i)</b>	<b>Oral</b>	17.29	16.66	0.52	1.56	2.44	4.86
		<b>Written</b>	25.36	26.46	1.49	0.83	4.29	5.66
	<b>Group (e)</b>	<b>Oral</b>	15.03	18.39	0.20	1.19	0.80	2.97
		<b>Written</b>	28.21	32.33	1.50	0.70	7.15	5.39
	<b>Group (i)</b>	<b>Oral</b>	17.00	17.17	0.00	1.18	1.95	1.57
		<b>Written</b>	26.94	25.39	0.71	1.49	6.03	7.77
	<b>Group (c)</b>	<b>Oral</b>	15.59	18.45	0.22	0.61	1.95	4.46
		<b>Written</b>	27.20	21.63	0.81	0.35	4.56	8.26
	<b>NS</b>	<b>Oral</b>		18.83		0.78		1.57
		<b>Written</b>		44.48		2.28		5.99

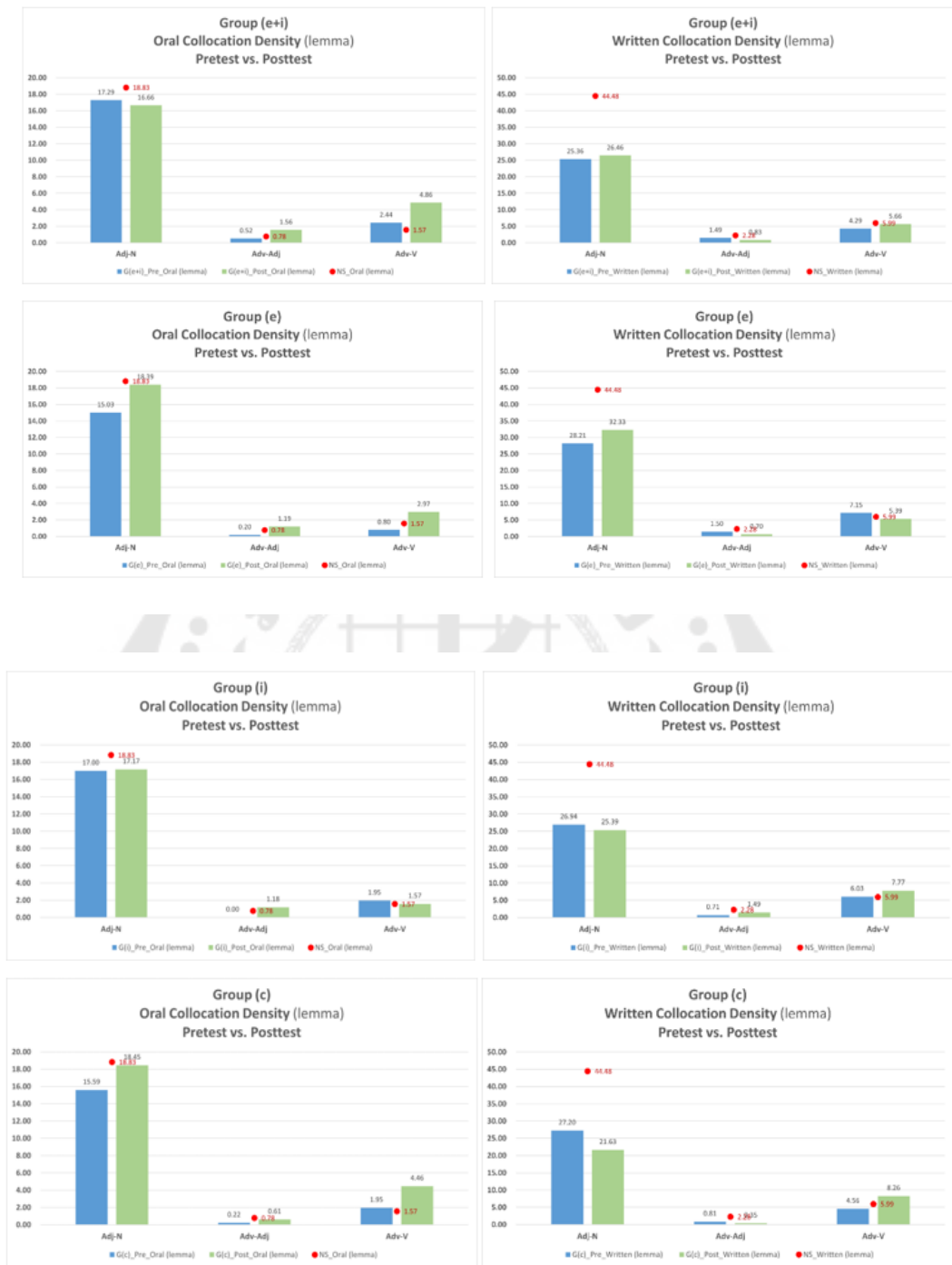


Figure 17 Collocation Density Development (lemma)  
(self-made)

Lemmawise, the trends are less typical but still similar, with a few exceptions. Again, we expect that for oral improvement, Group (e+i) > Group (i) > Group (e) > Group (c), and for written improvement, Group (e+i) > Group (e) > Group (i) > Group (c). The result is that, for oral part, everything meets our expectation except two cases: Group (e+i)'s Adj-N did not rise but fall, and Group (c)'s Adj-N rose significantly. These certainly pose a challenge for interpretation. The result for the written part is more consistent, only that the improvement of Group (e) is bigger than that of Group (e+i), which again is not easy to interpret. At this stage, the only thing we would suggest is improvement on the lemma level demands more effort (and probably longer period) than improvement on the token level.

## 2. Collocation Diversity Development

Table 21 Collocation Diversity Development

		Pretest			Posttest			
		LTR (Adj-N)	LTR (Adv-Adj)	LTR (Adv-V)	LTR (Adj-N)	LTR (Adv-Adj)	LTR (Adv-V)	
NNS	Group (e+i)	Oral	0.64	1.00	1.00	0.67	0.90	0.97
		Written	0.79	1.00	1.00	0.74	1.00	0.74
	Group (e)	Oral	0.79	0.50	0.67	0.71	1.00	1.00
		Written	0.78	1.00	1.00	0.81	1.00	0.51
	Group (i)	Oral	0.67	NA	1.00	0.58	1.00	1.00
		Written	0.82	1.00	0.94	0.77	0.83	0.57
Group (c)	Oral	0.73	1.00	0.82	0.62	1.00	0.88	
	Written	0.83	1.00	0.82	0.72	1.00	0.62	
NS	Oral	/	/	/	0.91	1.00	0.86	
	Written	/	/	/	0.95	1.00	1.00	

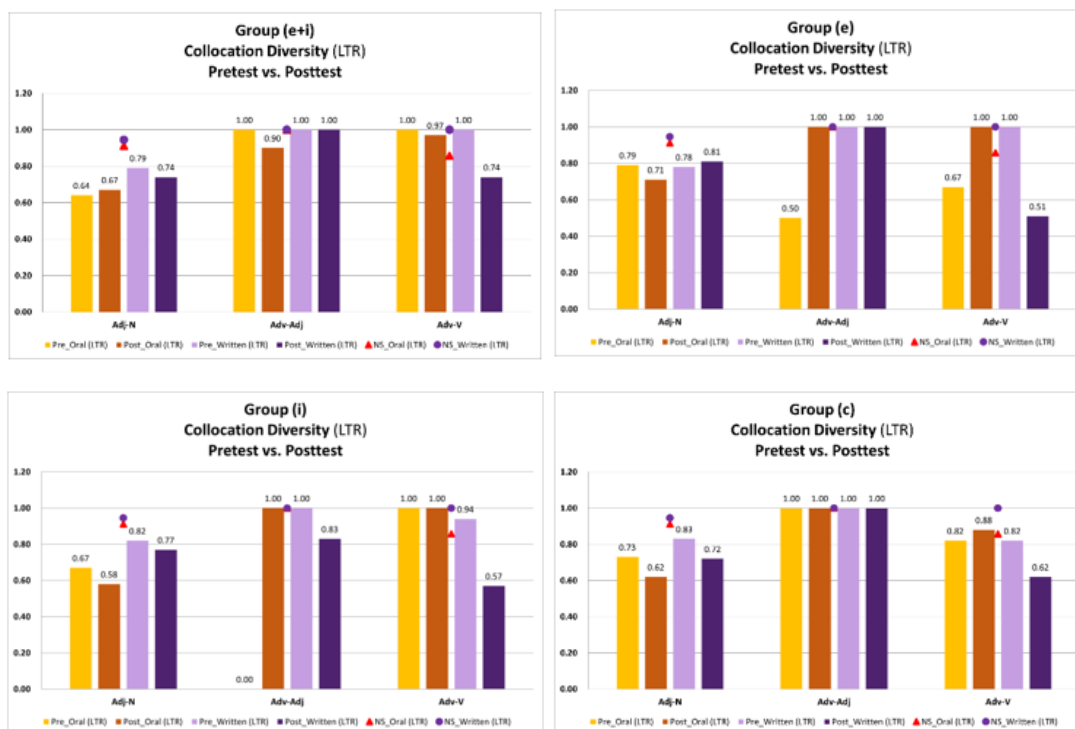


Figure 18 Collocation Diversity Development (self-made)

This is how one should read the diversity charts. In each one, there are four kinds of columns and two kinds of dots. The light orange column denotes NNS\_Pre\_Oral while the dark orange denotes NNS\_Post\_Oral. The little red triangle denotes NS\_Oral which serves as a reference point. So, we need to look at these three together. Then we have the light purple column which denotes NNS\_Pre\_Written and dark purple which denotes NNS\_Post\_Written. And the dark purple dot denotes NS\_Written which serves as a reference point. We need to observe these three together.

We can see that, in general, LTR is falling in most cases, reflecting a considerable increase in tokens and not much increase in lemmas. In other words, learners are using collocation more often in posttest than in pretest. And again, it is harder to make improvements on collocation lemmas. Interestingly, there are a few exceptions, including Group (e+i)\_Oral\_Adj-N, Group (e+i)\_Oral\_Adv-V, Group (e)\_Written\_Adj-N, and Group (e)\_Oral\_Adv-V, and Group (c)\_Oral\_Adv-V. Since four of

these five cases belong to the first two groups, it seems that, to have an improvement in LTR, explicit treatment may be crucial.

By the way, NS's LTR is almost always higher than NNS's. Thus, this area warrants efforts in one's learning of the language.

### 3. Collocation Accuracy Rate Development

Table 22 Collocation Accuracy Rate Development (lemma)

		Pretest			Posttest			
		C (%)	FC (%)	UC (%)	C (%)	FC (%)	UC (%)	
NNS	Group (e+i)	Oral	36.21%	53.45%	10.34%	36.09%	55.64%	8.27%
		Written	46.71%	38.32%	14.97%	40.40%	47.47%	12.12%
	Group (e)	Oral	43.75%	47.50%	8.75%	42.11%	47.37%	10.53%
		Written	42.35%	40.31%	17.35%	43.44%	41.18%	15.38%
	Group (i)	Oral	29.90%	59.79%	10.31%	39.53%	47.67%	12.79%
		Written	33.16%	47.89%	18.95%	37.93%	44.83%	17.24%
Group (c)	Oral	39.02%	52.44%	8.54%	35.34%	55.17%	9.48%	
	Written	38.00%	47.00%	15.00%	37.79%	45.93%	16.28%	
NS	Oral	/	/	/	56.79%	43.21%	NA	
	Written	/	/	/	51.35%	48.65%	NA	



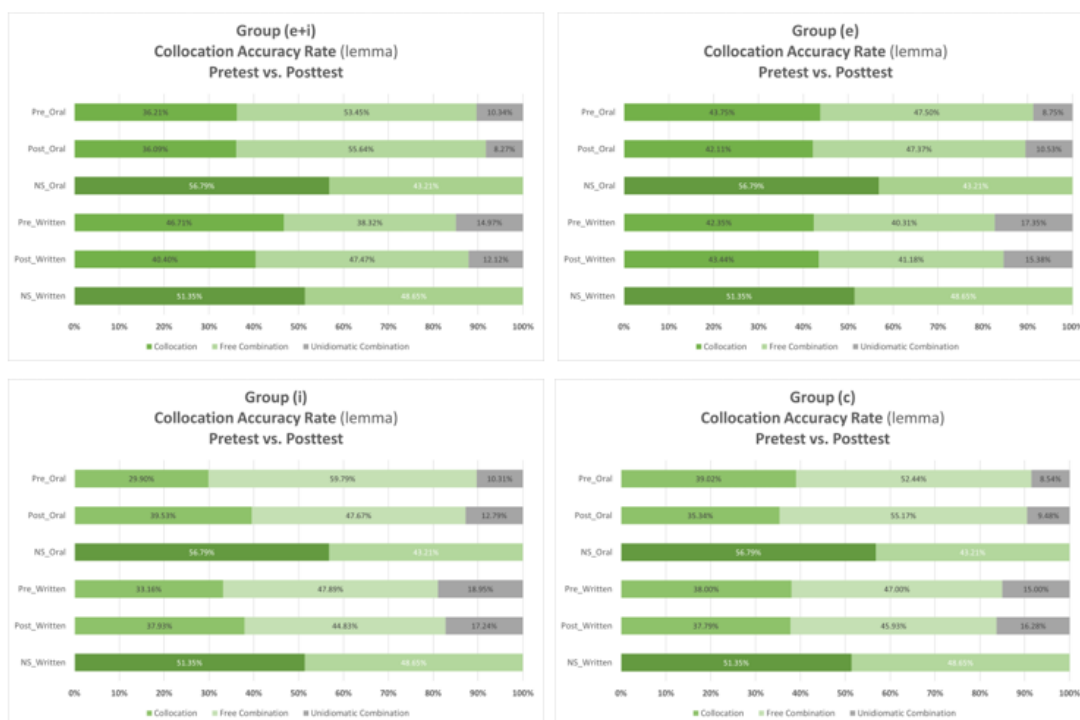


Figure 19 Collocation Accuracy Rate Development (self-made)

In this section, we shall examine the effects of different treatments on learners' collocation accuracy rate. Again, NS's serves as the benchmark. If we look at the percentage of unidiomatic combinations, the trend is quite straightforward: Group (e+i) is the only group whose oral and written error rates both dropped (10.34% to 8.27%, and 14.97% to 12.12%, respectively). For Group (e) and Group (i), the error rate for oral rose while the one for written dropped. For Group (c), both error rates rose (8.54% to 9.48%, and 15.00% to 16.28%, respectively)! This could well indicate that to reduce error rate, we need both explicit and implicit treatments to achieve the best result. If we look at the percentage for collocation, the trend is less clear. While Group (c) did not improve in either oral or written, which is what we expected, Group (e+i) also had no improvement in either genre. For Group (e), the oral collocation percentage dropped while the written collocation percentage rose. Group (i) enjoys the best result, with collocation percentage rising both for oral and for written genres. It is hard to draw

any conclusion here, not even for a preliminary one. More investigations (and probably more data too) are needed.

When it comes to NNS and NS comparison, NNS are far behind NS in terms of collocation percentage, for every single case. They indeed have a big gap to close in this area.

#### 4. Collocation Strength Development

Table 23 Collocation Strength Development (lemma)

		Pretest			Posttest			
		$\overline{MI}$ (Adj-N)	$\overline{MI}$ (Adv-Adj)	$\overline{MI}$ (Adv-V)	$\overline{MI}$ (Adj-N)	$\overline{MI}$ (Adv-Adj)	$\overline{MI}$ (Adv-V)	
NNS	Group (e+i)	Oral	3.76	2.62	3.40	3.74	3.07	3.99
		Written	3.80	2.63	3.74	3.60	5.06	3.52
	Group (e)	Oral	3.72	3.89	2.53	3.90	3.14	2.93
		Written	3.89	3.23	3.97	4.04	3.86	3.59
	Group (i)	Oral	3.84	NA	5.45	3.78	2.21	3.12
		Written	3.81	NA	4.52	4.01	4.86	4.11
	Group (c)	Oral	3.45	2.57	3.13	3.78	3.89	3.05
		Written	3.67	3.00	2.98	4.12	NA	3.78
NS	Oral	/	/	/	4.11	5.53	5.54	
	Written	/	/	/	4.43	4.79	4.68	

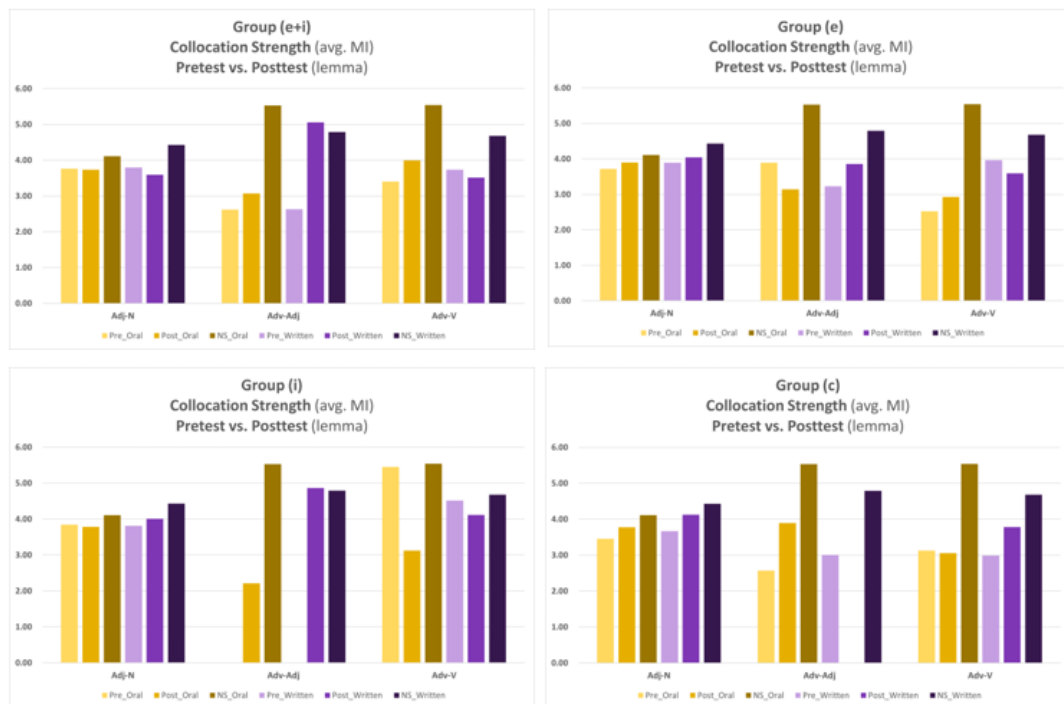


Figure 20 Collocation Strength Development (self-made)

To explain the result for collocation strength development is truly a challenge, because it is an extremely mixed one. The only sure thing is that NS's collocation strength is stronger than NNS's, in every single case. For example, while NNS only produced two collocations whose MI is above 9 (i.e. *literal translation* and *rote memorization*), NS did four (i.e. *oxidative stress*, *coronary arteriosclerosis*, *crusty bread*, and *set-top box*). This is even more remarkable if we consider the fact that the size of the NS corpus is only 10% of that of the NNS corpus. This means that there indeed is a gap for NNS to close. Other than this, we cannot draw any firm conclusion yet due to mixed results. It is possible to argue that the current treatments are not effective for collocation strength development, or their effect takes longer period to manifest.

#### 4.2.2 Qualitative Analysis

For the qualitative analysis, we will dig deeper into what subjects actually said in their speeches or essays that is relevant to our treatments. Some subjects' feedbacks indicated efficacy of extensive reading. Examples include:

*"I am cultivating ... I'm forming a habit of reading gradually every day."*

- Subject No. 12 of Group (i), in posttest speech

*"After mastering a certain amount of vocabulary, we can choose some interesting and moderately difficult reading materials to read. Read for 20 minutes every day, keep reading every day and learn the collocation of words."*

- Subject No. 7 of Group (e), in posttest essay

Reading indeed is a long-term habit and its real influence can only be felt as time passes by. It is good to see that some students start to recognize its importance and put it into practice. On the other hand, extensive reading task received some feedbacks which call for teacher's attention. These mainly come from students' weekly reading logs. For example, a handful of students expressed confusion during the first week or two, saying that they do not know how to proceed or keep it up with such minimal support and guidance from the teacher. For many, if not most, students, this may well be their first extensive reading experience on an English book. Therefore, they do need more guidance than a general introductory word and a set of principles. Another example is that some students expressed their discouragement because the book they chose turns out to be more difficult than they thought. This is a point that must be more properly handled in future, and teacher and students need more collaboration to make sure the book chosen is in line with Krashen's  $i+1$  principle. To achieve this, teachers could recommend some specific books rather than just referring them to some general sources. One last thing worth noticing is that some students chose an English book whose Chinese translation they have already read. And based on such prior knowledge about the content/plot of the book, the English original work seems more enjoyable and easier to follow. This could be passed on as a best practice to future

students. Now that you are largely familiar with the content, just enjoy the beauty of the English language!

Furthermore, some subjects talked about the metalinguistic and linguistic collocation knowledge they acquired. Below are some examples:

*"[My teacher] said we should accumulate some collocation to memory words. And it's good for us to learn English more. And so, then, I pay attention to accumulate a collocation. And to my surprise, accumulating collocations turns out to be one of my favorite things to do."*

- Subject No. 4 of Group (e+i), in posttest speech

*"It is undeniable that the more vocabulary the better, but we have to learn how to use these words flexibly and correctly, because it is not as simple as one-to-one correspondence between English and Chinese, and sometimes there will be some fixed collocation. More importantly, we need to match some appropriate phrases through context, so in the process of learning language, we can memorize relevant phrases to improve the learning of a language."*

- Subject No. 10 of Group (e), in posttest essay

*"Second, memorize lexical chunks. Provide that we want to realize and be able to translate English passage, we must memory more lexical chunks. Especially for writing, it helps to write correct collocation and reduce errors."*

- Subject No. 14 of Group (e), in posttest essay

*"In addition, you should know how to collocation truly or more advanced. So you need to accumulate lots of chunks. It will have a strong influence on your language expression."*

- Subject No. 18 of Group (e), in posttest essay

*"Lastly, focusing on chunks of language, also known as collocations, is a powerful learning strategy. These are groups of words that often go together, like "make a decision" or "take a shower". Learning these phrases as a whole helps us sound more natural and fluent. It also makes language learning more manageable as we're not just memorizing individual words, but meaningful units of language."*

- Subject No. 15 of Group (e), in posttest essay

*"It is necessary to grasp the foundation patiently, so, what we should do is that if we take a few hours every day to learn basic vocabulary, basic phrases or chunks and basic sentences or sentence patterns purposefully."*

- Subject No. 6 of Group (i), in posttest essay

*"The lexical chunks are a lot idiomatic combinations of words, for example, adverbs modify verbs, adverbs modify adjectives and nouns modify nouns, etc. It can help us have better understandings on the specific usage of words."*

- Subject No. 19 of Group (i), in posttest essay

It is encouraging to see that some students truly grasped the idea of top-down mode and started to give heed to larger units of language, such as chunks or collocations. This surely shows that their metaknowledge (if not explicit knowledge yet) about collocations is increasing. What is somewhat surprising is that two subjects from Group (i) also expressed similar idea, although they did not receive the explicit training. It seems that some autonomous learners could deduce the patterns from the texts and understand their significance all by themselves. It would be an interesting follow-up to approach these two students and get to know how they come to such a realization, in order to promote their strategy among all learners.

## CHAPTER 5

### CONCLUSION AND IMPLICATIONS

With all data analyses performed, now it is appropriate to draw some final conclusions. Any pattern identified in chapter 4, major or minor, will be considered here. Besides, I will have a discussion on the limitations of this study and the direction to move forward in future. Before getting into any detail, here I present the two research questions once again to refresh our memory:

*RQ1. What is the collocational competence profile of intermediate Chinese EFL learners?*

*RQ2. How can different learning modes (ER only, explicit teaching only, and the combination of the two) affect the development of their collocational competence?*

Next, we piece together every evidence corresponding to these two RQs.

#### 5.1 Conclusion

##### 5.1.1 Collocational Competence Profiling

Collocation, despite its seemingly straightforward concept, is indeed mysterious and tricky to profile. With clear criteria for collocation counting and judging (please refer to the beginning part of Chapter 4 for details), this study has strived to draw a picture of learners' collocational competence that is as comprehensive, accurate and enlightening as possible. All together six aspects are investigated: general lexical level, collocation lexical level, collocation density, collocation diversity, collocation accuracy rate, and collocation strength. For each aspect, I make two kinds of comparisons, namely, NNS\_Oral vs NNS\_Written, and NNS vs NS. The former is within the subjects based on the two genres, the result of which should reflect the difference between their implicit knowledge and the combination of implicit and explicit knowledge. The latter is between subjects and their native speaker counterparts, the difference of which should cast light on those areas that need improvement for NNS.

In terms of general lexical level, NNS\_Oral and NNS\_Written are pretty much on the same level, with the latter being slightly better by token. Therefore, overall, learners use the same level of words in their oral and written performance. When comparing NNS and NS, the gap is much more significant. NS's lexical level is apparently higher than NNS's, both in oral and in written genres. This is expected, considering the fact that the subjects in this study are at intermediate level. Undoubtedly, their lexical proficiency needs further development.

When it comes to collocation lexical level, the result is in line with the general lexical level for NS vs NNS comparison, with the former using much more advanced words to form collocations than the latter. However, the internal comparison between NNS\_Oral and NNS\_Written tells a different story. Instead of being virtually equal, as what we saw in general lexical level comparison, NNS make use of more advanced words in forming collocation in writing than in speaking. This is quite interesting. Since writing is based on the combination of one's implicit and explicit knowledge, this should mean that word knowledge and collocation knowledge are two related yet separate entities, with the latter being more difficult to develop, especially with regards to the implicit type. This calls for attention to the development of one's collocation knowledge in particular.

The third aspect is collocation density. Here I examined the normalized token and lemma of collocations in each NNS corpus and NS corpus. The internal comparison between NNS\_Oral and NNS\_Written shows a really clear trend: learners produce more collocations when they have access to their explicit knowledge across all collocation types (i.e. Adj-N, Adv-Adj, and Adv-V), both in terms of token and in terms of lemma. This highlights the importance of one's explicit collocation knowledge. When it comes to the comparison between NNS and NS, the trends are different for oral and written genres: for oral data, NNS outperform NS in token while NS outperform NNS in lemma; for written data, NS outperform NNS in both ways with a big margin. This suggests that learners should strive to expand their implicit collocation repertoire and make more diversified collocations in speaking, rather than over-repeating some



common ones. For written genre, learners produce significantly fewer collocations than their NS counterpart.

The next aspect is collocation diversity, which is measured by collocation's lemma/token ratio (LTR). For internal comparison, NNS's oral and written do not differ much, with the latter being slightly better. For comparison between NNS and NS, NS outperformed NNS in every collocation type, both for oral and for written genres, and by a wide margin, as illustrated in Adj-N type. This indicates that learners really need to diversify their collocation production and refrain from making repetitive use of the common ones. However, such an interpretation should be accepted with caution, since the overall size of NNS corpus and that of NS corpus are quite different (about 10:1), thus compromising the comparability of their LTRs.

The fifth aspect is collocation accuracy rate. This really lies in the heart of many teachers and students alike, since it directly pertains to the degree of being idiomatic. When it comes to internal comparison, very interestingly, NNS produce both more collocation and more unidiomatic combinations in writing than in speaking (in terms of lemma particularly). This indicates that by having access to their explicit knowledge, learners produce more right collocations, and at the same time, probably due to imperfect nature of such explicit knowledge (and probably combined with a willingness to take risk), they also produce more unidiomatic combinations. This is not necessarily a bad thing, because attempt and exposure usually bring about progress. When it comes to comparison between NNS and NS, NS produce higher percentage of collocation in both genres, either by token or by lemma. This is expected, and surely reveals the gap between the two groups. Learners need to improve and catch up in this area!

The last aspect is collocation strength which is measured by average MI of the identified collocations. If we look at NNS\_Oral and NNS\_Written, learners produce stronger collocations in writing than in speaking, indicating that their superior explicit collocation knowledge is at work. However, the difference is not much. When it comes to the comparison between NNS and NS, the latter outperforms the former in every

category, for both genres, both by token and by lemma. This shows that NS has a keener sense about which words should come together, and this sense needs to be developed in learners through various ways, with extensive reading being one of them.

Besides the above six aspects, some conclusions are also to be drawn based on qualitative analysis of the details of NNS's actual utterances. This is mainly to categorize the errors and to investigate the reason for making them. For overt errors, some are due to inadequate word knowledge, and some are due to inadequate collocation knowledge. Besides, there also exist some covert errors, ones that are acceptable in form but used incorrectly in context. This means that in learning collocations, students not only need to memorize the forms with all their details, but also need to learn them in real context to avoid (or at least minimize) misuse of them. The last thing is about L1 influence. Students should be made aware of the facilitative and the interfering effects of their L1, so as to keep an eye on it in their learning.

#### 5.1.2 Collocational Competence Development

Now I conclude what has been discovered regarding collocational competence development. In this experiment, I provided two treatments, namely, extensive reading and in-class training. The former is implicit in nature while the latter is explicit. The first group, Group (e+i), received both treatments; the second group, Group (e), only went through in-class training; the third group, Group (i), only did extensive reading; and the fourth group, Group (c), is the control group which received neither treatment.

The effects of both treatments have been investigated based on four parameters, namely, collocation density, collocation diversity, collocation accuracy rate, and collocation strength. An improvement is deemed to have occurred, if a group's performance in posttest is closer to NS benchmark.

In terms of collocation density, the overall trend, especially tokenwise, confirms that implicit treatment results in implicit gain (i.e. improvement in speaking) and explicit treatment results in explicit gain (i.e. improvement in writing). Also, there exists some cross-effect, meaning explicit treatment could cause moderate implicit gain and

vice versa. This bears out the interface position on implicit / explicit knowledge. Both types of treatment are valuable.

In terms of collocation diversity, learners' LTR largely dropped in posttest, being further off from NS benchmark. This is mainly due to the fact that more tokens are produced while little progress has taken place for lemma acquisition. This indicates that, compared to increase in lemma, increase in token is easier and quicker for learners. Another point worth noticing is that, the majority of LTR improvements lie in the first two experiment groups, suggesting that explicit treatment is crucial to develop one's collocation diversity.

In terms of collocation accuracy rate, the trend is very clear for unidiomatic combinations. Group (e+i) has the best result with a double drop in oral error and written error; Group (e) and Group (i) have an error rate drop in written genre but a rise in oral genre; Group (c) has a double rise, the worst result in all groups. This could mean that in order to reduce learner's collocation error, we need both types of treatment to produce the optimum result.

In terms of collocation strength, the trend is least clear. I cannot draw any conclusions at this moment, except that NS produce moderately stronger collocation than NNS in both genres and for all three collocation categories. It seems collocation strength is an area that really poses a challenge for learners, because progress in this area is difficult and/or very slow (so not showing much in seven weeks).

Lastly, from the result of qualitative analysis, we learned about the effectiveness of both treatments based on students' feedback. For extensive reading, one should provide more guidance to the students in need, especially in the early stage, to help them get on track as fast as possible. We could provide more specific recommendations for books rather than just pointing them to some general sources, so that they have a better chance of picking the right book. We could also advise students to choose a book for which they have read a translation in L1 already. This way they could better enjoy the beauty of the English language, which will probably generate bigger gain in collocation. For in-class training, a number of students demonstrated an

increase in their metaknowledge about top-down mode and importance of different levels of linguistic units. Besides, for some autonomous and talented learners, they could discover the significance of larger units of language (e.g. collocations, phrases, sentences, etc.) on their own, which is admirable.

## 5.2 Implications

In terms collocational competence profiling, I believe that the current study offers a framework that is of reference value. One should investigate at least these six aspects to assess a learner's collocational competence, and both oral and written data should be gathered, so as to get to know her respective implicit and explicit knowledge. Besides, some qualitative analysis is also beneficial to discover what numbers cannot tell. These details include learner's word knowledge, collocation knowledge, as well as how to put a collocation in actual use and avoid covert errors.

In terms of collocational competence development, this study largely bears out the interface position as well as the matching between treatment type and resultant knowledge type (Loewen, 2020:32-36). For three of the four parameters (i.e. collocation density, collocation diversity, and collocation accuracy rate), it seems that the combination of explicit treatment and implicit treatment produces the best result (i.e. biggest gain), which is in line with some scholars' suggestion of supplementing explicit instruction with implicit learning opportunities (e.g. Hulstijn, 2002). However, it is worth noting that either treatment alone is also beneficial, except implicit treatment for collocation diversity development. For collocation strength, based on the mixed result, none of the treatment seems to really work or their influence is still yet to come. How to improve in this area remains a mystery.

One of the unique strengths of this study is that, compared to most existing collocation studies, I examined learner's freely produced utterances rather than having them take an exam or fill out a questionnaire. This kind of data should best resemble what they actually do in real life, thus cast most light on their genuine collocational competence.

Lastly, I would also like to address the limitations. There are mainly three of them. The first is related to the reference corpus (i.e. COCA). COCA's collocation list for a node, unfortunately, does not provide much opportunity for users to tweak, and I have to accept it as it is. The collocates are identified based on the traditional 4:4 span, which has a discrepancy with my criteria, especially for Adj-N type (I only considered adjectives in attributive position, not predicative.) What's more, COCA does not differentiate various senses of a word, so an established collocation in the list may not apply to every sense of a polysemous node, thus making it possible to overlook a covert error. The second limitation is that the size of the NS corpora is relatively small (about 3000 tokens for oral and written genres, respectively). Therefore, collocation patterns identified from these NS corpora, especially those related to the rare Adv-Adj type, remain to be confirmed when larger NS corpora could be composed and examined. The last limitation is related to the data which were used to build the corpora. Although I made every effort to design a procedure that is as rigorous as possible, there was no control on NS's topics. In other words, the utterances of NS could be of any topics, not matching those of NNS's. This compromises the comparability of the two sets of data (i.e. NS data and NNS data), and future researchers could make improvement on this aspect.

### 5.3 Final Thought

Language is both magical and mysterious. To a large extent, it defines who we are and how we perceive the cosmos with everything in it. Linguistic barriers, along with cultural barriers associated with it, are arguably one of the most challenging to overcome. For any second language learner, she has embarked on a truly rewarding journey, and researchers and teachers alike bear the responsibility to help her develop faster and further. Engaging in this great endeavor, the current study, with its focus on an indispensable aspect of language (i.e. collocation), aims to make its own feeble contribution to the conquering of that mighty mountain. I welcome anyone to follow, to criticize, to join, and to inspire, so that we can march forward together until mastering a second language becomes truly attainable for any aspiring human being.

## REFERENCES



- Alangari, M. A. (n.d.). A corpus-based study of verb-noun collocations and verb complementation clause structures in the writing of advanced Saudi learners of English (dissertation).
- Anthony, L. (2022). AntConc (Version 4.2.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Anthony, L. (2023). AntWordProfiler (Version 2.1.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Anthony, L. (2022). TagAnt (Version 2.0.5) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, 26, 28–41. <https://doi.org/10.1016/j.jslw.2014.09.004>
- Bialystok, E. (1978). A theoretical model of Second language learning1. *Language Learning*, 28(1), 69–83. <https://doi.org/10.1111/j.1467-1770.1978.tb00305.x>
- Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics*, 14(3), 275–311. <https://doi.org/10.1075/ijcl.14.3.08bib>
- Biber, D., Conrad, S. (1999). Lexical bundles in conversation and academic prose. In: Hasselgard, H., Oksefjell, S. (eds) *Out of Corpus: Studies in Honour of Stig Johanson*. Netherlands: Rodopi B. V. 181-89.
- Boers, F., & Lindstromberg, S. (2005). Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System*, 33(2), 225–238. <https://doi.org/10.1016/j.system.2004.12.007>
- Bolinger, D. (1975). *Aspects of language* (2nd edition). New York: Harcourt Brace Jovanovich.
- Carroll, S., Swain, M., & Roberge, Y. (1992). The role of feedback in Adult Second Language Acquisition: Error Correction and morphological generalizations. *Applied Psycholinguistics*, 13(2), 173–198. <https://doi.org/10.1017/s0142716400005555>

- Celce-Murcia, M., Brinton, D., & Snow, M. A. (2014). Teaching English as a second or foreign language. National Geographic Learning.
- Chang, Y.-C., Chang, J. S., Chen, H.-J., & Liou, H.-C. (2008). An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology. *Computer Assisted Language Learning*, 21(3), 283–299.  
<https://doi.org/10.1080/09588220802090337>
- Chen, A. C.-H. (2021). Acquisition of L2 Collocational competence: A Corpus Analysis of Exclusivity, Directionality, Dispersion and Novel Usage. *Taiwan Journal of TESOL*, 18(1), 29–61. [https://doi.org/10.30397/TJTESOL.202104\\_18\(1\).0002](https://doi.org/10.30397/TJTESOL.202104_18(1).0002)
- Chen, W. (2017). Profiling collocations in EFL writing of Chinese tertiary learners. *RELC Journal*, 50(1), 53–70. <https://doi.org/10.1177/0033688217716507>
- Chen, Y. (2016). Dictionary use for collocation production and retention: A call-based study. *International Journal of Lexicography*. <https://doi.org/10.1093/ijl/ecw005>
- Chomsky, N. (1965). *Aspects of the theory of syntax*. The M.I.T. Press.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use*. Praeger.
- Corder, S. P. (1974). 'Error analysis' in Allen and Corder (eds.) 1974.
- Cowie, A. P. (2013). *Semantics*. Oxford University Press.
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238  
<https://www.jstor.org/stable/3587951>
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2014). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*.  
<https://doi.org/10.1093/applin/amt056>
- Darvishi, S. (2011). The investigation of collocational errors in university students' writing majoring in English. *Proceedings of International Conference on Education, Research and Innovation*, 18, 52–56.
- Davies, M. (2008-) *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Day, R. R. (2015). Extending extensive reading. *Reading in a Foreign Language*, 27(2), 294-301.



- Day, R. R., & Bamford, J. (2002). To ten principles for teaching extensive reading. *Reading in a Foreign Language*, 14, 136-141.
- DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–112). Mahwah, NJ: Lawrence Erlbaum Associates.
- Dellar, H., & Walkley, A. (2017). *Teaching lexically principles and Practice*. Ernst Klett Sprachen GmbH.
- Doughty, C. (1991). Second language instruction does make a difference. *Studies in Second Language Acquisition*, 13(4), 431–469.  
<https://doi.org/10.1017/s0272263100010287>
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL - International Review of Applied Linguistics in Language Teaching*, 47(2). <https://doi.org/10.1515/iral.2009.007>
- EI-Dakhs, D. (2015). Collocational competence in English language teaching: An overview. *Arab World English Journal*, 6(1), 68–82.  
<https://doi.org/10.24093/awej/vol6no1.5>
- Ellis, N. C. (2007). The weak interface, consciousness, and form-focused instruction: Mind the doors. In H. Nassaji & S. Fotos (Eds.), *Form-focused instruction and teacher education: Studies in honour of Rod Ellis* (pp. 17–34). Oxford: Oxford University Press.
- Ellis, R. (1990). A response to Gregg. *Applied Linguistics*, 11(4), 384–391.  
<https://doi.org/10.1093/applin/11.4.384>
- Ellis, R. (1996). *The study of Second language acquisition*. Oxford University Press.
- Ellis, R. (2005). Principles of instructed language learning. *System*, 33(2), 209–224.  
<https://doi.org/10.1016/j.system.2004.12.006>
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28, 339–368. doi:10.1017/S0272263106060141.

- Evert, S. (n.d.). *The statistics of word Cooccurrences word pairs and collocations* (thesis).
- Fan, M. (2009). An exploratory study of collocational use by ESL students – a task-based approach. *System*, 37(1), 110–123.  
<https://doi.org/10.1016/j.system.2008.06.004>
- Ferraro, G., Nazar, R., Alonso Ramos, M., & Wanner, L. (2013). Towards advanced collocation error correction in Spanish learner corpora. *Language Resources and Evaluation*, 48(1), 45–64. <https://doi.org/10.1007/s10579-013-9242-3>
- Firth, J. R. (1969). *Papers in linguistics: 1934-1951*. Oxford University Press.
- Foster, P. (2001). Rules & routines: a consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan & M. Swain (eds.) *Researching pedagogic tasks: second language learning, teaching and testing*. London, New York: Longman, 75-94.
- Friginal, E. (2018). *Corpus linguistics for English teachers: New tools, online resources, and classroom activities*. Routledge.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(S1), 155–179. <https://doi.org/10.1111/lang.12225>
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. *Phraseology: Theory, analysis, and applications*, 145, 160-175.
- Harmer, J. (2015). *The practice of English language teaching*. Pearson/Longman.
- Henriksen, B. (2013). Research on L2 learners' collocational competence and development - A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *Vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis* (pp. 29–56). essay, Eurosla Mnographs Series 2.
- Hill, D. R. (2013). Survey review: Graded readers. *ELT Journal*, 51(1), 57–81.  
<https://doi.org/10.1093/elt/51.1.57>
- Howarth, P. (1998). Phraseology and Second language proficiency. *Applied Linguistics*, 19(1), 24–44. <https://doi.org/10.1093/applin/19.1.24>

- Hsu, J. (2002). Development in collocational proficiency in a workshop on English for general business purposes for Taiwanese college students (dissertation).
- Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, 18, 193-223. <https://doi.org/10.1191/0267658302sr207oa>.
- Hymes, D. (1971). *On Communicative Competence*. Philadelphia, P.A.: University of Pennsylvania Press.
- James, C. (1998). Errors in language learning and use: Exploring error analysis. New York: Addison Wesley Longman Limited.
- Jiang, N., & Nekrasova, T. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91(3), 433–445. <https://doi.org/10.1111/j.1540-4781.2007.00589.x>
- Kim, S. H., & Kim, J. H. (2012). Frequency effects in L2 multiword unit processing: Evidence from self-paced reading. *TESOL Quarterly*, 46(4), 831–841. <https://doi.org/10.1002/tesq.66>
- Krashen, S. D. (1982). *Principles and practice in Second language acquisition*. Pergamon Press.
- Krashen, S. D. (2003). Explorations in language acquisition and use: The Taipei lectures. Heinemann.
- Langacker, R. W. (2000). 'A dynamic usage-based model' in M. Barlow and S. Kemmer (eds): *Usage-Based Models of Language*. The Center for the Study of Language and Information Publications, pp. 1-64.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of tasked-induced involvement load. *Applied Linguistics*, 22(1), 1–26. doi:10.1093/applin/22.1.1.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' english. *Language Learning*, 61(2), 647–672. <https://doi.org/10.1111/j.1467-9922.2010.00621.x>
- Lennon, P. (1998). Approaches to the teaching of idiomatic language. *International*

- Review of Applied Linguistics in Language Teaching*, 36(1).  
<https://doi.org/10.1515/iral.1998.36.1.11>
- Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Language Training Publications.
- Lewis, M. (1997). *Implementing the lexical approach putting theory into practice*. LTP.
- Liu, L. (2002). *A corpus-based lexical semantic investigation of verb-noun miscolllocations in Taiwan Learners' English* (thesis). Tamkang University, Graduate Institute of Western Languages and Literature, Taipei.
- Loewen, S., Erlam, R., & Ellis, R. (2009). The incidental acquisition of third person *s* as implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & R. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 262–280). Bristol, UK: Multilingual Matters.
- Loewen, S. (2020). *Introduction to instructed Second language acquisition*. Routledge, Taylor & Francis Group.
- Long, M. (1991). Focus on Form. *Studies in Bilingualism*, 39.  
<https://doi.org/10.1075/sibil.2.07lon>
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.
- Long, M. (2017). Instructed second language acquisition (ISLA): Geopolitics, methodological issues, and some major research questions. *Instructed Second Language Acquisition*, 1(1), 7–44. doi:10.1558/isla.33314.
- Lyutaya, T. (2011). Reading logs: Integrating ER with writing tasks. *English Teaching Forum*, 49, 26-34.
- McCarthy, M., & O'Dell, F. (2017). *English collocations in use: How words work together for fluent and natural English: Self-study and classroom use*. Cambridge University Press.
- Nakanishi, T. (2014). A meta-analysis of extensive reading research. *TESOL Quarterly*, 49(1), 6–37. <https://doi.org/10.1002/tesq.157>

- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nesselhauf, N. (2003). The use of collocations by Advanced Learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223–242.  
<https://doi.org/10.1093/applin/24.2.223>
- O'Donnell, M. B., Römer, U., & Ellis, N. C. (2013). The development of formulaic sequences in first and second language writing. *Current Issues in Phraseology*, 18(1), 83–108. <https://doi.org/10.1075/ijcl.18.1.07odo>
- Peng, X., Butler, Y. G., Frei, C., & Hornberger, N. H. (n.d.). *Use of verb-noun collocations by advanced learners of Chinese* (dissertation).
- Pirmoradian, M., & Tabatabaei, O. (2012). The enhancement of lexical collocation learning through concordancing: A case of Iranian EFL learners. *The Modern Journal of Applied Linguistics*, 4(4), 185–200.
- Pu, J. (2003). Colligation, collocation, and chunk in ESL vocabulary teaching and learning. *Foreign Language Teaching and Research*, 35(6), 438–445.
- Reber A. (1967) Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81(1) : 115-119.
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, 63(3), 595–626. doi:10.1111/lang.12010.
- Saeed, J. I. (1997). *Semantics*. Blackwell.
- Schachter, J. (1974). 'An error in error analysis'. *Language Learning* 27:205-14
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? *Language Learning & Language Teaching*, 127–151.  
<https://doi.org/10.1075/llt.9.08sch>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics* 11/2.
- Sinclair, J. (1991). *Corpus, concordance, Collocation*. Oxford Univ. Press.

- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of Collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64(3), 429–458. <https://doi.org/10.3138/cmlr.64.3.429>
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase “Time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 776–784. <https://doi.org/10.1037/a0022531>
- Siyanova-Chanturia, A. (2015). On the ‘holistic’ nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, 11(2). <https://doi.org/10.1515/cllt-2014-0016>
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263–308. doi:10.1111/j.1467-9922.2010.00562.x.
- Sun, H. (2004). A corpus-based study of semantic characteristics of adjective collocations in CLEC. *Modern Foreign Languages*. 27 (4). 410-418
- Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, 38(5), 1229–1261. doi:10.1017/S014271641700011X.
- Swan, M. (2006). Chunks in the classroom: let's not go overboard. *The Teacher Trainer*, 20(3).
- Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, 37(4), 529–540. <https://doi.org/10.3758/mc.37.4.529>
- Tarone, E. E. (1990). On variation in Interlanguage: A response to Gregg. *Applied Linguistics*, 11(4), 392–400. <https://doi.org/10.1093/applin/11.4.392>
- Taylor, D. S. (1988). The meaning and use of the term 'competence' in linguistics and applied linguistics. *Applied Linguistics*, 9(2), 148–168. <https://doi.org/10.1093/applin/9.2.148>
- Timmis, I. (2008). The lexical approach is dead: long live the lexical dimension. *Modern English Teacher* 17/3.

- Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, 69(2), 405–439. <https://doi.org/10.1111/lang.12335>
- Trahey, M., & White, L. (1993). Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition*, 15(2), 181–204. doi:10.1017/S0272263100011955.
- Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, 7(2), 215–244. <https://doi.org/10.1075/ijcl.7.2.04stu>
- Underwood, G., N. Schmitt & A. Galpin. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (ed.), *Formulaic sequences*, 153-172. Amsterdam, the Netherlands: John Benjamins.
- Ur, P. (2011). Grammar teaching: Research, theory, and practice. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning: Volume II* (pp. 507–522). New York, NY: Routledge.
- Vilkaitė, L., & Schmitt, N. (2017). Reading collocations in an L2: Do collocation processing benefits extend to non-adjacent collocations? *Applied Linguistics*, 40(2), 329–354. <https://doi.org/10.1093/applin/amx030>
- Vilkaitė, L., & Schmitt, N. (2019). Frequency as a Guide for Vocabulary Usefulness. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*. essay, Routledge.
- Vu, D. V., & Peters, E. (2022). A longitudinal study on the effect of mode of reading on incidental collocation learning and predictors of learning gains. *TESOL Quarterly*, 57(1), 5–32. <https://doi.org/10.1002/tesq.3111>
- Walker, C. P. (2011). A corpus-based study of the linguistic features and processes which influence the way collocations are formed: Some implications for the learning of collocations. *TESOL Quarterly*, 45(2), 291–312. <https://doi.org/10.5054/tq.2011.247710>
- Wang, Y. & Shaw, P. (2008). Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME journal*, 32, 201-232.

- Webb, S., & Chang, A. C.-S. (2020). How does mode of input affect the incidental learning of collocations? *Studies in Second Language Acquisition*, 44(1), 35–56. <https://doi.org/10.1017/s0272263120000297>
- Wen, Q. (1995). On strategy differences between a successful learner and a not-so successful learner. *Foreign Language Teaching and Research*, 61-66.
- West, M. (1953). *A General Service List of English Words*. London: Longman, Green and Co.
- Widdowson, H. G. (1983). *Learning purpose and language use*. Oxford University Press.
- Willis, D. (1990). *The lexical syllabus: A new approach to language teaching*. Harper Collins.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Xu, Z. (2019). Practices of Teaching Englishes: Pedagogical Reflections and Implications. In F. Fang & H. P. Widodo (Eds.), *Critical Perspectives on Global Englishes in Asia: Language policy, curriculum, pedagogy and assessment*. essay, *Multilingual Matters*.
- Zaabalawi, R. S., & Gould, A. M. (2017). English collocations: A novel approach to teaching the language's last bastion. *Ampersand*, 4, 21–29. <https://doi.org/10.1016/j.amper.2017.03.002>
- Zhang, W., & Li, X. (2016). Research on the Application of Explicit and Implicit Reinforcement Theory in the Teaching of College English Lexical Chunks. *Journal of Hubei Correspondence University*, 29(1), 167–168.





APPENDIX



APPENDIX A  
EXPLICIT TRAINING PLAN

This appendix explains the explicit treatment (i.e. in-class collocation trainings) in detail. Please note that only Group (e+i) and Group (e) receive such treatment according to the design of this study.

### < Week 1 >

#### Theme:

Resource Introduction

#### Content:

In this session, the researcher introduces some quality resources related to collocation learning. This is to help them foster a self-driven learning style and know where to seek help when problems arise.

##### a) Paper-based resources:

Dictionaries are the first and foremost resource a learner needs. As Harmer (2015) suggests, a good learner's dictionary should indicate how frequent a word is and provide its related collocation information. Therefore, I would recommend two dictionaries that are particularly helpful to the participants, especially when it comes to collocation learning. The first is *Oxford Advanced Learner's Dictionary*, the best seller in the world for English learners. Its entries are concise and with great clarity. Besides, it contains sporadic column for common yet confusing collocations for learners, such as the one on page 764 about housing moving related ones.

Another dictionary of choice is *Oxford Collocations Dictionary*. This one is dedicated to providing typical collocates of common words. On the other hand, it does not render any definition, with the assumption that users already know what the word means, but just unsure about what its collocates are. Participants could conveniently search a word and scan through the list of its common collocates. Additionally, both dictionaries have respective apps, although they do not come free of charge. Participants are recommended to get these apps as long as they are affordable to them. This way it will be much more convenient to keeping learning on the go.



Figure 21 Recommended Resources – Dictionaries (self-made)

b) *Online resources:*

1. extensive reading website: [www.er-central.com](http://www.er-central.com)

This website is the major resource that I would recommend to the participants. Not only does it provide ample essays of various topics for reading, but it also allows users to self-gauge their reading levels. In this way, they can do a filter to select essays that are appropriate for their current level. This is a very useful function for them to receive bountiful comprehensible input.



Languages: English



Username

Password

Login

Register as a Student, or Teacher / Institution, or Contributor | How to Register? | Forgot Password

**STUDENT**

Welcome to Extensive Reading Central.  
Learn words, Read and Listen in English for FREE!

**Read**

- ✓ 1000's of easy texts
- ✓ 20 Levels
- ✓ Speed reading practice

**Learn Words**

- ✓ Word Learning Games
- ✓ Make your own wordlists

**TEACHER**

**Listen**

- ✓ Over 1000 texts
- ✓ Read and Listen
- ✓ Comprehension check

**Test Helper**

**ER-Central guide**

**Word Learner**

Welcome to Word Learner  
- the fastest way to learn words

- ▶ Learn Words quickly using modern scientific methods
- ▶ Learn Words set by your teacher
- ▶ Use the Dictionary to find words and save them to learn later
- ▶ Create your own wordlists
- ▶ Creates tests to check your learning
- ▶ Learn in English, Japanese, Thai

**LIBRARY**

Search the texts

By Title

Search by level

17.0

Search

Categories

Factual

Story

Children

Student writings

New Releases

Strange Friends

Broken Life

Meru

I miss you so much

First 1 2 3 >> Last

<p>Living with other people</p> <p>17.60</p>	<p>I like Music!</p> <p>17.50</p>	<p>Buddhism</p> <p>17.70</p>	<p>Things You Didn't Know About Diamonds</p> <p>17.10</p>	<p>The North Wind and the Sun</p> <p>17.50</p>
<p>Getting to Know More about Sushi</p> <p>17.10</p>	<p>The Gorilla and the Crocodile</p> <p>17.80</p>	<p>EQ</p> <p>17.60</p>	<p>The Misunderstood Tomato</p> <p>17.70</p>	<p>Love in the Hospital</p> <p>17.30</p>
<p>Lie Detectors</p> <p>17.90</p>	<p>American Breakfast</p> <p>17.40</p>	<p>Flower Power</p> <p>17.90</p>	<p>Butter</p> <p>17.80</p>	<p>The Day the Science Teacher Disappeared</p> <p>17.30</p>

**Reading Library**

Figure 22 Recommended Resources – ER Website (self-made)

The screenshot displays an educational platform interface. At the top left, there is a 'Search by level' dropdown menu set to '12.0' and a 'Search' button. Below this is a 'Categories' sidebar with options like 'Factual', 'Art / Music', 'Biography / People', 'Business / Economics', 'Cars and transportation', 'Crime', 'Culture', 'Entertainment', 'Food and recipes', 'Health and Body', 'History', and 'Language and education'. The main content area shows a grid of text cards. A pop-up window for the text 'Ready for College?' by CompassMedia is open. The pop-up displays the following information:

- Genre:** Factual
- Sub Genre:** Language and education
- Level:** 12.00
- Words:** 273
- Read:** 0%
- Comprehension:** 4 stars (out of 5)
- Difficulty:** 4 stars (out of 5)
- Enjoyability:** 4 stars (out of 5)
- Stats:** 495 Completions & 399 Downloads
- Buttons:** 'Bookmark it!' and 'Read Text'

The pop-up also includes a lead-in introductory text: 'In the US, high school students have the freedom to choose most of their courses. How has this freedom affect their readiness for college?' and a 'Quiz' button.

Once learners click one of the texts, a mini-window will pop up which shows its genre, level, length as well as an lead-in introductory word. They can bookmark it if they would like to get quick access in future. By clicking "Read Text", one can begin to digest the text.

Total hours reading: 0 hours 0 minutes 54 seconds

Select a word and right click to learn it.



## Ready for College?

- By CompassMedia

Year after year, results on college entrance exams point out a problem. The problem is with students today. It seems that more and more students in the US are not ready for college. That is, they are not ready by the time they finish high school.

Click to save this word

- **Base**

**Base** -> *noun* -> beɪs -> Origin or start from which something came -> *The base of my suggestions come from a careful analysis of the findings*

high school students do not have a good enough base to succeed in college.

Some believe the problem comes from too much freedom. In the US, high school students can choose many of their courses. Students are free to take easy courses while avoiding the hard ones. They can also sign up for fun courses rather than "boring" courses.

In the reading interface, the website offers several handy features. On the top, it automatically monitors how much time you have spent on reading this article, which helps the reader to track his own pace. The text itself is laid out plainly and clearly, preceded by the title and the name of the author. While reading, should the reader want to check the meaning of a certain word, it is just one-click away. Just move the mouse on the word and right click, then a quick note will pop up, with information such as part of speech, pronunciation, gloss, and an exemplary sentence. The reader could even save the word for future reference and review.

2. collocation dictionary online: [www.freecollocation.com](http://www.freecollocation.com)

This is an abridged online version of the aforementioned *Oxford Collocations Dictionary*. It does not provide all the details of the original version, but is still a valuable reference for quick search and learning. Plus, it is free!

### Online Collocation Dictionary

A completely new type of dictionary with word collocation that helps students and advanced learners **effectively study, write and speak natural-sounding English**. This *online dictionary* is very helpful for the education of the IELTS, TOEFL test.

**Level:** Upper-Intermediate to Advanced

**Key features** of oxford dictionary online

1. Collocations/collocation - common word combinations such as *'bright idea'* or *'talk freely'* - are the essential building blocks of natural-sounding English. The dictionary contains over 150,000 collocations for nearly 9,000 headwords.
2. The dictionary shows words commonly used in combination with each headword: nouns, verbs, adjectives, adverbs, and prepositions, common phrases.
3. The collocation dictionary is based on 100 million word British National Corpus.
4. Over 50,000 examples show how the collocation/collocations are used in context, with grammar and register information where helpful.
5. The clear page layout groups collocations according to part of speech and meaning, and helps users pinpoint speedily the headword, sense and collocation they need.
6. [Free Download -- OXFORD Collocations Dictionary](#)
7. [Document image tool](#) -- [Free photo repair](#)[Free document conversion](#)[More](#)

[Google Dictionary](#) | [Wordnet Online](#) | [English Test Info](#) | [IELTS English Test](#) | [Collins Dictionary](#) | [谷歌词典中英文](#) | [IELTS Lessons](#) | [Practice tests](#) | [English exams](#) | [Document image tool](#) | [This site is supported by Send me an email.](#)

Welcome to my site from China. Provided by [IPLocator](#)

Figure 23 Recommended Resources – Online Collocation Dictionary

For example, if a learner is not sure whether “strong engine” is a conventional collocation in English, he could use the noun “engine” as the node and do a search in the search box, and the following page will show up:

**Online OXFORD Collocation Dictionary**

engine

**engine** *noun*

<sup>1</sup> **part of a vehicle that produces power**

**ADJ.** **big, powerful** | **small** | **twin** ~s *a large plane with twin engines* | **1.4-litre, 1200cc, 20-valve, four-cylinder, two-stroke, etc.** | **diesel, internal-combustion, jet, outboard, petrol, piston, turbine, turbo/turbocharged** | **aircraft, car, rocket**

**VERB + ENGINE** **crank (up), start, switch on** | **cut** (*informal*), **kill** (*informal*), **switch off** *He pulled up under some trees and cut the engine.* | **rev (up), run** *She sat at the traffic lights revving the engine.* | **repair, service, tune** | **lubricate** | **fit (sth with)** *The new model is fitted with a more powerful engine.*

**ENGINE + VERB** **run** *She waited with the engine running while he bought a paper. The engine runs on unleaded petrol.* | **idle, tick over, turn over** *The engine was just ticking over.* | **catch, start** *I pressed the starter and the engine caught first time.* | **stop** | **fire** *The engine's firing on all four cylinders now.* | **break down, die, fail, misfire, overheat, seize up, stall** | **cough, splutter** *The engine coughed and died.* | **roar, scream** *The plane's engine roared as it prepared for take-off.* | **race, rev (up)** *He heard a car engine racing behind him.* | **power sth** *This model is powered by a 1.8-litre petrol engine.*

**ENGINE + NOUN** **capacity, power, speed** | **compartment, room** *the ship's engine room* | **component** | **failure, problems, trouble** *It looks as if we've got a spot of engine trouble.* | **noise**

**PREP** **in an/the** ~ *You need more oil in the engine.*

**PHRASES** **the noise/roar/sound of the engine**

<sup>2</sup> **vehicle that pulls a train**

**ADJ.** **large, powerful** | **diesel, electric, steam** | **railway, tank**

**VERB + ENGINE** **build**

**ENGINE + VERB** **break down, fail** *The engine broke down just outside the station.*

**ENGINE + NOUN** **driver** | **failure** | **speed** | **shed**

In the section of “ADJ.” list the common adjective collocates of the node, such as *big* and *powerful*. Thus, one should know by now that *powerful engine*, rather than *strong engine*, is the conventional way of saying it.

3. COCA (registration needed) [www.english-corpora.org/coca](http://www.english-corpora.org/coca)

In general, corpus as a tool is new to undergraduates. However, it is worthwhile to expose them to this valuable tool at this stage of their learning. In terms of collocation, participants will get to know the real occurrences of



the node and what collocates it generally keeps in a single click. It should be extremely informative for them to dig deeper into the nuances of each word they are interested in.

The screenshot displays the COCA interface for the word "render". The top navigation bar includes "SEARCH", "WORD", "CONTEXT", and "OVERVIEW". The "WORD" tab is active, showing the word "render" (VERB) with a frequency of #3502. A bar chart shows the distribution of "render" across various genres: BLOG, WEB, TV/M, SPOK, FIC, MAG, NEWS, and ACAD. The definition provided is: "1. cause to become 2. provide or furnish with 3. give an interpretation or rendition of". Below the definition are links for "YouGlish", "PlayPhrase", and "Yarn", along with "ZH-CN" links for "Baidu" and "WordRef".

**SYNONYMS** ( > CONCEPT)      **NEW:** DEFIN +SPEC +GENL

decide	decide, declare, decree, judge, render	melt down
condense	extract, purify, reduce, render, solidify	portray
execute	perform, portray, render, represent, translate	provide

bestow, deliver, give, provide, render, submit, supply

**CLUSTERS** (more)

render •	rendered in • render it • renders it • rendered by • render them • rendering it • rendering them • rendered it
• render	would render • services rendered • has rendered • have rendered • will render • can render • not render • had rendered
render ••	render a decision • render an opinion • render a verdict • rendered obsolete by • rendered useless by • rendered even more • render the words • rendered irrelevant by

**TOPICS** (more)  
namely, embody, ie, conception, consequently, inherent, correspond, explicit, aesthetic, eg, rendering, remark, testament, narrative, explicitly, regard, contrary, accordingly, attain, critique

**COLLOCATES** (more)

**NOUN** service, decision, judgment, verdict, image, fat, opinion, version

**VERB** moot, paralyze, parse, incapacitate, sculpt, unfit, engross, crisp

**ADJ** useless, obsolete, invisible, impossible, meaningless, incapable, irrelevant, unable

**ADV** less, thus, beautifully, thereby, effectively, virtually, essentially, carefully

**RELATED WORDS**  
rendering

Figure 24 Recommended Resources – COCA (self-made)

We illustrate by using the same example as above. If a learner wants to know whether *strong engine* is an established collocation, she could search the node *engine* under the tab "Word", to see a wealth of its relevant information.

Corpus of Contemporary American English

SEARCH WORD CONTEXT ACCOUNT

List Chart **Word** Browse +

engine [POS]?  
See detailed info for word Reset

(HIDE HELP) NO LICENSE

You can find a wealth of information for the top 60,000 words in the corpus. As the following examples with *bread* show, you can see:

- an [overview](#) of all of the information below
- related [topics](#) (words that co-occur *anywhere* on the web page)
- [collocates](#) (automatically grouped by part of speech)
- [clusters](#) (the most frequent 2, 3, and 4 word strings)
- a resortable [Keyword in Context](#) (concordance) display
- [related words](#) (synonyms and WordNet entries), and
- [websites](#) use that word the most (can use these to create Virtual Corpora).

In this central page for the word in question, one could find numerous valuable information, such as genre, meaning, pronunciation, topics, collocates, related words, clusters and so on. We shall explore more by clicking the COLLOCATES label.

Corpus of Contemporary American English

SEARCH **WORD** CONTEXT ACCOUNT

See in IWeb Collocates Clusters Topics Texts KWIC HELP

**engine** (NOUN) #1796

BLOG WEB TV/M SPOK FIC MAG NEWS ACAD

1. motor that converts thermal energy to mechanical work 2. something used to achieve a purpose 3. wheeled vehicle consisting of a self-propelled engine used to draw trains along railway tracks  
D M O C G E

YouGlish PlayPhrase Yarn  
ZH-CN: Google WordRef Reverso Linguee

**SYNONYMS** ( ▶ CONCEPT) **NEW:** DEFIN +SPEC +GENL  
locomotive locomotive, train

**CLUSTERS** (more)

engine •	engine room • engine for • engine in • engine optimization • engines in • engine with • engines for • engine on
• engine	search engine • search engines • steam engine • combustion engine • jet engine • diesel engine • diesel engine • jet engines
engine ••	engine that could • engine of growth • engine that drives • engine of economic • engines such as • engine and transmission • engine does n't • engine that can
•• engine	in the engine • internal combustion engine • started the engine • to the engine • off the engine • with the engine • on the engine • from the engine

**TOPICS** (more)  
fuel, rear, vehicle, brake, car, motor, search, wheel, suspension, speed, automobile, cargo, aluminum, transmission, passenger, engineer, hood, electric, steer, diesel

**COLLOCATES** (more)

NOUN search, car, diesel, jet, steam, fire, combustion, fuel  
VERB shut, power, search, fire, rev, roar, idle, gun  
ADJ internal, economic, powerful, main, four-cylinder, electric, efficient, standard  
ADV efficiently, smoothly, continuously, aft, eg, astern, electronically, horizontally

**RELATED WORDS**  
engineer, engineer, single-engine, twin-engine, twin-engined

Within a few seconds, several lists of possible collocates of *engine* will show up, each of which corresponds to a certain part of speech category.

Corpus of Contemporary American English

SEARCH WORD CONTEXT ACCOUNT

COLLOCATES ENGINE NOUN

Advanced options

+ NOUN	NEW WORD	?	+ ADJ	NEW WORD	?	+ VERB	NEW WORD	?	+ ADV	NEW WORD	?
6444	7.83	search	593	4.99	internal	1371	2.54	start	39	2.23	loud
1265	3.30	car	512	2.86	economic	806	2.05	turn	38	2.22	automatically
934	8.39	diesel	387	3.62	powerful	794	2.06	run	25	3.06	efficiently
892	2.61	room	375	3.14	main	460	2.53	drive	19	2.82	smoothly
892	6.21	jet	277	9.86	four-cylinder	409	3.55	shut	18	2.41	loudly
837	7.15	steam	240	4.22	electric	401	2.09	build	18	2.74	continuously
817	3.79	fire	155	3.95	efficient	391	2.22	cut	18	5.54	aft
767	9.31	combustion	122	2.66	standard	333	6.51	power	15	2.54	ie
657	2.19	power	101	9.72	turbocharged	318	3.84	search	15	3.71	eg
547	4.63	fuel	99	3.22	marine	280	3.23	fire	13	6.99	astern
530	3.79	gas	90	10.81	internal-combustion	267	9.28	rev	12	3.32	electronically
454	8.54	optimization	81	2.85	advanced	262	6.22	roar	12	3.97	horizontally
449	6.41	gasoline	81	9.81	six-cylinder	173	2.11	design	12	4.19	offline
380	5.15	rocket	79	2.97	conventional	170	2.50	blow	11	2.35	front
377	2.95	sound	74	6.22	unreal	164	2.59	burn	8	2.02	specially
376	2.61	oil	73	3.00	loud	162	2.39	replace	8	2.46	manually
355	2.99	growth	70	3.52	automatic	155	7.68	idle	7	2.19	commercially
350	3.60	plane	70	3.89	rear	144	6.85	gun	7	6.57	amidships
341	4.34	noise	70	9.55	two-stroke	117	2.47	generate	5	2.18	furiously
321	5.47	transmission	68	2.52	electronic	114	2.19	operate	5	2.55	technologically
319	3.18	web	68	3.98	mighty	102	2.94	switch	5	2.80	madly
315	3.52	truck	66	4.58	hybrid	98	5.81	crank	5	2.84	eerily
273	4.92	aircraft	62	2.49	smooth	97	3.49	tire	4	2.32	effortlessly
270	3.38	speed	55	4.05	optional	91	3.42	cool	4	2.38	alongsid

As we scan the list of +ADJ, the collocate *strong* is nowhere to be found while *powerful* is positioned among the top choices. If we scan further down, *mighty* is another accepted word to modify *engine*. Therefore, the learner should know that \**strong engine* is not the usual expression, but *powerful engine* and *mighty engine* are.

## &lt; Week 2 &gt;

## Theme:

Schema Modification

## Content:

The researcher would first have the students reflect on their own English learning experience and ask them what their beliefs are in learning an L2. Based on this, I would present and comment on three common myths among Chinese learners of English:

- ◆ *Myth #1 - "Memorizing more words will lead to higher proficiency."*

*Comment:* In fact, the relationship between two languages is much more than a system of word-word correspondences. Any correspondence, at best, is a rough approximation. Also, the mastery of vocabulary has multiple levels: while active words are always readily available, the passive ones are accessible only in reception.

- ◆ *Myth #2 - "Grammatically correct equals perfectly idiomatic."*

*Comment:* grammars, which in essence are rules, only accounts for half of the story of language acquisition. The other half which could hardly be explained by any rule is the "feel" about the language. And it is this "feel" that is the key to idiomaticity.

- ◆ *Myth #3 - "Words are the only building blocks of a language."*

*Comment:* According to Sinclair (1991), there are two modes of language processing, namely, the open-choice principle and the idiom principle. In order to acquire a language properly, both principles need to develop in a learner's mind. Therefore, students are advised to pay attention to all kinds of units, especially the bigger ones, such as chunks (including collocations). Some examples from the textbook will be identified and explained.

This would be followed by a list of several typical “Chinglish” expressions from their own essays, such as: “*I very like it*” / “*We should pay more time to read*” / “*Books are the most constant friends*”. All of these illustrate one point: merely doing rote memorization of words and adopting mental translation will result in unidiomatic even incorrect expressions. As such experience is common among the learners, their way of English acquisition is clearly the undesired bottom-up mode.

Based on the above elaboration, two contrastive formulae are introduced:

- #1: word memorization + grammar learning + Chinese thinking = fancy Chinglish
- #2: chunk memorization + grammar learning + English thinking = fancy English

Apparently, the second is to be followed. Therefore, the participants are advised to switch their learning style from bottom-up to top-down mode. They will be better off if they habitually acquire bigger units such as collocations, avoid mental translation, and think directly in English.

## &lt; Week 3 &gt;

**Theme:**

Collocation Awareness Raising – Pervasiveness

**Content:**

A lead-in activity of collocation identification would be conducted with a certain paragraph to make the participants aware that collocations are pervasive. As Nation (2001) claims, one's linguistic knowledge is essentially his collocational knowledge. Therefore, it is vital for any serious learner to develop his or her collocational competence.

To make its importance crystal clear, three benefits of learning collocations are explained: your English will be more versatile, more accurate, and potentially receive higher scores in tests. In a word, you will sound more idiomatic. That means learning collocations well can partly help you cross the gap between intermediate learners and truly advanced users like native speakers.

Additionally, four typical types of collocations are introduced, namely, verb + noun, adjective + noun, verb + adverb, and adverb + adjective. Participants are provided with examples of each type from their course textbook for illustration. As a result, starting from this week, they are required to pay attention to collocations in their after-class extensive reading activity, and note down at least one example for each type they encounter. This is to help them form a daily habit of noticing and acquiring collocations. Also, this practice should reinforce their impression of the pervasive nature of this linguistic aspect.

## &lt; Week 4 &gt;

## Theme:

Collocation Awareness Raising – Arbitrary Convention

## Content:

Arbitrariness is an outstanding property of collocations (Nesselhauf, 2003) and needs to be made clear to the participants. It means that how words combine to form collocations is, to a certain extent, language specific. In week 2, we mentioned that one cannot learn English well through word memorization plus mental translation. Through this session, participants would have a deeper understanding of the reason: the way how words combine with each other is largely language-specific.

Here are some examples for the sake of illustration:

写 / 作业	→	写 / 作业
*write / homework		do / homework
上 / 学	→	上 / 学
*go up / school		go to / school
打 / 篮球	→	打 / 篮球
*beat / basketball		play / basketball
踢 / 足球	→	踢 / 足球
*kick / football		play / football
皇 / 室	→	皇 / 室
*royal / room		royal / family

Since collocations are largely arbitrary and language-specific, one would be in vain if he looks for fixed rules to learn them in a super-efficient manner, like how we learn grammar. The takeaway is clear: we should gradually get rid of the habit of mental translation and pay attention to not only words but also collocations we encounter in reading and listening, in order to expand and build up our collocational knowledge step by step, day by day.

## &lt; Week 5 &gt;

**Theme:**

Collocation Awareness Raising – The Trap of Synonyms

**Content:**

It is a common misunderstanding that synonyms can be freely replaced with each other without affecting the idiomaticity of the expression. This is simply not true. As Cowie (2013) argues, words which are “synonymous as wholes ... are extremely rare.” One of the best ways to make distinction between synonyms is through collocation learning.

Through this session, I would rectify participants' perception about synonyms and have them realize that synonyms, though largely overlapping in their semantic range, usually behave differently in use. In other words, when it comes to actual use in a specific context, synonyms are generally NOT freely replaceable by each other. This, again, testifies that learning collocations well can make our expression more accurate and appropriate.

Based on the examples from the Cambridge textbook and those from COCA, I would ask the participants to differentiate three pairs of synonyms which, as my experience suggests, will probably be considered as almost identical by most participants:

- quick / fast
- big / large
- desk / table

The participants first try to explain how one is different from the other in terms of meaning (and they will have a hard time!) Then, they are asked to do a matching exercise, match each word with its appropriate collocate as follows:



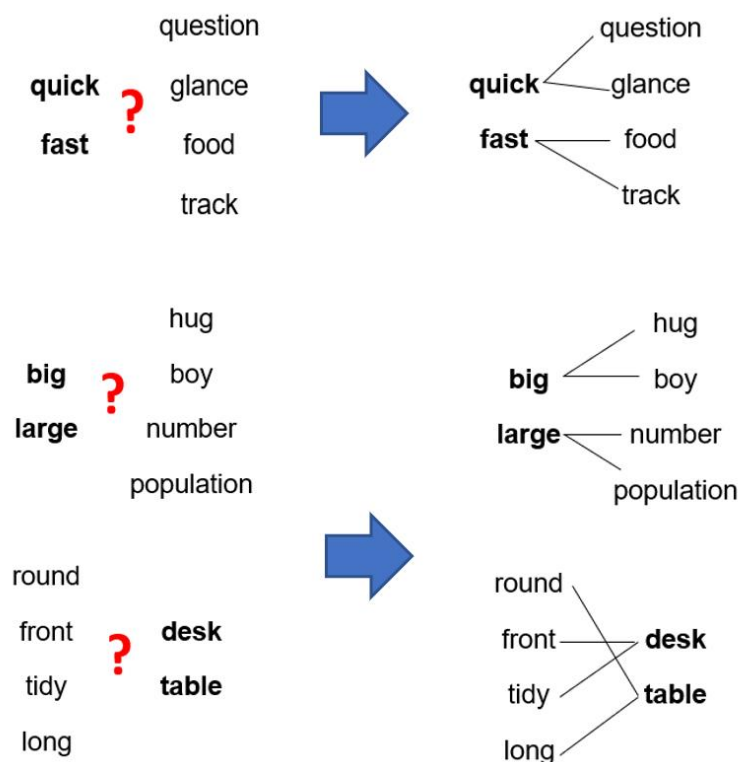


Figure 25 Collocate Matching Practice (self-made)

Through this practice, participants would realize that so-called synonyms often attract different “company” when forming collocations, and this in turn fine-tunes our understanding of each synonym’s nuanced meaning (one of the benefits of learning collocations). For example, from the last group, participants will get to know that a desk is often used by one person while a table can be routinely used by a group of people. And while a table can be used in various settings, a desk is normally related to the office.

All in all, participants now should realize the trap of synonyms: although similar in meaning, they could behave quite differently in use. Each word has its specific company to go together, and this is exactly the information participants need to learn in order to master the language well.

## &lt; Week 6 &gt;

## Theme:

Collocation Awareness Raising – The Rule of Rhythm

## Content:

Rhythm is a common figure of speech in English. Participants are asked to observe and comment on the following examples:

- I have stood still and stopped the sound of feet. (**alliteration**)
- There is a white kite high in the sky. (**assonance**)
- Don't do the crime if you can't do the time. (**tail rhyme**)

Such choice of words makes the language sound colorful and elegant.

The same principle, to a certain extent, applies to collocation formation. Although collocations are largely arbitrary, rules such as rhythm still exist, which is to be learned to enhance the learning efficiency. Here are some examples from last training:

- ✧ quick question
- ✧ fast food
- ✧ big boy

All of them exemplify the principle of alliteration. If one is keen to observe in her study, many other cases could be identified. Here just to list a few:

*run the risk / sold one's soul / do the deeds / a strong stroke / pale ale / deep depression*

As demonstrated in the above, words with similar soundings (e.g. alliteration) tend to attract each other. A word of caution: this is far from a hard and fast rule and provides no more than a little hint to the learners when acquiring collocations. However, they will do well to pay notice to such rules/patterns on their way to idiomaticity.

## &lt; Week 7 &gt;

## Theme:

Collocation Learning Demo

## Content:

As Harmer (2015) points out, knowledge to be learned will be far more memorable for learners if they do the work themselves. In this session, the researcher would do a demo to show participants how to develop their collocational competence. The intensifying adverbs are chosen as the focus. First, a short passage as follows will be presented:

### Intensifying adverbs (表程度强的副词)

Tom is a very shy person. Nevertheless, he has a very brilliant mind. Although he is very skillful at drawing pictures, Tom is very reluctant to share his works with others. His parents are very concerned about his future.



As participants read through the lines, most likely they would notice the excessively-used adverb "very". The researcher then points out that this problem of lacking versatility is very common among Chinese learners. Next, the researcher asks some volunteers to think of some adverbs that share the same meaning as "very", through which process they will get more engaged in class and learn from each other.

Then, the researcher show a list of 24 intensifying adverbs as follows:

## Intensifying adverbs (表程度强的副词)

### Any other words to express “very”?

- extremely
- really
- highly
- utterly
- absolutely
- quite
- fully
- totally
- completely
- deeply
- ridiculously
- strongly
- bitterly
- greatly
- exceedingly
- severely
- fairly
- thoroughly
- perfectly
- entirely
- dead
- wholly
- badly
- immensely

Figure 26 A List of Intensifying Adverbial Synonyms (self-made)

This exposure is to make participants aware that there exist many choices in English when it comes to conveying a specific meaning, and they need to commit themselves to learning them well. Also, the researcher would remind them that although these adverbs seem to be synonymous to each other, they behave differently in use. The next step is to pick out three for a closer look: *highly*, *absolutely*, *deeply*.

The researcher would refer to a collocation dictionary (i.e. the Oxford Collocations Dictionary) to find eight collocations of each adverb for participants to observe and identify the patterns:

## Pattern Identification Practice

- |                                                     |                                                     |                                                       |
|-----------------------------------------------------|-----------------------------------------------------|-------------------------------------------------------|
| <input type="checkbox"/> highly <b>likely</b>       | <input type="checkbox"/> deeply <b>moved</b>        | <input type="checkbox"/> absolutely <b>marvelous</b>  |
| <input type="checkbox"/> highly <b>successful</b>   | <input type="checkbox"/> deeply <b>concerned</b>    | <input type="checkbox"/> absolutely <b>impossible</b> |
| <input type="checkbox"/> highly <b>skillful</b>     | <input type="checkbox"/> deeply <b>ashamed</b>      | <input type="checkbox"/> absolutely <b>wrong</b>      |
| <input type="checkbox"/> highly <b>effective</b>    | <input type="checkbox"/> deeply <b>shocked</b>      | <input type="checkbox"/> absolutely <b>convinced</b>  |
| <input type="checkbox"/> highly <b>recommended</b>  | <input type="checkbox"/> deeply <b>hurt</b>         | <input type="checkbox"/> absolutely <b>phenomenal</b> |
| <input type="checkbox"/> highly <b>competitive</b>  | <input type="checkbox"/> deeply <b>unhappy</b>      | <input type="checkbox"/> absolutely <b>devastated</b> |
| <input type="checkbox"/> highly <b>confident</b>    | <input type="checkbox"/> deeply <b>disappointed</b> | <input type="checkbox"/> absolutely <b>miserable</b>  |
| <input type="checkbox"/> highly <b>entertaining</b> | <input type="checkbox"/> deeply <b>religious</b>    | <input type="checkbox"/> absolutely <b>unbearable</b> |

Figure 27 Collocation Pattern Identification Practice (self-made)

After listening to the findings of the participants, the researcher renders the answer as follows:

- ✧ *highly* generally has a positive prosody, meaning it tends to modify positive adjectives, though not always
- ✧ *deeply* is often used to modify adjectives that are emotional in nature
- ✧ *absolutely* is a rather strong adverb, often associated with adjectives that lack gradability

Then the participants are asked to do an exercise to test their understanding as follows:

### ***Identify the incorrect one in each group:***

A. highly educated B. highly profitable  
C. highly unusual D. highly exhausted ~~X~~

A. absolutely convinced B. absolutely tired ~~X~~  
C. absolutely devastated D. absolutely absurd

A. deeply unhappy B. deeply religious  
C. deeply successful ~~X~~ D. deeply committed

Figure 28 Test of Understanding on Collocational Patterns (self-made)

If they indeed get all of them correct, that means progress has been made with regard to these three adverbs. With such knowledge, participants are asked to read again the original passage and make an effort to improve the style. One possible improved version is:

## Intensifying adverbs (表程度强的副词)

Tom is a **fairly** shy person. Nevertheless, he has a **absolutely** brilliant mind. Although he is **highly** skillful at drawing pictures, Tom is **quite** reluctant to share his works with others. His parents are **deeply** concerned about his future.



Through this demo, the participants would a) understand the benefits of learning collocations better, b) accumulate knowledge about some common intensifying adverbs, and c) learn how to acquire collocations in general with firsthand experiences.



APPENDIX B

COLLOCATIONS FROM NNS CORPORA

This appendix is an exhaustive list of all the potential collocations participants produced, along with their number of occurrence (R), judgement result (Q), and Freq. & MI in COCA (only for correct collocations). Please note that in the judgement result (Q) column, *C* denotes *collocation*, *FC* denotes *free combination*, and *UC* denotes *unidiomatic combination*.

Form	R	Q	Freq.	MI	Form	R	Q	Freq.	MI
<b>ADJ-N</b>					ancient saying	1	C	22	2.46
absolute fairness	1	UC			ancient times	2	FC		
absolute music	2	UC			appropriate audios	1	UC		
abundant cases	1	FC			appropriate phrases	1	FC		
abundant experience	1	UC			arbitrary people	1	UC		
academic performance	2	C	2384	5.33	arduous process	1	C	216	5.97
academic success	1	C	1287	4.65	arduous task	1	C	200	7.37
aching heart	1	C	72	4.75	associative memory	1	C	26	5.98
active learning	1	C	714	4.49	audio books	1	C	502	3.68
active process	1	FC			authentic expressions	5	C	61	3.49
advanced expressions	2	UC			authentic language	1	FC		
all-round environment	1	FC			authentic videos	1	FC		
amazing environment	1	FC			awkward situation	1	C	403	4.33
amazing views	1	FC			back content	1	UC		
American book	1	FC			bad book	1	FC		
American cat	1	FC			bad emotions	1	UC		
American drama	2	FC			bad influence	1	FC		
American employer	1	FC			bad mood	5	C	1098	3.4
American shows	1	FC			bad pen	1	FC		
American TV show	1	FC			bad side	1	FC		
ancient books	1	FC			basic expression	1	FC		
ancient era	1	FC			basic expressions	1	FC		
ancient events	1	FC			basic grammar	1	C	81	4.3
ancient Greece	2	FC			basic knowledge	3	C	769	3.48



ancient people	2	FC		
ancient Roma	2	FC		
basic reading	1	C	286	2.87
basic requirement	1	C	428	3.83
basic sentences	1	FC		
basic situation	1	FC		
basic skill	3	C	2621	5.27
basic things	1	FC		
basic unit	1	C	356	2.82
basic usage	1	FC		
basic vocabulary	2	C	61	3.1
beautiful articles	1	UC		
beautiful attractions	1	FC		
beautiful city	10	FC		
beautiful culture	1	FC		
beautiful face	1	FC		
beautiful flowers	1	C	617	3.7
beautiful house	1	FC		
beautiful island	1	FC		
beautiful life	2	FC		
beautiful place	3	FC		
beautiful sceneries	2	C	244	5.88
beautiful scenery	2	C	244	5.88
beautiful sea	1	FC		
beautiful sentences	3	FC		
beautiful things	2	FC		
beautiful women	1	C	5924	3
beneficial activity	1	C	105	2.74
beneficial books	1	UC		
beneficial habits	1	FC		

basic need	1	C	1880	4.16
basic phrases	1	FC		
big fish	1	C	2331	2.63
big girl	1	FC		
big house	1	FC		
big results	1	FC		
big school	2	FC		
big things	1	FC		
big work	1	UC		
black soul	1	FC		
blue girls	1	UC		
bonny cars	2	UC		
boring learning	1	UC		
boring process	2	FC		
boring thing	2	FC		
boring time	1	FC		
bright color	1	C	2186	4.93
bright future	1	C	2121	4.9
bright light	1	C	4037	5.51
bright places	1	FC		
bright way	1	UC		
British dramas	1	FC		
British people	1	FC		
broken vase	1	C	27	4.64
busy day	1	C	1976	2.19
busy life	2	FC		
calm word	1	UC		
careful observation	1	C	178	3.55
catalytic role	1	C	30	4.23
central role	1	C	1876	3.6

beneficial thing	1	FC		
big building	1	FC		
big city	2	UC		
big concert	1	FC		
big data	1	FC		
big deal	2	C	20166	4.87
big dinner	1	FC		
big family	2	FC		
certain number	1	FC		
certain phrase	1	FC		
certain point	1	C	2763	2.2
certain practice	1	FC		
certain skills	1	FC		
certain techniques	1	FC		
certain time	2	FC		
certain times	1	FC		
certain vocabulary	3	FC		
certain work	1	UC		
challenging thing	2	FC		
childlike innocence	1	C	56	8.74
Chinese characters	1	C	359	2.15
Chinese culture	3	C	722	3.27
Chinese history	1	FC		
Chinese language	1	C	395	2.41
Chinese literature	2	FC		
Chinese lychee	1	UC		
Chinese masterpieces	1	FC		
Chinese meaning	1	FC		
Chinese music	1	FC		
Chinese people	2	FC		
Chinese prosperity	1	UC		

certain ability	1	FC		
certain accumulation	1	UC		
certain amount	5	C	4357	4.78
certain country	1	FC		
certain degree	1	C	1173	3.06
certain extent	3	C	1510	4.76
certain foundation	4	UC		
certain kind	1	C	2836	2.64
cognitive skills	1	C	859	5.42
colloquial language	1	C	46	6.18
comfortable way	1	UC		
comic books	1	C	381	3.04
common behaviors	1	FC		
common fallacy	1	C	60	4.03
common goal	1	C	1333	3.19
common grammar	1	UC		
common misconceptions	1	C	460	6.75
common phenomenon	3	C	222	2.9
common phrases	1	C	233	2.78
common practice	1	C	1882	3.39
common question	1	FC		
common saying	1	FC		
common situation	1	FC		
common view	1	FC		
common vocabulary	1	C	86	2.78
common word blocks	1	UC		
communist leader	1	C	627	4.39
complete understanding	1	C	407	3.39
complex network	1	C	340	2.88
complex world	1	FC		
comprehensive application	1	UC		

Chinese saying	1	C	60	3.24
Chinese strengths	1	UC		
Chinese students	1	FC		
Chinese subtitles	2	FC		
Chinese version	1	FC		
civilized citizen	1	C	18	2.05
civilized language	1	C	45	2.81
classical literature	1	C	196	4.02
clean place	1	FC		
clear direction	1	FC		
clear plan	1	FC		
close relationship	1	C	2800	3.7
cloudy day	1	C	502	4.22
continuous exchanges	1	UC		
continuous process	1	C	341	3.51
contrary writing	1	UC		
convenient tool	1	C	62	2.94
convictive way	1	UC		
cool autumn	1	C	95	3.68
core methods	1	FC		
correct attitude	1	FC		
correct collocation	1	FC		
correct habits	1	FC		
correct ideas	1	UC		
correct outlook	1	FC		
correct pronunciation	1	C	147	7.12
correct usage	2	C	103	4.26
correct use	2	FC		
correct view	1	FC		
correct way	2	FC		
corresponding meaning	1	FC		

comprehensive lesson	1	FC		
concrete tactics	1	UC		
confident citizen	1	UC		
confident girl	2	FC		
confident person	1	FC		
constant dedication	1	UC		
constant dropping	1	UC		
constant effort	1	FC		
constant persistence	1	UC		
constant practice	1	FC		
contemporary world	1	UC		
continuous accumulation	1	C	4	2.55
continuous efforts	2	C	102	2.14
cute dogs	1	C	204	2.6
daily classes	1	FC		
daily communication	2	FC		
daily English	1	FC		
daily life	57	C	6215	3.89
daily lives	2	C	6215	3.89
daily practice	1	C	496	2.56
daily reading	2	C	137	2.09
daily routine	1	C	1464	6.77
daily study	1	FC		
daily writing	1	FC		
dark place	1	FC		
dark trade	1	UC		
dear friends	1	C	3297	5.06
decreasing significance	1	C	4	3.18
deep attainment	1	UC		
deep breath	1	C	9431	7.9
deep communication	1	UC		

corresponding pictures	1	FC		
countless sustenance	1	UC		
critical process	1	FC		
critical thinking	2	C	4214	6.97
cross-cultural communication	2	C	151	6.88
crucial question	1	C	562	2.43
cultural ability	1	UC		
cultural background	1	C	883	4.52
cultural backgrounds	1	C	883	4.52
cultural customs	2	UC		
cultural differences	2	C	2073	4.15
cultural events	1	C	626	2.35
cultural information	1	UC		
cultural knowledge	1	C	435	2.58
cultural level	1	UC		
cultural transmission	1	C	146	3.71
cultural treasures	1	C	111	3.51
current events	1	C	1791	2.94
digital era	1	C	234	3.76
direct help	1	UC		
dirty words	1	C	721	2.86
diverse perspective	1	C	222	4.03
double reduction	1	UC		
dubbing movies	1	UC		
dumb English	1	UC		
dumb language	1	UC		
dynamic city	1	FC		
early age	3	C	3746	3.18
easy code	1	FC		
easy event	1	UC		
easy films	1	UC		

deep comprehension	1	C	48	2.74
deep interest	1	FC		
deeper understanding	1	C	1710	4.96
delicious dishes	1	C	244	5.7
delicious food	2	C	708	4.43
delicious foods	3	C	708	4.43
deserted place	1	C	80	2.9
developed city	1	FC		
difficult book	1	FC		
difficult books	1	FC		
difficult materials	1	FC		
difficult part	1	FC		
difficult point	1	FC		
difficult process	1	FC		
difficult question	1	FC		
difficult task	1	C	1827	4.33
difficult thing	1	FC		
difficult words	2	FC		
electronic products	2	C	279	2.88
elegant people	1	FC		
elegant sentence	3	FC		
elegant sentiment	1	UC		
emotional aspect	1	C	271	3.36
emotional experience	1	C	713	2.89
emotional intelligence	1	C	762	4.81
emotional state	1	C	1162	2.14
emotional well-being	1	C	497	6.53
emotional world	1	UC		
endless benefits	1	FC		
endless possibilities	1	C	604	5.98
endless presentation	1	UC		

easy task	1	C	1588	3.36
easy thing	3	FC		
easy things	1	FC		
easy way	2	FC		
easy ways	1	FC		
economic development	1	C	9349	5.19
economic globalization	1	C	301	5.09
educational experts	1	FC		
effective absorption	1	UC		
effective approach	1	C	698	3.23
effective combination	1	UC		
effective learning	1	C	512	3.58
effective methods	2	C	1126	4.06
effective skill	1	C	423	2.53
effective strategy	1	C	1838	4.75
effective tool	1	C	989	4.11
effective way	6	C	3522	2.12
effective ways	1	C	3522	2.12
efficient method	1	C	339	4.02
electric bicycle	5	C	51	4.02
electric bike	2	C	104	3.39
electric cars	6	C	2470	4.95
electronic devices	3	C	1666	6.47
English speech	1	FC		
English subtitles	3	C	613	8.42
English things	1	FC		
English vocabularies	1	C	191	5.39
English words	2	C	1350	3.13
enjoyable process	1	FC		
enjoyable thing	1	FC		
enormous significance	1	C	66	3.56

endless sea	1	C	121	3.34
English articles	1	FC		
English book	2	FC		
English books	3	FC		
English cartoons	1	FC		
English culture	1	FC		
English drama	1	C	49	2.11
English environment	2	FC		
English film	1	FC		
English films	1	FC		
English logic	1	UC		
English magazines	1	FC		
English movie	7	FC		
English movies	4	FC		
English music	6	FC		
English news	1	FC		
English newspaper	1	FC		
English novels	2	C	147	2.76
English phonology	1	FC		
English pronunciation	1	C	68	6.24
English song	6	FC		
English songs	6	FC		
English speakers	1	C	798	5.57
experienced friends	1	UC		
expressive ability	1	C	47	3.33
extensive reading	20	C	76	2.65
external factors	1	C	807	5.01
external help	1	FC		
external pressure	1	C	503	4.6
extra courses	1	FC		
extracurricular books	3	UC		

enormous strain	1	C	78	4.12
enough courage	1	C	153	2.98
enough interest	1	UC		
enough love	1	FC		
enough practice	1	FC		
enough time	3	FC		
enough vocabulary	3	FC		
enough waste	1	UC		
enough water	1	C	1578	2.31
entertaining way	1	FC		
entire life	1	C	6962	3.05
environmental exhibition	1	UC		
ephemeral thrills	1	UC		
essential aspects	1	C	264	3.6
essential part	3	C	1518	3.11
essential points	1	FC		
essential role	1	C	587	3.23
essential step	1	C	276	2.42
essential subject	1	FC		
European church	1	UC		
excellent people	2	UC		
excellent person	3	FC		
excellent reader	1	FC		
excellent teacher	2	FC		
excellent understanding	1	FC		
excellent way	1	FC		
executive force	1	FC		
exotic moment	1	UC		
favorite books	1	C	1839	3.04
favorite character	1	C	821	3.42
favorite class	1	FC		

extraordinary temperament	1	UC		
extreme world	1	FC		
extrovert classmates	1	UC		
extrovert people	1	UC		
faithful friend	1	C	193	2.21
fake news	1	C	1563	4.96
familiar vocabulary	1	FC		
famous books	1	FC		
famous food	1	UC		
famous foods	1	UC		
famous fruit	1	FC		
famous interest	1	UC		
famous people	1	FC		
famous places	1	FC		
famous Portuguese	1	FC		
famous quotes	1	C	275	4.61
famous saying	3	C	89	4.14
famous writer	1	C	410	3.3
famous writers	1	C	410	3.3
fantastic adventure	1	C	39	2.87
fantastic way	1	FC		
fast way	1	FC		
fast-paced age	1	UC		
fast-paced society	2	C	21	3.52
fast-paced times	1	UC		
favorable impression	1	C	192	6.36
favorite anime	1	C	15	4.24
favorite book	1	C	1839	3.04
first class	2	FC		
first college	1	FC		
first content	1	UC		

favorite course	1	FC		
favorite food	3	C	1321	3.29
favorite fruit	1	C	117	2.3
favorite hobby	2	C	166	5.19
favorite language	1	FC		
favorite lesson	1	FC		
favorite movie	2	C	1548	3.98
favorite movies	1	C	1548	3.98
favorite music	1	FC		
favorite place	3	C	1298	2.31
favorite shop	1	C	137	2.13
favorite singers	1	C	146	3.56
favorite songs	2	C	1668	4.65
favorite teachers	1	FC		
favorite things	2	FC		
favorite time	1	FC		
fierce competition	1	C	598	6.85
fierce wheels	1	UC		
fifth grade	1	FC		
final decision	1	C	8	2.5
final exam	6	C	849	5.97
final examination	1	C	216	3.62
final goal	1	FC		
final task	1	FC		
fine listening	2	UC		
fine teachers	1	FC		
firm belief	1	C	337	5.03
first answer	1	FC		
first beach	1	UC		
first benefit	2	FC		
first birthday	2	FC		

first day	2	FC		
first dream	1	FC		
first exam	1	FC		
first examination	1	FC		
first experience	1	FC		
first floor	1	FC		
first foreigner	1	FC		
first grade	1	C	3060	2.51
first grades	1	C	3060	2.51
first impression	1	C	2357	3.2
first interview	3	FC		
first language	2	FC		
first month	1	FC		
first person	1	FC		
first place	4	C	28143	2.64
first point	1	FC		
first position	1	FC		
first presentation	1	FC		
first prize	2	FC		
first semester	3	C	897	3.28
first step	2	C	13137	3.35
first teacher	1	FC		
first thing	5	FC		
first time	44	C	100644	2.49
first weekends	1	FC		
first year	1	FC		
fixed collocation	1	FC		
flexible methods	1	UC		
fluent English	2	C	363	8.58
focused state	1	FC		
following points	4	FC		

first book	1	FC		
first choice	5	FC		
foreign classics	1	FC		
foreign companies	1	C	1643	2.08
foreign company	2	C	1643	2.08
foreign countries	1	C	4536	3.34
foreign country	1	C	4536	3.34
foreign debate	1	C	371	2.12
foreign film	1	C	674	2.1
foreign films	1	C	674	2.1
foreign friend	1	FC		
foreign institution	1	FC		
foreign language	38	C	5053	5.37
foreign movies	1	FC		
foreign music	1	FC		
foreign novels	1	FC		
foreign people	2	FC		
foreign radios	1	UC		
foreign songs	1	FC		
foreign speeches	1	UC		
foreign teacher	12	FC		
foreign teachers	1	FC		
foreign trade	2	C	1111	3.71
former college	2	FC		
former one	1	FC		
former things	1	UC		
former writer	1	FC		
fragmented reading	2	UC		
free reading	1	FC		
free talk	14	FC		
free talks	1	FC		

following solutions	1	FC		
foreign class	1	UC		
fried foods	1	C	512	4.78
friendly communication	1	FC		
friendly teacher	1	FC		
front content	2	UC		
full class	1	FC		
full man	1	UC		
full schedule	2	C	441	2.82
full sentence	1	FC		
full use	2	FC		
fun time	1	FC		
fun way	1	FC		
fundamental elements	2	C	216	3.38
fundamental purpose	1	C	149	2.48
funny guys	1	C	1305	2.13
funny joke	1	C	728	4.55
further education	2	FC		
future career	1	C	477	2.64
future days	1	UC		
future job	1	FC		
future plan	1	C	979	2.76
future study	2	C	2121	3.22
general activity	1	FC		
general idea	2	FC		
general meaning	3	FC		
general principle	1	C	1085	3.73
gentle girl	1	FC		
gigantic microwave	1	FC		
global village	1	C	402	3.18
golden house	4	UC		



free time	11	FC		
free topic	1	UC		
French movies	1	FC		
French songs	1	FC		
frequent exchanges	1	FC		
fresh air	1	C	8749	6.06
fresh water	3	C	2835	3.73
good command	5	FC		
good common	1	UC		
good conditions	1	FC		
good course	1	FC		
good days	1	FC		
good environment	3	FC		
good exercise	1	FC		
good experience	1	FC		
good friend	13	FC		
good friends	6	FC		
good grade	1	FC		
good grades	2	FC		
good guessers	1	FC		
good guy	1	FC		
good habit	6	FC		
good habits	1	FC		
good health	1	FC		
good hobbies	1	FC		
good idea	2	C	25277	2.98
good impression	1	FC		
good input	1	FC		
good interest	1	UC		
good interpretation	1	FC		
good job	4	C	21483	2.48

good ability	1	FC		
good attitude	3	FC		
good book	12	FC		
good books	1	FC		
good care	1	FC		
good character	3	FC		
good choice	3	FC		
good personality	1	FC		
good place	1	FC		
good ppt	1	FC		
good practice	1	FC		
good preparation	1	FC		
good preview	1	UC		
good profession	1	UC		
good question	1	FC		
good reading	2	FC		
good results	1	FC		
good roommates	1	FC		
good saying	1	FC		
good school	1	FC		
good score	1	FC		
good sentences	1	FC		
good skills	2	FC		
good sleep	1	FC		
good student	1	FC		
good students	1	FC		
good teacher	3	FC		
good teachers	1	FC		
good thing	5	FC		
good things	1	FC		
good time	1	FC		

good man	1	FC		
good manner	1	FC		
good memory	1	FC		
good method	1	UC		
good methods	1	UC		
good mind	1	UC		
good mood	3	C	1381	2.35
good movies	2	FC		
good need	1	FC		
good news	2	C	24814	3.26
good opportunity	1	FC		
good output	1	FC		
great benefit	2	FC		
great benefits	1	FC		
great book	1	FC		
great charm	2	FC		
great contributions	1	FC		
great danger	1	C	1500	2.38
great deal	2	C	13462	4.07
great difference	1	FC		
great difficulties	1	C	1344	2.46
great discoveries	1	FC		
great effort	1	FC		
great harvest	1	FC		
great help	4	FC		
great importance	3	C	1648	2.51
great influence	1	FC		
great interest	4	FC		
great leader	1	FC		
great meaning	1	UC		
great men	1	FC		

good tips	1	FC		
good use	2	FC		
good values	1	FC		
good video	1	FC		
good vocabulary	1	FC		
good way	19	FC		
good ways	2	FC		
good words	3	FC		
good work	1	FC		
graceful words	1	UC		
grammatical meaning	1	C	10	3.64
grammatical well-formedness	1	UC		
happy experience	1	FC		
happy life	1	FC		
happy meeting	1	UC		
happy thing	1	UC		
happy throngs	1	UC		
happy time	3	FC		
hard job	1	FC		
hard work	4	C	12648	3.91
hard working	1	C	12651	3.91
hard-working people	1	C	586	3.26
harsh language	1	C	119	2.61
healthy body	2	C	921	2.52
healthy hobbies	1	UC		
healthy mind	1	FC		
heavy luggage	1	C	30	3.08
heavy music	1	FC		
heavy rain	1	C	2445	6.27
helpful books	1	FC		
helpful friend	1	FC		

great method	1	FC		
great minds	1	FC		
great necessity	1	UC		
great oaks	1	UC		
great people	1	FC		
great person	1	FC		
great pressure	1	FC		
great progress	3	FC		
great significance	8	C	621	2.34
great talent	1	FC		
great test	1	FC		
great thing	1	FC		
greatest effort	1	FC		
green book	1	FC		
green lake	1	UC		
green trees	1	C	737	2.81
handy tips	1	C	66	4.34
huge world	1	FC		
human being	5	C	31166	8.08
human beings	7	C	31166	8.08
human civilization	3	C	590	3.96
human communication	2	FC		
human consciousness	1	C	637	3.99
human friends	1	UC		
human progress	29	C	566	2.41
human thought	1	FC		
human wisdom	1	FC		
humble advice	1	FC		
ideal effect	1	FC		
idiomatic combinations	1	UC		
idiomatic English	1	FC		

high achievement	1	C	1024	2.6
high character	1	UC		
high mark	1	C	1470	3.06
high proficiency	1	C	126	2.8
high rate	1	C	22010	4.58
high salary	1	C	1140	3.06
high school	65	C	109049	5.17
high score	1	C	6121	4.02
high speed	2	C	3585	3.63
high time	1	FC		
historical books	1	FC		
historical events	1	C	1508	4.13
historical stage	1	UC		
hot day	2	FC		
hot pot	1	C	309	3.23
hot weather	1	C	1143	4.32
huge number	1	C	2159	2.45
important question	1	FC		
important role	8	C	6671	3.76
important sources	1	FC		
important step	1	C	2334	2.53
important subject	1	FC		
important thing	17	C	16964	2.4
important things	1	FC		
important tool	3	C	972	2.03
important topic	1	C	857	2.46
important way	6	FC		
important ways	1	FC		
impossible things	1	FC		
impressive words	1	FC		
inconvenient role	1	UC		

idiomatic expressions	1	C	45	8.53
idiomatic sentences	1	C	3	4.93
imaginary space	1	UC		
immediate benefits	1	C	310	3.29
immersive environment	1	C	93	6.76
impeccable grammar	1	C	11	6.6
impetuous emotions	1	UC		
impetuous mood	1	C	2	4.05
important amusement	2	UC		
important benefit	1	FC		
important condition	1	FC		
important form	1	FC		
important guarantee	1	UC		
important learning	1	FC		
important meaning	1	FC		
important meanings	1	FC		
important means	3	FC		
important meetings	1	FC		
important method	1	FC		
important organization	1	FC		
important part	7	C	6963	2.33
important point	2	FC		
intensive reading	8	C	81	3.97
intercultural communication	1	C	153	8.26
interesting book	1	FC		
interesting experience	1	FC		
interesting feeling	1	FC		
interesting film	1	FC		
interesting games	1	FC		
interesting materials	1	UC		
interesting place	2	FC		

incredible power	1	FC		
independent learners	1	C	96	3.92
independent thinking	4	C	235	3.08
indispensable necessity	1	C	2	2.17
indispensable part	3	C	107	2.9
indispensable process	1	FC		
indispensable role	1	C	81	3.99
indispensable tool	1	C	107	5.44
indispensable way	1	UC		
indispensable ways	1	UC		
individual level	1	C	1427	2.82
individual words	1	FC		
inferior reality	1	UC		
infinite space	1	C	224	4.13
inherent grammar	1	UC		
initial intention	1	C	50	2.17
inner feelings	1	C	209	3.05
inner self	1	C	505	5.62
inner world	1	FC		
innovative city	1	FC		
innovative spirit	1	C	45	2.32
intellectual ability	1	C	425	4.09
last point	1	FC		
last semester	2	C	469	3.01
last suggestion	1	FC		
last Sunday	1	FC		
last term	3	FC		
last time	12	FC		
last week	3	C	73423	4.76
last weekend	2	C	4610	3.32
last year	3	C	161431	4.06

interesting places	1	FC		
interesting stories	1	C	2128	2.39
interesting teacher	1	FC		
interesting thing	2	C	6004	2.41
interesting things	2	C	6004	2.41
interesting video	1	FC		
interesting way	1	FC		
international vision	1	UC		
interpersonal communication	1	C	309	6.65
interpersonal conflicts	1	C	194	6.11
invisible ruler	1	C	14	3.4
invisible way	1	UC		
irreplaceable role	1	C	25	4.04
key point	1	C	2390	2.86
kind people	1	FC		
Korean dance	1	FC		
large amount	1	C	5762	3.98
large number	3	C	16966	3.74
large numbers	1	C	16966	3.74
large vocabulary	1	FC		
last class	2	FC		
last day	1	FC		
last habits	1	UC		
last holiday	1	FC		
last misunderstanding	1	UC		
last moment	1	FC		
last month	1	C	35709	4.07
local food	3	C	1774	2.06
local friends	1	FC		
local group	2	FC		
local history	1	FC		

later class	1	UC		
later time	1	FC		
lexical chunks	4	FC		
liberal arts	2	C	2705	5.12
lifelong partners	1	C	34	2.41
lifelong thing	1	UC		
like-minded friends	1	C	83	3.55
like-minded people	1	C	387	3.68
limited vocabularies	1	C	113	4.8
linguistic lesson	1	UC		
literal translation	1	C	540	9.13
literary works	1	C	504	2.56
literary world	1	FC		
little brother	1	C	4146	2.93
little car	1	FC		
little chance	1	FC		
little child	1	FC		
little girl	2	C	24009	4.53
little strawberry	1	FC		
little stress	1	FC		
little strokes	1	UC		
little thing	1	FC		
local conditions	1	FC		
local culture	2	FC		
local cultures	1	FC		
local customs	2	C	307	2.95
local expression	2	FC		
main reasons	1	C	4405	4.16
main way	3	FC		
main ways	2	FC		
mature people	1	FC		

local location	1	UC		
local netizen	1	FC		
local people	3	FC		
logical ability	1	UC		
logical system	1	FC		
logical thinking	2	C	182	4.69
long distance	2	C	3546	4.31
long holiday	1	FC		
long journey	1	C	1957	4.21
long jump	1	C	523	3.67
long process	12	FC		
long ride	1	C	1053	3.09
long term	1	C	13302	4.09
long time	23	C	71811	3.68
long vacation	1	C	379	2.23
long walk	1	C	1556	3.9
long way	1	C	16000	2.19
long-term accumulation	1	C	17	3.17
long-term efforts	1	FC		
long-term learning	1	UC		
long-term practice	1	FC		
long-term process	3	FC		
lovely bird	1	FC		
lucky money	1	UC		
lyric books	1	FC		
magical language	2	UC		
main assignment	1	FC		
main character	1	C	4925	5.31
main characters	1	C	4925	5.31
main idea	1	FC		
main parts	1	FC		

maximum way	1	UC		
meaningful experience	1	C	397	3.36
meaningful project	1	FC		
meaningful sentence	1	C	44	2.5
meaningful thing	5	FC		
meaningful units	1	FC		
memorable experience	6	C	275	4.08
memorable things	3	FC		
mental abilities	1	C	495	2.95
mental conditions	1	C	599	3.12
mental fun	1	UC		
mental health	3	C	19431	7.07
mental journey	1	FC		
mental life	1	FC		
mental pressure	1	FC		
mental quality	1	UC		
mental well-being	2	C	383	5.93
mere conversation	1	FC		
middle school	14	C	1988	2.36
mini models	1	C	32	3.14
mobile phone	9	C	3432	6.44
mobile phones	2	C	3432	6.44
modern era	1	C	1188	5.34
modern society	1	C	1888	4.07
modern times	1	FC		
momentous role	1	FC		
momentous way	1	UC		
monolingual environment	1	FC		
moral accomplishment	1	UC		
moral character	1	C	790	3.51
multilingual website	1	C	3	3.21

main point	1	C	2403	2.57
mutual benefit	1	C	429	4.47
mutual understanding	1	C	550	5.51
mysterious culture	1	FC		
national day	4	FC		
national holiday	1	C	615	2.4
national holidays	1	C	615	2.4
national language	1	FC		
national level	1	C	3413	2.23
national stories	1	FC		
native English	1	UC		
native expressions	2	UC		
native language	5	C	1665	4.84
native speaker	8	C	1050	5.81
native speakers	3	C	1050	2.81
native way	1	UC		
natural beauty	1	C	1168	4.64
natural resources	1	C	8505	6.2
negative emotions	1	C	856	5.36
negative side	2	FC		
nervous body	1	UC		
nervous mood	1	FC		
new city	1	FC		
new college	1	FC		
new diction	1	UC		
new door	1	FC		
new dormitory	1	FC		
new environments	1	FC		
new friend	1	FC		
new friends	1	FC		
new grammar	1	FC		

multiple views	1	UC		
new mindset	1	FC		
new phrase	1	FC		
new phrases	1	FC		
new plan	1	FC		
new progress	1	UC		
new requirements	1	FC		
new semester	1	FC		
new skills	1	FC		
new team	2	FC		
new technology	1	C	12843	3.28
new things	1	FC		
new vocabulary	1	FC		
new word	10	FC		
new words	19	FC		
new year	2	FC		
next day	1	C	33068	3.15
next life	1	FC		
next month	1	C	16329	3.61
next place	1	FC		
next step	3	C	9029	4.09
next term	2	FC		
next thing	1	FC		
next time	7	FC		
next topic	1	FC		
next week	2	C	30886	4.19
next year	1	C	60238	3.32
nice city	1	FC		
nice day	1	FC		
nice person	1	FC		
nice teacher	2	FC		

new information	1	FC		
new interpretations	1	FC		
new knowledge	1	FC		
new language	34	FC		
new life	4	FC		
new member	1	FC		
non-profit makings	1	UC		
normal conditions	1	C	644	2.76
normal life	1	C	4360	2.95
normal person	1	C	1341	2.62
normal phenomenon	1	FC		
normal way	1	FC		
novel stories	1	C	1088	2.46
nutritional hobbies	1	UC		
obvious function	1	FC		
old man	8	C	35520	2.91
old people	1	FC		
old quote	1	FC		
old saying	9	C	1461	4.86
old school	2	FC		
old street	1	FC		
old streets	1	FC		
old subjects	1	FC		
old vocabularies	1	UC		
old word	1	FC		
online chatroom	1	C	13	7.35
online classes	3	C	546	2.28
online education	1	C	965	3.03
online videos	1	C	958	3.61
only method	1	FC		
only thing	2	C	29585	3.23

nice thing	1	FC		
nice way	1	FC		
nice woman	1	FC		
noble character	1	C	58	2.15
noble person	1	FC		
noisy society	1	UC		
organized way	1	FC		
original movie	1	FC		
original works	3	FC		
ostensible meaning	1	FC		
outside world	4	C	4059	4.27
outstanding scores	1	FC		
own conclusion	1	FC		
own culture	1	FC		
own efforts	1	FC		
own experience	1	C	5953	2.46
own feature	2	FC		
own feelings	1	FC		
own grammar	3	FC		
own ideas	1	FC		
own job	1	FC		
own language	1	FC		
own list	1	UC		
own mind	1	FC		
own opinions	1	C	2335	2.27
own perspective	1	FC		
own plans	1	FC		
own responsibility	1	FC		
own room	1	FC		
own show	1	FC		
own study	1	FC		



only way	1	C	27629	3.06
oral class	1	FC		
oral English	8	FC		
oral expression	1	C	35	2.24
oral expressions	1	C	35	2.24
oral language	1	C	48	4.78
oral skill	1	C	220	3.92
oral speaking	8	UC		
ordinary methods	1	FC		
organizational ability	1	C	87	2.92
organizational role	1	C	82	2.25
past time	1	FC		
past weeks	1	C	9102	4.18
patriotic heart	1	FC		
peaceful city	1	FC		
peaceful life	1	FC		
perfect method	1	FC		
perfect way	1	FC		
personal culture	1	UC		
personal experience	1	C	5003	4.02
personal opinion	1	C	1461	3.41
personal thinking	1	UC		
pessimistic world	1	UC		
phonetic symbol	4	C	14	6.3
physical education	1	C	11710	5.84
physical exam	4	C	506	5.17
physical exercise	2	C	695	3.98
physical fitness	1	C	1817	6.94
physical health	1	C	3084	3.54
physical life	1	UC		
physical tests	1	FC		

own style	1	FC		
own system	1	FC		
own things	2	FC		
own tongue	1	FC		
own views	2	FC		
own way	3	FC		
pale face	1	C	1428	5.04
pan listening	2	UC		
particular country	1	FC		
part-time job	6	C	1644	5.73
past experience	1	C	1692	2.81
precious product	1	UC		
precious things	1	FC		
pretty birds	1	C	165	2.81
previous answer	1	FC		
previous group	1	FC		
previous major	1	FC		
previous school	1	FC		
priceless things	1	FC		
primary school	18	C	1977	2.66
primary student	1	FC		
professional interviews	1	FC		
professional knowledge	3	C	302	2.01
professional knowledges	1	C	302	2.01
professional language	1	FC		
professional teachers	1	C	1290	3.08
professional vocabulary	1	FC		
profound truth	2	C	134	2.92
promising student	1	FC		
proper expressions	1	FC		
proper language	1	FC		

physical well-being	1	C	486	5.4
planned way	1	FC		
pleasant atmosphere	1	C	45	3.38
pleased activity	1	UC		
polite citizen	1	UC		
poor city	1	FC		
poor conditions	1	C	1302	3.1
poor eyesight	1	C	149	6.48
poor place	1	FC		
poor places	1	FC		
poor speaking	1	UC		
popular attractions	1	C	280	4.65
positive energy	1	C	644	2.4
positive impact	2	C	2033	4.7
powerful strategy	1	FC		
practical experience	3	C	649	3.4
real child	1	C		
real context	1	C		
real life	2	C	12996	2.44
real meaning	1	C		
real passion	1	C		
recent situations	1	FC		
recent survey	1	C	2271	4.77
recent years	1	C	20017	3.24
red book	2	FC		
regular activities	1	C	510	2.6
regular breakfast	1	UC		
regular time	1	FC		
related cultures	1	FC		
related movies	1	UC		
relaxing way	1	FC		

proper means	1	FC		
proper practice	1	FC		
proper use	1	C	667	3.18
proper ways	2	FC		
public party	1	FC		
public places	1	FC		
pure expression	1	C	155	3.25
qualified citizen	1	C	41	2.46
qualitative change	2	C	121	2.84
quantitative change	2	FC		
quiet atmosphere	1	FC		
quiet environment	2	FC		
quiet girl	1	FC		
rainy day	1	C	1649	5.16
rapid development	4	C	640	4.39
rational aspect	1	UC		
romantic novels	1	C	155	3.9
romantic words	1	FC		
rote learning	1	C	117	7.72
rote memorization	1	C	114	14.25
round table	1	C	1343	4.89
royal road	1	UC		
rural areas	1	C	6183	6.11
rural places	1	FC		
scientific community	1	C	2291	4.09
scientific knowledge	1	C	1446	4.76
scientific system	1	UC		
second benefit	1	FC		
second canteen	2	FC		
second choice	2	FC		
second dream	1	FC		

relevant audio	1	UC		
relevant books	1	FC		
relevant language	2	FC		
relevant materials	1	C	277	2.84
relevant phrases	1	FC		
responsible teacher	1	FC		
rich experience	2	FC		
rich perspective	1	UC		
rich vocabulary	1	C	84	3.28
right answers	1	C	1903	2.27
right channel	1	FC		
right method	1	FC		
right order	1	FC		
right resources	1	FC		
right sentence	1	FC		
right strategies	1	FC		
right syllables	1	FC		
right things	1	C	19820	2.45
right way	4	FC		
right words	2	FC		
rigorous system	1	FC		
sheer quality	1	FC		
short movie	1	FC		
short passages	1	C	168	2.07
short period	2	C	4049	4.62
short time	2	C	12882	2.45
short video	1	C	736	2.04
short videos	5	C	736	2.04
short-term happiness	1	C	9	2.01
shy girl	2	C	279	3
significant boost	1	C	103	3.71

second group	1	FC		
second hobby	1	FC		
second interview	5	FC		
second language	1	C	1719	2.69
second myth	1	FC		
second part	1	FC		
second place	1	FC		
second reason	1	FC		
second semester	1	C	340	4.05
second thing	1	FC		
second time	6	FC		
secret garden	1	C	339	3.72
senior boy	1	FC		
senior brothers	4	UC		
senior school	3	C	2159	2.42
senior schoolmate	1	FC		
senior sister	5	UC		
senior students	1	FC		
sensitive state	1	UC		
serious meeting	1	FC		
several weeks	1	FC		
small boat	1	C	1332	2.75
small business	1	C	14905	3.69
small city	2	FC		
small clinic	1	FC		
small company	1	FC		
small goal	1	FC		
small plan	1	UC		
small surprise	1	FC		
small things	2	FC		
small town	1	C	12041	4.49

significant components	1	C	255	2.46
significant factor	1	C	1483	3.59
significant function	1	FC		
significant part	1	FC		
significant parts	1	FC		
significant portion	1	C	1247	5.12
significant process	1	FC		
significant sentence	1	FC		
silent friend	1	UC		
silent reading	1	C	433	4.6
simple articles	1	FC		
simple books	2	FC		
simple example	1	C	668	2.32
simple fruit	1	UC		
simple matter	1	C	1022	2.52
simple reading	1	FC		
simple thing	1	FC		
simple title	1	FC		
simple views	1	UC		
simple way	1	FC		
simple word	1	FC		
simple words	1	FC		
simple work	1	FC		
sincere friends	1	UC		
single practice	1	UC		
small action	1	FC		
special food	1	FC		
special mountains	1	UC		
special present	1	FC		
special reason	1	FC		
special tea	1	FC		

smart person	1	FC		
smooth running	1	C	55	4.13
social circle	1	C	859	2.86
social city	1	UC		
social contact	1	C	659	2.27
social contradictions	1	UC		
social culture	1	FC		
social department	2	UC		
social elites	2	UC		
social environment	1	C	1513	2.67
social harmony	1	C	205	3.02
social levels	1	UC		
social media	1	C	22469	5.89
social phenomena	1	C	490	2.89
social practice	2	C	2324	2.54
social problem	1	FC		
social skill	1	C	4474	4.07
social tool	1	FC		
soft bed	1	C	307	2.53
solid foundation	2	C	687	4.99
sound personality	1	UC		
Soviet writer	1	FC		
spacious places	1	UC		
spare time	9	C	1946	3.75
special attention	1	C	1616	2.96
special dish	1	FC		
subsequent grammar	1	UC		
successful elites	2	UC		
successful employer	1	FC		
successful learners	3	C	74	3.25
successful learning	2	C	180	2.1

special tour	1	FC		
specific context	1	C	676	3.59
specific environment	1	FC		
specific measure	1	C	403	2.4
specific usage	1	FC		
spiritual core	1	UC		
spiritual energy	1	FC		
spiritual enjoyment	1	C	8	2.19
spiritual lair	1	UC		
spiritual life	3	C	1579	2.69
spiritual strength	1	C	149	2.92
spiritual sustenance	2	C	49	6.81
spiritual wealth	2	UC		
spiritual world	5	FC		
spoken English	18	C	114	6.7
spoken language	3	C	583	7.43
standard audio	1	FC		
steady flow	1	C	382	5.77
strange language	2	FC		
strange place	1	FC		
strange thing	1	C	3497	2.62
strange things	1	C	3497	2.62
strange words	1	FC		
strong determination	1	FC		
strong efforts	1	UC		
strong influence	1	C	765	3.04
strong pillar	1	FC		
strong practicality	1	UC		
strong skills	1	FC		
strong will	2	FC		
strong-willed people	1	FC		

sudden illness	1	C	106	4.18
sufficient input	1	FC		
sufficient sayings	1	UC		
suitable method	2	C	84	3.07
suitable place	1	FC		
suitable time	1	FC		
sunny day	1	C	2149	4.39
superior strength	1	C	104	3.19
sustainable use	1	C	330	3.52
systematic exercises	1	FC		
talented people	1	FC		
tall person	1	FC		
tasty food	3	C	251	4.43
technical progress	1	C	166	3.24
temporary member	1	FC		
terrible thing	1	C	4023	2.99
third grade	5	C	2679	5.17
third thing	1	FC		
third year	4	FC		
tired day	1	UC		
total subjects	1	FC		
tough girls	1	FC		
tough task	1	C	271	2.18
traditional memorization	1	C	7	3.93
tranquil world	1	UC		
true meaning	1	C	883	3.08
true person	1	FC		
true refrigerator	1	UC		
trusted person	1	FC		
typical skills	1	FC		
typical views	1	FC		

ultimate purpose	2	C	240	3.38
unaffluent time	1	UC		
uncomfortable experience	6	FC		
unexpected surprises	1	C	149	4.2
unfamiliar knowledge	1	FC		
unfamiliar words	1	C	222	3.2
unfinished plan	1	FC		
unfinished works	1	C	153	3.25
unforgettable course	1	FC		
unforgettable experience	28	C	130	4.94
unforgettable lesson	1	C	16	3.73
unforgettable memory	7	C	30	4.04
unforgettable thing	13	FC		
unique atmosphere	1	FC		
unique expression	1	C	3	2.87
unique method	1	FC		
unique values	1	FC		
unknown knowledge	1	UC		
unknown words	1	UC		
unknown world	1	UC		
unlimited knowledge	1	FC		
unlimited motivation	1	UC		
unsolvable problem	1	C	113	6.61
unstable period	1	FC		
urban area	1	C	5056	5.37
useful advice	1	C	195	3.01
useful books	1	FC		
useful knowledge	2	C	326	2.83
useful materials	1	FC		
useful methods	1	C	201	2.26
useful skill	1	C	338	2.9

vague impression	1	C	62	4.57
valuable resources	1	C	944	5
valuable role	1	FC		
valuable time	2	FC		
various expressions	2	FC		
various fields	1	FC		
various figures	1	FC		
various languages	1	FC		
various lives	1	UC		
various themes	1	FC		
various way	1	FC		
various ways	1	FC		
vast number	2	C	978	2.81
vast ocean	1	C	312	4.38
vast sea	2	C	241	3.15
visual angle	1	C	51	2.53
visual effects	1	C	952	3.86
vital component	1	C	262	4.87
vital element	1	C	179	3.56
vital methods	1	UC		
vital part	1	C	865	3.13
vital points	1	FC		
vital role	1	C	1038	4.88
vivid environment	1	UC		
voluntary activities	1	C	92	2.71
warm weather	1	C	1610	5.87
well-defined plans	1	C	11	2.09
Western classics	1	C	60	2.72
western countries	1	C	2726	3.34
western culture	7	C	1767	4.58
western diseases	1	UC		

useful skills	1	C	338	2.9
useful things	1	FC		
useful way	1	FC		
useful websites	1	FC		
usual environment	1	FC		
whole daytime	1	UC		
whole holiday	1	FC		
whole life	8	C	12611	2.98
whole meaning	1	UC		
whole nations	1	FC		
whole person	1	UC		
whole process	1	C	2659	2.59
whole story	1	C	4593	2.5
whole world	2	C	10741	2.96
wide range	4	C	9963	7.4
wide variety	1	C	5524	6.96
wide vocabulary	1	FC		
wide world	1	C	4295	3.14
willing heart	1	FC		
wise friends	1	FC		
wise man	2	C	2432	3.44
wonderful effects	1	FC		
wonderful experience	1	C	641	2.25
wonderful family	1	FC		
wonderful life	1	FC		
wonderful night	1	FC		
wonderful scenery	1	C	15	2.61
wonderful sentences	1	FC		
wonderful thing	1	C	3249	2.22
wonderful world	1	FC		
world-famous writer	1	C	9	3.55

western stories	1	UC		
western wars	1	FC		
western wire	1	UC		
white sky	1	UC		
whole course	1	FC		
big enough	1	C	6281	2.21
closely native	1	UC		
commonly used	1	C	1162	7.4
deeply afraid	1	UC		
deeply bored	1	UC		
deeply fast	1	UC		
deeply frustrated	1	FC		
deeply shy	1	UC		
deeply upset	1	C	94	3.76
difficult enough	1	FC		
easily comfortable	1	UC		
easily influenced	1	C	125	2.16
especially true	1	C	2165	2.62
ever young	1	UC		
exceedingly crucial	1	UC		
exceedingly significant	1	UC		
extremely difficult	1	C	2344	5.36
extremely important	7	C	2468	4.16
extremely interesting	1	C	282	2.53
extremely meaningful	1	FC		
extremely necessary	1	UC		
extremely normal	1	UC		
extremely significant	2	FC		
firmly convinced	1	C	119	6.08
fully convinced	1	C	97	3.59
good enough	3	C	14896	2.57

wrong pronunciation	1	C	2	4.66
wrong understanding	1	FC		
young age	3	C	6904	3.46
young boy	1	C	6743	3.08
young folk	1	FC		
young people	2	C	29305	2.57
young teenager	1	C	762	3.3
<b>ADV-ADJ</b>				
almost perfect	1	FC		
awfully essential	1	UC		
increasingly focused	1	C	5	2.03
increasingly impatient	1	C	1197	3.12
increasingly important	1	C	1197	3.32
moderately difficult	1	C	51	3.56
outstanding enough	1	UC		
particularly familiar	1	FC		
particularly heavy	1	FC		
particularly important	2	FC		
pretty good	3	C	24448	3.89
pretty vital	1	FC		
psychologically strong	1	FC		
quite difficult	1	FC		
quite full	1	FC		
quite important	1	FC		
quite nervous	1	FC		
relatively difficult	1	FC		
seemingly impossible	1	C	270	4.94
somewhat encouraged	1	FC		
spiritually empty	1	C	13	3.39
spiritually sublimating	1	UC		
super happy	1	FC		

greatly beneficial	1	C	12	2.64
greatly encouraged	1	C	60	2.35
greatly useful	1	UC		
half full	1	FC		
hard enough	1	C	4167	3.83
highly challenging	1	FC		
highly representative	1	UC		
hugely good	1	UC		
immediately inspired	1	FC		
increasingly aware	2	C	260	3.5
advance rapidly	1	C	118	4.39
answer fluently	1	UC		
appear frequently	1	C	480	2.75
basically consist	1	FC		
believe firmly	1	C	1524	4.77
better understand	1	C	6026	3.15
briefly meet	1	C	304	2.55
briefly talk	1	FC		
calmly accept	1	C	33	2.72
calmly read	1	UC		
carefully read	1	C	1201	3.71
carefully select	1	C	484	4.78
certainly believe	1	FC		
certainly found	1	FC		
chatted occasionally	1	FC		
check carefully	1	C	335	2.63
clearly know	2	FC		
clearly understand	1	C	1414	2.25
closely link	1	C	811	5.73
commonly believe	1	FC		
communicate easily	1	C	95	2.31



totally absorbed	1	C	2	2.2
undoubtedly significant	1	C	44	2.34
universally known	1	FC		
way hotter	1	C	86	2.09
well known	1	FC		
well prepared	1	FC		
widely involved	1	UC		
<b>ADV-V</b>				
accurately express	2	C	36	2.34
absolutely love	1	C	1781	2.43
absorb abundantly	1	C	2	2.94
accomplish overnight	2	UC		
accumulate slowly	1	C	69	3.81
accurately judge	1	C	89	4.08
actively think	1	FC		
constantly increase	1	UC		
constantly reflect	1	UC		
constantly speak	1	UC		
constantly think	1	FC		
constantly update	1	C	185	4.69
contact happily	1	UC		
correctly hold	1	FC		
cried secretly	1	FC		
dabble widely	1	UC		
deeply realize	2	FC		
deeply think	1	FC		
definitely believe	1	FC		
definitely find	1	FC		
definitely give	1	FC		
develop rapidly	2	C	359	3.4
devote wholeheartedly	1	C	3	3.38

communicate effectively	1	C	820	6.71
communicate fluently	1	C	6	5.18
communicate frequently	1	C	62	2.63
communicate fully	1	FC		
communicate often	1	FC		
completely summarize	1	UC		
completely understand	1	C	1571	2.43
consciously build	1	FC		
consciously play	1	UC		
constantly broaden	1	FC		
constantly change	1	C	928	3.41
constantly discover	1	UC		
constantly encourage	1	UC		
constantly enter	1	UC		
constantly improve	3	C	216	3.09
face scornfully	1	UC		
finally applied	1	FC		
firmly believe	2	C	1524	4.77
fix timely	1	UC		
flexibly utilize	1	FC		
fluently communicate	1	C	6	5.18
fluently summarize	1	UC		
follow quickly	1	FC		
follow tightly	1	UC		
form automatically	1	FC		
form gradually	1	C	100	2.84
frequently encounter	1	C	189	4.48
fully enrich	1	FC		
fully read	1	FC		
fully understand	1	C	3787	4.41
further improve	1	FC		

directly reflect	1	FC		
directly refused	1	UC		
do better	1	FC		
do quickly	1	FC		
do successfully	1	FC		
do well	8	FC		
easily know	1	UC		
easily master	1	C	32	2.34
easily solve	1	C	185	2.58
effectively enhance	1	UC		
enjoy easily	1	UC		
enlarge further	1	UC		
eventually find	1	FC		
exactly know	1	C	16163	2.16
exercise regularly	1	C	504	6.31
explain further	1	C	19	2.03
express clearly	1	C	478	2.93
face actively	1	UC		
face bravely	1	C	51	4.6
face positively	3	UC		
greatly enhance	2	C	449	6.52
greatly expand	1	C	537	5.99
greatly improve	5	C	925	5.93
guide well	1	UC		
happened lately	1	FC		
hardly have	1	FC		
hardly know	1	FC		
highly comprehend	1	UC		
hurt deeply	1	C	313	3.3
imitate constantly	1	FC		
improve accordingly	1	FC		

further understand	1	UC		
generally believe	1	FC		
generally consider	1	C	1202	3.03
generally speaking	3	C	2253	3.66
get slowly	1	FC		
get smoothly	1	FC		
get suddenly	1	FC		
get well	1	FC		
go abroad	1	FC		
go early	1	FC		
go repeatedly	1	UC		
gradually cultivate	1	FC		
gradually dislike	1	UC		
gradually enrich	2	UC		
gradually improve	1	C	144	3.32
gradually learn	1	C	190	2
gradually lose	1	C	292	2.45
gradually noticed	1	UC		
gradually show	1	UC		
grasp patiently	1	UC		
learn well	104	FC		
like best	1	FC		
listen carefully	2	C	2170	4.91
listen closely	1	C	891	3.91
listening carefully	1	C	2170	4.91
listening clearly	1	FC		
literally melt	1	C	34	2.5
live happily	1	C	1410	4.96
live long	1	C	7829	2.64
live well	1	FC		
lose lively	1	UC		

improve greatly	1	C	925	5.93
improve quickly	1	FC		
increase greatly	1	C	1124	5.67
inevitably appear	1	UC		
inevitably encounter	1	C	21	3.14
involuntarily use	1	FC		
know conveniently	1	UC		
know curiously	1	UC		
know exactly	1	C	16163	2.16
know truly	1	FC		
learn adequately	1	FC		
learn better	1	FC		
learn comprehensively	1	FC		
learn directly	1	FC		
learn easily	1	FC		
learn effectively	1	FC		
learn efficiently	1	FC		
learn flexibly	1	FC		
learn online	1	FC		
learn purposefully	1	UC		
learn quickly	3	C	1831	2.6
learn seriously	1	UC		
learn skillfully	1	FC		
learn slowly	2	FC		
learn systematically	3	FC		
practice frequently	1	FC		
practice nervously	1	UC		
practice regularly	1	C	56	3.51
practice repeatedly	1	UC		
practice seriously	1	UC		
practice slowly	1	FC		

mainly listen	2	FC		
master firmly	1	UC		
master quickly	1	C	76	2.94
master well	2	UC		
meet constantly	1	UC		
memorize associatively	1	UC		
naturally arise	1	C	142	4.49
naturally associate	1	FC		
naturally improve	1	FC		
naturally move	1	FC		
naturally understand	1	FC		
nearly pass	1	FC		
necessarily mean	1	C	2627	3.28
observe happily	1	FC		
officially learn	1	UC		
open accordingly	1	FC		
operate correctly	1	C	46	2.39
organize quietly	1	UC		
participate individually	1	FC		
pass easily	1	FC		
pass properly	1	FC		
pass smoothly	1	FC		
passed eventually	1	FC		
personally speaking	1	FC		
play well	2	FC		
read meticulously	1	FC		
read patiently	1	FC		
read quickly	2	FC		
read quietly	1	UC		
read well	2	FC		
realize deeply	1	UC		

prepared well	1	FC		
profoundly believe	1	UC		
progress constantly	1	UC		
progress quickly	1	C	144	3.69
properly eradicate	1	UC		
put first	1	FC		
quickly catch	1	FC		
quickly find	1	FC		
quickly read	1	FC		
quickly soured	1	C	28	4.13
quickly understand	2	FC		
quite agree	2	FC		
rarely learn	1	FC		
rarely see	1	C	2088	2.06
read carefully	2	C	1201	3.71
read consistently	1	FC		
read constantly	1	FC		
read correctly	1	C	670	3.6
read extensively	8	C	81	2.08
read fluently	1	C	108	5.71
read frequently	1	FC		
read happily	1	UC		
read hard	1	FC		
read independently	1	C	154	2.34
read intensively	1	FC		
read loudly	1	UC		
sincerely hope	2	C	758	6.26
sleep early	1	FC		
slowly develop	1	FC		
slowly form	2	UC		
slowly grow	1	C	935	2.94

recall efficiently	1	UC		
recite regularly	1	FC		
reinvent constantly	1	C	50	5.34
relax mentally	1	C	15	2.02
relax physically	1	FC		
remember easily	1	FC		
remember firmly	3	UC		
remember mechanically	1	UC		
remember well	2	FC		
repeat consistently	1	UC		
run slowly	1	FC		
run tiringly	1	UC		
say confidently	1	FC		
say loudly	1	FC		
see clearly	1	FC		
see repeatedly	1	UC		
seldom come	1	FC		
seldom have	1	FC		
seldom yield	1	FC		
seldomly go	1	FC		
selectively listen	1	UC		
show bravely	1	FC		
significantly improve	1	C	1168	5.04
simply copy	1	C	143	2.55
simply put	1	C	3847	2.22
simply send	1	FC		
suddenly told	1	FC		
summarize regularly	1	UC		
surely gain	1	FC		
talk confidently	1	UC		
think highly	1	FC		

slowly integrate	1	UC		
slowly know	1	UC		
slowly learn	1	FC		
slowly precipitate	1	UC		
slowly promote	1	UC		
smartly utilize	1	C	3	4.23
socially speaking	1	FC		
speak bravely	1	C	20	2.33
speak confidently	1	C	87	3.87
speak fluently	4	C	330	7.71
speak frequently	2	FC		
speak idiomatically	1	FC		
speak loudly	2	C	665	5.08
speak properly	3	C	278	2.01
speak repeatedly	1	UC		
speak well	3	FC		
speaking generally	1	C	2253	3.66
spell well	2	FC		
spontaneously improve	1	UC		
studied hard	1	C	547	2.5
study carefully	1	C	542	4.09
study hard	11	C	547	2.5
subconsciously learn	1	FC		
successfully become	2	FC		
successfully elect	1	FC		
successfully participate	1	C	59	2.53
suddenly agree	1	FC		
suddenly bring	1	FC		
suddenly heard	1	FC		
suddenly let	1	FC		
suddenly saw	1	FC		

think quietly	1	FC		
think seriously	1	FC		
think slowly	1	FC		
think twice	1	FC		
thoroughly learn	1	FC		
told fluently	1	UC		
totally overcome	1	FC		
transform appropriately	1	UC		
travel domestically	1	C	18	3.68
treat well	1	FC		
truly absorb	1	FC		
truly believe	1	C	2889	3.4
truly learn	1	FC		
truly love	1	C	1254	2.46
try hard	1	C	7901	3.25
turned quickly	1	C	2185	2.24
understand accurately	1	UC		
understand deeply	2	FC		
understand internally	1	UC		
understand naturally	1	FC		
understand well	1	FC		
use actively	1	UC		
use appropriately	1	C	284	2.1
use authentically	1	UC		
use correctly	3	C	745	2.21
use expertly	1	FC		
use flexibly	1	FC		
use fluently	1	UC		
use frequently	1	C	1918	2.4
use properly	3	FC		
view calmly	1	UC		

talk differently	1	FC		
teach formally	1	UC		
teach systematically	1	C	39	2.38
think actively	1	FC		
work constantly	1	UC		
work hard	11	C	25386	4.83

vividly experience	1	C	13	2.65
waste carelessly	1	FC		
widely consider	1	C	765	3.59
wind fiercely	1	UC		



## VITA

NAME MISTER ZHILIANG YUE

DATE OF BIRTH 30 April 1985

PLACE OF BIRTH Chinese

INSTITUTIONS ATTENDED MS in Petroleum Engineering from University of Southern  
California  
BS in Biological Science from Fudan University

HOME ADDRESS No. 2 Xuezhai Rd., Mazhang District, Zhanjiang, Guangdong,  
China

