國立臺灣師範大學英語學系

碩　士　論　文

Master's Thesis

Department of English
National Taiwan Normal University

# 不同母語之英語學習者動名搭配詞錯誤分析

# Verb Noun Collocation Errors Made by 11 Groups of EFL Learners

指導教授：陳浩然博士

Advisor: Hao-Jan Chen, Ph.D.

研 究 生：陳玟均

Wen-Chun Chen

中華民國 105 年 6 月

June, 2016

# 中文摘要

在第二語言學習之搭配詞研究中，動詞與名詞的搭配是十分重要的。先前研究主要是探討英語學習者在動名搭配詞的過度使用、過少使用與錯誤使用。在錯誤使用的研究中，研究者大部分以人工檢查，或者是結合語料搜索工具和人工檢查的方式，在學習者的英語作文中找尋錯誤搭配詞，這些方式需要相當大量的人力與時間。此外，這些研究大部分探討一到兩個非英語系國家之搭配詞錯誤，較少研究討論多國學習者的動名搭配詞錯誤，此現象可能和現有語料搜索工具的限制有關。因此，本研究藉由 Sketch Engine 中半自動化搜索搭配詞錯誤的功能，以此相對節省時間與人力之方式，來探討十一種不同母語之英語學習者的搭配詞錯誤，並且列出普遍型錯誤（六個以上不同母語之英語學習者共同犯的錯誤）和特定型錯誤（六種母語以下或特定幾種母語人士才會犯的錯誤）。本研究討論的學習者語料為 TOEFL 11，此語料含有十一種母語之英語學習者的托福作文，這些英語學習者的母語有阿拉伯文、中文、法文、德文、印地語、義大利語、日文、韓文、西班牙文、泰盧固語（印度東部德拉維人語言）和土耳其文等十一種語言。

本研究回答以下三個研究問題：(1)對於這十一種不同母語之英語學習者，最頻繁的動名搭配詞錯誤有哪些？ (2) 這些動名搭配詞錯誤可以分為哪些類別？(3) 在這些搭配詞錯誤當中，哪些錯誤是普遍型錯誤，而哪些錯誤是特定型錯誤？ 針對第一個問題，本研究找出了 109 類之錯誤，共發生了 807 次，在這些錯誤中，較常發生的錯誤包括 agree, travel, specialize, loose 和 do。agree、travel 和 specialize 的錯誤和省略後面介系詞有關，而學習者也常搞混 loose 和 lose，此現象可能與此兩字字形和字義有某些程

度的相似有關。另外 do 屬於虛義動詞（delexical verb），學習者傾向於在虛義動詞後接任何一個名詞，因為虛義動詞較無明確意義，因而造成學習者之搭配詞誤用。

而關於第二個研究問題，本研究的搭配詞錯誤可以分為四類，分別是錯誤動詞搭配、介系詞錯誤、錯誤名詞搭配和其他錯誤等等。最後，本研究發現 61 種普遍性錯誤分別是錯誤動詞搭配(*do, *loose, *experiment, *try and *reach), 介系詞錯誤(*agree, *travel, *specialize, *listen, *search, *go, *participate), 錯誤名詞搭配 (*sale) 和其他錯誤。這些錯誤顯示出學習者對於虛義動詞、字義和字型相似之動詞、使用特定動詞時省略介系詞等共同問題。

另一方面，此研究也發現 48 種特定型錯誤，這些錯誤以他們的語言系列(language family)做分類，有十五類之錯誤屬於一個語系，二十六類錯誤為兩個語系所共享，七類錯誤為三個語系所共有，印歐語系（Indo-European）傾向產生錯誤動詞搭配，而阿爾泰語語系(Altaic)傾向犯較多介系詞錯誤。另外，韓國學習者和日本學習者常常犯相同之錯誤，此現象可能與他們皆屬於同一語系有關。

關鍵字：動名搭配詞; 半自動化; 跨語言錯誤; Sketch Engine; Sketch-Diff.

# ABSTRACT

The dimension of verb-noun collocations has been important in studies of EFL learners' collocation use. Much research has investigated the overuse, underuse or misuse of verb-noun collocations in different EFL learner groups. In terms of misuse of verb-noun collocations, these studies usually examined collocation errors in learner production with manual inspection or the combination of corpus tool and manual inspection, which needs a lot of time and efforts. In addition, they mostly observed errors made by one or two nonnative groups. Relatively few studies discussed the cross-linguistic use of collocation errors made by learners of several L1s. The possible reason might be due to the limited function of corpus tools. Therefore, the present study, introducing a semi-automated method of a query system, Sketch Engine, aims to explore the cross-linguistic use of collocation errors in eleven EFL learner groups in a more time-saving and less laborious way. The universal errors, defined as errors shared by more than six L1 groups, and specific errors, shared by certain L1 groups, will be revealed. The learner corpus examined in the study is TOEFL 11, which includes essays written by learners of eleven mother tongues, containing Arabic, Chinese, French, German, Hindi, Italian, Japanese, Korean, Spanish, Telugu, and Turkish.

Three research questions will be answered in the present study. These questions are (1) to find out the frequent verb noun collocation errors made by eleven L1 groups, (2) to explore what the error types of these miscollocations are and (3) to discover universal collocation errors shared by more than six L1 groups and specific errors made by certain L1 groups.

First, 109 types of miscollocations were found in the present research, with with 807 occurrences in total. Among these miscollocations, the frequent miscollocations containing the deviant use of *agree*, *travel*, *specialize*, *loose* and *do*, were shown in the finding. The incorrect

use of *agree, travel* and *specialize* is related to missing prepositions. The misuse of the verb *loose* is attributed to learners' confusion of *lose* with *loose* because they have similar forms and a certain degree of shared meanings. Finally, the misuse of the *do*, which is a delexical verb, might be due to the lack of specific meanings of delexical verbs.

As for the second research question, four groups of error types were found, including deviant verb usages, deviant prepositions of verbs, deviant noun usages and other types. Both deviant verb usages and deviant prepositions contain 46 error types, followed by deviant noun usages, comprising 13 types and the others containing 4 types.

Finally, there are 61 types of universal errors shared by over six L1 groups in TOEFL 11, including deviant verb usages (**do, *loose, *experiment, *try* and **reach*), deviant prepositions of verbs (**agree, *travel, *specialize, *listen, *search, *go* and **participate*), deviant noun usages (**sale*) and others. These results showed the learners' problems with delexical verbs, verbs sharing similar meanings and forms, and lack of prepositions when using certain verbs.

On the other hand, 48 specific errors shared by less than six L1 groups were found. These errors were categorized by language families with 15 types made by one language family, 26 by two language families and 7 by more than three language families. Indo-European language family tends to make more errors of deviant verbs while Altaic makes more errors of missing prepositions. Besides, Korean and Japanese learners tend to make the same mistakes quite often in TOEFL 11, attributed to their similar L1 background.

Keywords: Verb-noun collocations; semi-automated; cross-linguistic errors; Sketch Engine, Sketch-Diff.

# ACKNOWLEDGMENT

I would like to thank Prof. Howard genuinely for inspiring me to explore this research topic and giving me many suggestions to complete the present research. With the professor's encouragement and advices, I started to explore the features of *The Sketch Engine* and fortunately discovered the new method of utilizing the platform, leading to the completion of the study. Without the professor's guidance, I might not have found such an interesting topic to further investigate. My sincere thankfulness also goes to the committee members, Prof. Tseng and Prof. Chang whose valuable advices make the thesis more complete. I thank professors for reminding me of some neglected details.

On the way to the completion of my thesis, there are a lot of people that support me. Thank you, my partner, Willie, for listening to my complaints all the time and supporting me when I struggle with both thesis and teacher examination. I also express my deep thankfulness to my friends, Terri, Ignace, Christine and Boa for chatting with me when I feel impatient and frustrated. Every time after chatting with you, I feel more energetic to go on the journey. Finally, I would like to show my deepest gratefulness to my family, who financially and spiritually supports me in my graduate school study and the teacher examination. Thanks to people who love me and whom I love, this thesis was accomplished on time and without a lot of difficulties.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

## 1.1 Background and Motivation

The importance of collocations has long been discussed and emphasized in the field of second language learning for many decades by previous research (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Granger, 1998; Lewis & Conzett, 2000; Schmitt, 2004; Schmitt, Ng, & Garras, 2011; Wray, 2002). Collocations, as word combinations in *make a decision* or *commit suicide*, play an essential role to help learners think more efficiently and quickly in listening, speaking, reading and writing (Hill, 2000). Collocation knowledge is also significant in vocabulary learning because, as Woolard (2000) has stated, "learning more vocabulary is not just learning new words; it is often learning familiar words in new combination." Once learners get familiar with these word combinations, their language usages can become more accurate and fluent (Wray, 2002).

Three kinds of word combinations, including free combinations, collocations and idioms, have been discussed in studies of collocations. Collocations located in the middle of the continuum between free combinations, which are phrases freely combined depending on rules of syntax, and idioms, which are used in restricted senses and cannot be determined its meanings according to its literal meanings (Howarth, 1998; Nattinger & DeCarrico, 1992). Collocations, therefore, fall around the middle ground of the continuum and are regarded as "habitually occurring lexical combinations that are characterized by restricted co-occurrence of elements and relative transparency of meaning" (Laufer & Waldman, 2011).

Many researchers have found that foreign language learners have difficulties in the usage of collocations. Some learners may doubt that why people cannot say “conduct a diary” or “maintain a diary”; instead, they should say “keep a diary.” Similar instances occur in everyday life and Benson, Benson, and Ilson (1997) also provided some good examples of difficulties that learners might face:

> Collocations are arbitrary and non-predictable. Non-native speakers cannot cope with them; they must have a guide. They have no way of knowing that one says in English *make an estimate*, (but not **make an estimation*), *commit treason* (but not **commit treachery*). In English, one says *commit fraud* and *perpetrate fraud*. However, only the collocation *commit suicide* is possible; one does not say **perpetrate suicide*. One says *bake a cake*, but *make pancakes* (p. 258).

Previous research showed that EFL learners usually make collocation errors due to the lack of collocation knowledge in their writing or translation tasks (Bahns & Eldaw, 1993; Laufer & Waldman, 2011; Nesselhauf, 2003; Zinkgräf, 2008). Due to the avoidance of using unfamiliar collocations, many students tend to ‘rely on larger, rarer, and clumsier words which make their language sound stilted and awkward’ (Sinclair, 1991, p. 79). Cobb (2003) also mentioned that learners would produce grammatically acceptable sentences that are unidiomatic and strange to native speakers because of their unfamiliarity of collocations. Likewise, Laufer (2010) found that students had trouble finding the right verbs of collocations, and they thought they knew the collocations without a need to consult dictionaries but actually they did not. When learners face difficulties in finding the right collocations, they tend to rely on lexical simplification strategies, such as avoidance, synonyms, paraphrasing, and L1 equivalence (Mohammed Farghal & Obiedat, 1995). However, learners sometimes find it difficult to paraphrase collocations that they cannot produce (Bahns & Eldaw, 1993).

Sometimes, learners' misuse of collocations are originated from overusing certain collocations with core verbs (make, do, have, etc.) without utilizing native-like expressions due to their deficiency in collocation knowledge (Altenberg & Granger, 2001; De Cock, 1998; Granger, 1998). The overuse may be attributed to their overdependence on those high frequent verbs, with which they feel comfortable since they learn them early and use them widely (Hasselgren, 1994). Furthermore, since these verbs lack specific meanings, learners tend to overgeneralize them and combine them freely with other nouns (Zinkgräf, 2008), which also explain why learners overuse the high frequent verbs. These phenomena reveal that overusing collocations with core verbs sometimes results in learners' misuse of collocations, and learning collocations are challengeable for language learners.

Learning collocations is a long and tiring process for not only low proficiency learners, but also advanced learners. Several studies have revealed that even for advanced learners, they still make a variety of collocation mistakes in their writing tasks (Bahns & Eldaw, 1993; Mohammed Farghal & Obiedat, 1995; Nesselhauf, 2003). The finding of Gitsaki (1996) might somehow explain the difficulties among advanced learners. Gitsaki (1996) observed that Greek junior high school students' collocation knowledge did not increase as their grammatical competence improved, which showed that learners' language competence did not match collocation competence. This implies that even though advanced learners' grammatical competence is well developed, they could still have trouble dealing with collocations.

According to Benson et al. (1997), collocations can be divided into two categories: grammatical and lexical collocations. Grammatical collocations are composed of a dominant word—noun, verb, adjective, participle, and a preposition or a grammatical construction. On the other hand, lexical collocations do not contain a dominant word and they have structures, like verb + noun, adjective + noun, noun + verb, noun + noun, adverb + adjective, adverb +

verb, etc. Studies have revealed that EFL learners have relatively more problems with lexical collocations rather than grammatical ones (Bahns & Eldaw, 1993; Gitsaki, 1996, 1999). Especially, much research revealed that learners have difficulties in the usages of verb noun lexical collocations (Bazzaz & Samad, 2011; Mohammed Farghal & Obiedat, 1995; Granger, 1998; Juknevičienė, 2008; Laufer & Waldman, 2011; Liu, 2002; Nesselhauf, 2003; Todirascu & Gledhill, 2008; Zinkgräf, 2008). Therefore, it seems that verb noun collocations can be a potential area to further explore.

Previous research tended to explore verb noun collocations from two different perspectives, conducting collocation tests (Bahns & Eldaw, 1993; Bazzaz & Samad, 2011; Boers, Demecheleer, Coxhead, & Webb, 2013; Jaén, 2007) and exploring learner corpora. In terms of the latter one, most previous studies have extracted EFL learners' use of verb noun collocations from learner corpora through the method of manual extraction (Mohammed Farghal & Obiedat, 1995; Liu, 1999; Nesselhauf, 2003; Zinkgräf, 2008). The manual extraction of miscollocations is mostly time-consuming and needs substantial humor labor. For instance, Bazzaz and Samad (2011) observed Iranian EFL learners' verb noun collocations in the writing tasks and the number of verb noun collocations were counted manually. Nesselhauf (2003) extracted German advanced learners' use of verb noun collocation errors in their free writing manually. Likewise, Zinkgräf (2008) manually searched for instances of verb noun miscollocations in the writing production of 102 Spanish students.

On the other hand, some studies attempted to resort to the assistance of concordancer for the purpose of accelerating the process of extraction but researchers still needed to check each concordance to find collocations. Altenberg and Granger (2001), for example, compared French learners and Swedish learners' lexical patterning of *make* with native speakers' usages through the concordancer, *Wordsmith Tools* and manual examination. Moreover,

Juknevičienė (2008) manually selected the verb noun collocations of five core verbs (have, do, make, take, give) through the aid of *Wordsmith Tools* as well. From studies above, it can be found that many studies chose to manually search for miscollocations, which means they need to go through learners' production to achieve the purpose, which is really time-consuming. Even though some studies indeed resorted to some tools, such as *Wordsmith Tools,* to search for miscollocations, which is more efficient than pure manual extraction but it is still necessary for them to look at each concordance to find the collocation errors after they typed in certain target words, Accordingly, it seems that a more efficient way to retrieve collocation errors is needed.

Studies that discussed collocations mostly attributed learners' collocation errors to L1 interference (Altenberg & Granger, 2001; Biskup, 1992; Nesselhauf, 2003, 2005; Zinkgräf, 2008). Since these studies investigated one or two non-native learner groups, the L1 influence would be more obvious or emphasized more often. However, it seems that those collocation errors may not be totally due to the influence of learners' mother tongues. Other strategies, such as using synonyms or extension of analogy, which is the process of substituting one lexical item for another known element, might play important roles as well. For example, Källkvist (1999) found that Swedish and Norwegian learners tend to overuse delexical verbs, referring to verbs without specific meanings, such as *make*, because of L1 influence since these two languages are both Germanic languages. However, French learners also overuse those delexical verbs (Granger, 1996), which revealed that this finding is not totally L1-related. There is a need to further compare collocation errors among different L1 backgrounds to have more insight in the causes of these errors.

Although previous studies provided fruitful insight of learners' misuse of verb noun collocations, two aspects of verb noun collocations seemed to be neglected and unsolved. First, it can be observed that these studies only focused on collocation use in one or two

nonnative groups. Few studies investigated learners' collocation use across different L1 groups; in this way, a wide range of cross-linguistic differences remain unstudied. Second, most research adopted a manual method to extract the verb noun collocations, which is laborious, tiring and cannot confirm the reliable cover of all miscollocations in learner corpora. While some studies indeed utilize the corpus tools, such as *Wordsmith tools*, those researchers still need to look at each concordance carefully to find the collocations, which could be time-consuming.

These phenomena show that there still remains a need to further explore the cross-linguistic different use of verb-noun collocations in order to discover what kinds of verb noun miscollocations are universal for learners of different first languages (L1s) and what kinds of errors tend to be made only by learners of certain L1s. In order to explore the cross-linguistic differences, the innovative tool of extracting collocations is needed to save time and energy while researchers examine verb noun collocations.

Therefore, the present study aims to investigate EFL learners' verb noun miscollocations across eleven L1 groups in the learner corpus of TOEFL 11 through its comparison with British National Corpus (BNC) and Corpus of Contemporary American English (COCA) in a platform—*The Sketch Engine* to innovatively and efficiently save time and human labor.

## 1.2 Purposes of the Study

A query system—*The Sketch Engine* provides a new direction to explore collocations in different learner groups more easily and effectively. *The Sketch Engine* can allow users to view word sketches, similar words, and compare differences in corpora. Users can search for

a keyword in the word sketch so that the concordance of its collocates will show up to provide more information (Kilgarriff, Rychly, Smrz, & Tugwell, 2004).

Two special functions, *Corpus creating* and *Sketch Diff* make *The Sketch Engine* different from other concordancers (Kilgarriff et al., 2004). The *Corpus creating* function allows researchers to upload their own corpus data to *The Sketch Engine* and they can explore a keyword's collocates through making use of the functions of the query system in their corpora. On the other hand, the *Sketch Diff* function can enable users to find out the differences in a keyword's collocates of various part of speech in two different corpora, such as one native speaker corpus and one learner corpus. Users can compare the frequency of the keyword's collocates in two different corpora at the same time.

Therefore, researchers can create their corpora through the function of *Corpus creating*, and then use the *Sketch Diff* function to observe the differences of certain keywords' collocates between their corpora and other corpora. Moreover, *The Sketch Engine* can track the sources of the concordances after users zip the data that share the same background as the same file, and upload them to the platform, which can help users know the source of the data when observing each concordance. The efficiency of searching for verb noun collocations can be enhanced through the use of *Corpus creating* and *Sketch Diff*, which can be a reliable way to replace the traditional method of extracting collocations manually.

Through the use of the Sketch Engine, the present study attempts to compare EFL learners' verb noun collocation errors in different L1 groups through comparing them with native speakers' usages in a semi-automated method. The corpus of TOEFL 11, which contains learners' English production from eleven L1 groups, would be analyzed and compared with BNC and COCA so as to observe the misuse of collocations from eleven L1 groups, including Arabic, Chinese, French, German, Hindi, Italian, Japan, Korean, Spanish, Telugu and Turkish learners of English. Collocations that appear in TOEFL 11 rather than the

BNC and COCA will be possible miscollocations. The possible collocation errors will be checked by native speakers again to make sure that they are unacceptable. When the researcher searches for a keyword's verb noun collocations in TOEFL 11, the concordances of those collocations will appear by clicking on the collocates. Each concordance has a file name before it; thus, the researcher can click on the file name and then the original source (e.g. languages) will pop up so that the researcher can know learners of which L1 produced the collocations. The source of collocation errors will be found out through this way. The collocation errors will also be divided into four types as deviations in verbs, deviations in noun phrases, deviations in prepositional phrases and others based on Nesselhauf's (2003; 2005) study, which will be explained in Chapter 2.

Three dimensions will be investigated in the present study. One deals with the frequent verb noun collocation errors made by eleven L1 groups. Another is about what the error types of these verb noun miscollocations are. The other is concerned about what universal collocation errors are made by at least six L1 groups, and what errors are only made by certain L1 groups.

## 1.3 Research Questions of the Study

Based on the purposes of the present study, the following three research questions are proposed by the researcher:

1. What are the verb noun miscollocations made frequently by eleven L1 groups of EFL learners?
2. What are the error types of these miscollocations?
3. Among these miscollocations, what are the universal collocation errors shared by at least six L1 groups, and what are collocation errors made only by specific L1 groups?

## 1.4 Significance of the Study

Previous studies tend to neglect the cross-linguistic differences and similarities of verb noun miscollocations among EFL learners. Also, most researchers extracted verb noun collocations manually in previous research, which is time-consuming and laborious (Bazzaz & Samad, 2011; Mohammed Farghal & Obiedat, 1995; Granger, 1998; Juknevičienė, 2008; Laufer & Waldman, 2011; Nesselhauf, 2003; Todirascu & Gledhill, 2008; Zinkgräf, 2008). Therefore, the present study aims to use a semi-automated way to more effectively investigate EFL learners' universal and specific errors of verb noun collocations among eleven L1 groups through the corpus query system, *The Sketch Engine*.

The present study can offer some insights in four perspectives. Firstly, the study will reveal the frequently misused verb noun collocations from these eleven L1 groups. Instructors can observe what kinds of collocations that learners tend to make mistakes and pay attention to those collocations while teaching them. Secondly, among these errors, the researcher will divide these miscollocations into different error types and discuss whether the result of error types is similar to previous research so as to expand the findings of error types in previous studies. Thirdly, the methodology used by previous researchers are usually time-consuming, the present study thus used an innovative way to extract miscollcations, which is *The Sketch Engine*. This way can save efforts and time, and at the same time, avoid human misjudgment to search for miscollocations. Last but not least, due to the Sketch-Diff, the researcher can find out the sources of those collocation errors more easily, which can let the researcher discover the universal errors shared by different L1 groups. In that way, the instructors and researchers can know what errors are really problematic for most learners no matter what languages they use. Also, specific errors only made by certain L1 groups can be revealed as well.

## 1.5 Definitions of Terms

1. Verb noun collocations: According to Benson et al. (1997), verb noun collocations belong to one type of lexical collocations, which consist of verbs and nouns, e.g. *make a decision* and *keep a diary*. In the present study, verb noun collocations will be extracted by *The Sketch Engine* semi-automatically.
2. Learner corpus: A learner corpus contains learners' written or spoken production. Researchers usually compared a learner corpus with a native speaker corpus in order to find learners' underuse, overuse or misuse of language production. In the present study, the learner corpus—TOEFL 11 will be compared with the native speaker corpora—the BNC and COCA in order to find the misuse of verb noun collocations.
3. The Sketch Engine: It is a corpus query system, which allows users to observe word sketches, thesaurally similar words, sketch differences and create a corpus. The word sketch can provide one list of collocates consisting of various grammatical relations (Kilgarriff et al., 2004). The special functions of creating corpora and *Sketch Diff* in *The Sketch Engine* are utilized in the present study to extract verb noun collocations in the eleven L1 groups.
4. Semi-Automated: Traditionally, researchers need to extract verb noun collocations manually by looking at the plain text or each concordance carefully and then checking their pragmatic and semantic appropriateness to determine whether they are miscollocations or not. *The Sketch Engine*, in the present study, eases the process of extracting and determining the appropriateness of collocations by doing the part of human labor, which is considered as the semi-automated way.

# CHAPTER 2

# REVIEW OF LITERATURE

## 2.1 Collocations

In this section, the definitions of collocations will be firstly reviewed. One specific type of collocations, verb noun collocations, will be defined later. Then, the theories of collocation studies will be discussed.

### 2.1.1 Definitions and V-N types of collocations

Collocations are firstly defined as "the company words keep together" by Firth (1957, p.196). Researchers proposed different terms to refer to collocations. For instance, Alexander (1984) considered collocations "fixed expressions" and Granger (1998) termed collocations as "prefabricated patterns." The notion of collocations are related to different types of lexical expression, containing "catch-phrases, clichés, fixed expressions, formulae, free and bound collocations, idioms, lexical phrases, turns-of-phrase and so on" (C. J. Gledhill, 2000, p. 7). The distinction between these category types spurred a lot of debates and discussion among different researchers.

Various researchers have their own interpretation about the definition of collocations in previous studies. According to Frath and Gledhill (2005), there are three different perspectives of collocations: (a) co-occurrence, which refer to collocations as textual clusters commonly co-occurring together from a statistical view; (b) construction, in which collocations are considered "obeying inherent semantic and syntactic patterns of particular lexical items" (Frath & Gledhill, 2005, p. 3); (c) expression, a pragmatic view of collocations on the basis of a

relation between a sign and a textual function. However, neither of them can fully describe the notion of collocations since traditionally speaking, collocations are considered as a continuum from free combination to idioms (Howarth, 1998; Laufer & Waldman, 2011; Nesselhauf, 2003; Todirascu & Gledhill, 2008):

'Free Combination' ↔ 'Bound Collocation' ↔ 'Frozen Idiom'

Many researchers who hold the view of collocations as a continuum believe that free combinations and idioms are on the two opposite sides of the continuum. Laufer and Waldman (2011) claimed that the elements of restricted co-occurrence and relative semantic transparency differentiate collocations from free combinations and idioms. Free combinations (e.g. want a car), in which the individual words can be replaceable easily following the rules of grammar, differ from collocations based on the concept of restricted co-occurrence. On the other hand, opaque idioms (e.g. kick the bucket) are different from collocations because people cannot determine their meanings from the individual words that compose them, which is concerned about relative semantic transparency. However, sometimes it is hard to differentiate free combinations and collocations. It seems that free combinations occur without restriction but actually they are still bounded in some cases. Take the free combination "accept murder" for example, the word "accept" cannot be replaced by any synonym like "take." The model proposed by Benson et al. (1997), therefore, might solve the problem of defining collocations. Benson et al. (1997) considered collocations as the following:

> In any language, certain words regularly combine with certain other words or grammatical constructions. These recurrent, semi-fixed combinations, or collocations, can be divided into two groups: grammatical collocations and lexical collocations. Grammatical collocations consist of a dominant word—noun, adjective/participle,

verb—and a preposition or a grammatical construction. Lexical collocations, on the other hand, do not have a dominant word; they have structures such as the following: verb+noun, adjective+noun, noun+verb, noun+noun, adverb+adjective, adverb+verb. (p.xv)

Therefore, in the present study, we follow Benson et al. (1997) definition considering collocations as "fixed, identifiable, non-idiomatic phrases and constructions" (p. xv), such as *commit suicide* and *make a mistake*. Some free combinations might be referred to either grammatical or lexical collocations, based on the features they have (Wei, 1999). In order to understand the collocation types more explicitly, the two broad categories of grammatical and lexical collocations are presented in Table 2.1 and Table 2.2.

Table 2.1 Grammatical collocations adapted from Benson et al. (1997)

| **Rules** | **Examples** |
|---|---|
| (1) noun + preposition | *The blockade of enemy ports* by our navy. |
| (2) noun + to + infinitive | We made *an attempt to do* it. |
| (3) noun + that-clause | She took *an oath that* he would do his duty. |
| (4) preposition + noun | We discovered it *by accident.* |
| (5) adjective + preposition | She was *angry at* my friends. |
| (6) adjective + to + infinitive | She is *ready to go*. |
| (7) adjective + that-clause | It is *necessary that* he be replaced immediately. |
| (8) 19 verb patterns, including: verb + to + infinitive, verb + gerund, verb + object 1 + object 2, and others | She *continued to write.*<br>They *enjoy watching* TV.<br>The police *fined her fifty dollars.* |

Table 2.2 Lexical collocations adapted from Benson et al. (1997)

| **Rules** | **Examples** |
|---|---|
| (1) verb (creation/activation) + noun | She *does the laundry* every day. |
| (2) verb (eradication/nullification) + noun | The teacher *declined* our *invitation.* |
| (3) adjective + noun / noun + noun | The room has a *sour smell*. |
| (4) noun + verb (action) | *Bombs exploded* across Bangladesh. |
| (5) noun (unit) + of + noun | David gave Sandy *a bouquet of flowers*. |
| (6) adverb + adjective | They are *closely acquainted.* |
| (7) verb + adverb | They *argued heatedly* in that debate. |

Bahns and Eldaw (1993) and Gitsaki (1999) stated that EFL learners have relatively more problems with lexical collocations than grammatical ones in their language production since they tend to make more errors in lexical word combinations. Among various types of lexical collocations, many previous studies found that EFL learners face difficulties especially in verb noun lexical collocations (Bazzaz & Samad, 2011; Mohammed Farghal & Obiedat, 1995; Granger, 1998; Juknevičienė, 2008; Laufer & Waldman, 2011; Nesselhauf, 2003; Todirascu & Gledhill, 2008; Zinkgräf, 2008). According to Cowie (1992), verbs can include three kinds of features in verb noun collocations:

1. Figurative: in "*deliver a speech*," the keyword *deliver* has the figurative or abstract meaning of conveying certain ideas to the listeners, which goes beyond its original meaning of sending an object to others.
2. Delexical: in "*make recommendations*," the verb *make* does not have a specific meaning but is rather grammatical and vague. In this type of collocations, their meanings depend on the nouns since delexical verbs do not contain special meanings.

3. Technical: in "*try a case*," the action *try* whose meaning is constrained, narrow and specific, exhibits the meaning of the collocation.

Since verb noun collocations were found problematic for L2 learners from previous research, the present study aims to focus on this specific type of collocations in learner corpora. Previous studies in verb noun collocations tend to explore learners' overuse, underuse or misuse of collocations. The use of verb noun collocations and also other types of collocations is usually explained by theories of the Open Principle and the Idiom Principle, which will be further explained in the following part.

### 2.1.2 Theories of Collocations

Sinclair (1991) suggested that both spoken and written human discourse are governed by two major principles: the Open Choice Principle (OP) and the Idiom Principle (IP), which are two contrary camps of thoughts. The OP, based on Chomskyan School of linguistics and the theory of Universal Grammar, claims that people have the ability to grammaticalize meaning in natural language in accordance with limitation rules of sub-categorization and selection in a language while there can be parametric dissimilarities between different languages. People construct the meanings of collocations by adding up the meaning of each individual word according to the OP. The IP, on the other hand, is related to a grouping of socio-lexical conventions in any language, which reveals that collocation meaning is constructed through the "whole phrase" of a collocation. The similar idea is also revealed in the statement of Sinclair (1991, p. 110), "the principle of idiom is that a language user has available to him a larger number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments."

For instance, in English greeting, “good morning” resorts to the IP rather than the OP since the expression of “wonderful morning” or “excellent morning” as greeting is unavailable. The lexeme *morning* can be freely combined with adjectives, such as *nice*, *terrible, wonderful* and *excellent* based on the OP; however, in the greeting, the only acceptable and frozen expression is “good morning,” which can be explained by the idea of the IP instead of the OP (M Farghal & Al-Hamly, 2007).

Bazzaz and Samad (2011) examined Iranian learners’ use of collocations and supported the importance of the idiom principle in their finding, indicating that the role of collocations should be highlighted since collocation knowledge is more beneficial than knowledge of individual words for EFL learners. According to Granger (2011), the IP considers each word form as having its phraseology, which explains learners’ difficulties since learners are not capable of resorting to this principle in their language production. This phenomenon was also shown in the finding of Zinkgräf (2008), who stated that learners tended to rely on the OP rather than the IP and ignored some restrictions of word combinations when they used collocations. The consequence of resorting to the OP for learners is that collocation errors add some foreign flavor to learners’ speech and writing, and make them differ from native speakers’ production. Mohammed Farghal and Obiedat (1995), who explored Arabic learners’ use of collocations, also reported that the participants were highly dependent on the OP for item selection, leading to wrong and incorrect collocations. These studies supported Sinclair’s (1991) claim that learners tend to rely on the OP while native speakers actually depend on the IP, which leads to EFL learners’ difficulties in collocation usages.

## 2.2 Difficulties Faced by EFL learners

Regarding learners’ use of verb noun collocations, most studies have indicated EFL learners’ deficiency in collocation knowledge, such as the overuse, underuse or misuse of

certain collocations in their language production (Altenberg & Granger, 2001; Hasselgren, 1994; Juknevičienė, 2008). Some studies indicated that learners' knowledge of lexical items is better than knowledge of collocations comprised of the same lexical items. For instance, Barfield (2006) found that Japanese learners had better knowledge of single nouns and verbs than the collocations which consist of the same nouns and verbs. Additionally, not only low proficiency learners but also advanced learners have difficulties in the use of collocations (Altenberg & Granger, 2001; Bahns & Eldaw, 1993; Zinkgräf, 2008). These phenomena showed learners' difficulties in collocation usages; thus, many researchers utilized different methods to conduct collocation studies for the purposes of providing more insight in learners' use of collocations and offering pedagogical implications. Therefore, in the following two sections, the methods of exploring collocation knowledge will be firstly reviewed and the findings, which especially indicated the error types and causes of miscollocations, will be demonstrated later.

### 2.2.1 Methods to Explore Collocation Knowledge

Researchers found learners' problems in verb noun collocation use through three methods: collocation tests, elicitation studies of collocations and error analyses of collocations. The latter two methods are usually related to exploring learner corpora while the first method is to design collocation assessment for learners, which will be reviewed in the following three parts.

#### I. Collocation Tests

Previous studies used a variety of collocation tests, such as translation tasks or cloze tests to examine learners' collocation competence, and found that learners' collocation competence was still limited. For example, in a study of Bazzaz and Samad (2011), they found that there is a linear relationship between Iranian EFL learners' collocation knowledge and their use of

collocations in writing stories by statistically comparing the result of collocation c-test with manually extracted verb noun collocations from the stories. Additionally, the work of Bahns and Eldaw (1993) developed two instruments, a translation task and a cloze test, to test German advanced learners' productive knowledge of verb noun collocations. The result found that the number of collocation errors was twice more than the number of errors about single lexical items in the translation tasks among these EFL learners. Jaén (2007) not only assessed productive collocational knowledge but also receptive knowledge by an 80-item test consisted of multiple choice questions and cloze questions among Spanish university students. Expectedly, the result revealed that the score of productive knowledge items are significantly lower than test items of receptive knowledge, which shows that EFL learners' collocation productive knowledge still needs to be improved.

Boers et al. (2013), on the other hand, conducted four trials of pretests and posttests, which are fill-in-the-blank tests, in order to compare the effectiveness of collocation matching exercises in textbooks with exercises containing collocations as intact wholes for the purpose of finding out which exercises can better improve learners' collocation competence. The finding indicated that the negative side effects were found more often in the matching exercises than the exercises where collocations are intact since the matching exercises might lead to error connections of collocations in learners' memory. In addition to collocation tests, other studies examined learners' collocation knowledge in their actual usage of collocations through learner corpora, including elicitation of certain collocations and error analyses of all collocations in learner production.

## II. Elicitation Studies of Collocations

In terms of elicitation of certain collocations, researchers searched for the pre-determined set of collocations in learner production to examine learners' collocational

competence. These pre-determined set of collocations is usually the collocations with delexical verbs or high frequency verbs, such as *have*, *do*, *make*, *take* and *give*. High-frequency verbs contain some characteristics that make them specifically interesting from a cross-linguistic perspective (Viberg, 1996):

1. They express basic meanings and often prevail various semantic fields;
2. They have equivalents that are frequently used in most languages;
3. They contain a high degree of polysemy, resulted from two types of meaning extension:

   --one universal tendency producing more abstract, general, grammaticalized or delexicalized usages,

   --a variety of language-specific tendencies leading to specialized meanings, collocations, and idiomatic usages;
4. They are usually problematic for EFL learners.

Zinkgräf (2008, p. 108) stated, "under the assumption that these verbs lack a specific meaning, learners overgeneralize and combine them with any noun under the illusion that there is no restriction to the way they can be used." Boers et al. (2013), similarly, mentioned that it is usually the verb in collocations that learners replace by an unconventional choice (*do mistakes), and such replacement might be originated from L1 interference when equivalent L1 nouns can collocate with a different verb (Nesselhauf, 2005; Yamashita & Jiang, 2010). Learners' dependence on L1 and their tendency to ignore restriction on word combinations could imply that learners construct meanings from individual words instead of from prefabricated patterns, unlike native speakers (Kjellmer, 1991; Wray, 2002), which corresponds to the Open Choice Principle. Thus, learners tend to consider delexical verbs as having unrestricted collocations, which leads to learners' misuse of verb noun collocations.

Two contradictory findings of delexical structures have been discussed by previous studies. Some studies pointed out the overuse of collocations with delexical verbs by EFL

learners. For example, Hasselgren (1994) observed that core words, such as high frequency verbs, are widely overused by Norwegian learners who tend to rely on them, like "lexical teddy bears." She explained, "core words—learnt early, widely usable, and above all safe (because they do not show up as errors) are usually overused, even among learners sufficiently advanced to have been weaned off them"(Hasselgren, 1994, p. 250). A similar finding is also revealed in the study of Källkvist (1999), which showed the overuse of collocations with delexical verbs among Norwegian and Swedish learners. Granger (1996) found French learners also overuse those high frequency verbs in their collocation usages.

On the contrary, other studies revealed that learners tend to underuse collocations with delexical verbs. Sinclair (1991, p. 79) reported, "many learners avoid the common verbs as much as possible, and especially where they make up idiomatic phrases. Instead of using them, they rely on larger, rarer, and clumsier words which make their language sound stilted and awkward." For instance, Altenberg and Granger (2001) explored French and Swedish advanced learners' use of grammatical and lexical collocations of MAKE in International Corpus of Learner English (ICLE) through Wordsmith Tools, and found that both Swedish and French learners underuse the delexical combinations, which correspond to Sinclair's underuse hypothesis. These advanced learners not only underused delexical structures but also made several mistakes in collocates with a high frequency verb, *make*.

Similarly, Juknevičienė (2008) conducted a pilot study to examine Lithuanian learners' collocations with high frequency verbs, including HAVE, DO, MAKE, TAKE and GIVE in comparison with native speakers' usages quantitatively and qualitatively by the assistance of Wordsmith Tools. The result showed that Lithuanian learners underused the typical collocations with delexical verbs in academic writing compared to native speakers, which can be explained through their deficiency of collocation knowledge. To compensate for the lack of knowledge, learners tended to resort to their L1 to create their own collocations, which

were rarely used by native speakers. Nesselhauf (2004), on the other hand, analyzed German learners' miscollocations of delexical verbs, such as *have*, *take*, *make* and *give*. Although she did not discuss the dimension of underuse and overuse of miscollocations, she raised some typical learners' difficulties of verb noun combinations and divided them into different error types, such as wrong verbs (*take changes) and wrong nouns (*make a trial) in the finding.

From these studies, although the overuse or underuse of delexical verbs in collocations is still paradoxical and inconclusive, these phenomena can sometimes lead to the misuse of verbs in verb noun collocations, which can explain the partial causes of verb noun collocation errors. Besides, the importance of delexical verbs in collocation usages can be revealed in these studies.

## III. Error Analyses of Collocations

In addition to the studies eliciting certain collocations with delexical structures, other studies adopted a broader approach to examine the overall collocations of complete texts created by students. Some of them resorted to the manual extraction of collocation errors in learner corpora while others observed learners' collocation usages through the assistance of corpus tools and manual inspection.

Traditionally speaking, researchers tend to investigate learners' use of collocations manually. For example, Liu (1999) collected 94 learners' writings and 127 exam papers among Taiwanese EFL learners and manually checked learners' miscollocations. The finding showed that 63 collocation errors were found in their written production and most of them are verb noun collocation errors. According to Chen (2002), she manually checked learners' collocation errors in 90 compositions from 30 Taiwanese senior high school students through consulting the BNC. If a collocation appeared in the BNC, it would be considered as the correct one; however, if no occurrence of the collocation appeared in the BNC, the

collocation would be identified as the wrong one. In the end, 147 grammatical collocation errors and 125 lexical errors were found in her study.

Nesselhauf (2003) manually counted the verb noun combinations in German advanced learners' 32 essays from The International Corpus of Learner English (ICLE) and categorized the verb noun combination mistakes into three types as free combination errors, collocation errors and idiom errors through looking up the Oxford Advanced Learner's Dictionary (OALD 2000) and the Collins COBUILD English Dictionary (CCED 1995). She found that among collocation mistakes, the most frequent type of mistakes is the wrong choice of verb.

Two years later, Nesselhauf (2005) went on to analyze a larger German corpus from ICLE. She searched for verb noun miscollocations manually by using the BNC, dictionaries and consultation from native speakers. First of all, the collocations in learner production would be considered acceptable if they appeared in the following dictionary: Oxford Advanced Learner's Dictionary 2000, the Collins COBUILD English Dictionary 1995, The BBI Dictionary of English Word Combinations 1997, and Oxford Dictionary of English Idioms 1993. The BNC would also be consulted to assist the researcher to determine the acceptability of the collocations. If the researcher still could not decide whether the collocation was right or wrong, she would consult native speakers to find out the answers. Finally, her result revealed that among 2078 verb noun collocations, 744 of them were misused by German learners.

Zinkgräf (2008) also extracted the number of verb noun miscollocations manually from 13 written assignments by 102 high-intermediate Spanish students through consulting several dictionaries, and he found that the wrong choice of verbs and nouns accounted for 70% of the corpus. These studies checked for miscollocations manually in learner production, which is time-consuming and laborious.

Recently, more and more researchers attempted to make use of some corpus tools, such as Wordsmith Tools or Part-of-Speech (POS) taggers to ease the process of searching verb noun collocations. Laufer and Waldman (2011), for instance, investigated the collocation use in both the Israeli Learner Corpus of Written English and Louvain Corpus of Native English Essays (LOCNESS). They used the word list function of Wordsmith Tools to retrieve the list of frequent nouns in the LOCNESS and then analyzed the list in the Vocabulary Profile to produce 220 most frequent nouns in English. Later, the researchers created the concordances for each noun in both corpora so as to extract verb noun collocations through checking each of them in The BBI Dictionary of English Word Combinations and The LTP Dictionary of Selected Collocations. Combinations that occurred in the two dictionaries would be considered as collocations. The collocation errors and the proficiency differences in the learner corpus would be analyzed as well. Their result indicated that L2 learners underused verb noun collocations in their essays compared to native speakers. While advanced learners used more collocations than low and intermediate level students, they made more mistakes as well, which implies the more collocations they produced, the more errors they made.

Todirascu and Gledhill (2008) attempted to develop a semi-auto method to extract collocations through analyzing the Acquis Communautaire Corpus (ACC), a parallel corpus available in all European languages (Steinberger, Pouliquen, Widiger, & Ignat, 2006), through conducting various POS taggers, such as TreeTagger and TTL – a complex tool to pre-process texts (Ion, 2006). After tagging the corpus, the researchers applied a statistical module, using Log-likelihood (LL) scores, from the studies of Steinberger et al. (2006) and Todirascu and Gledhill (2008) in order to identify verb noun constructions. A list of verb noun co-occurrences was created through the statistical module. The finding, however, showed that the statistical methods alone were not enough to identify verb noun collocations

and a manual inspection was needed to sort the data of the result into verb noun collocations and irrelevant combinations.

Lin (2010) also explored verb noun collocations in tagged corpora, including a Taiwanese learners' corpus, a Chinese learners' tagged corpus and the BNC, through the tools of *Antconc*, *Monoconc* and *Perl*. The researcher firstly generated a list of thirty-three verb noun collocation patterns in the corpora through the *Antconc* program. Later, *Monoconc* was used to retrieve these patterns of verb noun collocations in the three corpora. Then, she compared and deleted the same patterns between two learner corpora and the BNC by *Perl*. In this way, only the potential miscollocations were remained since these usages did not appear in a native speaker corpus, the BNC. The potential miscollocations were inspected manually to delete the acceptable and correct ones and the real miscollocations were found in her study. The result listed several collocates that were frequently misused by Taiwanese and Chinese learners. Most miscollocations were attributed to L1 interference and the wrong choice of verb could be found in most of the collocation errors.

More innovatively, Liu (2013) utilized *The Sketch Engine*, a platform capable of performing the usages of *Antconc*, *Monoconc* and *Perl* in Lin's (2010) study, to explore Taiwanese and Chinese verb noun collocation errors and their causes through the comparison with the BNC. *The Sketch Engine* enabled the researcher to create his own corpora, provide a keyword's collocates and spot the difference of collocates between two corpora. The researcher found the potential miscollocations through recording collocations that only appeared in the learner corpus instead of the BNC through the function of *Sketch Diff*, which can spot differences of collocates in two corpora. In the end, a total of 134 types of miscollocations were found in the finding and its possible causes were described by checking them in several dictionaries and online resources.

From error analyses of collocations in learner corpora, one can observe that most previous research investigated collocations in a time-consuming and laborious way. Even though some research examined collocations through corpus tools and tagged corpora, some methodological limitations can be revealed in those studies. For example, although Laufer and Waldman (2011) utilized Wordsmith Tools to ease the process of searching for collocations, they still needed to observe a concordance after a concordance in the corpus. Todirascu and Gledhill (2008) can retrieve a list of collocations based on the tagged corpus and LL scores semi-automatically but this method cannot help researchers identify possible collocation errors. Lin (2010) indeed explored collocation errors in learner corpora through the comparison with the BNC; however, her procedure was really complicated and a lot of corpus tools, such as *Antconc*, *Monoconc* and *Perl* were needed. Therefore, *The Sketch Engine*, which can save time and efforts in the work of Liu (2013), will be a more optimal option for the present study to further discuss learners' cross-linguistic differences and similarities of collocation usages.

From the results of previous studies, some researchers categorized the collocation errors into different types and attempted to describe the possible causes of the miscollocations, which will be further elaborated in the following part.

### 2.2.2 Error Types and Causes of Miscollocations

Nesselhauf (2003) categorized German learners' verb noun combination errors into three types as free combination errors, collocation errors and idiom errors. Regarding collocation errors, he divided verb noun miscollocations into nine types:

1. Deviant verbs: **carry out races* (hold races)
2. Deviant nouns: **close lacks* (close gaps)
3. Inappropriate combinations: **take notice* (to notice)

4. Not existed combinations: **hold children within bounds* (show children where boundaries lie)
5. Inappropriate prepositions of verbs: **fail in one's exam* (fail one's exam)
6. Inappropriate prepositions of nouns: **raise the question about* (raise the question of)
7. Determiners (lack of or extra determiners): **get the permission* (get permission)
8. Numbers (nouns used in the singular instead of the plural): **pass one's judgments* (pass judgment)
9. Wrong syntactic structures: **make sb. friends* (make friends with sb.)

Nesselhauf (2005) went on to analyze a larger German corpus from ICLE and categorized the miscollocations into deviations in verbs, deviations in noun phrases or prepositional phrases and more global deviations, which contain stretched verb construction instead of the corresponding verb, whole collocation inappropriate and deviations in the structure of the collocation. Most of the types are similar to her study in 2003. Only the categorization of more global deviations is somewhat different. Summarized from Nesselhauf (2003; 2005), error types which appear in both studies will be the standard of error types in the present study since these types are more common. Therefore, in the present research, the researcher will categorize the collocation errors into deviations in verbs, deviations in noun phrases, deviations in prepositional phrases and others to understand the general distribution of the error types.

The reasons why learners misuse verb noun collocations can be attributed to two main aspects, interlingual errors or intralingual errors. Many researchers found learners' atypical collocations are mostly interlingual, that is, their errors are influenced by their mother tongues (Altenberg & Granger, 2001; Mohammed Farghal & Obiedat, 1995; Granger, 1998; Laufer & Waldman, 2011; Nesselhauf, 2003; Zinkgräf, 2008).

For instance, Nesselhauf (2003, 2005) suggested that German learners' verb noun miscollocations were mostly attributed to their mother tongue. Altenberg and Granger (2001) concluded that collocation errors with delexical verbs made by French and Swedish learners were due to the influence of their mother tongues. Likewise, Laufer and Waldman (2011) searched for verb noun miscollocations in Israeli Learner Corpus of Written English and demonstrated that more than half of the collocation errors were influenced by learners' L1. From the result of Juknevičienė (2008), the researcher explained that Lithuanian learners resorted to their L1 to compensate for their deficient collocation knowledge, which usually resulted in the misuse of collocations. In the study of Zinkgräf (2008), the finding showed that 61% of Spanish learners' atypical collocations were generated from the interference of L1. Biskup (1992) indicated that the wider the meaning scope of a lexical item, the more the L1 interference in the collocations that lexical items took part in. Likewise, the more synonyms a lexical item had, the more difficulties learners faced when producing a restricted collocation.

In addition to L1 interference, when learners fail to find appropriate verbs or nouns of collocations, they will resort to some lexical simplification strategies, such as the use of synonyms, avoidance, paraphrasing, extension by analogy or reliance on the Open Choice Principle (Mohammed Farghal & Obiedat, 1995; Zinkgräf, 2008), which can explain learners' intralingual errors shared by learners from different L1 backgrounds. Howarth (1998) analyzed several studies and summarized a few strategies that learners frequently use when facing difficulties in collocation knowledge:

1. Avoidance: When learners are unable to produce collocations or they have difficulties in paraphrasing, they tend to avoid using those collocations. For instance, Biskup (1992) found that Polish students were more inclined to avoid answering questions while German students tended to paraphrase the unknown collocations.

2. Experimentation: From the study of Biskup (1992), the author concluded that German learners were risk-takers and when they did not know the correct form of collocations, they could find some synonyms to replace them. This kind of learners' experimentation is also found in the results of Zinkgräf (2008) and Mohammed Farghal and Obiedat (1995).
3. Transfer: Learners' risk-taking behavior might lead to L1 transfer even though learners may not be aware of problems in using L1 equivalence in L2 collocations. **State a record* (set the record) by Polish learners and **lead a bookshop* (run a bookshop) by German learners from Biskup's (1992) study could be good examples.
4. Analogy: The process of substituting one lexical item for another known element in a collocation can be considered as L2 intralingual transfer. As Howarth (1996) indicated, a nonnative speaker can use the following verbs, including *adopt*, *base on*, *bring in*, *discuss, implement* and *modify* as the collocates of *method*, which can lead to a great degree of variation. However, analogy may also give rise to overgeneralization of collocation usages, such as **adopt a way* (adopt an approach).
5. Repetition: A nonnative writer would repeatedly use a limited number of collocations when they lack confidence to produce collocations with analogies. Granger (1998), for example, concluded that French learners in her study tended to overuse certain adjective collocations with *very,* and some constructions like *deeply-rooted* appeared several times in the finding.

These learner strategies could evolve into the causes of leaners' atypical collocations. Based on the discussion of Liu (1999), investigating miscollocations from 94 copies of general writings and 127 exam papers written by college students, and Chang and Yang (2009), exploring Chinese learners' causes of collocation errors modified from the research of James (1998), several possible causes of EFL learners' verb noun collocation errors were summarized as follows:

1. Overgeneralization: If a word contains more than one usage, learners may overuse the word. For example, since "worry" can be a verb or a noun, some learners may use the wrong sentence, **I am worry about you* instead of "*I am worried about you*."
2. False analogy: The incorrect usage **ask you a favor* can be an extension from the usage of Verb + Noun + O.C.
3. Erroneous assumption: it can be found in delexical verbs, such as **do plans* (the correct one as *make plans*). EFL learners tend to make wrong guesses of a delexical verb's meaning.
4. Erroneous use of synonyms: The phrase **broaden your eyesight* (*broaden your horizon* as the right one) can be an "straightforward application of the Open Choice Principle" by learners (Mohammed Farghal & Obiedat, 1995).
5. Negative transfers: Negative influences originate from learners' L1. **Eat medicine* instead of *take medicine* can be a good instance among Chinese learners.
6. Erroneous coinage: Due to vocabulary deficiency or a lack of collocational knowledge, learners may make a mistake of **see sun-up* rather than *see sunrise*.
7. Approximation: When two words share the similar forms, learners may feel confused and make an error, like **release my pressure* while it should be "*relieve my pressure."* In learners' memory, the word "release" could be quite similar as "relieve."
8. Ignorance of TL rules: Learners ignore the rules and the structures of the target language, such as **go school* (go to school).
9. Ignorance of semantic constraints of collocates: Learners ignore the sematic constraints of words, like **cost several days* (take several days) and **eat tea* (drink tea). Learners ignore the semantic constraints of *cost* and *eat*.
10. Incomplete rule application: It means the under-generalization of target language rules, like **catch others* (catch up with others).

11. Hypercorrection: Learners made collocation errors originating from learners' over-monitoring their L2 production and deliberately suppressing possible L1 negative transfer, such as **establish an evaluation body* (make evaluation).

From studies discussed above, it can be concluded that both interlingual and intralingual errors were explored in one or two nonnative groups in each study. Though various sources of collocation errors were summarized by previous research, the causes of collocation errors might be limited to the scope of learners' mother tongue, which might not be true since the same collocation errors might be made by learners from different L1 backgrounds as well. It can be observed that the causes of collocation errors concluded by previous studies might be wrong. Sometimes, the causes of miscollocations are too complex to explain or there is more than one cause of the same miscollocation. It is hardly possible for researchers to explain every cause of a miscollocation since the cause of the miscollocation made by different learners might vary from case to case. Therefore, on one hand, the intralingual errors, which are not only related to L1, need more studies that focus on learners of several L1s to be conducted for gaining more insight. On the other hand, studies that tended to explore the causes of miscollocations should not convey the strong claim about the causes of these errors. Accordingly, the present study aims to investigate the cross-linguistic use of collocation errors, categorize them into different error types and observe whether there are some possible causes of miscollocations instead of explaining the causes of miscollocations certainly. That is, the researcher will put more emphases on cross-linguistic use of collocation errors and error types, and at the same time, some insights about possible causes of miscollocations might be revealed as well.

## 2.3 Cross-linguistic Use of Collocations

Relatively few studies pointed out the wide range of cross-linguistic similar or different use of collocations in previous research since corpus data in most studies were learner production from the same country. These studies tended to attribute collocation errors to learners' mother tongues but sometimes it was not simply due to L1 influence. For example, Källkvist (1999) found that Swedish and Norwegian learners tend to overuse delexical verbs due to L1 influence because the learners' L1s, like Swedish and Norwegian, are Germanic languages. However, French learners also overuse those delexical verbs (Granger, 1996), which showed that this phenomenon is not completely L1-related. Nesselhauf (2003) suggested that collocations that are not congruent with learners' mother tongue should be paid more attention based on her finding. Nevertheless, she mentioned the mistakes were also made in the congruent collocations so future studies should explore what congruent collocations that learners tend to make mistakes. These phenomena imply that factors other than L1 influences should be further investigated. The optimal method of exploring factors other than L1 transfer is observing errors shared by learners of several L1s.

Studies that discussed cross-linguistic use of collocations were usually limited to two or three languages at most. Källkvist (1999), for instance, examined the use of collocations among Swedish and Norwegian learners. Altenberg and Granger (2001) compared the collocation use among French learners, Swedish learners and native speakers. Wang and Shaw (2008) examined Swedish and Chinese learners' use of collocations and found both groups made a similar number of collocation errors, which cannot be explained by their mother tongues. Biskup (1992), on the other hand, observed Polish and German learners' use of collocations, and found that Polish learners, on the basis of L1 equivalents, extended the meaning of L2 collocations successfully while German learners made more collocation mistakes when they switched between L1 and L2.

These studies showed that it is unclear about collocation errors shared by many nonnative groups, that is, what kinds of collocation errors are universal for language learners. Also, among these learner groups, what collocation errors are only made by learners of certain L1s, which might be due to their L1 transfer or other reasons. As Ringbom (1998, p. 198) stated, " the relationship between transfer and possible universal learner language characteristics is too complex to allow more than a few tentative points to be made on the material so far available." Therefore, this relationship should be "further investigated by error analysis, discourse analysis and by studying the frequency differences against the background of the contexts where the words are used" (Ringbom, 1998, p. 198). This implies that there still remains a need to explore collocation errors shared by learners from different L1 backgrounds, and errors specific to learners of certain L1s through error analyses in learner corpora accordingly.

## Summary of Chapter Two and Research Gaps

In the sections of 2.1 to 2.3, the definitions of collocations and verb noun collocations were reviewed carefully. Then, learners' difficulties in collocation use, including the underuse, overuse and misuse of collocations, were discussed as well. Researchers tended to explore learners' collocation knowledge through conducting collocation tests and examining learner corpora, which were reviewed separately. Then, different methods of retrieving collocations in learner corpora were also demonstrated. The researcher later summarized the error types and causes of verb noun collocations in different studies. Finally, the cross-linguistic use of collocations was raised up to bring out the main research gap in the field of collocation studies. From the discussion above, three issues can be raised from previous research.

First of all, acquisition of verb–noun collocations has shown to be particularly problematic for L2 learners. Not only low proficiency learners but also advanced learners have difficulties in verb noun collocation usages. This indicates that collocation knowledge may not gradually develop with language competence. Therefore, there remains a need to conduct studies to help EFL learners and teachers understand what collocations would be more problematic for language learners.

Secondly, the methodologies of previous studies were usually time-consuming and laborious. Relatively few studies, like Lin (2010) and Liu (2013), could use a semi-automatic ways to extract collocations without the need to see a concordance after a concordance. Thus, there remains a need to further explore learners' collocations through the innovative way conducted by Liu (2013). Also, the source tracing function of the Sketch Engine, which will be introduced in the methodology, provides another direction of explore cross-linguistic differences.

Thirdly, previous studies, no matter what dimension they emphasized, such as collocation tests, elicitation of certain collocations and error analysis of all miscollocations in learner corpora, focused on one or two nonnative-speaker groups. Few studies indeed explored the collocation errors made by several nonnative groups, which makes the cross-linguistic difference and similarity of collocation errors a potential area to further investigate. Furthermore, previous studies observed that learners' collocation errors were mostly from L1 interference, which might not be totally true since there is no comparison of learner production from other mother tongues. Other strategies, such as avoidance, synonyms, paraphrasing, are thus neglected or less discussed. Through the cross-linguistic observation, more insight may appear in these aspects.

Therefore, the present study aims to explore nonnative speakers' collocation errors to observe the cross-linguistic similarities and differences of collocation use. The universal

errors and errors specific to certain L1 groups can be found through the innovative and special function of the query system, *The Sketch Engine*, which will be introduced in the following chapter.

# CHAPTER 3

# METHODOLOGY

In this session, the corpora, the query system, the procedure of data collection and analysis in the present study will be introduced. First, the two native speaker corpora and one learner corpus will be presented. Second, the detailed function of the query system, the Sketch Engine, will be elaborated. Finally, the process of collecting and analyzing data will be represented.

## 3.1 Corpora

### 3.1.1 BNC

The British National Corpus (BNC) is 100 million words of written and spoken language from a wide range of sources, developed by Oxford University Press to represent a wide cross-section of British English from the late 20th century. The BNC contains 90 percent of written language and 10 percent of spoken language. The written and spoken production contains various sources from university essays and business meetings to informal interviews, newspapers and radio shows, etc. In the present study, only BNC written corpus is included as a native speaker's corpus since the learner corpus, TOEFL 11, is a written corpus as well.

The BNC is composed of four distinct and essential features, including monolingual, synchronic, general and sampling. First, the BNC is a monolingual corpus, which deals with modern British English instead of other languages used in Britain. Second, it is synchronic corpus since it copes with late twentieth English rather than old English usages. Third, because it covers different styles and various genres of English without restrictions to any specific field, register or genre, it can be the general corpus used widely. Finally, from different parts of

single-author texts, 45,000 words are extracted for written language. Shorter texts, which contain the maximum of 45,000 words and muti-author texts, like magazines and newspapers, are included in full. This kind of sampling enables the corpus to have a wide range of texts within the 100 million words limit, and prevents from over-representing certain texts.

### 3.1.2 COCA

The Corpus of Contemporary American English (COCA) is the only large and balanced corpus of American English, which is the largest freely available corpus of English. The corpus includes more than 450 million words of texts comprised of 20 million words each year from 1990-2012, and was updated regularly. The corpus text is equally divided among fiction, newspapers, popular magazines, spoken and academic texts. Due to its updating design, it is probably the only corpus that can investigate the ongoing and current changes of the language. The present research only includes the academic texts of COCA as the reference of learner corpus.

The BNC written corpus and the academic texts in COCA will be combined as a large native speaker corpus, labeled as BNCCOCA, containing 222,422,474 word tokens and uploaded to the query system in order to have the whole coverage of both British and American English usages, and also strike a balance between British English and American English.

### 3.1.3 TOEFL 11

TOEFL 11, developed by Educational Testing Service (ETS) (Blanchard, Tetreault, Higgins, Cahill, & Chodorow, 2013), contains 12,100 essays produced by EFL learners of eleven L1s, including Arabic, Chinese, French, German, Hindi, Italian, Japanese, Korean, Spanish, Telugu, and Turkish in 2006—2007. The *TOEFL iBT*® test measures learners'

ability of using and understanding English at the university level, and evaluates learners skills of listening, speaking, reading and writing to perform academic tasks. Universities where English is the language of instruction usually resort to the outcome of students' TOEFL tests as one of the standards for entrance exams. The test, administered via the Internet, includes the sections of listening, speaking, reading and writing, and lasts for 4 hours to complete the test.

The TOEFL test consists of two writing tasks, independent and integrated tasks, in the writing section. The independent task requires test-takers to respond to a brief writing topic to write an essay while the integrated task asks them to first read a short passage presenting one perspective, and then listen to a lecture that presents another perspective. Test-takers need to summarize and synthesize the perspectives that they have read and listened in an essay. TOEFL 11 is composed of essays from the independent task. Since the goal of TOEFL 11 is to create a corpus with a great number of essays per L1, which were evenly distributed across languages and prompts, only the L1s that have essays above 1,100 essays would be collected. In the end, only eleven languages mentioned above meet the criterion. In addition, only essays that the authors provided permission for research would be included. The number of essays of each L1 from each topic is shown in Figure 3.1.

Figure 3.1: Number of Essays per Language per Prompt from Blanchard et al. (2013)

| Language | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 |
|---|---|---|---|---|---|---|---|---|
| Arabic | 138 | 137 | 138 | 139 | 136 | 133 | 138 | 141 |
| Chinese | 140 | 141 | 126 | 140 | 134 | 141 | 139 | 139 |
| French | 158 | 160 | 87 | 156 | 160 | 68 | 151 | 160 |
| German | 155 | 154 | 157 | 151 | 150 | 28 | 152 | 153 |
| Hindi | 161 | 162 | 163 | 86 | 156 | 53 | 158 | 161 |
| Italian | 173 | 89 | 138 | 187 | 187 | 12 | 173 | 141 |
| Japanese | 116 | 142 | 140 | 138 | 138 | 142 | 141 | 143 |
| Korean | 140 | 133 | 136 | 128 | 137 | 142 | 141 | 143 |
| Spanish | 141 | 133 | 54 | 159 | 134 | 157 | 160 | 162 |
| Telugu | 165 | 166 | 167 | 55 | 169 | 41 | 166 | 171 |
| Turkish | 169 | 145 | 90 | 170 | 147 | 43 | 167 | 169 |
| Total | 1,656 | 1,562 | 1,396 | 1,509 | 16,48 | 960 | 1,686 | 1,683 |

The corpus contains both high-scoring (usually long essays) and low-scoring essays (usually short essays) so the essay length of each essay varies a lot, ranging from 2 to 876 words. Most essays fall around the middle of the length distribution, which can be observed in Figure 3.2. The average of word tokens in TOEFL 11 is 348 per essay. The overall word tokens in TOEFL 11 are 3,846,185, shown in Table 3.1.

Figure 3.2**:** Histogram of essay lengths for all essays (Blanchard et al., 2013).

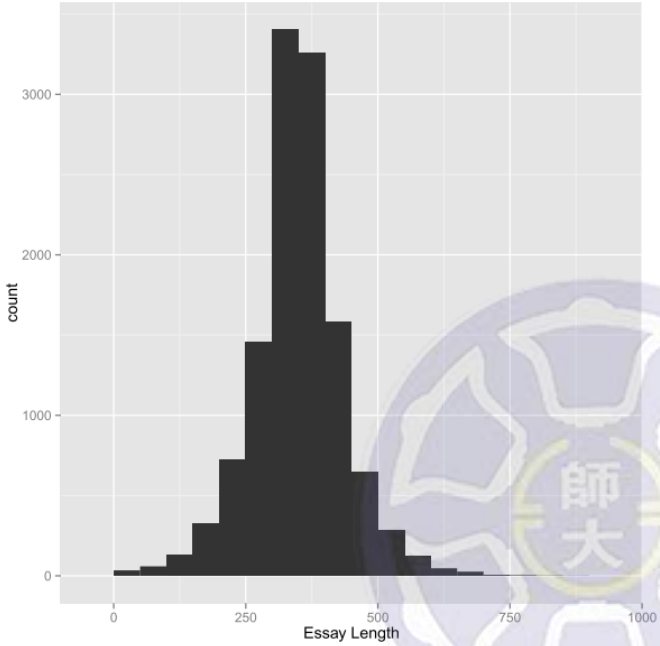
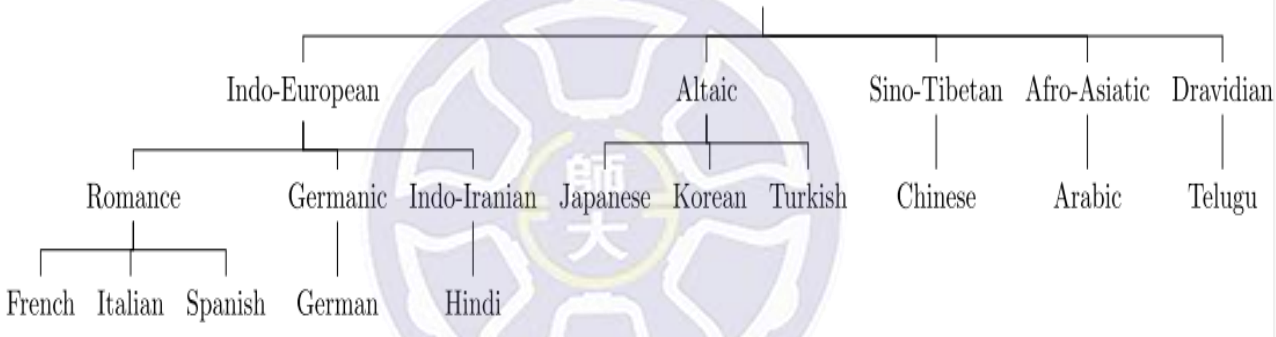



Table 3.1 Word Tokens in Each L1 and Word Tokens in Total.

| **Language** | **Word Tokens** |
|---|---|
| Arabic | 314,232 |
| Chinese | 358,252 |
| French | 353,945 |
| German | 380,147 |
| Hindi | 391,142 |
| Italian | 326,088 |
| Japanese | 306,263 |
| Korean | 332,277 |
| Spanish | 363,087 |
| Telugu | 368,114 |
| Turkish | 352,638 |
| **Total** | 3,846,185 |

According to Blanchard et al. (2013), the eleven languages cover seven language families, including Romance (French, Italian, Spanish), Germanic (German), Indo-Iranian (Hindi), Altaic (Japanese, Korean, Turkish), Sino-Tibetan (Chinese), Afro-Asiatic (Arabic), and Dravidian (Telugu). It should be noted that the Romance, Germanic and Indo-Iranian languages all belong to the larger Indo-European group of languages. On the other hand, Altaic, Sino-Tibetan, Afro-Asiatic, and Dravidian locate at the top level of language family taxonomy as the Indo-European group; thus, they share less common language features with each other. The language family taxonomy can be revealed in Figure 2.

Figure 3.3: Language families in the corpus (Blanchard et al., 2013)




## 3.2 The Query System—*The Sketch Engine*

The Sketch Engine (SKE), also called as Word Sketch Engine, is a corpus query system, which allows users to observe concordances, word sketches and thesaurally similar words, and more innovatively, sketch differences and create a corpus. Except for the basic function of observing concordances, word sketches, as the main function of the SKE, can provide a keyword's list of collocates comprising various grammatical relations (Kilgarriff et al., 2004).

The word sketches were firstly conducted in the production of Macmillan English Dictionary (Rundell & Fox, 2002) and presented at Euralex 2002 (Kilgarriff & Rundell, 2003). The development of word sketch function was traced back to the fact that the traditional interface of concordancer, providing the reading lines of co-occurrences, could not satisfy the

need of researchers, instructors and learners. As a result, providing the more systematic arrangement of concordances, word sketches were developed to offer various grammatical contexts through a State-of-the-art Corpus Query System (CQS), *Manatee*. According to Kilgarriff et al. (2004), word sketches can provide 27 grammatical relations in English so far. Take the noun "contribution" for instance, the modifiers, the verbs before it, prepositions and prepositional objects, the conjoined nouns (e.g. tax and contribution), etc. are all presented in one page of different lists in word sketches, which can be illustrated in Figure 3.4.

Figure 3.4: An example of the word sketch function on the Sketch Engine

**contribution** *(noun)*
British National Corpus freq = 8,129 (72.5 per million)

| modifier | 4,600 | 1.5 |
|---|---|---|
| **insurance** | 293 | 9.31 |
| Ni | 71 | 8.8 |
| **significant** | 200 | 8.58 |
| valuable | 88 | 8.4 |
| substantial | 74 | 7.74 |
| voluntary | 56 | 7.67 |
| **major** | 192 | 7.62 |
| positive | 77 | 7.56 |
| **national** | 261 | 7.48 |
| outstanding | 40 | 7.37 |
| **important** | 202 | 7.21 |
| relative | 38 | 7.16 |
| pension | 61 | 7.07 |
| greatest | 38 | 7.04 |
| distinctive | 24 | 6.78 |
| parental | 20 | 6.77 |
| useful | 47 | 6.7 |
| additional | 39 | 6.66 |
| reduced | 16 | 6.41 |
| individual | 46 | 6.33 |
| unique | 23 | 6.33 |
| financial | 54 | 6.2 |
| notable | 14 | 6.19 |
| vital | 22 | 6.13 |
| largest | 21 | 6.13 |

| object_of | 3,020 | 2.8 |
|---|---|---|
| **pay** | 225 | 7.1 |
| **make** | 1,468 | 7.02 |
| assess | 31 | 6.26 |
| acknowledge | 20 | 6.18 |
| evaluate | 12 | 5.87 |
| value | 11 | 5.85 |
| levy | 7 | 5.75 |
| recognise | 26 | 5.67 |
| underestimate | 6 | 5.53 |
| maximize | 5 | 5.49 |
| increase | 42 | 5.33 |
| calculate | 12 | 5.29 |
| feature | 10 | 5.12 |
| reduce | 31 | 4.92 |
| double | 6 | 4.9 |
| highlight | 7 | 4.83 |
| collect | 13 | 4.82 |
| abolish | 5 | 4.76 |
| identify | 19 | 4.67 |
| invest | 6 | 4.66 |
| welcome | 8 | 4.55 |
| afford | 7 | 4.51 |
| ignore | 8 | 4.46 |
| exclude | 7 | 4.44 |
| appreciate | 6 | 4.44 |

| pp_to-p | 2,255 | 14.9 |
|---|---|---|
| debate | 76 | 7.27 |
| warming | 7 | 6.11 |
| understanding | 32 | 5.99 |
| budget | 32 | 5.88 |
| fund | 45 | 5.81 |
| exchequer | 6 | 5.81 |
| knowledge | 40 | 5.54 |
| well-being | 5 | 5.5 |
| literature | 15 | 5.46 |
| critique | 5 | 5.33 |
| theory | 36 | 5.19 |
| economy | 26 | 5.13 |
| welfare | 10 | 5.1 |
| conservation | 6 | 5.07 |
| physics | 5 | 5.02 |
| stability | 6 | 5.0 |
| success | 22 | 4.8 |
| scheme | 32 | 4.78 |
| maintenance | 7 | 4.76 |
| science | 18 | 4.75 |
| journal | 7 | 4.72 |
| philosophy | 7 | 4.66 |
| safety | 11 | 4.63 |
| development | 52 | 4.63 |
| cause | 14 | 4.61 |

| and/or | 690 | 0.5 |
|---|---|---|
| indemnity | 7 | 7.03 |
| levy | 7 | 6.87 |
| tax | 48 | 5.63 |
| achievement | 6 | 4.42 |
| benefit | 17 | 4.36 |
| payment | 10 | 4.15 |
| fee | 6 | 4.08 |
| potential | 5 | 4.05 |
| charge | 11 | 3.79 |
| role | 5 | 1.96 |
| cent | 7 | 1.47 |
| work | 5 | 0.4 |

| possessor | 676 | 6.0 |
|---|---|---|
| ex-husband | 5 | 7.65 |
| employer | 35 | 6.58 |
| Keynes | 5 | 6.02 |
| employee | 29 | 5.96 |
| Gloucester | 6 | 5.47 |
| husband | 18 | 4.78 |
| Britain | 16 | 4.08 |
| partner | 5 | 3.26 |
| parent | 6 | 2.62 |
| wife | 6 | 2.53 |
| woman | 13 | 2.32 |
| worker | 5 | 2.17 |
| member | 10 | 1.94 |
| year | 16 | 1.7 |
| company | 9 | 1.52 |
| country | 6 | 1.41 |
| government | 6 | 0.93 |
| child | 5 | 0.78 |

Additionally, the word sketches are completely combined with the concordancing, that is, by clicking on a collocate of a word in the word sketch, the user will be taken to the concordances of the corpus evidence, which leads to that collocate in the grammatical relation. For instance, as revealed in Figure 3.4, if the user clicks on the modifier "insurance" of the word "contribution", then the concordances of "insurance contribution" will appear.

With well-developed part of speech (POS) tagging, *The Sketch Engine*, providing various functions of word sketches, thesaurus searching, sketch differences and creating a corpus, save a lot of efforts and time for humans by listing a variety of grammatical relations of a keyword, and accelerate the process of searching for collocations as well. Two special functions in the SKE, *Corpora Creating* and *Sketch Diff*, are fully utilized in the present study to extract verb noun collocations in different EFL learner groups, which will be elaborated in the following parts. Most importantly, the source tracing function in *Sketch Diff* will also be introduced.

### 3.2.1 Corpus Creating Function

The researcher can create the corpora through the function of *Corpus creating.* Once the corpus files are arranged and uploaded to the platform, the researcher can utilize the functions of the SKE, including *corpus querying*, *wordlist compiling*, *word sketches*, *thesaurus* and *sketch-diff* in her own corpora.

Therefore, the corpus of TOEFL 11 and the native speaker corpus, BNCCOCA, will be uploaded to the platform for the researcher to further make use of the function of the SKE.

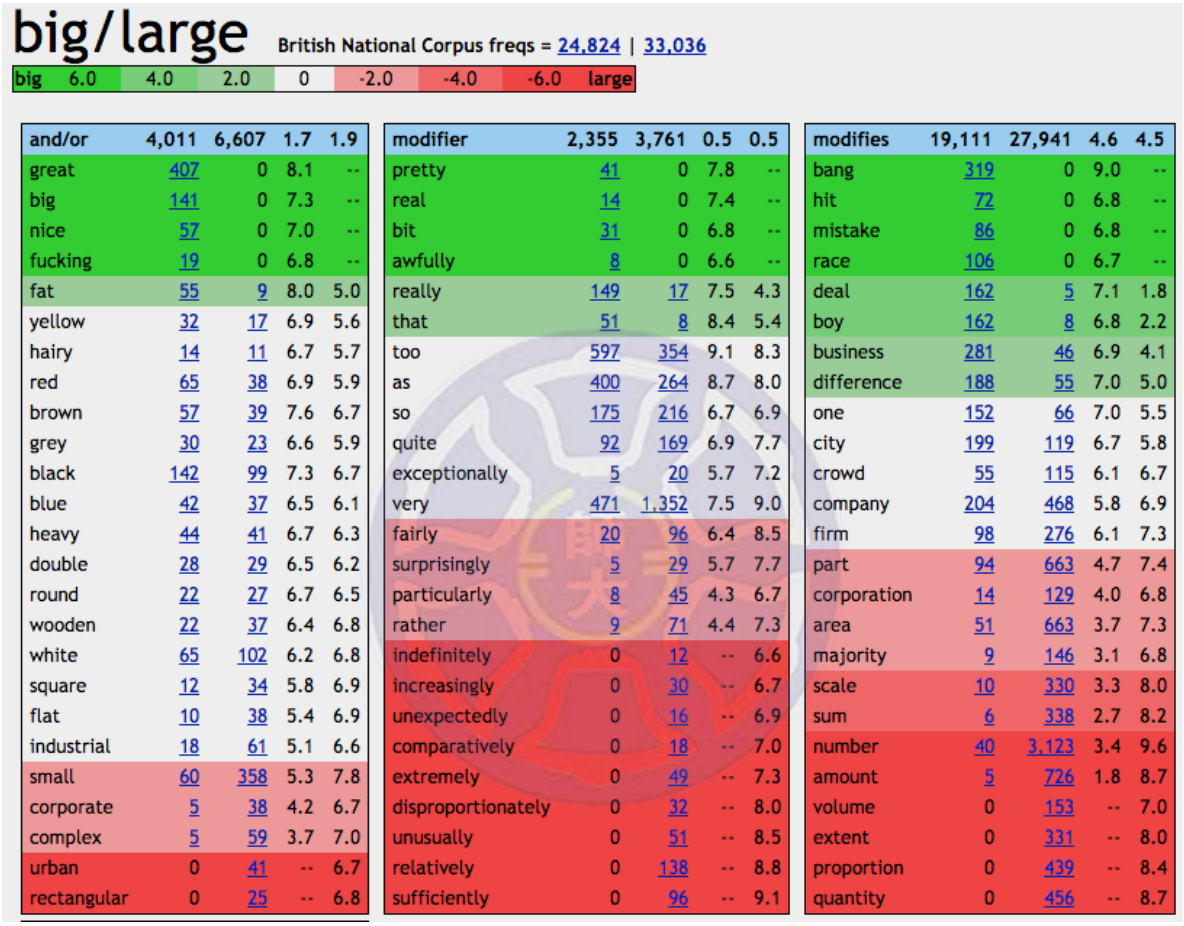
### 3.2.2 Sketch Diff Function

The *Sketch Diff* function, developed by Kilgarriff et al. (2004), can enable users to find out the collocation differences between two synonyms. Generally, when students confront two synonyms, they might have questions like, “what makes the words so similar?” and “how do they differ from each other?” In response to answering those questions, the function of *Sketch Diff* was created for learners to observe the different collocates of the two synonyms. For instance, the differences between the two synonyms *big* and *large* can be shown in Figure 3.5.

As shown in Figure 3.5, the green ones are collocates that can only be combined with *big* while the red ones are collocates that can only be connected with *large*. As for the white ones,

they are combined with both *big* and *large* evenly. Thus, take the modifies of the words for example, it can be observed that *big* can be followed by *hit*, *mistake* and *race* while *large* is usually followed by *population* and *quantity*, which can indicate the different usages of the two synonyms.

Figure 3.5: The *Sketch Diff* interface showing *big* and *large* in the BNC



**big/large** British National Corpus freqs = 24,824 | 33,036

big 6.0 4.0 2.0 0 -2.0 -4.0 -6.0 large

| and/or | 4,011 | 6,607 | 1.7 | 1.9 |
|---|---|---|---|---|
| great | 407 | 0 | 8.1 | -- |
| big | 141 | 0 | 7.3 | -- |
| nice | 57 | 0 | 7.0 | -- |
| fucking | 19 | 0 | 6.8 | -- |
| fat | 55 | 9 | 8.0 | 5.0 |
| yellow | 32 | 17 | 6.9 | 5.6 |
| hairy | 14 | 11 | 6.7 | 5.7 |
| red | 65 | 38 | 6.9 | 5.9 |
| brown | 57 | 39 | 7.6 | 6.7 |
| grey | 30 | 23 | 6.6 | 5.9 |
| black | 142 | 99 | 7.3 | 6.7 |
| blue | 42 | 37 | 6.5 | 6.1 |
| heavy | 44 | 41 | 6.7 | 6.3 |
| double | 28 | 29 | 6.5 | 6.2 |
| round | 22 | 27 | 6.7 | 6.5 |
| wooden | 22 | 37 | 6.4 | 6.8 |
| white | 65 | 102 | 6.2 | 6.8 |
| square | 12 | 34 | 5.8 | 6.9 |
| flat | 10 | 38 | 5.4 | 6.9 |
| industrial | 18 | 61 | 5.1 | 6.6 |
| small | 60 | 358 | 5.3 | 7.8 |
| corporate | 5 | 38 | 4.2 | 6.7 |
| complex | 5 | 59 | 3.7 | 7.0 |
| urban | 0 | 41 | -- | 6.7 |
| rectangular | 0 | 25 | -- | 6.8 |

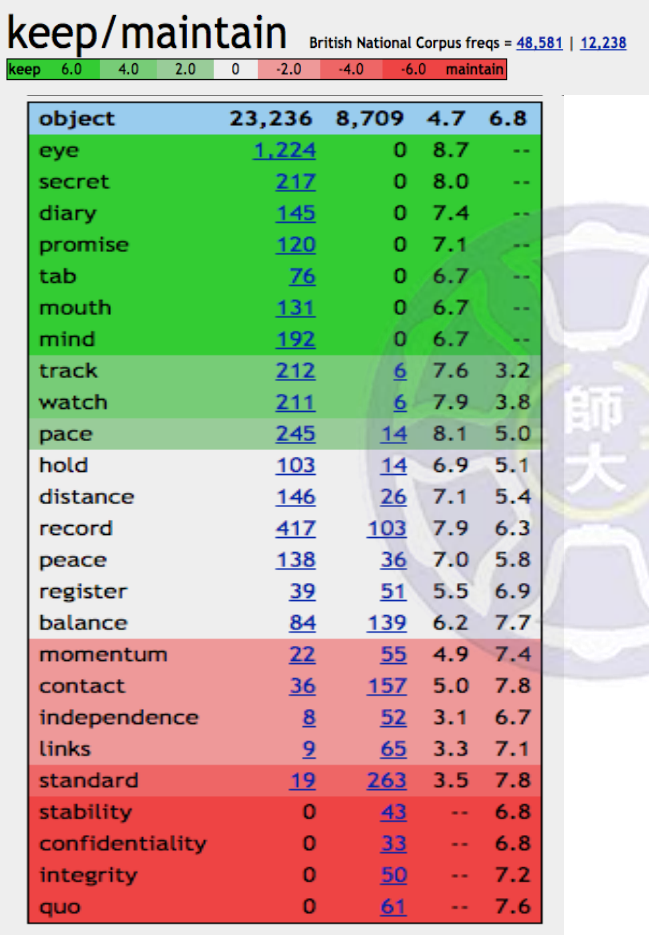
| modifier | 2,355 | 3,761 | 0.5 | 0.5 |
|---|---|---|---|---|
| pretty | 41 | 0 | 7.8 | -- |
| real | 14 | 0 | 7.4 | -- |
| bit | 31 | 0 | 6.8 | -- |
| awfully | 8 | 0 | 6.6 | -- |
| really | 149 | 17 | 7.5 | 4.3 |
| that | 51 | 8 | 8.4 | 5.4 |
| too | 597 | 354 | 9.1 | 8.3 |
| as | 400 | 264 | 8.7 | 8.0 |
| so | 175 | 216 | 6.7 | 6.9 |
| quite | 92 | 169 | 6.9 | 7.7 |
| exceptionally | 5 | 20 | 5.7 | 7.2 |
| very | 471 | 1,352 | 7.5 | 9.0 |
| fairly | 20 | 96 | 6.4 | 8.5 |
| surprisingly | 5 | 29 | 5.7 | 7.7 |
| particularly | 8 | 45 | 4.3 | 6.7 |
| rather | 9 | 71 | 4.4 | 7.3 |
| indefinitely | 0 | 12 | -- | 6.6 |
| increasingly | 0 | 30 | -- | 6.7 |
| unexpectedly | 0 | 16 | -- | 6.9 |
| comparatively | 0 | 18 | -- | 7.0 |
| extremely | 0 | 49 | -- | 7.3 |
| disproportionately | 0 | 32 | -- | 8.0 |
| unusually | 0 | 51 | -- | 8.5 |
| relatively | 0 | 138 | -- | 8.8 |
| sufficiently | 0 | 96 | -- | 9.1 |

| modifies | 19,111 | 27,941 | 4.6 | 4.5 |
|---|---|---|---|---|
| bang | 319 | 0 | 9.0 | -- |
| hit | 72 | 0 | 6.8 | -- |
| mistake | 86 | 0 | 6.8 | -- |
| race | 106 | 0 | 6.7 | -- |
| deal | 162 | 5 | 7.1 | 1.8 |
| boy | 162 | 8 | 6.8 | 2.2 |
| business | 281 | 46 | 6.9 | 4.1 |
| difference | 188 | 55 | 7.0 | 5.0 |
| one | 152 | 66 | 7.0 | 5.5 |
| city | 199 | 119 | 6.7 | 5.8 |
| crowd | 55 | 115 | 6.1 | 6.7 |
| company | 204 | 468 | 5.8 | 6.9 |
| firm | 98 | 276 | 6.1 | 7.3 |
| part | 94 | 663 | 4.7 | 7.4 |
| corporation | 14 | 129 | 4.0 | 6.8 |
| area | 51 | 663 | 3.7 | 7.3 |
| majority | 9 | 146 | 3.1 | 6.8 |
| scale | 10 | 330 | 3.3 | 8.0 |
| sum | 6 | 338 | 2.7 | 8.2 |
| number | 40 | 3,123 | 3.4 | 9.6 |
| amount | 5 | 726 | 1.8 | 8.7 |
| volume | 0 | 153 | -- | 7.0 |
| extent | 0 | 331 | -- | 8.0 |
| proportion | 0 | 439 | -- | 8.4 |
| quantity | 0 | 456 | -- | 8.7 |

As for verb noun collocations, take the verbs *keep* and *maintain* for instance, it can be observed that words accompanying *keep* and *maintain* as objects are clearly presented in Figure 3.6. The four numbers next to every collocate are the frequency of a collocate and salience scores with the two keywords, e.g. *keep* and *maintain*. A quick observation of two keywords' differences can be revealed through looking at the frequency, salience scores or colors. As revealed in Figure 3.6, *Eye* usually follows *keep*, with the frequency of 1224 times while 0 times for *maintain*. Similarly, *keep a diary* (the 3rd one) occurs for 145 times while

**maintain a diary* occurs for 0 times. *Maintain*, on the other hand, is usually followed by *integrity* (the 2nd one from the bottom), with the occurrence of 50 times while *keep* is never followed by *integrity* in the BNC. In other words, *keep an eye/ a secret/ a diary* are strong collocates while strong collocates for *maintain* are *maintain a status quo/ the integrity/ the confidentiality*. Thus, a quick observation for comparison and contrast for two words can be revealed in the function of Sketch-Diff.

Figure 3.6: The objects of *keep* and *maintain* as collocates in the BNC.




| object | 23,236 | 8,709 | 4.7 | 6.8 |
|---|---|---|---|---|
| eye | 1,224 | 0 | 8.7 | -- |
| secret | 217 | 0 | 8.0 | -- |
| diary | 145 | 0 | 7.4 | -- |
| promise | 120 | 0 | 7.1 | -- |
| tab | 76 | 0 | 6.7 | -- |
| mouth | 131 | 0 | 6.7 | -- |
| mind | 192 | 0 | 6.7 | -- |
| track | 212 | 6 | 7.6 | 3.2 |
| watch | 211 | 6 | 7.9 | 3.8 |
| pace | 245 | 14 | 8.1 | 5.0 |
| hold | 103 | 14 | 6.9 | 5.1 |
| distance | 146 | 26 | 7.1 | 5.4 |
| record | 417 | 103 | 7.9 | 6.3 |
| peace | 138 | 36 | 7.0 | 5.8 |
| register | 39 | 51 | 5.5 | 6.9 |
| balance | 84 | 139 | 6.2 | 7.7 |
| momentum | 22 | 55 | 4.9 | 7.4 |
| contact | 36 | 157 | 5.0 | 7.8 |
| independence | 8 | 52 | 3.1 | 6.7 |
| links | 9 | 65 | 3.3 | 7.1 |
| standard | 19 | 263 | 3.5 | 7.8 |
| stability | 0 | 43 | -- | 6.8 |
| confidentiality | 0 | 33 | -- | 6.8 |
| integrity | 0 | 50 | -- | 7.2 |
| quo | 0 | 61 | -- | 7.6 |

However, in the present study, the function of Sketch-Diff is not used for comparing two synonyms but used for comparing the same collocates in two corpora, e.g. the native speaker corpus and the learner corpus. Through this way, the differences of collocation usages in two corpora can be revealed. A collocation showing up in the learner corpus instead of the native speaker corpus would be considered as a potential miscollocation. Time and efforts can be

saved since the researcher does not need to read learners' written production or observe each concordance to extract collocations manually. Instead, a more comprehensive overview and systematic comparison of verb noun collocations can be provided with Sketch-Diff to avoid human misjudgments and time-consuming problems.

### 3.2.3 Source Tracing Function

In order to observe the source of the collocations, that is, learners from which L1s produce the collocations, the source tracing function is needed. As mentioned before, the word sketch is completely combined with the concordancing, so is Sketch-Diff. When the user clicks on the frequency number after collocates, the concordance of collocates will appear to offer more corpus evidence, as shown in Figure 3.7.

Figure 3.7: Part of concordance of *maintain the integrity* in the BNC

Word sketch item **50** (0.4 per million)

| | | | |
|---|---|---|---|
| HTE | one level physiologists study how cells | maintain | their ***integrity*** and respond to stimuli |
| HJ5 | transfer of files to the printer's system will | maintain | the ***integrity*** of the document and be cost |
| HJ1 | pressures from within the minority community for | maintaining | cultural ***integrity*** ; (2) continuing discrimination |
| HLG | opposition and the Iraqi Kurds had agreed to | maintain | the territorial ***integrity*** of Iraq.  |
| HLA | like Germany had hitherto strongly favoured | maintaining | Yugoslav ***integrity*** , made it clear on July |
| HHX | on that is clear from our amendment. We | maintain | the ***integrity*** of our nation as we promote |
| HHX | trust that we as a nation will continue to | maintain | our ***integrity*** and influence.  Madam Deputy |
| HU2 | , the luminal route of action of EGF for | maintaining | oesophageal mucosal ***integrity*** seems unlikely |
| HU3 | averages 84% (range 74-93%). The leukocytes | maintain | their ***integrity*** and function during this |
| HU4 | averaged 86% (range 79%-95%). The leukocytes | maintain | their ***integrity*** and function during the |
| HWF | the LIFESPAN manager.  In order to | maintain | the ***integrity*** of the offlined files, each |
| HWF | any movement of software its ***integrity*** is | maintained | intact by performing a Cyclic Redundancy |
| HWW | medical practice, to operate successfully, to | maintain | its financial ***integrity*** , to attract capital |
| HY7 | subsequently Prussia) had agreed not only to | maintain | the ***integrity*** of the Sultan's possessions |
| H91 | is not on the side of those who wish to | maintain | the ***integrity*** of the nation state. Left |
| HEP | Unix and ensure that our application will | maintain | the ***integrity*** of data within the system |

As can be seen, before each concordance, there is a file name before it. If the user clicks on the file name, the source of the concordance will appear to offer more information about the concordance. For example, if the user clicks on the file name *HTE* of the first concordance, then the source of it, such as the author, publish year and titles will pop up in the bottom of the interface, which is revealed in Figure 3.8.

Figure 3.8: The function of tracing source of the concordance


Therefore, in the present study, the function of tracing source would be fully utilized in order to find the original source of the concordance, that is, in order to find the learner of which L1 produce that collocation. In this way, once the collocation errors are found, we can also trace each collocation error's source so as to observe what universal errors are shared by at least six EFL learner groups, and what collocation errors are only made by learners of certain L1s. The more detailed procedure of extracting miscollocations and tracing the sources will be elaborated in the following part.

## 3.3 Data Collection

TOEFL 11 is attained from the Linguistic Data Consortium (LDC), which is an open consortium of libraries, universities, corporations and government research laboratories. LDC, hosted by the University of Pennsylvania, is a center within the University's School of Arts and Sciences. It is formed to deal with the critical data shortage faced by language technology research and development. It originally served as a repository and distribution point for

language resources and later is developed to create and allocate a wide range of language resources. In addition, it provides resources and organizational expertise to assist sponsored research programs and language-based technology evaluations. The data of TOEFL 11 developed by ETS can be obtained through the LDC website (https://www.ldc.upenn.edu/).

## 3.4 Data Analysis

To efficiently extract verb noun collocations, the corpus creating and *Sketch Diff* functions will be completely utilized. In order to understand the source of a concordance, 12,100 text files in TOEFL 11 will be firstly divided into 11 files based on the L1 languages and each file is named by the country codes according to the source of its country. The country codes are shown in Table 3.1. The 11 files will be zipped into one zip file and uploaded onto the Sketch Engine, named as TOEFL 11. On the other hand, the native speaker corpora, the BNC written corpus and academic texts of COCA, will be combined together as one large native speaker corpus and uploaded to the platform as well, named as BNCCOCA.

Table 3.2 The language codes of 11 languages

| Language | Arabic | Chinese | French | German | Hindi | Italian |
|---|---|---|---|---|---|---|
| Language codes | ARA | ZHO | FRA | DEU | HIN | ITA |
| Language | Japanese | Korean | Spanish | Telugu | Turkish | |
| Language codes | JPA | KOR | SPA | TEL | TUR | |

Then, as for the target items searched in the interface, the researcher will utilize the function of compiling wordlist in the Sketch Engine to create word lists of both nouns and verbs, which are ranked by the frequency and retrieved from the learner corpus. Both nouns and verbs in the wordlist will be typed in the interface to search for miscollocations. The production of word lists from learner corpora to retrieve collocations was also adopted in the

studies of Laufer (2011) and Liu (2013) although both studies only generated a word list based on frequent nouns since nouns express the semantic process of the predicates and serve as crucial indicators of verb noun collocations (C. Gledhill, 2007; Juknevičienė, 2008; Liu, 2002). However, in order to ensure the full coverage of verb noun collocations, the present study will produce the wordlists of both verbs and nouns and consider frequent verbs and nouns as the target items that are tested with the *Sketch Diff*.

The threshold of frequency count is 200 in the present research. Both verbs and nouns that occur above 200 times in the learner corpus will be the keywords that are typed in the interface of the *Sketch Diff*. After the researcher types each noun and verb from the word lists in the interface of *Sketch Diff*, she will observe the column of *object_of* for nouns as the target items and *object* for verbs as the target items, meaning that the noun is the *object of* the verb or the verb's *object* is the noun, to search for full coverage of verb noun collocations, as shown in Figure 3.9. Take *time* for example, *time* is the object of *allow*, comprising *allow_ time* as presented in the first line of the following figure. It should be noted that some miscollocations will be searched repeatedly if the researcher searches for both nouns and verbs. For instance, if the researcher searches for both *do* and *mistake*, the wrong collocate *do_ mistake*, which occur 26 times, will be searched twice. Thus, the researcher will check the miscollocations and delete the repeated ones.

Figure 3.9 The columns of *object of* and *object* from the noun "time" and the verb "do."

time (noun) Sketch diff for subcorpora BNCCOCA/TOEFL11

BNCCOCA 6.0 4.0 2.0 0 -2.0 -4.0 -6.0 TOEFL11

| object_of | 63,114 | 8,081 | 12.60 | 11.50 |
|---|---|---|---|---|
| allow | 617 | 0 | 7.2 | -- |
| allot | 138 | 0 | 6.1 | -- |
| elapse | 137 | 0 | 6.1 | -- |
| bide | 126 | 0 | 6.0 | -- |
| mark | 154 | 0 | 5.9 | -- |
| measure | 174 | 0 | 5.9 | -- |
| record | 138 | 0 | 5.6 | -- |
| buy | 149 | 0 | 5.6 | -- |
| provide | 291 | 17 | 5.1 | 1.4 |
| happen | 174 | 6 | 5.8 | 2.0 |
| see | 175 | 11 | 4.0 | 0.2 |
| go | 233 | 15 | 4.3 | 0.5 |
| reduce | 424 | 18 | 6.8 | 3.1 |
| do | 444 | 33 | 4.0 | 0.3 |
| take | 4,253 | 296 | 8.4 | 4.7 |
| spare | 112 | 35 | 5.8 | 6.8 |
| commute | 10 | 3 | 2.4 | 3.5 |
| sleep | 7 | 4 | 1.6 | 2.8 |
| sacrifice | 17 | 7 | 3.1 | 4.4 |
| order | 7 | 5 | 1.6 | 2.9 |
| relax | 9 | 6 | 2.1 | 3.8 |
| dedicate | 43 | 35 | 4.3 | 6.3 |
| enjoy | 84 | 162 | 4.7 | 6.7 |
| utilise | 5 | 3 | 1.3 | 3.4 |
| spended | 0 | 3 | -- | 3.6 |
| wast | 0 | 7 | -- | 4.8 |
| loose | 0 | 9 | -- | 5.1 |

do (verb) Sketch diff for subcorpora BNCCOCA/TOEFL11

BNCCOCA 6.0 4.0 2.0 0 -2.0 -4.0 -6.0 TOEFL11

| object | 142,013 | 7,536 | 12.40 | 10.40 |
|---|---|---|---|---|
| show | 193 | 0 | 5.2 | -- |
| battle | 176 | 0 | 5.2 | -- |
| man | 427 | 0 | 5.1 | -- |
| share | 198 | 0 | 5.1 | -- |
| Yeah | 164 | 0 | 5.1 | -- |
| look | 179 | 0 | 5.0 | -- |
| ya | 144 | 0 | 5.0 | -- |
| hair | 177 | 0 | 5.0 | -- |
| woman | 357 | 0 | 5.0 | -- |
| round | 177 | 0 | 5.0 | -- |
| right | 263 | 0 | 5.0 | -- |
| interview | 163 | 0 | 4.9 | -- |
| counterpart | 145 | 0 | 4.9 | -- |
| week | 201 | 0 | 4.9 | -- |
| child | 383 | 0 | 4.9 | -- |
| compare | 35 | 4 | 3.0 | 4.0 |
| travel | 29 | 13 | 2.6 | 4.1 |
| master | 20 | 10 | 2.0 | 3.8 |
| favor | 11 | 4 | 1.3 | 3.2 |
| stunt | 14 | 5 | 1.7 | 4.3 |
| advertising | 14 | 10 | 1.6 | 4.2 |
| reseach | 0 | 7 | -- | 4.9 |
| reaserch | 0 | 7 | -- | 4.9 |
| somthing | 0 | 8 | -- | 5.1 |
| mistake | 0 | 26 | -- | 5.4 |
| somethings | 0 | 10 | -- | 5.4 |
| someting | 0 | 12 | -- | 5.7 |
| specialization | 0 | 20 | -- | 5.9 |

As indicated above, the frequency of collocations in two corpora will be viewed to find out the potential miscollocations. Verb noun collocations, which occur at least 3 times in TOEFL 11 while do not occur in the native speaker corpus, will be considered as possible collocation errors. Since the corpus of TOEFL 11 is not large enough, the present study sets the low criterion of occurrence, which is three times. In addition, previous research, such as the study of Liu (2013), also set the standard of three-time occurrence in learner production and none in native corpora as the miscollocations analyzed in the research. The present study thus adopts the same standard of miscollocation definition.

The repeated searched miscollocations will be deleted as mentioned. The researcher will later clicks on those collocates to see the concordances, and clicks on the filename before the

concordance in order to check the original source of the concordance, which is the country code, as illustrated in Figure 3.10. Then, the potential collocation errors will be double-checked by a native speaker to delete the acceptable collocations. In the end, both verb noun miscollocations and their sources of L1s will be recorded. The whole data extraction process is summarized in Figure 3.11.

Figure 3.10 The source tracing function (After a user clicked on *file2263628* before the concordance, the country code, *ARA*, in the yellow note will pop up).

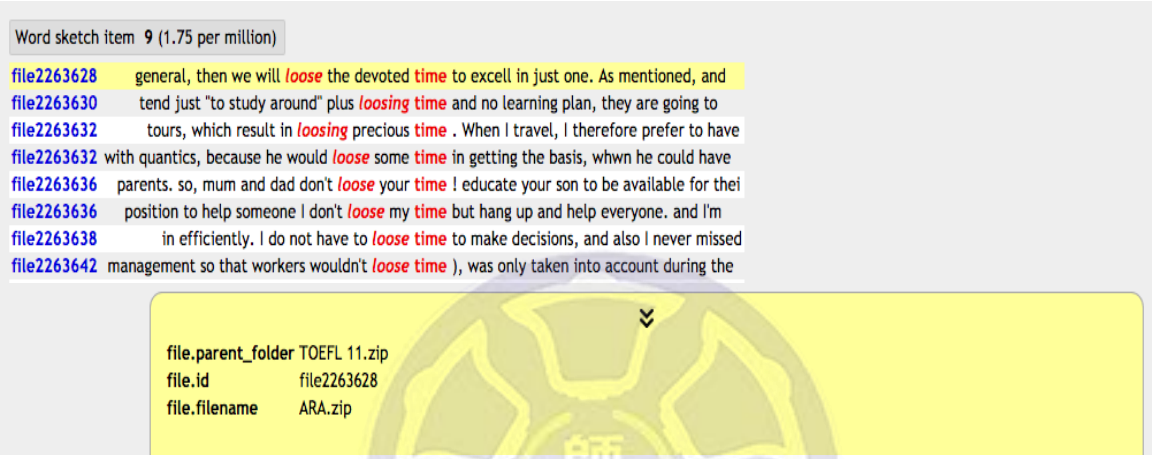
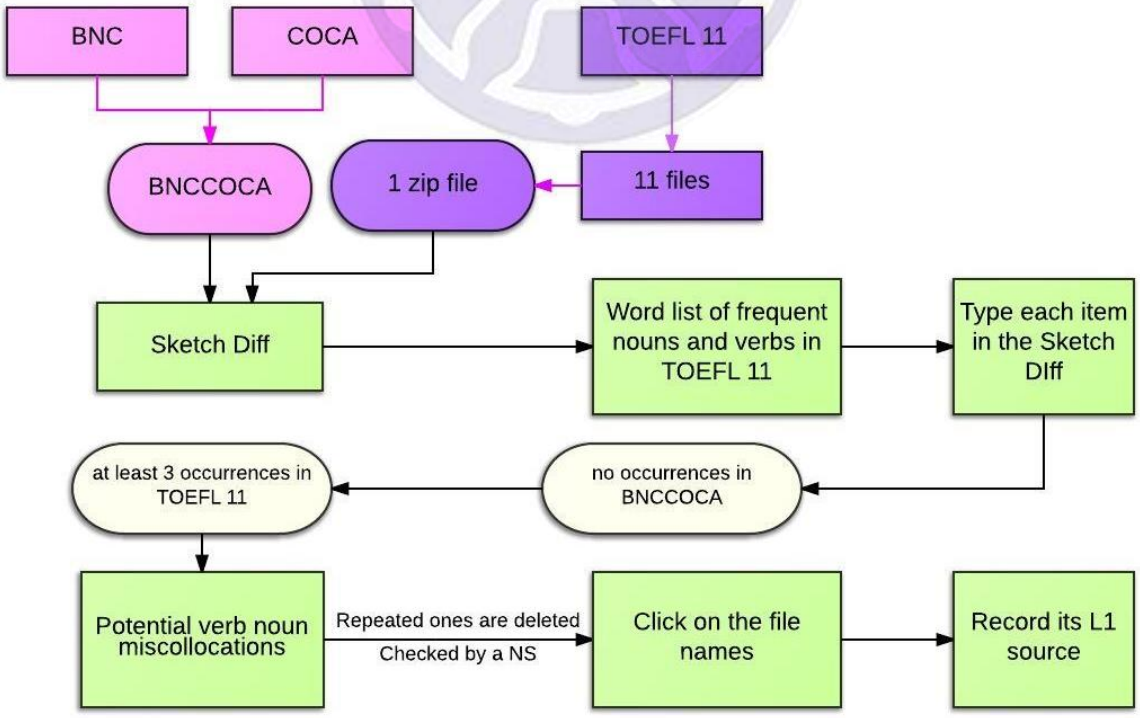



Figure 3.11 The procedure of data extraction.

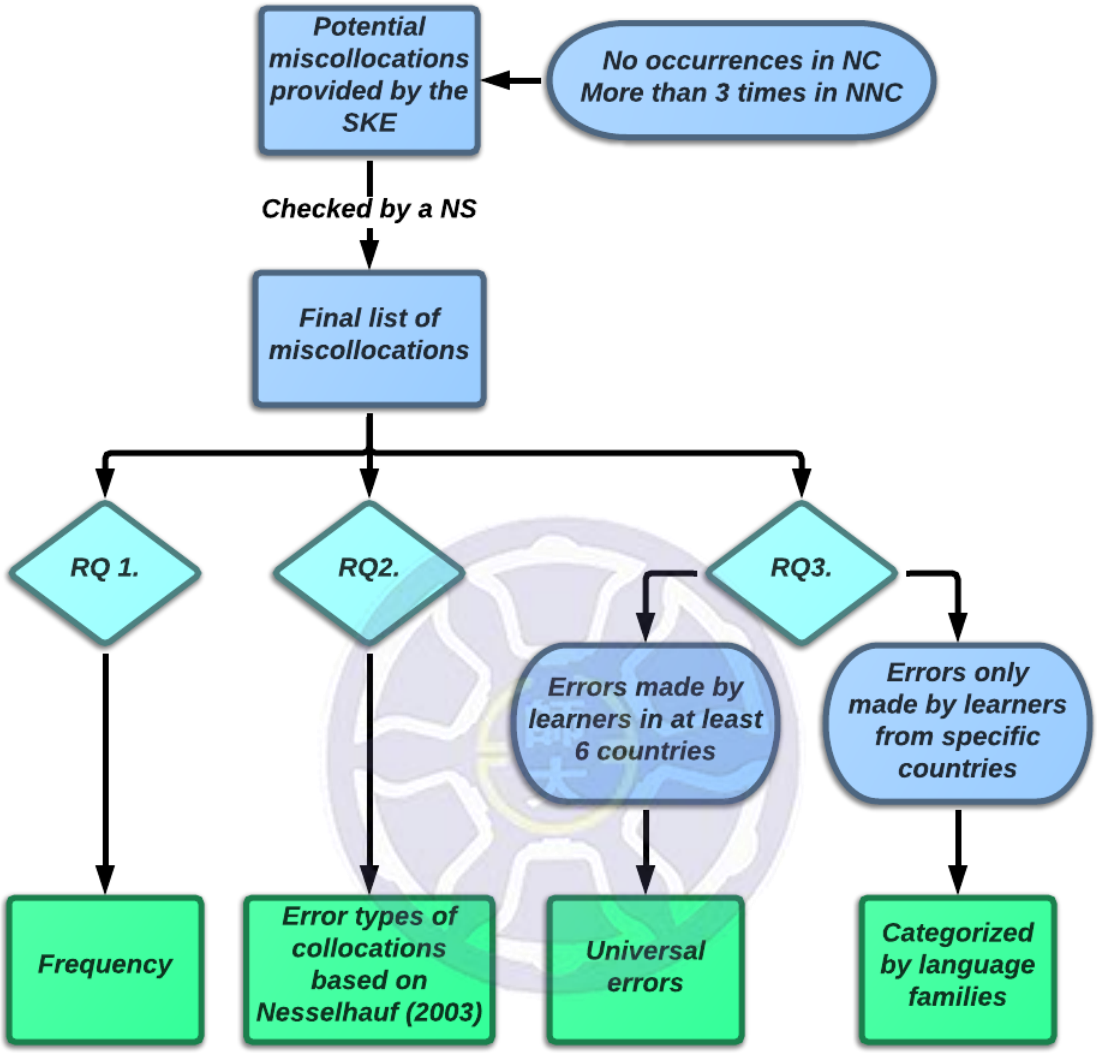


To sum up, in order to answer the three research questions, the researcher will analyze the data in the following steps. First, verb noun collocations will be considered as possible errors when they occur in the learner corpus instead of the native speaker corpus. Additionally, only the collocations that occur more than three times in the learner corpus will remain in the finding. The collocations that occur less than three times will be ignored in the result. The final list of miscollocations will be checked by a native speaker to delete the acceptable collocations.

Second, the collocation errors will be presented based on their frequency to observe the frequent collocation errors made by EFL learners of eleven L1s in response to the first research question. Third, in terms of the types of miscollocations, the present study will categorize verb noun miscollocations into different types, such as deviant verbs, deviant nouns, deviant preposition usages and others , based on the error types summarized by Nesselhauf (2003) in the section 2.2.2.

Fourth, in order to answer the third research question, which is the main purpose of the present study, the collocation errors, which occur in six out of eleven L1 groups, will be considered as universal errors that shared by most EFL learners. These errors are universal and intralingual errors that are difficult for most EFL learners, no matter what L1 language they use. As for errors only made by learners of specific L1s, the researcher will attempt to categorize those collocation errors based on language families and draw a general conclusion. Since the possible causes of these errors are difficult to explain, the present study will try to offer some insight in the possible causes of these errors but will not explicitly conclude the causes of these errors.

Figure 3.12 The data analysis in the present study



# CHAPTER 4

# RESULTS AND DISCUSSION

## 4.1 Frequency

After the researcher deleted the collocation errors searched repeatedly and a native speaker deleted the acceptable collocations, 109 types of verb noun collocation errors, which occurred at least three times in TOEFL 11 and none in BNCCOCA, were found in TOEFL 11, with 807 occurrences in total. The types of collocation errors made by 11 groups of EFL learners were categorized into four main groups, including deviant prepositions of verbs, deviant verbs, deviant nouns and others by consulting the error type categorizations from Nesselhauf (2003), as shown in Table 4.1. The complete concordances of these miscollocations are presented in Appendix 1.

Table 4.1 The collocation errors in TOEFL 11 (ranked by frequency)

| **Types** | | **Patterns of collocation errors (occurrences)** | **The correct forms** | **Frequency of the categories** | **Total frequency** |
|---|---|---|---|---|---|
| **(1)Deviant prepositions of verbs** | Missing prepositions | *agree ideas (40)<br>*agree the opinion (31)<br>*disagree the idea (14)<br>*agree the statement (13)<br>*disagree the opinion (9)<br>*agree the topic (6)<br>*agree sentences (4)<br>*disagree the topic (3) | agree with | 120 | 395 |

| | | | | | |
|---|---|---|---|---|---|
| | | *travel places (43)<br>*travel Japan (4)<br>*travel a park (3)<br>*travel lands (3) | travel to | 53 | |
| | | *specialize subjects (46)<br>*specialize themselves (3) | specialize in | 49 | |
| | | *listen music (20)<br>*listen a lecture (5)<br>*listen their story (3) | listen to | 28 | |
| | | *participate activities (17)<br>*participate events (3) | participate in | 20 | |
| | | *go vacation (4)<br>*go Egypt (3)<br>*going a picnic (3)<br>*go concerts (2)<br>*go shopping malls (2)<br>*go the Tower (2)<br>*go the workplace (2) | go to | 18 | |
| | | *stay the place (12)<br>*stay a hotel (4) | stay at | 16 | |
| | | *look the products (8)<br>*look the advertisement (4)<br>*look the surface (3) | look at | 15 | |
| | | *search the information (8)<br>*search a job (5) | search for | 13 | |
| | | *graduate university | graduate from | 9 | |
| | | *concentrate subjects | concentrate on | 8 | |
| | | *prepare the travel<br>*prepare an advertisement | prepare for | 8 | |
| | | *arrive the place | arrive at | 7 | |
| | | *care their community | care about | 7 | |

| | | | | | |
|---|---|---|---|---|---|
| | | *contribute their communities | contribute to | 6 | |
| | | *adapt the society | adapt to | 4 | |
| | | *contract a guide | contract with | 4 | |
| | | *major subjects | major in | 4 | |
| | | *succeed the business | succeed in | 3 | |
| | | *learn hardware | learn about | 3 | |
| **(2)Deviant verbs** | similar forms | *loose money (14)<br>*loose time (10)<br>*loose everything (7)<br>*loose the chance (5)<br>*loose clients (4)<br>*loose customers (4)<br>*loose something (4)<br>*loose your interest (4)<br>*loose anything (3)<br>*loose importance (3)<br>*loose patience (3)<br>*loose reputation (3)<br>*loose sense (3)<br>*loose the ability (3)<br>*loose weight (3) | lose | 73 | 295 |
| | | *experiment things (25)<br>*experiment a culture (3) | experience | 28 | |
| | delexical verbs | *do specialization (31)<br>*do mistakes (26)<br>*do graduation (8)<br>*do adventure (6)<br>*make internships (6)<br>*take a decision (5)<br>*do enjoyment (4) | have specialization<br>make mistakes<br>have graduation<br>experience adventure<br>have internships<br>make a decision<br>experience enjoyment | 116 | |

| | | | | | |
|---|---|---|---|---|---|
| | simple verbs | *try risks (17)<br>*try challenges (5)<br>*try ventures (4) | take | 26 | |
| | | *retire their jobs | quit | 15 | |
| | | *use celebrities | hire celebrities | 6 | |
| | | *face accidents | have | 5 | |
| | | *study majors | take | 4 | |
| | | *underline opinion | defend/underpin opinion | 4 | |
| | | *eat medicine | take | 3 | |
| | | *invent the energy | find | 3 | |
| | | *leave my routine | change | 3 | |
| | | *meet accidents | have | 3 | |
| | | *realize chances | take | 3 | |
| | | *see hardships | experience/ encounter | 3 | |
| | | *sharp your skills | sharpen | 3 | |
| | | *stand testimony | bear | 3 | |
| | | *travel a tour | go on | 2 | |
| | synonymous verbs | *reach the dream (7)<br>*reach fame (4)<br>*reach happiness (3) | achieve | 14 | |
| | | *earn knowledge (5)<br>*earn the facts (3) | acquire knowledge<br>learn the facts | 8 | |
| **(3)Deviant nouns** | | *take a travel | trip | 15 | 73 |
| | | *earn their life | money | 11 | |
| | | *increase selling (5)<br>*increase the sell (5) | sales | 10 | |
| | | *try thinks (3)<br>*know the thinks (3)<br>*learn thinks (3) | thoughts | 9 | |
| | | *watch advertising (4) | advertisement | 7 | |

| | | | | |
|---|---|---|---|---|
| | *create advertising (3) | | | |
| | *help neighborhood | neighbors | 5 | |
| | *manipulate customers | customers' desire | 5 | |
| | *reach the destiny | destination | 5 | |
| | *outnumber the decrease | outnumber the (number of cars…) | 4 | |
| | *learn the intuition | fact | 2 | |
| **(4)Others** | *have a travel | take a trip | 30 | 44 |
| | *conclude ideas | make conclusions | 7 | |
| | *conclude the fact | make conclusions | 4 | |
| | *ride a ski | go skiing | 3 | |
| **Total frequency** | | | | 807 |

Among the four types, the most and second frequent error types are the deviant preposition of verbs and deviant verbs. In terms of deviant prepositions of verbs, learners tend to neglect prepositions when they use certain verbs, such as *agree, travel, specialize,* etc. In this type, the most frequent pattern is the incorrect usage of "agree," comprising 8 miscollocation types out of 109 types, such as **agree ideas*, **agree the opinion* and **disagree the topic*, with 120 occurrences in TOEFL 11. The suggested form should be *agree with,* like *agree with ideas* and *agree with the opinion,* which reveals that learners seemed to ignore the preposition "with" when they used *agree* in their writing.

The second and third frequent patterns in the deviant prepositions of verbs are *travel* and *specialize.* Four examples of collocations containing the incorrect usage of *travel* were found, with 53 occurrences in TOEFL 11. Learners seemed to connect the verb *travel* directly with any places, such as **travel places*, **travel Japan* and **travel a park.* The preposition *to* was neglected by those learners, as in *travel to places, Japan* and *a park.* The similar prepositional errors were found in the usage of *specialize,* such as **specialize subjects* and **specialize themselves*, which occur 49 times in the learner corpus. The correct usage should

be *specialize in subjects* and *specialize in* respectively. The following are the examples of *travel* and *specialize*:

[1] *We have to travel more places to get more information (ARA).*

[2] *Last winter, I traveled Japan alone, and I studied the Japan to plan my trip (KOR).*

[3] *I strongly believe that specializing one subject is more appropriate than knowing entire knowledge base (HIN).*

[4] *Other people prefer to specialize themselves in one specific subject (FRA).*

The second frequent type of collocation errors is the deviant verbs, in which the most frequent pattern is the the incorrect usages of "loose[1]," with 15 examples of collocation errors and 73 occurrences. *Loose* means *"to untie someone or something"* while *lose* refers to *"to stop having a particular attitude, quality, ability etc., or to gradually have less of it"* based on the definitions of Longman Dictionary of Contemporary English. *Loose* is to let go of something and *lose* is to stop having something, which indicates that the meaning of *loose* shares a certain degree of meaning with *lose*, giving rise to learners' misuse and confusion of these two words. Additionally, the forms of these words are similar, contributing to learners' misuse of collocations. When the forms of the words are similar, learners may feel confused and make an error, leading to learners' approximation of the two words' meanings (Liu, 1999). The more detailed concordance examples are revealed as follows.

[5] *First of all, all advertising companies create new product and see how we react to it even though they might loose some money for producing it (JPN).*

[6] *I do not have to loose time to make decisions, and also I never missed famous places to visit (JPN).*

[1] Some researchers may doubt that the use of *loose* might be due to wrong spelling instead of learners' confusion of word meanings and forms. However, EFL learners' intention of using these collocations remains unknown; thus, the present study will not discuss this debatable aspect.

The second frequent pattern in deviant verbs is related to the incorrect use of delexical verbs, including *do, have, make* and *take*, which comprise the errors, like **do mistakes*, **take a decision* and **do enjoyment.* This corresponds to previous studies that since delexical verbs do not have specific meanings, learners tend to consider delexical verbs as unrestricted combinations with collocations (Boers et al., 2013; Liu, 1999; Zinkgräf, 2008), which would give rise to the collocation errors.

The above result is discussed based on the frequency of miscollocations. After discussing the more frequent miscollocation categories, the researcher will further indicate the number of error types among collocation errors in TOEFL 11 based on Table 4.1 in the following section.

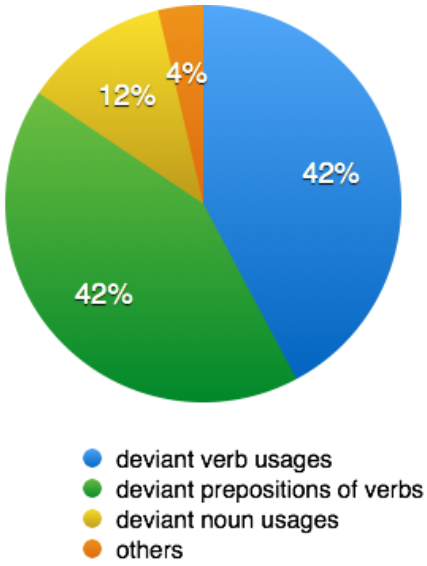
## 4.2 Error Types

Among 109 miscollocation types, both deviant verb usages and deviant prepositions of verbs contained 46 types of miscollocations. 13 miscollocation types belonged to the deviant noun usages and 4 types were others. Although the number of error types between deviant verb usages and deviant prepositions of verbs were the same, the frequency of deviant prepositions of verbs was more than that of verb usages, which indicates that learners in TOEFL 11 seemed to have more problems with the usage of prepositions in verb noun collocations. The error types and ratio of error types are illustrated in Table 4.2 and Figure 4.1.

Table 4.2 Error types of 109 collocation errors in TOEFL 11.

| Error Types | The numbers of error types | Frequency |
|---|---|---|
| Deviant verb usages | 46 | 295 |
| Deviant prepositions of verbs | 46 | 395 |
| Deviant noun usages | 13 | 73 |
| Others | 4 | 44 |
| Total | 109 | 807 |

Figure 4.1 The ratio of error types in TOEFL 11




The frequent miscollocations and the error types of all collocations were presented above. And it would be our interests to observe among these collocation errors, which collocations belonged to universal errors shared by learners of most L1s (at least 6 L1s) and which miscollocations only appeared in certain learner groups. The two dimensions will be revealed in the following part.

## 4.3 Universal Errors

Among 109 types of collocation errors, 61 types of miscollocations were universal errors, with 28 deviant verb usages, 27 deviant prepositions of verbs, 2 deviant noun usages and 3 other deviant usages, as shown in Table 4.3. The frequency of universal errors, the same as the frequent errors, comprised the most deviant prepositions of verbs, followed by deviant verbs and deviant nouns.

Table 4.3 The error types of universal errors.

| Deviant usages | V | P | N | O | Total |
|---|---|---|---|---|---|
| Error types | 28 | 27 | 3 | 3 | 61 |
| Frequency | 216 | 298 | 25 | 41 | 580 |

*V= deviant verb usages, P= deviant prepositions of verbs, N= deviant noun usages, O=others.

### 4.3.1 Deviant verb usages

28 types of deviant verb usages were found among universal errors. The error types were categorized into five groups based on the same incorrect usage of verbs, which would be the similar types of mistakes that can be further elaborated. The five groups of universal errors are the incorrect usage of *do, loose, experiment, try* and *reach*, ranked by their total frequency in the learner corpus.

#### 1. The deviant use of *do* in collocations

Group 1 contains the incorrect usage of a delexical verb *"do",* such as **do mistakes/ specialization/ many adventures/graduation* and *such enjoyment,* with the total frequency of 75 times in TOEFL 11. Learners of eight L1s made the errors regarding *do,* containing Arabic, German, French, Spanish, Hindi, Japanese, Korean, Turkish and Telugu EFL learners, as revealed in Table 4.4.

Table 4.4 The deviant use of *do*

| NO. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | *do mistakes | make | 6 | 2 | 4 | 0 | 1 | 1 | 0 | 2 | 1 | 9 | 0 | 26 | 8 |
| | *do specialization | have | 0 | 0 | 1 | 0 | 0 | 4 | 0 | 0 | 0 | 26 | 0 | 31 | |
| | *do many adventures | experience | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 5 | 0 | 6 | |
| | *do graduation | have | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 5 | 0 | 8 | |
| | *do such enjoyment | experience | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 4 | |
| **Total Frequency** | | | 6 | 2 | 5 | 0 | 1 | 12 | 0 | 3 | 1 | 45 | 0 | 75 | |

In terms of the suggested verbs above, the suggested forms are the frequent and more appropriate verb collocates of the noun from consulting BNCCOCA, including the frequent usages of verb noun collocations in both British and American English. **Do mistakes,* occurring 26 times in TOEFL 11*,* should be corrected as *make mistakes*. The correct verb for **do specialization* and **do graduation* is *have*. And the more appropriate verb for **do many adventures* and **do such enjoyment* can be *experience*.

The possible explanation for learners to make these errors is because the verb *do* is a delexical verb, which does not have a specific meaning or special meaning compared to other lexical verbs. Thus, learners in TOEFL tended to overgeneralize the meaning of *do* and combine it with any nouns (Zinkgräf, 2008). In addition, *do* is the high-frequency vocabulary, which often includes a high degree of polysemy (Viberg, 1996), leading to learners' erroneous assumption of a delexical verb's meaning (Liu, 1999) and causing collocation errors.

## 2. The deviant use of *loose* in collocations

In Group 2, there are 15 types of the incorrect usage of *loose,* such as **loose time, *loose some money, *loose everything, *loose something* and **loose your interest,* with 73 occurrences in total. Except Chinese learners, learners of the other L1s all made this kind of mistakes, as revealed in Table 4.5. The correct usages of collocations should be *lose time, money, everything, something* and *interest*. As explained in Section 4.1, learners confused about these two words probably because these two words share a certain extent of meanings and their forms and pronunciations are similar.

Table 4.5 The deviant use of *loose*

| NO. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | *loose time | waste | 1 | 1 | 2 | 2 | 2 | 0 | 1 | 0 | 1 | 0 | 0 | 10 | **10** |
| | *loose some money | lose | 0 | 2 | 2 | 0 | 2 | 2 | 1 | 0 | 5 | 0 | 0 | 14 | |
| | *loose everything | lose | 0 | 2 | 3 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 7 | |
| | *loose something | lose | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 4 | |
| | *loose your interest | lose | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 4 | |
| | * loose anything | lose | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 3 | |
| | *loose clients | lose | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 4 | |
| | *loose their present customers | lose | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | |
| | *loose importance | lose | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 3 | |
| | *loose this practical ability | lose | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 3 | |
| | *loose the chance | lose | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 2 | 0 | 0 | 5 | |
| | *loose weight | lose | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 3 | |
| | *loose the sense | lose | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 3 | |
| | *loose his | lose | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 3 | |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | reputation | | | | | | | | | | | | | |
| | *loose patience | lose | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Total Frequency | | | 1 | 10 | 9 | 8 | 7 | 7 | 6 | 3 | 17 | 5 | 0 | 73 |

## 3. The deviant use of *experiment* in collocations

Group 3 contains the errors of *experiment*, such as **experiment new things* (F=25) and **experiment a different culture* (F=3), which are shared by learners of seven L1s, including French, Spanish, Hindi, Korean, Turkish, Telugu and Chinese EFL learners, as illustrated in Table 4.6. Observed from the concordance *"Success depends on experimenting new things even though they are risky (TEL),"* the verb *experiment* should be corrected as *experience*, like *experience new things* and *experience a different culture*.

The causes of these errors could be discussed through their meanings and forms. According to Longman Dictionary of Contemporary English, *experiment* is an intransitive verb and its meaning is *"to try using various ideas, methods etc. to find out how good or effective they are"* while *experience* is a transitive verb and *"if you experience a problem, event, or situation, it happens to you or affects you."* Both words convey the meanings of trying something while the former one is emphasizing whether those things are effective and the latter is focusing on how people are affected by the things they have tried. Therefore, the meanings might be confusing for learners. On the other hand, from the forms of these two words, it can be seen that the forms of these words are similar, which causes the confusion and learners' approximation for those two words (Liu, 1999), which is similar to the case of "*loose* and *lose*" as discussed above.

Table 4.6 The deviant use of *experiment*

| NO. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | *experiment new things | experience | 0 | 0 | 3 | 0 | 4 | 3 | 0 | 2 | 2 | 8 | 3 | 25 | 7 |
| | *experiment a different culture | experience | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | |
| **Total frequency** | | | 0 | 0 | 3 | 0 | 7 | 3 | 0 | 2 | 2 | 8 | 3 | 28 | |

## 4. The deviant use of *try* in collocations

In Group 4, the incorrect use of *try* in collocations, with 26 occurrences, were produced by learners of nine L1s, including German, French, Italian, Spanish, Hindi, Japanese, Korean, Turkish, Telugu and Chinese learners. Errors such as **try new risks, *try new challenges* and **try new ventures* were made by these learners. The most frequent type of errors is **try new risks,* which occurs 17 times in TOEFL 11, as shown in Table 4.7.

Table 4.7 The deviant use of *try*

| NO. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *try new risks | take | 0 | 5 | 1 | 3 | 2 | 1 | 1 | 1 | 0 | 2 | 1 | 17 | 9 |
| | *try new challenges | take | 0 | 0 | 1 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 5 | |
| | *try new ventures | undertake | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 4 | |
| **Total Frequency** | | | 0 | 5 | 2 | 7 | 2 | 3 | 1 | 3 | 0 | 2 | 1 | 26 | |

The more correct verb for *risk* and *challenge* is *take,* as *take risks* and *challenges.* For **try new ventures,* the more appropriate verb would be *undertake*. *Take risks* occurs 1160 times, *take challenges* occurs 41 times and *undertake ventures* occurs 16 times in BNCCOCA while none of them combined with *try* occurs in BNCCOCA. The nouns here, *risks, challenges* and *ventures,* means something that is uncertain, dangerous or novel. EFL learners tends to combine these uncertain and novel things with the verb, *try,* which might be influenced by the combination "try something new." The phenomenon of these errors could be explained with the ideas of false analogy (Liu, 1999), that is, this kind of errors might be learners' false analogy of "try something," making learners feel that it is acceptable to add any nouns, especially something uncertain or new, after *try*.

## 5. The deviant use of *reach* in collocations

The universal collocation errors of *reach* were found in Group 5, including **reach the dream, *reach happiness* and **reach fame,* with 14 occurrences in total. Arabic, German, French, Spanish, Hindi, Japanese and Korean learners made these errors in their TOEFL writing, as shown in Table 4.8. The more appropriate verb for them should be *achieve,* such as *achieve the dream, happiness* and *fame* according to BNCCOCA. *Reach* and *achieve* are synonyms, which makes learners tend to use these two words interchangeably with collocations since learners usually resort to Open Choice Principle (Sinclair, 1991) where learners will add up the meanings of each individual word. In this way, substituting one synonym with another in collocations will make no differences for learners. This phenomenon corresponds to one learner strategy, analogy, proposed by P. Howarth (1998), which is the process of substituting one lexical item for another known element in a collocation. This kind of L2 intralingual error is the erroneous use of synonyms in collocation errors pointed out by Liu (1999).

Table 4.8 The deviant use of *reach*

| NO. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | *reach the dream | achieve | 1 | 1 | 1 | 0 | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 7 | 7 |
| | *reach happiness | achieve | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 3 | |
| | *reach fame | achieve | 1 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | |
| **Total frequency** | | | 2 | 1 | 2 | 0 | 6 | 1 | 1 | 1 | 0 | 0 | 0 | 14 | |

After the discussion of the universal errors concerning deviant verb usages, the following section is the deviant prepositions of verbs among universal errors.

## 4.3.2 Deviant prepositions of verbs

27 types of collocation errors out of 61 universal errors were deviant prepositions of verbs, categorized as seven groups sharing similar kinds of errors in each group and containing the incorrect usages of *agree, specialize, travel, listen, search, go* and *participate.* The causes of these errors are all due to the lack of prepositions after verbs.

### 1. <u>The deviant use of *agree* in collocations</u>

The first deviant prepositions of verbs are the use of *agree* and *disagree* (Frequency =120). 9 groups of EFL learners made the errors, such as **agree this idea, *agree the opinion, *disagree the opinion* and **disagree that topic*, as shown in Table 4.9. The preposition *with* is neglected by learners. The incorrect usage of *agree/disagree* was also found in the studies of Lin (2010) and Liu (2013). Both studies analyzed Chinese and Taiwanese learners'

collocation errors. After finding the misuse of *agree/disagree,* both Lin (2010) and Liu (2013) attributed these errors to L1 interferences, indicating that since in Chinese syntax, there are no structures for *agree+prep,* learners may have more troubles in using the verb *agree* or *disagree*. However, the present study found that these errors are not only made by Chinese learners. Learners, such as French, Italian, Spanish, Japanese and Korean, also made the errors in their written production, revealing that the misuse of *agree/disagree* may be also influenced by intralingual factors.

Table 4.9 The deviant use of *agree*

| No. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Learner corpus Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *agree this idea | agree with | 1 | 0 | 0 | 0 | 0 | 0 | 20 | 7 | 11 | 0 | 1 | 40 | 9 |
| | *agree the opinion | agree with | 0 | 0 | 1 | 0 | 0 | 0 | 17 | 10 | 0 | 0 | 3 | 31 | |
| | *disagree the idea | disagree with | 1 | 0 | 0 | 0 | 0 | 0 | 8 | 3 | 1 | 0 | 1 | 14 | |
| | *agree the statement | agree with | 1 | 0 | 0 | 2 | 0 | 0 | 5 | 0 | 0 | 2 | 3 | 13 | |
| | *disagree the opinion | disagree with | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 4 | 0 | 0 | 0 | 9 | |
| | *agree the topic | agree with | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 0 | 2 | 6 | |
| | *agree this sentence | agree with | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 4 | |
| | *disagree that topic | disagree with | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 3 | |
| | Total | | 3 | 0 | 1 | 2 | 1 | 0 | 56 | 26 | 15 | 2 | 11 | 120 | |

## 2. The deviant use of *travel* in collocations

The second group of deviant prepositions of verb is the use of *travel* (Frequency=53), produced by learners of 9 L1s as indicated in Table 4.10. Learners in TOEFL 11 tended to combine *travel* directly with some places, making errors like **travel places, *travel lands, *travel a park* and **travel Japan* while ignoring the preposition *to*. From the concordances in TOEFL 11, as revealed in Section 4.1, learners' expression for *travel* means having a journey. In that way, the verb *travel* is an intransitive verb and the prepositions, such as *to,* are needed.

Table 4.10 The deviant use of *travel*

| NO. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Learner corpus Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | *travel places | travel to | 1 | 1 | 0 | 0 | 0 | 1 | 7 | 16 | 1 | 0 | 16 | 43 | 9 |
| | *travel lands | travel to | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 3 | |
| | *travel a park | travel to | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | |
| | *travel Japan | travel to | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 4 | |
| | Total | | 1 | 1 | 0 | 1 | 1 | 1 | 7 | 21 | 1 | 0 | 19 | 53 | |

## 3. The deviant use of *specialize* in collocations

The third deviant prepositions of verbs among universal errors are the incorrect use of *specialize*, as illustrated in Table 4.11. **Specialize one subject* was produced by learners of 11 L1s with the occurrences of 46 times. The correct form should be *specialize in one subject.* The preposition *in* seems to be neglected by most learners, leading to the collocation errors.

Table 4.11: The deviant use of *specialize*

| NO. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Learner corpus Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | *specialize one subject | specialize in | 1 | 2 | 1 | 1 | 2 | 2 | 8 | 6 | 10 | 8 | 5 | 46 | 11 |

## 4. The deviant use of *listen* and *search* in collocations

Group 4 and 5 contain the errors of *listen* (Frequency=28) and *search* (Frequency=13), comprising the errors like **listen music* and **search the information,* both produced by learners of 8 L1s. *Listen to music* and *search for the information* are the correct forms. Like **agree,* both **listen* and **search* were also found in Lin's (2010) and Liu's (2013) studies, investigating Chinese and Taiwanese learners. Lin (2010) concluded that the cause for **listen* is attributed to the ignorance of TL rules while **search* is due to the L1 influences since Chinese and Taiwanese learners shared the error of **search*. The result of the present research pointed out that **search* was not only produced by Chinese learners, it was also made by German, French, Italian, Spanish learners and so on. This reveals that not only *agree,* but also **listen* and **search* might be related to intraligual factors, which may result from learners' ignorance of TL rules.

Table 4.12 The deviant use of *listen* and *search*

| NO. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Learner corpus Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | *listen music | listen to | 0 | 0 | 3 | 1 | 0 | 0 | 1 | 2 | 8 | 4 | 1 | 20 | 8 |
| | *listen a lecture | listen to | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 5 | |
| | *listen their story | listen to | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 3 | |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | | 0 | 0 | 3 | 3 | 0 | 1 | 2 | 3 | 8 | 5 | 3 | 28 | |
| 5 | *search the information | search for | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 3 | 1 | 0 | 2 | 8 | 8 |
| | *search a job | search for | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 5 | |
| | Total | | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 3 | 3 | 1 | 2 | 13 | |

## 5. The deviant use of *go* and *participate* in collocations

Group 6 includes two wrong types of **go,* such as **go concerts* and **go vacation*. The correct forms for both are *go to concerts* and *go on vacation* respectively. The lack of *to* occurred 11 times and that of *on* occurs 7 times in TOEFL 11. The similar errors of missing *to* in the use of *go* were also shown in Lin's (2010) study, which attributed the errors to learners' ignorance of the TL rules when constructing collocations.

Table 4.13 The deviant use of *go*

| NO. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Learner corpus Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | *go concerts | go to | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 6 |
| | *go Egypt | go to | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 3 | |
| | *go shopping malls | go to | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 2 | |
| | *go the Tower | go to | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | |
| | *go his workplace | go to | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | |
| | *go vacation | go on | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 4 | |
| | *go a picnic | go on | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 3 | |
| | Total | | 0 | 0 | 0 | 0 | 2 | 1 | 5 | 7 | 2 | 0 | 1 | 18 | |

Similarly, **participate* in Group 7 was erroneous probably because of learners' ignorance of TL rules, resulting in the errors like **participate activities, **participate events* (participate in activities and events), which occurs 20 times, shared by learners of 6 L1s. Though these errors did not appear in the research of Lin (2010), they occurred in Chinese and Taiwanese learner corpus of Liu's (2013) study. Liu (2013) attributed these errors to negative L1 transfer; however, as can be seen in Table 4.14, these errors were also produced by learners of different L1s in addition to Chinese learners, such as French, Japanese, Korean and Turkish learners. This corresponds to the phenomenon that the errors might result from learners' ignorance of TL rules rather than simply L1 interferences.

Table 4.14 The deviant use of *participate*

| NO. | Miscollocations | Suggested verbs | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Learner corpus Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | *participate activities | participate in | 0 | 1 | 0 | 0 | 0 | 0 | 3 | 4 | 0 | 1 | 8 | 17 | 6 |
| | *participate events | participate in | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 3 | |
| | Total | | 0 | 1 | 0 | 0 | 0 | 0 | 4 | 5 | 1 | 1 | 8 | 20 | |

## 4.3.3 Deviant noun usages and others

Three types of deviant noun usages were found in 61 types of universal errors. The two types were related to the incorrect use of *sell* and *selling,* such as **increase the sell* (F=5) and **increase selling* (F=5), shared by learners of seven L1s. The other type of errors is **take a travel,* occurring 15 times and shared by six L1s, as illustrated in Table 4.15. The proper form for the first two incorrect usages should be i*ncrease sales*. The errors may be related to learners' confusion about part of speech. *Sell* represents a verb while *sale* is a noun. The

word following increase should be a noun. Although *selling* is a noun, which refers to *"the job and skill of persuading people to buy things,"* it does not correspond to the meaning of the collocations in concordances. The plural form *sales*, on the other hand, meaning *"the total number of products that are sold during a particular period of time"* according to Longman Dictionary of Contemporary English, corresponds to the meaning of concordances in TOEFL 11. Thus, the suggested noun for the collocations should be *sales*.

Table 4.15 Deviant noun usages

| NO. | Miscollocations | Suggested nouns | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Learner corpus Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *increase the sell | sales | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 5 | 5 |
| | *increase selling | sales | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 5 | 4 |
| | Total | | 2 | 1 | 1 | 1 | 0 | 1 | 0 | 2 | 2 | 0 | 0 | 10 | 7 |
| 2 | *take a travel | trip | 1 | 1 | 0 | 0 | 1 | 0 | 2 | 6 | 0 | 0 | 4 | 15 | 6 |

Furthermore, in terms of other error types, three error types were found among 61 types of universal errors. Two of them were concerned about the incorrect use of *conclude* and the other one was related to the usage of *travel*, as presented in Table 4.16. All of them are a kind of inappropriate combinations. The correct expression for **conclude ideas* and **conclude the fact* should be *conclude* or *make conclusions*. The verb *conclude* here means *"to decide that something is true after considering all the information you have."* Thus, *conclude* already contains the meaning of true information or ideas. *Ideas* or *the fact* that follows *conclude* would be unnecessary and repetitive. The incorrect use of *conclude* might be resulted from learners' ignorance of TL rules. On the other hand, the other inappropriate combination was

*have a travel*, whose correct usage should be *have/take a trip.* The miscollocation **have a travel* might be the false analogy of *take a trip* (Liu, 1999).

Table 4.16 Other types of errors

| NO. | Miscollocations | Suggested forms | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Learner corpus Frequency | Shared by how many L1 groups |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *conclude ideas | make conclusions | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 2 | 0 | 7 | 6 |
| | *conclude the fact | make conclusions | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 3 |
| | Total | | 2 | 0 | 2 | 1 | 1 | 0 | 1 | 0 | 1 | 2 | 1 | 11 | 8 |
| 2 | *have a travel | take a trip | 3 | 0 | 0 | 0 | 3 | 0 | 7 | 8 | 1 | 0 | 8 | 30 | 6 |

## 4.4 Errors Specific to Certain L1 Groups

48 out of 109 types of miscollocations belonged to errors specific to certain L1 groups, with 227 occurrences in total. These 48 types of errors were produced by one to five L1 groups. The errors were categorized based on learners' native language belonging to which language family, including Indo-European, Altaic, Sino-Tibetan, Afro-Asiatic and Dravidian language families. These Errors were divided by how many language families making the mistakes. Among 48 types of errors, 15 types of mistakes were only made by one language family. 26 types of mistakes were shared by two language families and 7 types of errors were made by more than three language families, as indicated in Table 4.17. The following parts will discuss the errors made by certain L1 groups according to language families.

Table 4.17 Specific errors categorized by language families

| Errors shared by… | No. | Language Families | Types | Frequency |
|---|---|---|---|---|
| **One language family** | 1 | Indo-European | 6 | 21 |
| | 2 | Altaic | 7 | 28 |
| | 3 | Sino-Tibetan | 1 | 3 |
| | 4 | Dravidian | 1 | 3 |
| | Total | | 15 | 55 |
| **Two language families** | 1 | Indo-European/Altaic | 12 | 70 |
| | 2 | Altaic/Sino-Tibetan | 5 | 33 |
| | 3 | Indo-European/ Dravidian | 6 | 21 |
| | 4 | Altaic/Dravidian | 3 | 13 |
| | Total | | 26 | 137 |
| **More than three language families** | 1 | Others | 7 | 35 |
| **Total Frequency** | | | 48 | 227 |

### 4.4.1 Errors made by one language family

Among 15 errors made by one language family, 6 of them were produced by learners of Indo-European, 7 by Altaic learners, 1 by Sino-Tibetan learners and the other one by Dravidian learners, occurring 55 times overall in TOEFL 11.

**1. Errors made by learners of Indo-European**

Six types of errors were from Indo-European language family. Two of them were made by learners of more than one Indo-European language, including **manipulate customers* (manipulate customers' desire) and **sharp your skills* (sharpen your skills). Four were only from one of the Indo-European languages, containing **underline opinions* (defend/underpin opinions), **realize chances* (take your chances), **stand testimony* (bear testimony), and

**specialize themselves* (specialize in), as shown in Table 4.18. Among these errors, the most frequent type of errors is deviant verbs, accounting for 67 %.

Table 4.18 Errors made by learners of Indo-European language family

| No. | Language family | L1s | Miscollocations | Suggested forms | Learner corpus Frequency |
|---|---|---|---|---|---|
| 1 | Indo-European | DEU/FRA/ ITA | *manipulate customers | manipulate customers' mind/ desire | 5 |
| | | DEU/HIN | *sharp your skills | sharpen your skills | 3 |
| | | DEU | *underline opinions | express /defend opinions | 4 |
| | | | *realize chances | take your chances | 3 |
| | | HIN | *stand testimony | bear testimony | 3 |
| | | FRA | *specialize themselves | specialize in | 3 |
| | | Total | | | 21 |

*The underlined ones are deviant verb usages.

The most frequent errors are **manipulate customers,* shared by German, French and Italian learners, with no occurrences in BNCCOA while 5 occurrences in TOEFL 11. In one of the concordances of *manipulate, "To manipulate the customer by giving him the illusion life cannot continue without buying this product is a very common procedure nowadays (DEU),"* it can reveal that the more appropriate form would be "manipulate customers' mind or desire."

From Table 4.18, it can also be observed that sometimes learners would confuse the vocabulary's part of speech. For instance, both German and Hindi learners confused the adjective "sharp" and its verb "sharpen" as revealed in the collocation error **sharp your skills.*

On the other hand, some mistakes were only made by learners of one Indo-European language. For example, German learners made the mistakes like **underline opinions* (Frequency=4) and **realize chances* (Frequency=3). From the concordances "*Another*

*argument to underline my opinion is that you have better conversations with people by knowing many things briefly*" and *"It is often impossible to realize your chances with an absolute certainty that everything will turn out all right"* in TOEFL 11, the proper forms would be "express my opinion" and "take your chances." Also, Hindi learners made the errors, as **stand testimony* (bear testimony) while French learners produced **specialize themselves* (specialize in).

## 2. Errors made by learners of Altaic

Seven types of errors were produced by Altaic learners, including Japanese, Korean and Turkish learners. Among these errors, one of them was made by learners of all Altaic languages and three of them were shared by learners from both Korea and Japan while the rest of them were made by learners of one specific L1, as given in Table 4.19. The most frequent type of errors is deviant prepositions of verbs, accounting for 57 % of the errors.

Table 4.19 Errors made by learners of Altaic language family

| No. | Language family | L1s | Miscollocations | Suggested forms | Learner corpus Frequency |
|---|---|---|---|---|---|
| 2 | Altaic | JPN/KOR/TUR | *graduate university | graduate from university | 9 |
| | | JPN/KOR | *major subjects | major in subjects | 4 |
| | | | *invent the energy | find the energy | 3 |
| | | | *succeed the business | succeed in the business | 3 |
| | | JPN | *stay a hotel | stay at a hotel | 4 |
| | | KOR | *ride a ski | go skiing | 3 |
| | | TUR | *learn the intuition | learn the fact | 2 |
| | Total Frequency | | | | 28 |

*The underlined ones are deviant prepositions of verbs.

The most frequent errors made by Altaic learners were **graduate university (graduate from university),* which is a deviant prepositional error, with 9 occurrences in TOEFL 11. Turkish, Japanese and Korean learners all made this error. This result corresponds to the study of J. Lee and Seneff (2008), who analyzed the the Japanese Learners of English corpus (JLE) to find the grammatical errors in non-native speakers' speech in English and found that the misuse of articles and prepositions are the most frequent errors in JLE, including the error **graduate university*.

In addition, both Japanese and Korean learners of English shared three collocation errors. It seems that Korean and Japanese learners tend to make the same mistakes quite often. 2 out of 3 errors are deviant prepositional errors, such as **major subjects* (Frequency=4) and **succeed the business* (Frequency=3) and one is deviant verb usage, as **invent the energy* (Frequency=3). From the concordance *"For example, suppose there is a graduate who majored three subjects, History, Business, and Math,"* indicating that *major* here is a verb instead of an adjective; thus, it should be *major in subjects* instead of **major subjects*. Likewise, the preposition *in* was also neglected by learners in **succeed the business*. As for the deviant verb usage, from the concordance *"What people want is clean and cheap energy. and twenty years is a long time to discover or invent the alternative energy (KOR),"* it can be seen that the energy here means "*power that is used to provide heat, operate machines, etc"* according to Longman Dictionary. Then this energy should already exist instead of being invented by people; therefore, *find the energy* is more proper than **invent the energy*.

Finally, the rest of collocation errors were only made by learners of one Altaic language, containing **stay a hotel (stay at a hotel)* produced by Japanese learners, **ride a ski (go skiing)* written by Korean learners and **learn the intuition (learn the fact)* created by Turkish learners.

## 3. Errors made by learners of Sino-Tibetan

Table 4.20 indicates that the error **eat medicine,* with 3 occurrences in TOEFL 11, was only produced by Chinese learners, which further confirms the fact that this error is attributed to L1 negative interferences according to previous studies. Chinese learners of English tend to directly translate Chinese into English, which leads to the error **eat medicine* (take medicine). The similar results were also shown in the studies of Liu (2013),Lin (2010) and Y.-C. Chang, Chang, Chen, and Liou (2008).

Table 4.20 Errors made by learners of Sino-Tibetan language family

| No. | Language families | L1s | Miscollocations | Suggested forms | Learner corpus Frequency |
|---|---|---|---|---|---|
| 3 | Sino-Tibetan | ZHO | *eat medicine | take medicine | 3 |

## 4. Errors made by learners of Dravidian

Table 4.21 reveals that only Telugu learners made the mistakes **learn hardware.* The verb *learn,* sometimes is transitive verb, as *learn a language,* while it can be intransitive verb as well, such as *learn about music*, according to Longman Dictionary. Here *learn* is more likely to be an intransitive verb so the proper form should be *learn about the hardware* rather than **learn hardware.*

Table 4.21 Errors made by learners of Dravidian language family

| No. | Language families | L1s | Miscollocations | Suggested forms | Learner corpus Frequency |
|---|---|---|---|---|---|
| 4 | Dravidian | TEL | *learn hardware | learn about the computer hardware | 3 |

### 4.4.2 Errors shared by two language families

As for 28 types of errors shared by two language families, 12 types were from Indo-European/Altaic, 5 from Altaic/Sino-Tibetan, 6 from Indo-European/ Dravidian and 3 from Altaic/Dravidian, as discussed below.

#### 1. Errors made by learners of Indo-European/Altaic

There are 12 types of errors produced by Indo-European and Altaic language family, occurring 70 times in total. The errors are ranked by the number of the L1s making the errors. 6 of them were made by five L1 groups, 1 from four L1 groups, 4 from three L1 groups and 1 from two L1 groups, as can be seen in Table 4.22. Half of them is deviant prepositions of verbs.

Table 4.22 Errors made by learners of Indo-European/Altaic language family

| No. | Language families | L1s | Miscollocations | Suggested forms | Learner corpus Frequency |
|---|---|---|---|---|---|
| 1 | Indo-European/Altaic | FRA/ SPA/ JPN/ KOR/ TUR | *look your products | look at your products | 15 |
| 2 | | | *look the advertisement | look at the advertisement | |
| 3 | | | *look the surface | look at the surface | |
| 4 | | FRA/JPN/KOR/ ITA/TUR | *prepare the travel | prepare for the travel | 8 |
| 5 | | | *prepare an advertisement | prepare for an advertisement | |
| 6 | | DEU/ FRA/ JPN/ KOR/ TUR | *earn knowledge | acquire knowledge | 5 |
| 7 | | FRA/HIN/JPN/TUR | *earn their life | earn money | 11 |
| 8 | | DEU/FRA/TUR | *make internships | have internships | 6 |
| 9 | | ITA/SPA/KOR | *contract a guide | contract with a guide | 4 |

| | | | | | |
|---|---|---|---|---|---|
| 10 | | FRA/SPA/TUR | *leave my routine | change my routine | 3 |
| 11 | | FRA/ KOR/ TUR | *earn the facts | learn the facts | 3 |
| 12 | | DEU/JPN | *retire their jobs | retire | 15 |
| Total Frequency | | | | | 70 |

*The underlined ones are deviant prepositions of verbs.

The most frequent types of errors are related to the usage of *look.* French, Spanish and Altaic learners produced the errors, such as **look your products, *look the advertisement* and **look the surface,* with 15 occurrences in TOEFL 11, and neglected the preposition *at*. Like some deviant prepositions of verbs among universal errors, such as *listen, search, go* and *participate* discussed before, the possible reason for learners to make these mistakes might be attributed to the ignorance of TL rules. The second frequent errors, **prepare the travel/ an advertisement (prepare for…),* made by learners of five L1s, were probably also caused by the same reason.

Among the rest of eight errors, 3 of them are concerned about the usage of *earn*, referring to **earn knowledge*, **earn their life* and **earn the facts.* The miscollocations may be attributed to two reasons for the misuse of *earn* for *acquire* and *earn* for *learn*. As can be seen, *earn* and *acquire* are synonyms, which makes learners resort to the OP principle and use these two words interchangeably in collocations, leading to the collocation errors. On the other hand, the misuse of *earn* for *learn* might be due to their similar forms and the certain degree of shared meanings, resulting in the approximation of the two words' meanings for learners (Liu, 1999).

As for the most frequent error **make internships* shared by learners of three L1s, the more appropriate verb is *have.* Since both *make* and *have* are delexical verbs, which do not have very specific meanings, thus make learners confuse the two words' usage and feel that it

is acceptable to add any nouns after delexical verbs (Zinkgräf, 2008). The internship is not created by learners; instead, they own this position. Therefore, the more appropriate usage is *have internships* rather than **make internships.*

Lastly, the error **retire their jobs (retire)* was made by learners from two countries (Germany/ Japan), occurring 15 times in the learner corpus. Since *retire* is an intransitive verb, learners should simply write *retire,* which may result from learners' ignorance of TL rules.

## 2. Errors made by learners of Altaic/Sino-Tibetan

5 out of 26 error types with 33 occurrences were produced by Altaic and Sino-Tibetan learners. According to Table 4.23, three errors are deviant prepositions of verbs which might originate from learners' ignorance of TL rules, such as **stay the place (stay at the place), *adapt the society (adapt to the society) and *arrive the place (arrive at the place)*. Japanese, Korean and Chinese learners shared two types of errors regarding deviant verb or noun usages, which are **face accidents (have accidents)* and **help neighborhood (help neighbors).*

Table 4.23 Errors made by learners of Altaic/Sino-Tibetan language family

| No. | Language families | L1s | Miscollocations | Suggested forms | Learner corpus Frequency |
|---|---|---|---|---|---|
| 1 | Altaic/Sino-Tibetan | JPN/KOR/TUR/ZHO | *stay the place | stay at the place | 12 |
| 2 | | JPN/KOR/ZHO | *face accidents | have accidents | 5 |
| 3 | | | *help neighborhood | help neighbors | 5 |
| 4 | | KOR/TUR/ZHO | *adapt the society | adapt to the society | 4 |
| 5 | | KOR/ZHO | *arrive the place | arrive at the place | 7 |
| | | | Total Frequency | | 33 |

## 3. Errors made by learners of Indo-European/Dravidian

Among errors made by Indo-European and Dravidian language family, three are concerned about deviant verbs as follows. **Reach the destiny (reach the destination)* and **see hardships (experience/encounter hardships)* were made by Hindi and Telugu learners while German and Telugu learners made the errors of **outnumber the decrease (exceed the decrease).* Three types of errors are related to the misuse of *think*, with 21 occurrences overall, produced by learners of five L1s, as indicated in Table 4.24.

Table 4.24 Errors made by learners of Indo-European/ Dravidian language family

| No. | Language families | L1s | Miscollocations | Suggested forms | Learner corpus Frequency |
|---|---|---|---|---|---|
| 1 | Indo-European/ Dravidian | HIN/TEL | *reach the destiny | reach the destination | 5 |
| 2 | | | *see hardships | experience/encounter hardships | 3 |
| 3 | | DEU/ TEL | * outnumber the decrease | exceed the decrease | 4 |
| 4 | | DEU/ HIN/ ITA/ SPA/ TEL | *learn thinks | learn thoughts | 3 |
| 5 | | | *try thinks | try thoughts | 3 |
| 6 | | | *know the thinks | know the thoughts | 3 |
| | | Total | | | 21 |

In terms of **reach the destiny,* it can be seen that learners confuse *destiny* with *destination*, which can be attributed to learners' approximation of these two words since the two words share a similar form. As for **outnumber the decrease,* the more proper verb would be *exceed.* Since *outnumber* and *exceed* are synonyms, the reason that learners made this error might originate from the erroneous use of synonyms.

Regarding the misuse of *think*, identical to the cause of the misuse of *sharp/sharpen,*

learners tended to confuse the verb "think" and the noun "thought." Thus, they made the mistakes, like **learn thinks, *try thinks* and **know the thinks*. This points out that not only learners of Indo-European, but also Dravidian learners, would confuse with the part of speech of *think* and *thought*.

## 4. Errors made by learners of Altaic/Dravidian

Three types of errors, **concentrate subjects, *meet accidents* and **travel a tour*, occurring 13 times in total, were made by Altaic and Dravidian learners, as presented in Table 4.25. Japanese, Korean and Telugu learners neglected *on* when using *concentrate* and ignored *with* when using *meet with accidents,* which are deviant prepositions of verb usages, resulting from the ignorance of TL rules as mentioned previously. In addition, Japanese and Telugu learners produced one deviant verb usage, **travel a tour*, whose correct usage should be *go on a tour*.

Table 4.25 Errors made by learners of Altaic/Dravidian language family

| No. | Language families | L1s | Miscollocations | Suggested forms | Learner corpus Frequency |
|---|---|---|---|---|---|
| 1 | Altaic/Dravidian | JPN/KOR/TEL | *concentrate subjects | concentrate on subjects | 8 |
| 2 | | JPN/TEL | *meet accidents | meet with accidents | 3 |
| 3 | | | *travel a tour | go on a tour | 2 |
| | | Total | | | 13 |

## 4.4.3 Errors shared by more than two language families

Some errors are made by learners from three or four language families. 5 types of errors were made by Indo-European/ Altaic/ Afro-Asiatic and one was produced by Indo-European/ Altaic/ Sino-Tibetan while the other one was made by Indo-European/ Altaic/ Afro-Asiatic/

Dravidian language families, with 35 occurrences in total, as presented in Table 4.26. There are three deviant verbs, two deviant nouns and two deviant prepositions of verbs.

Table 4.26 Errors made by learners of Indo-European/Altaic /Afro-Asiatic language family

| | **Language families** | **L1s** | **Miscollocations** | **Suggested forms** | **Learner corpus Frequency** |
|---|---|---|---|---|---|
| 1 | Indo-European/Altaic /Afro-Asiatic | ARA/ ITA/ SPA DEU/ KOR | *create advertising | create advertisements | 3 |
| 2 | | | *watch advertising | watch advertisements | 4 |
| 3 | | ARA/HIN/JPN | *contribute their communities | contribute to their communities | 6 |
| 4 | | ARA/SPA/TUR | *take a decision | make a decision | 5 |
| 5 | | ARA/KOR/TUR | *study majors | take majors | 4 |
| 6 | Indo-European/ Altaic/Sino-Tibetan | ITA/HIN/TUR/ZHO | *care their community | care about their community | 7 |
| 7 | Indo-European/Altaic/ Afro-Asiatic/ Dravidian | ARA/ HIN/ JPN/ KOR/ TEL | *use celebrities | hire celebrities | 6 |
| | Total | | | | 35 |

The most frequent type of errors made by Indo-European, Altaic and Afro-Asiatic learners is related to the use of *advertising* and *advertisement,* including **create advertising* and **watch advertising.* The gerund *advertising* refers to *"the activity or business of advertising things on television, in newspapers, etc."* from Longman Dictionary of Contemporary English while an *advertisement* is a picture, set of words, or a short film that persuades people to buy things. *Advertisement* is more appropriate in the collocations from the meaning of the concordances. These errors are not only concerned about part of speech

but also result from learners' unawareness of the subtle differences of meanings between a gerund and a noun of the same verb.

The type of errors made by Indo-European, Altaic and Sino-Tibetan language families is **care their community*. Here *care* is an intransitive verb and the preposition *about* is needed. Learners' neglecting the combination of *care about* might be due to their ignorance of TL rules (Liu, 1999), as some deviant prepositions of verbs discussed above.

Finally, errors made by Indo-European/ Altaic/ Afro-Asiatic/ Dravidian language families are concerned about the miscollocations with *use,* like **use celebrities.* From the concordance "*another technique that many business firms use is using celebrities to persuade people for buying their goods (HIN),*" it can be revealed that learners' intended meanings should be *hire celebrities*. From the definition of Longman Dictionary of Contemporary English, *using people* refers to *"treat somebody unfairly"* in order to gain something you desire. The verb *hire* would be more proper for the collocation since these firms do not treat celebrities unfairly. The reason why learners made this error might be due to the lack of collocation knowledge or vocabulary deficiency, thus choosing the wrong known word for the collocations according to Liu (1999).

## 4.5 Discussion

### 4.5.1 Universal errors

The 61 types of universal errors shared by over six L1 groups can be the collocations that are problematic for learners. EFL teachers could pay attention to these collocation types while teaching related materials. Three main dimensions can be discussed from the result above.

First of all, the incorrect usage of delexical verbs found in the present study corresponds to the fact that the delexical verbs are problematic for EFL learners shown in many previous studies (Altenberg & Granger, 2001; Boers et al., 2013; Juknevičienė, 2008; Nesselhauf, 2004, 2005; Yamashita & Jiang, 2010; Zinkgräf, 2008). The incorrect use of delexical verb *do* in the present research might be due to two main reasons. One is dealt with the phenomenon that delexical verbs, such as *do,* lack specific meanings. Because of the unspecific meanings of the verbs, these verbs contain a high degree of polysemy originating from their universal tendency of producing more abstract, general, grammaticalized or delexicalized usages (Viberg, 1996). Therefore, learners tend to overgeneralize these verbs and combine them with any noun under the assumption that these verb usages have no restrictions (Zinkgräf, 2008), causing the misuse of *do* in collocations.

The other reason is concerned about learners' overuse of delexical verbs leading to the misuse of them. The delexical verb *do* is the core word learned early and widely usable, and usually makes learners feel comfortable to use, resulting in the overuse of the verb, even for advanced learners (Hasselgren, 1994). Sometimes, when learners do not have enough vocabulary or collocation knowledge, they would depend on these core words as well without utilizing native-like expressions (Altenberg & Granger, 2001; De Cock, 1998; Granger, 1998). The overuse of delexical verbs was also presented in the studies of Källkvist (1999) and Granger (1996). The above possible explanations imply that learners resort to the Open Choice Principle, constructing meanings from individual words instead of from prefabricated patterns, unlike native speakers (Kjellmer, 1991; Wray, 2002).

Second, the result of deviant verbs in universal errors indicated that EFL learners tend to make collocation errors of two verbs sharing certain extent of meanings or similar forms, giving rise to the confusion of them. For instance, the cases of misusing *loose/lose, experiment/experience* and *reach/achieve* can show that each pair shares meanings to certain

extent, which makes learners perplexed. Besides, the forms of *loose/lose* and *experiment/experience* are similar, also bringing about learners' approximation of the two words in both instances (Liu, 1999).

Finally, the deviant prepositions of verbs among universal errors is all related to lack of prepositions in the present study, including *agree, travel, specialize, listen, search, go* and *participate*. Although some previous studies concluded that the deviant prepositions of verbs are L1-related (Chen, 2002; Lin, 2010; Liu, 2013), these errors are shared by more than six EFL learner groups in the present study, indicating that L1 interference might not be the only reason that cause the errors. Instead, the intralingual causes, such as the ignorance of TL rules, might be also influential on the prepositional mistakes in the use of verb noun collocations. Although most studies emphasized the importance of interlingual factors that lead to collocation errors, in the meta-analysis study of Heydari and Bagheri (2012), they reviewed many studies of error analysis and found that a great number of researchers in the field have considered intralingual factors a common cause of learner errors. While a few researchers found that EFL learners with the beginning level tend to produce many interligual errors, more and more intralingual errors were demonstrated in their production as they make progress in the norms of target language (Lee, 2001). This offers some insights in the result of the present study since the corpus data in TOEFL 11 is most likely from learners of advanced or high-intermediate level, which might explain why intralingual errors were manifested in the present research.

### 4.5.2 Errors specific to certain L1s

The researcher categorized the errors from specific L1 groups into different categorizations based on their language family. From the observation of the result above, two main phenomena, regarding L1 interferences, will be discussed in the following part.

Firstly, if we compare the frequent error types among these language families, it can be observed that Indo-European learners produced more deviant verbs (67%) while Altaic learners produced more deviant prepositions of verbs (57%) in the use of verb noun collocations, as provided in Table 4.27.

The deviant verbs made by Indo-European learners can be discussed from learners' native languages. According to Kochmar (2011), the different lexical systems in these Indo-European languages lead to the misuse of certain verbs. Different verbs in English may have the same translation in other Indo-European languages. For instance, *tell* and *say* are both *sagen* in German, and *decir* in Spanish. In addition, *Make* and *do* are both *machen* in German. In this way, when these EFL learners produce verb noun collocations, it is very likely that they made deviant verb usages in collocation use.

As for learners of Altaic, it can be seen that they tend to make prepositional mistakes which usually are related to lack of prepositions in verb noun collocations. Explained from the perspective of L1, many verbs are not followed by prepositions in Japanese or Korean and thus learners are influenced by the rules in their native language and neglect the prepositions in certain verb noun collocates in English, which is also shown in the study of Tanimura, Takeuchi, and Isahara (2004). According to Tanimura et al. (2004), Japanese learners sometimes neglected prepositions after verbs because of L1 transfer. Take the verb *go* in English for instance. The verb has certain meanings when followed by prepositions, such as *go to*, *go for*, *go on*, *go out,* etc. in English while Japanese *"iku"* meaning *go* in English, has meaning without case markers, which can explain the phenomenon that Japanese learners neglect prepositions when using verb noun collocations.

Table 4.27 The most frequent types of collocation errors by different language families

| **Language Families** | **The most frequent types** (the number of frequent error types/ total error types) | **Percentage(%)** |
|---|---|---|
| Indo-European | Deviant verbs (4/6) | 67% |
| Altaic | Deviant prepositions of verbs (4/7) | 57% |
| Indo-European/Altaic | Deviant prepositions of verbs (6/12) | 50% |
| Altaic/Sino-Tibetan | Deviant prepositions of verbs (3/5) | 60% |
| Indo-European/Dravidian | Deviant verbs (3/6) | 50% |
| Altaic/Dravidian | Deviant prepositions of verbs (2/3) | 67% |
| Indo-European/Altaic/ Afro-Asiatic | Deviant verbs (2/5); Deviant nouns (2/5) | 40% |

*Language families that only contain one error type are neglected.

Secondly, Koreans and Japanese seem to make the same mistakes quite often in the finding. Among 71 types of miscollocations made by Koreans, Japanese and both of the learners, 40.8% is shared by learners of both L1s (Appendix 2). It seems that Koreans and Japanese tend to make similar collocation errors quite often in English, which can be explained from the fact that they all belong to Altaic language and share certain elements in common in both languages.

Table 4.28 The number of errors made by Korean, Japanese or both learners

| Error types made by Koreans, Japanese or both | Error types shared by both | Percentage |
|---|---|---|
| 71 | 29 | 40.8% |

# CHAPTER 5

# CONCLUSION

## 5.1 Summary of major findings

The present study discusses the cross-linguistic collocation errors shared by over six EFL learner groups and also specific errors shared by less than six groups, which are seldom explored by previous research. In addition, the innovative function of Sketch Engine, Sketch-Diff, is used to extract miscollocations among 11 languages in TOEFL 11. The source-tracing function also allows the researcher to discover the error distribution among these languages in order to find out universal and specific errors. With Sketch-Diff and its source-tracing function, it can save a lot of time and efforts to collect collocation errors in learner corpus, compared to the manual way and using concordance tool with manual inspection. The present research answers the following three research questions:

Q1: What are the verb noun miscollocations made frequently by eleven L1 groups of EFL learners?

Among 109 types of miscollocations, the frequent types of errors are divided into four categories, in which the frequent collocation errors contain the deviant use of *agree, travel, specialize, loose* and *do.* The incorrect use of *agree, travel* and *specialize* is related to missing prepositions. As for the verb *loose,* learners tend to confuse *lose* with *loose* because of their similar forms and a certain degree of shared meanings. Finally, concerning the delexical verb *do,* learners tend to misuse the delexical verb, and have wrong assumption that they can freely combine delexical verbs with any noun, which is attributed to the lack of specific meanings of delexical verbs.

Q2: What are the error types of these miscollocations?

The error types can be divided into four groups, including deviant verb usages, deviant prepositions of verbs, deviant noun usages and other types. Both deviant verb usages and deviant prepositions contain 46 error types, followed by deviant noun usages, comprising 13 types and 4 other types.

Q3: Among these miscollocations, what are the universal collocation errors shared by at least six L1 groups, and what are collocation errors made only by specific L1 groups?

There are 61 types of universal errors shared by over six L1 groups in TOEFL 11, including deviant verb usages (**do, *loose, *experiment, *try* and **reach*), deviant prepositions of verbs (**agree, *travel, *specialize, *listen, *search, *go* and **participate*), deviant noun usages (**sale*) and others. These results showed the learners' problems with delexical verbs, verbs sharing similar meaning and forms and lack of prepositions when using certain verbs.

Additionally, 48 specific errors shared by less than six L1 groups can be found with 15 types made by one language family, 26 by two language families and 7 by more than three language families. Indo-European language family tends to make more errors of deviant verbs while Altaic makes more errors of deviant prepositions. Besides, Korean and Japanese learners tend to make the same mistakes quite often in TOEFL 11, attributed to their similar L1 background.

## 5.2 Pedagogical Implications

From observing universal errors and specific errors found in TOEFL 11, the present study provides several pedagogical implications for teachers in EFL countries and especially for

teachers who might need to teach English to learners from various countries in the same class, such as sheltered English program. Several suggestions for learning and teaching verb noun collocations are illustrated as follows. First, the deviant use of delexical verb *do* seem universal for EFL learners in language learning. Thus, while teachers instruct verb noun collocations with delexical verbs, they should raise learners' awareness toward the delexical verbs and remind them that delexical verbs can not be replaceable in collocation usages. For instance, we cannot replace *make* with *do* in *make mistakes.*

Second, the verbs sharing similar form and a certain degree of meaning seem confusing for EFL learners, such as "*loose* and *lose*" and "*experience* and *experiment,*" which are universal errors for EFL learners. Therefore, when teachers teach these vocabulary, they should supplement the other similar form and explain their differences in meanings, forms and usages, with L1 explanation if needed, which can decrease the possibility that they misuse the words in collocation use.

Third, sometimes learners making collocation mistakes is because they feel that they can interchange one word with its synonym and also the other way around. Take the universal errors, *reach* and *achieve,* in the present study for instance. They are synonyms which are used by learners interchangeably with collocations in TOEFL 11 since learners usually resort to Open Choice Principle (Sinclair, 1991). Teachers, thus, should pinpoint the fact that we should not interchange synonyms of word usages in some cases. If students have difficulty determining whether they can substitute the word with it synonyms, several online collocation-checking tools besides dictionaries can be introduced to them, such as *Just the Word* and *ForBetterEnglish*.

Fourth, a great number of universal errors revealed in the present research are concerned about deviant prepositions of verbs, which are due to lack of prepositions in the result. Most previous studies attributed the mistakes to L1 interferences (Chen, 2002; Lin, 2010; Liu,

2013). However, since more than six L1 groups have made the errors, meaning that the possible reasons for the phenomenon is not totally L1-related. Instead, these might be the intralingual errors. Therefore, when instructors teach certain verbs, such as *agree, specialize, listen, search, participate,* etc., their following prepositions must be emphasized by teachers and can be practiced in drills to strengthen their memory.

Fifth, as can be seen in the result, learners sometimes confuse the words' part of speech as in *think/thought* and *sharp/sharpen.* Teachers should raise students' awareness of part of speech and provide various examples of words regarding different part of speech when teaching vocabulary. Besides, the meanings of gerunds and nouns are usually different, as in *advertising/advertisement.* Learners' unclear understanding of word meaning leads to collocation errors accordingly. Teachers, thus, should remind students that the meaning of gerund is extended from its verb, whose meaning has subtle differences with its noun.

Finally, in terms of specific errors, it is found that Korean and Japanese learners tend to make the similar collocation mistakes; thus, for teachers in sheltered English programs, if they have both Korean and Japanese students in class, they should notice that these learners might have difficulty in the same collocation usages because of their shared language family. In this way, when teachers instruct verb noun collocations to Japanese and Korean learners, they can take their L1 language background into account, and design and adjust their curriculums.

## 5.3 Limitations of the Present Study and Future Research

Although the present study provides some insights in universal and specific collocation errors and introduce an innovative method to utilize the Sketch-Diff, there are still some limitations of the present study. Some suggestions for future research are thus indicated.

First of all, since the definition of universal errors does not have a clear standard from previous research, the present study simply provides some insight in the universal errors shared by more than half L1 groups in TOEFL 11. It should be noticed that even though the present study sets the threshold of errors shared by six L1 groups as universal errors, errors made by five L1 groups can still possibly be universal errors in other kinds of context. Therefore, the definition of universal and specific errors should be seen as a continuum instead of the clear-cut polarity.

Second, the corpus of TOEFL 11 is not large enough, with approximately 3.8 million word tokens. Future research can replicate the present study at a larger scale in terms of word tokens and the number of L1s. In addition, other collocation types, such as adjective+noun, noun+verb, noun+noun, adverb+adjective and adverb+verb, can also be investigated in future studies to explore learner errors in different L1 groups.

Third, as indicated before, previous studies tend to explore collocation errors within one or two languages; thus, they usually attribute learners' errors to L1 interferences. However, in the present research, it can be seen that several errors are shared by various languages and intralingual factors play an important role here. Although the present research tends to resort to learners' ignorance of TL rules as an explanation of intralingual errors, as Ringbom (1998, p. 198) said, " the relationship between transfer and possible universal learner language characteristics is too complex to allow more than a few tentative points to be made on the material so far available." Therefore, the possible causes of these universal and intralingual errors still need more studies to further investigate.

Finally, with the plausible method to examine the cross-linguistic differences of collocation errors in the present study, future research can further adopt the source-tracing function of Sketch-Diff to investigate different variables, such as proficiency differences,

gender differences, age differences of learner production so as to offer pedagogical implications for both teachers and researchers.

# REFERENCES


Alexander, R. J. (1984). Fixed expressions in English: reference books and the teacher. *ELT Journal, 38*(2), 127-134.

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics, 22*(2), 173-195.

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System, 21*, 101-114. doi:10.1016/0346-251X(93)90010-E

Barfield, A. W. (2006). *An exploration of second language collocation knowledge and development.* University of Wales Swansea.

Bazzaz, F., & Samad, A. (2011). The Use of Verb Noun Collocations in Writing Stories among Iranian EFL Learners. *English Language Teaching, 4*, 158-163. doi:10.5539/elt.v4n3p158

Benson, M., Benson, E., & Ilson, R. (1997). The BBI dictionary of English word combinations (Rev. ed.). *Amsterdam: Benjamins*.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Grammar of spoken and written English. *Harlow: Longman*.

Biskup, D. (1992). L1 influence on learners' renderings of English collocations: A Polish/German empirical study. *Vocabulary and applied linguistics*, 85-93.

Blanchard, D., Tetreault, J., Higgins, D., Cahill, A., & Chodorow, M. (2013). TOEFL11: A CORPUS OF NON-NATIVE ENGLISH. *ETS Research Report Series, 2013*(2), i-15.

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2013). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research, 18*, 54-74. doi:10.1177/1362168813505389

Chang, W.-z., & Yang, S.-c. (2009). An Analysis of VN Collocation Errors in CLEC. *Journal of PLA University of Foreign Languages, 2*, 008.

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.

Chen, P. (2002). A corpus-based study of the collocational errors in the writings of the EFL learners in Taiwan. *Unpublished master's thesis, National Taiwan Normal University, Taipei*.

Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Quebec replications of three European studies. *Canadian Modern Language Review/La Revue canadienne des langues vivantes, 59*(3), 393-424.

Cowie, A. P. (1992). Multiword lexical units and communicative language teaching. *Vocabulary and applied linguistics*, 1-12.

De Cock, S. (1998). A Recurrent Word Combination Approachto the Study of Formulae in the Speech of Native and Non-Native Speakers of English. *International Journal of Corpus Linguistics, 3*(1), 59-80.

Farghal, M., & Al-Hamly, M. (2007). Lexical Collocations in EFL Writing. *The Journal of Asia TEFL, 4*(1), 269-294.

Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *IRAL. International review of applied linguistics in language teaching, 33*(4), 315-331.

Frath, P., & Gledhill, C. (2005). Free-range clusters or frozen chunks? Reference as a defining criterion for linguistic units. *RECHERCHES ANGLAISES ET NORD AMERICAINES, 38*, 25.

Gitsaki, C. (1996). The development of ESL collocational knowledge.

Gitsaki, C. (1999). *Second language lexical acquisition: A study of the development of collocational knowledge*: International Scholars Publications.

Gledhill, C. (2007). *La portée: seul dénominateur commun dans les constructions verbo-nominales.* Paper presented at the Actes du 1er colloque Res per nomen, Université de Reims.

Gledhill, C. J. (2000). *Collocations in science writing* (Vol. 22): Gunter Narr Verlag.

Granger, S. (1996). *Romance words in English: from history to pedagogy.* Paper presented at the International Symposium.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. *Phraseology: theory, analysis and applications. Oxford: OUP*.

Granger, S. (2011). From phraseology to pedagogy: Challenges and Prospects. *The Phraseological View of Language: A Tribute to John Sinclair*, 123-146. doi:10.1515/9783110257014

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*(2), 237-258.

Heydari, P., & Bagheri, M. S. (2012). Error Analysis: Sources of L2 Learners' Errors. *Theory and Practice in Language Studies, 2*, 1583-1589. doi:10.4304/tpls.2.8.1583-1589

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. *Teaching collocation: Further development in the lexical approach*, 47-69.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*, 24-44.

Howarth, P. A. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making* (Vol. 75): Walter de Gruyter.

Jaén, M. M. (2007). A corpus-driven design of a test for assessing the ESL collocational competence of university students. *IJES, International Journal of English Studies, 7*, 127-148.

James, C. (1998). Errors in language learning and use: Exploring error analysis. *London/NY: Longman*.

Juknevičienė, R. (2008). Collocations with high-frequency verbs in learner English: Lithuanian learners vs native speakers. *Kalbotyra, 59*(3), 119-127.

Källkvist, M. (1999). *Form-class and task-type effects in learner English: A study of advanced Swedish learners* (Vol. 95): Lund University Press.

Kilgarriff, A., & Rundell, M. (2003). *Lexical Profiling Software and its lexicographic applications: a case study.* Paper presented at the Proceedings of the Tenth EURALEX International Congress, EURALEX 2002: Copenhagen, Denmark, August 13-17, 2002.

Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). Itri-04-08 the sketch engine. *Information Technology, 105*, 116.

Kjellmer, G. (1991). *A mint of phrases*: na.

Kochmar, E. (2011). *Identification of a writer's native language by error analysis.* Master's thesis, University of Cambridge.

Laufer, B. (2010). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, ecq039.

Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography, 24*, 29-49. doi:10.1093/ijl/ecq039

Laufer, B., & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning, 61*(2), 647-672.

Lee, E. (2001). Error Analysis on medical students' writing. Retrieved October 21, 2008.

Lee, J., & Seneff, S. (2008). *An analysis of grammatical errors in non-native speech in English.* Paper presented at the Spoken Language Technology Workshop, 2008. SLT 2008. IEEE.

Lewis, M., & Conzett, J. (2000). *Teaching collocation: Further developments in the lexical approach*: Language Teaching Publications Hove.

Lin, M.-J. (2010). 根據台灣和大陸學習者語料庫的英語錯誤動名詞搭配詞研究. *臺灣師範大學英語學系學位論文*, 1-115.

Liu, C.-P. (1999). A Study of Chinese Culture University Freshmen's Collocational Competence. *Hwa Kang Journal of English Language & Literature*(5), 81-99.

Liu, L.-e. (2002). *A corpus-based lexical semantic investigation of verb-noun miscollocations in Taiwan learners' English.* Tamkang University, Graduate Institute of Western Languages and Literature.

Liu, Y. C. (2013). A Semi-Automated Corpus-Based Method of Comparing Verb-Noun Collocations Between Native and Non-Native English Speakers *Unpublished master's thesis, National Taiwan Normal University, Taiwan (R.O.C).*

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*: Oxford 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.

Nesselhauf, N. (2004). How learner corpus analysis can contribute to language teaching: A study of support verb constructions. *Corpora and language learners, 17*, 109.

Nesselhauf, N. (2005). *Collocations in a learner corpus* (Vol. 14): John Benjamins Publishing.

Ringbom, H. (1998). High-frequency verbs in the ICLE corpus. *LANGUAGE AND COMPUTERS, 23*, 191-200.

Rundell, M., & Fox, G. (2002). *Macmillan English dictionary for advanced learners*: Macmillan Oxford.

Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use* (Vol. 9): John Benjamins Publishing.

Schmitt, N., Ng, J. W. C., & Garras, J. (2011). The word associates format: Validation evidence. *Language Testing, 28*(1), 105-126.

Sinclair, J. (1991). *Corpus, concordance, collocation* (Vol. 1): Oxford University Press Oxford.

Steinberger, R., Pouliquen, B., Widiger, A., & Ignat, C. (2006). The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. *arXiv preprint cs/ …*.

Tanimura, M., Takeuchi, K., & Isahara, H. (2004). *From Learners' Corpora to Expert Knowledge Description: Analyzing Prepositions in the NICT JLE (Japanese Learner English) Corpus.* Paper presented at the Proceedings of IWLeL 2004: An Interactive Workshop on Language e-Learning.

Todirascu, A., & Gledhill, C. (2008). Extracting collocations in context: the case of verb-noun constructions in English and Romanian. *Recherches Anglaises et Nord-Américaines (RANAM), Université Marc Bloch Strasbourg*.

Viberg, Å. (1996). *Cross-linguistic lexicology: The case of English go and Swedish gå*.

Wang, Y., & Shaw, P. (2008). Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal, 32*, 201-232.

Wei, Y. (1999). Teaching Collocations for Productive Vocabulary Development.

Woolard, G. (2000). Collocation-encouraging learner independence. *Teaching collocation: Further developments in the lexical approach*, 28-46.

Wray, A. (2002). *Formulaic language and the lexicon* (Vol. 5): Cambridge University Press Cambridge.

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *Tesol Quarterly, 44*(4), 647-668.

Zinkgräf, M. (2008). V+ N miscollocations in the written production of university level students. *Elia, 8*, 91-116.

# APPENDIX 1

## Complete Verb Noun Collocation Errors (ranked alphabetically)

*Type A=Deviant verb usages; Type B= Deviant prepositions of verbs; Type C= Deviant noun usages; Type D=Others

| No. | Error Types | Collocation Errors | Suggested Forms | Concordance Examples in TOEFL 11 | ARA | DEU | FRA | ITA | SPA | HIN | JPN | KOR | TUR | TEL | ZHO | Learner Corpus Frequency |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | **A** | *do adventure | experience adventure | To do many adventures in life also we need youth. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 5 | 0 | 6 |
| **2** | **A** | *do enjoyment | experience enjoyment | This will allow him for greater success throughout the life as the elder have cross such limit they have no more spare time to do such enjoyment. | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 4 |
| **3** | **A** | *do graduation | have graduation | And the student who is doing graduation or undergraduation. He requires all the three methods as he has to submit research papers on every subject. | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 5 | 0 | 8 |
| **4** | **A** | *do mistakes | make mistakes | If the people having the capacity of understanding ideas there may be little chances to do the mistakes. | 6 | 2 | 4 | 0 | 1 | 1 | 0 | 2 | 1 | 9 | 0 | 26 |
| **5** | **A** | *do | have specialization | It is better to do specialization in one particular subject of | 0 | 0 | 1 | 0 | 0 | 4 | 0 | 0 | 0 | 26 | 0 | 31 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | specialization | | interest. | | | | | | | | | | | | | |
| **6** | **A** | *earn knowledge | acquire knowledge | In fact, it also could be fundamental to earn knowledge in politics, economics, art, music, every academics subjects that can help you to understand better the world. | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 5 |
| **7** | **A** | *earn the facts | learn the facts | It should be more important for students to understand ideas and concept 8than it is for them to earn the facts. | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 3 |
| **8** | **A** | *eat medicine | take medicine | Just eat this medicine ,everything will be OK . | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 |
| **9** | **A** | *experiment a culture | experience a culture | Therefore, I think that the best way to travel is without a tour guide because it helps you to experiment a different culture, | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| **10** | **A** | *experiment things | experience things | Experimenting new things is very useful to have success | 0 | 0 | 3 | 0 | 4 | 3 | 0 | 2 | 2 | 8 | 3 | 25 |
| **11** | **A** | *face accidents | have accidents | Moreover, when I face the traffic accident or disaster ,for example earthquake, My family get information immediately from the tour company. | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 1 | 5 |
| **12** | **A** | *invent the energy | find the energy | What people want is clean and cheap energy. and twenty years is long enough time to discover or invent the alternative energy. | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 3 |
| **13** | **A** | *leave my routine | change my routine | When I decide to leave my daily routine to enjoy the pleasures of a trip abroad, it is always to go somewhere I can be alone. | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 3 |
| **14** | **A** | * loose anything | lose anything | Sometimes, your risk may not have been a success, but you may not loose anything. | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 3 |
| **15** | **A** | *loose clients | lose clients | However , we should be careful about that because when we | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 4 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | loose clients , deceived them, you never be succeed . | | | | | | | | | | | | | |
| **16** | **A** | *loose customers | lose customers | People have so much choice these days, the manufacturers can't afford to loose their present customers . | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 4 |
| **17** | **A** | *loose everything | lose everything | Depending on how great the risk is, you may loose nearly everything you possess. | 0 | 2 | 3 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 7 |
| **18** | **A** | *loose importance | lose importance | Especially friends and family are sources of joy I expect to compensate for The loss of those other springs of bliss which will probably loose importance over time. | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 3 |
| **19** | **A** | *loose money | lose money | First of all, all advertising companies create new product and see how we react to it even though they might loose some money for producing it. | 0 | 2 | 2 | 0 | 2 | 2 | 1 | 0 | 5 | 0 | 0 | 14 |
| **20** | **A** | *loose patience | lose patience | Young people are the future of the world,they remain active in their young period but while they get older they loose patience, concentration and the body organs get weak, they may see inactive comparing with young people. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| **21** | **A** | *loose reputation | lose reputation | By publishing his theory of relativity he risked to loose his reputation in the scientific community | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 3 |
| **22** | **A** | *loose sense | lose sense | Now that the world is accessible to anyone and the young travel everywhere they loose the sense of home that someone that has always lived in only one village could feel. | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 3 |
| **23** | **A** | *loose | lose something | If you take risks, there will always be the opportunity to make | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 4 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | something | | it wrong, to loose something. | | | | | | | | | | | | | |
| **24** | **A** | *loose the ability | lose the ability | I often loose this practical ability to understand the real problems and the ideas that is hidden by the formula. | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 3 |
| **25** | **A** | *loose the chance | lose the chance | Although we do not need to consider about the schedule, we may loose the chance to talk with people who live in the tour place. | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 2 | 0 | 0 | 5 |
| **26** | **A** | *loose time | waste time | I do not have to loose time to make decisions | 1 | 1 | 2 | 2 | 2 | 0 | 1 | 0 | 1 | 0 | 0 | 10 |
| **27** | **A** | *loose weight | lose weight | Other commercials on television or posters hanged on walls in the city that make me mad are those about loosing weight. | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 3 |
| **28** | **A** | *loose your interest | lose your interest | After getting that thing, you loose your interest in it. | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 4 |
| **29** | **A** | *make internships | have internships | For example, most of my fellow students spend a lot of time on making internships, so there is no time left to voluntarily participate community activities. | 0 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 6 |
| **30** | **A** | *meet accidents | have accidents | In this case I have one bitter experience that is, few months back I met an accident to my leg. | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 3 |
| **31** | **A** | *reach fame | achieve fame | We have several examples of success stories of ordinary people that excelled in their traditional jobs and reach fame and wealth by offering their customers a usual products but with better quality. | 1 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| **32** | **A** | *reach | achieve happiness | It is just a matter of will to live, desire to reach happiness. | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 3 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | happiness | | | | | | | | | | | | | | | |
| **33** | **A** | *reach the dream | achieve the dream | I have been helped by my predecessors in regard to the homework which I could not solve by alone and the way to reach the dream I have held. | 1 | 1 | 1 | 0 | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 7 |
| **34** | **A** | *realize chances | take your chances | It is often impossible to realize your chances with an absolute certainty that everything will turn out all right. | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| **35** | **A** | *retire their jobs | retire /quit their jobs | Old people do not go to school, probably retired their job too. | 0 | 1 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 15 |
| **36** | **A** | *see hardships | experience/encounter hardships | They have not seen hardships of life and are capable to take on larger risks. | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 3 |
| **37** | **A** | *sharp your skills | sharpen your skills | Later you also got the chance to become a specialist in your subject, you can still sharp your skills. | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 3 |
| **38** | **A** | *stand testimony | bear testimony | The media might be in part responsible for this, as is evident from the interest in publishing the un-necessary crap like the celebrity gossip, than to highlight the problems of the community, and sensitizing the young generation about it. The news papers and magazines that we read these days stand testimony to this. | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 3 |
| **39** | **A** | *study majors | take majors | Firstly, for example, in my university eligible students are given an opportunity to study double majors. | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 4 |
| **40** | **A** | *take a | make a decision | But that it is not true, a well known case is the way that one | 1 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 5 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | decision | | must take a decision in a game | | | | | | | | | | | | | |
| **41** | **A** | *try challenges | take challenges | You may be less kind to try new challenges because you feel that you may not be able to do it. | 0 | 0 | 1 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 5 |
| **42** | **A** | *try risks | take risks | The capacity to try new risks, new things and new challenges rather than only doing what you already do well. | 0 | 5 | 1 | 3 | 2 | 1 | 1 | 1 | 0 | 2 | 1 | 17 |
| **43** | **A** | *try ventures | take ventures | Only when you reach that level, than you can think of trying new ventures by exploring new business opportunities. | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 4 |
| **44** | **A** | *underline opinion | defend/underpin opinion | Another argument to underline my opinion is that you have better conversations with people by knowing many things briefly. | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| **45** | **A** | *use celebrities | hire celebrities | Another negative thing that influence adults as well as teenagers is using celebrities to promote a certain product. | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 2 | 0 | 6 |
| **46** | **B** | *adapt the society | adapt to the society | First of all, the goal of education makes a good person who is adapted the society. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 1 | 4 |
| **47** | **B** | *agree ideas | agree with ideas | In a few aspects, I agree this idea, but generally I don't agree this idea. | 1 | 0 | 0 | 0 | 0 | 0 | 20 | 7 | 11 | 0 | 1 | 40 |
| **48** | **B** | *agree sentences | agree with sentences | I agree this sentence.Young people every time enjoy life more than older people. | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 4 |
| **49** | **B** | *agree the opinion | agree with the opinion | I do not agree the opinion and think that people getting a specialized knowledge will be beneficial to our society as follow reasons. | 0 | 0 | 1 | 0 | 0 | 0 | 17 | 10 | 0 | 0 | 3 | 31 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **50** | **B** | *agree the statement | agree with the statement | I agree the statement "successful people try new thing and take risk rather than only doing what they already know how to do well" | 1 | 0 | 0 | 2 | 0 | 0 | 5 | 0 | 0 | 2 | 3 | 13 |
| **51** | **B** | *agree the topic | agree with the topic | The second reason that I agree the topic is that. | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 0 | 2 | 6 |
| **52** | **B** | *arrive the place | arrive at the place | All I need to do is taking train and MRT, I can arrive the place. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 6 | 7 |
| **53** | **B** | *care their community | care about their community | Such kind of act from the young ones shows that they care their community and they devote their precious time for helping their community. | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 2 | 0 | 1 | 7 |
| **54** | **B** | *concentrate subjects | concentrate on subjects | It is often said that knowing a variety of subjects is better than concentrating only one subject. | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 0 | 3 | 0 | 8 |
| **55** | **B** | *contract a guide | contract with a guide | We invited them to come with us and we contracted a local tour guide even if we both speaks Spanish. | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| **56** | **B** | *contribute their communities | contribute to their communities | The major problem arises today and it is that young people nowadays do not give sufficient time to help in contributing their communities. | 1 | 0 | 0 | 0 | 0 | 1 | 4 | 0 | 0 | 0 | 0 | 6 |
| **57** | **B** | *disagree the topic | disagree with the topic | Because of these reason, I disagree that topic. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 3 |
| **58** | **B** | *disagree the idea | disagree with the idea | I disagree the idea the test had given me. | 1 | 0 | 0 | 0 | 0 | 0 | 8 | 3 | 1 | 0 | 1 | 14 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **59** | **B** | *disagree the opinion | disagree with the opinion | I disagree the opinion "The best way to travel is in a group led by a tour guide." | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 4 | 0 | 0 | 0 | 9 |
| **60** | **B** | *go vacation | go on vacation | Most of the people preferred to go vacation with their personal cars. | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 4 |
| **61** | **B** | *go concerts | go to concerts | I could enjoy eating out at good restaurant, going precious classic concert, buying good quality clothing. | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 |
| **62** | **B** | *go Egypt | go to Egypt | If we wish to go Egypt and we don't have any information of Egypt, maybe we can not find the way to go Egypt with own our self. | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 3 |
| **63** | **B** | *go shopping malls | go to shopping malls | Because of that, I was able to see what I wanted to see; the beaches to surf, rather than going shopping malls, which the tour focused. | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 2 |
| **64** | **B** | *go the Tower | go to the Tower | I prepared information before i went Haha Tower. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 |
| **65** | **B** | *go the workplace | go to the workplace | Father needs his car to go his workplace. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 |
| **66** | **B** | *going a picnic | going on a picnic | Going a picnic is a really good enjoyable thing for a child but not for an old person because he or she thinks many thinks in his or her head. | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 3 |
| **67** | **B** | *graduate university | graduate from university | If they graduated university , they can have a job. | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 2 | 0 | 0 | 9 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **68** | **B** | *learn hardware | learn about the computer hardware | In the same way knowledge of computer hardware was not learnt by many but now people attend special classes to learn computer hardware. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| **69** | **B** | *listen a lecture | listen to a lecture | In a class room when many students are listening a particular lecture. | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 5 |
| **70** | **B** | *listen music | listen to music | People can read a book in the train or can also listen music with their MP3 player. | 0 | 0 | 3 | 1 | 0 | 0 | 1 | 2 | 8 | 4 | 1 | 20 |
| **71** | **B** | *listen their story | listen to their story | I notice that the number of young people in this place is so small and there are a lot of elderly people that might have the help of a person who can be stay with them listening their story. | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 3 |
| **72** | **B** | *look the advertisement | look at the advertisement | When I buy the car, I always look the advertisement which are listed a lot of them. | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 4 |
| **73** | **B** | *look the products | look at the products | Of course they will go to the shopping centr and look your products and usually this looking will finished with buying the product. | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 2 | 1 | 0 | 0 | 8 |
| **74** | **B** | *look the surface | look at the surface | First of all, it is to look only surface of stiffs. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 3 |
| **75** | **B** | *major subjects | major in subjects | For example, suppose there is a graduate who majored three subjects , History, Business, and Math. | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 0 | 0 | 0 | 4 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **76** | **B** | *participate activities | participate in activities | For example, most of my fellow students spend a lot of time on making internships, so there is no time left to voluntarily participate community activities. | 0 | 1 | 0 | 0 | 0 | 0 | 3 | 4 | 0 | 1 | 8 | 17 |
| **77** | **B** | *participate events | participate in events | For instance, my friend have never participated events of her communities, such as Japanese festival, instead of only seeing it. | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 3 |
| **78** | **B** | *prepare an advertisement | prepare for an advertisement | I mean that people, who study the way to present an article or a product and finally to prepare an advertisement. | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 3 |
| **79** | **B** | *prepare the travel | prepare for the trip | I would prepare the travel before to go travel. | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 3 | 0 | 0 | 0 | 5 |
| **80** | **B** | *search a job | search for a job | When you have studied business for example I think it does not make a lot of sense to search a job where you have to apply the whole knowledge you have collected during your studies. | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 5 |
| **81** | **B** | *search the information | search for the information | So for me, a good student will generally curious and in the most cases, he will search the information by his owns capabilities. | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 3 | 1 | 0 | 2 | 8 |
| **82** | **B** | *specialize subjects | specialize in subjects | I strongly believe that specializing one subject is more appropriate than knowing entire knowledge base. | 1 | 2 | 1 | 1 | 2 | 2 | 8 | 6 | 10 | 8 | 5 | 46 |
| **83** | **B** | *specialize themselves | specialize in | Other people prefers to specialize themselves in one specific subject. | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| **84** | **B** | *stay a hotel | stay at a hotel | I stayed a excellent hotel and rent a four-door car, which made | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 4 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | reservations by myself. | | | | | | | | | | | | | |
| **85** | **B** | *stay the place | stay at the place | We can stay the place as we want. | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 2 | 1 | 0 | 1 | 12 |
| **86** | **B** | *succeed the business | succeed in the business | Now most company recognize that advertisement products should be the best products than other one to succeed the business. | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 3 |
| **87** | **B** | *travel a park | travel to a park | You wants to travel a small park in Beijing. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 |
| **88** | **B** | *travel a tour | go on a tour | Therefore, when we travel and visit a lot of people, we have to have some risks as travel the tour . | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 2 |
| **89** | **B** | *travel Japan | travel to Japan | Last winter, I traveled Japan alone, and I studied the Japan to plan my trip. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 4 |
| **90** | **B** | *travel lands | travel to lands | Moreover, these burglars are mostly caught for the safety of lonely travelers who even wants to travel undeveloped lands of China. | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 3 |
| **91** | **B** | *travel places | travel to places | We have to travel more places to get more information. | 1 | 1 | 0 | 0 | 0 | 1 | 7 | 16 | 1 | 0 | 16 | 43 |
| **92** | **C** | *create advertising | create advertisement | Advertising is created to make people buy products or services, that's the only reason why companies spend so much money advertising. | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| **93** | **C** | *earn their life | earn money | They earn their life in their own so they are able to plan it better than a young people because they do not have to rely onto their parents. | 0 | 0 | 4 | 0 | 0 | 2 | 1 | 0 | 4 | 0 | 0 | 11 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **94** | **C** | *help neighborhood | help neighbors | In my hometown, students only help their neighborhood when their schools ask them to do that. | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 2 | 5 |
| **95** | **C** | *increase selling | increase sales | Most of new product appear with their own advertising in order to announce the people product's information and get people's attention to increase selling. | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 5 |
| **96** | **C** | *increase the sell | increase the sales | This advertisement company will increase the sell and the consumers will not notice the bad quality. | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 5 |
| **97** | **C** | *know the thinks | know the thoughts | He must know more think as possible about the activity that he has chosen. | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| **98** | **C** | *learn the intuition | learn the fact | I mean that after understanding the concepts or ideas the other important process should be applied which is learning the intuition behind particular issue. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 2 |
| **99** | **C** | *learn thinks | learn thoughts | everyone can improve these skills and learn new thinks in order to achieve his aims. | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 3 |
| **100** | **C** | *manipulate customers | manipulate customers' desire | To manipulate the customer and giving him the illusion life cannot continue without buying this product is a very common procedure nowadays. | 0 | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| **101** | **C** | * outnumber the decrease | outnumber the (number of cars…) | During the next twenty years the increase of the number of cars in Asia will outnumber the decrease of the number of cars in Europe and America, | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 4 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **102** | **C** | *reach the destiny | reach the destination | When a person wants to learn the facts first he has to have some ideas how he can come to the end point .first of all he has to make clear about the concepts he wants to include to reach that destiny. | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 | 0 | 5 |
| **103** | **C** | *take a travel | take a trip | Such people tend to think that they feel more comfortable taking a travel alone as they do not have to follow the schedules and intentions of others. | 1 | 1 | 0 | 0 | 1 | 0 | 2 | 6 | 0 | 0 | 4 | 15 |
| **104** | **C** | *watch advertising | watch advertisement | According to this reasons, we can think about that when we read newspaper and watch the advertising , we need more caution to care the fact. | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 4 |
| **105** | **C** | *try thinks | try thoughts | In my opinion a successful person is someone who takes risks and tries new thinks. | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 3 |
| **106** | **D** | *have a travel | take a trip | I would like to have a travel with full of unpredicted events. | 3 | 0 | 0 | 0 | 3 | 0 | 7 | 8 | 1 | 0 | 8 | 30 |
| **107** | **D** | *conclude ideas | make conclusions | this is carried out in such a way to be able to perform certain actions and conclude ideas. | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 2 | 0 | 7 |
| **108** | **D** | *conclude the fact | make conclusions | Students gained the ability to conclude the facts themselves. | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 4 |
| **109** | **D** | *ride a ski | go skiing | About 3 years ago, I decide to learn snow boarding since I only know how 4to ride ski. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 3 |
| **Total frequency** | | | | | 27 | 48 | 47 | 34 | 44 | 46 | 154 | 143 | 86 | 97 | 81 | 807 |

# APPENDIX 2

## Errors Made by Japanese, Korean or Both leaners

| | Languages | Miscollocations | Suggested forms | JPN | KOR | Learner corpus Frequency |
|---|---|---|---|---|---|---|
| **1** | BOTH | *agree the opinion | agree with the opinion | 17 | 10 | 27 |
| **2** | BOTH | *agree ideas | agree with ideas | 20 | 7 | 27 |
| **3** | BOTH | *travel places | travel to places | 7 | 16 | 23 |
| **4** | BOTH | *have a travel | take a trip | 7 | 8 | 15 |
| **5** | BOTH | *specialize subjects | specialize in subjects | 8 | 6 | 14 |
| **6** | BOTH | *disagree the idea | disagree with the idea | 8 | 3 | 11 |
| **7** | BOTH | *stay the place | stay at the place | 8 | 2 | 10 |
| **8** | BOTH | *take a travel | take a trip | 2 | 6 | 8 |
| **9** | BOTH | *disagree the opinion | disagree with the opinion | 4 | 4 | 8 |
| **10** | BOTH | *graduate university | graduate from university | 3 | 4 | 7 |
| **11** | BOTH | *participate activities | participate in activities | 3 | 4 | 7 |
| **12** | BOTH | *look the products | look at the products | 4 | 2 | 6 |
| **13** | BOTH | *concentrate subjects | concentrate on subjects | 1 | 4 | 5 |
| **14** | BOTH | *major subjects | major in subjects | 1 | 3 | 4 |
| **15** | BOTH | *prepare the travel | prepare for the trip | 1 | 3 | 4 |
| **16** | BOTH | *face accidents | have accidents | 2 | 2 | 4 |
| **17** | BOTH | *invent the energy | find the energy | 1 | 2 | 3 |
| **18** | BOTH | *agree the topic | agree with the topic | 1 | 2 | 3 |
| **19** | BOTH | *listen music | listen to music | 1 | 2 | 3 |
| **20** | BOTH | *help neighborhood | help neighbors | 1 | 2 | 3 |
| **21** | BOTH | *loose the chance | lose the chance | 2 | 1 | 3 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **22** | BOTH | *succeed the business | succeed in the business | 2 | 1 | 3 |
| **23** | BOTH | *earn knowledge | acquire knowledge | 1 | 1 | 2 |
| **24** | BOTH | *try risks | take risks | 1 | 1 | 2 |
| **25** | BOTH | *use celebrities | hire celebrities | 1 | 1 | 2 |
| **26** | BOTH | *agree sentences | agree with sentences | 1 | 1 | 2 |
| **27** | BOTH | *go vacation | go on vacation | 1 | 1 | 2 |
| **28** | BOTH | *go concerts | go to concerts | 1 | 1 | 2 |
| **29** | BOTH | *participate events | participate in events | 1 | 1 | 2 |
| **30** | JPN | *retire their jobs | retire /quit their jobs | 14 | 0 | 14 |
| **31** | JPN | *agree the statement | agree with the statement | 5 | 0 | 5 |
| **32** | JPN | *contribute their communities | contribute to their communities | 4 | 0 | 4 |
| **33** | JPN | *stay a hotel | stay at a hotel | 4 | 0 | 4 |
| **34** | JPN | *meet accidents | have accidents | 2 | 0 | 2 |
| **35** | JPN | *go Egypt | go to Egypt | 2 | 0 | 2 |
| **36** | JPN | *look the advertisement | look at the advertisement | 2 | 0 | 2 |
| **37** | JPN | *loose money | lose money | 1 | 0 | 1 |
| **38** | JPN | *loose the ability | lose the ability | 1 | 0 | 1 |
| **39** | JPN | *loose time | waste time | 1 | 0 | 1 |
| **40** | JPN | *loose your interest | lose your interest | 1 | 0 | 1 |
| **41** | JPN | *reach happiness | achieve happiness | 1 | 0 | 1 |
| **42** | JPN | *go shopping malls | go to shopping malls | 1 | 0 | 1 |
| **43** | JPN | *listen a lecture | listen to a lecture | 1 | 0 | 1 |
| **44** | JPN | *travel a tour | go on a tour | 1 | 0 | 1 |
| **45** | JPN | *earn their life | earn money | 1 | 0 | 1 |
| **46** | JPN | *conclude ideas | make conclusions | 1 | 0 | 1 |
| **47** | KOR | *travel Japan | travel to Japan | 0 | 4 | 4 |
| **48** | KOR | *ride a ski | go skiing | 0 | 3 | 3 |
| **49** | KOR | *look the surface | look at the surface | 0 | 3 | 3 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **50** | KOR | *search the information | search for the information | 0 | 3 | 3 |
| **51** | KOR | *do mistakes | make mistakes | 0 | 2 | 2 |
| **52** | KOR | *experiment things | experience things | 0 | 2 | 2 |
| **53** | KOR | *adapt the society | adapt to the society | 0 | 2 | 2 |
| **54** | KOR | *disagree the topic | disagree with the topic | 0 | 2 | 2 |
| **55** | KOR | *go the Tower | go to the Tower | 0 | 2 | 2 |
| **56** | KOR | *go the workplace | go to the workplace | 0 | 2 | 2 |
| **57** | KOR | *increase selling | increase sales | 0 | 2 | 2 |
| **58** | KOR | *watch advertising | watch advertisement | 0 | 2 | 2 |
| **59** | KOR | *do adventure | experience adventure | 0 | 1 | 1 |
| **60** | KOR | *earn the facts | learn the facts | 0 | 1 | 1 |
| **61** | KOR | *loose clients | lose clients | 0 | 1 | 1 |
| **62** | KOR | *loose something | lose something | 0 | 1 | 1 |
| **63** | KOR | *reach the dream | achieve the dream | 0 | 1 | 1 |
| **64** | KOR | *study majors | take majors | 0 | 1 | 1 |
| **65** | KOR | *try challenges | take challenges | 0 | 1 | 1 |
| **66** | KOR | *try ventures | take ventures | 0 | 1 | 1 |
| **67** | KOR | *arrive the place | arrive at the place | 0 | 1 | 1 |
| **68** | KOR | *contract a guide | contract with a guide | 0 | 1 | 1 |
| **69** | KOR | *going a picnic | going on a picnic | 0 | 1 | 1 |
| **70** | KOR | *listen their story | listen to their story | 0 | 1 | 1 |
| **71** | KOR | *travel lands | travel to lands | 0 | 1 | 1 |