A Comparative Corpus–Based Analysis of Lexical Collocations used in EFL textbooks

Ahmad MOLAVI
Islamic Azad University, Khorasgan Branch
(Isfahan, Iran)

Mansour Koosha
Islamic Azad University, Khorasgan Branch
(Isfahan, Iran)

Hossein Hosseini
Islamic Azad University, Khorasgan Branch
(Isfahan, Iran)

Abstract
As collocations play an important role in second language learning, especially at intermediate-advanced levels, the present study not only seeks to examine distribution of lexical collocations in three selected series of general English textbooks, Interchange Third Edition by (Jack C. Richards and Jonathan Hull, 2005), American Headway by (Liz and John Soars, 2003), American File by (Clive Oxenden and Christina Latham Koeing, 2008), through analyzing face to face and telephone conversation scripts collected from intermediate and upper-intermediate level books of the aforementioned series, but also employs Open American National Corpus (OANC) available online and AntConc 3.2.1 concordancer program to compare lexical collocations from textbook to their real use by native speakers. The findings of present research suggested that special attention has been paid to specific types of lexical collocations noun + verb and adjective + noun while the frequency of collocations in series could not be affective on learners collocations learning and on the other hand comparing textbooks collocations to reference corpus (OANC) showed choice of collocations in these series did not have big refers to real use of language by native speakers.

Key Words: collocation; lexical collocation; corpus.

Resumen
Considerando que las colocaciones juegan un papel importante en el aprendizaje de un segundo idioma, especialmente en los niveles intermedio-avanzado, el presente estudio no sólo tiene por objeto examinar la distribución de las colocaciones léxicas en tres series seleccionadas de libros de texto de inglés general, Interchange Third Edition escrito por Jack C. Richards and Jonathan Hull, (2005), American Headway por Liz and John Soars, (2003), American File por Clive Oxenden and Christina Latham Koeing (2008) por medio del análisis de guiones de conversación cara a cara y por vía telefónica, recogidas de libros de nivel intermedio y superior-intermedio de las serie mencionadas, sino también emplea Open American National Corpus (OANC) disponible en línea y AntConc 3.2.1 el cual es un programa de concordancias para comparar colocaciones léxicas encontradas en los textos con su uso real por hablantes nativos. Los resultados del presente estudio sugieren que se ha prestado una atención especial a los tipos específicos de colocaciones léxicas nombre+verbo y adjetivo+nombre mientras que la frecuencia de colocaciones en serie puede no afectar el proceso de aprendizaje de colocaciones; por otro lado la comparación de colocaciones existentes en textos con referencia al corpus (OANC) mostró que la selección de colocaciones en estas series no era un referente significativo al uso real del idioma por hablantes nativos.

Palabras Claves: colocación; colocación lexic; corpus.
INTRODUCTION

Learning vocabulary was seen as a key element to achieve a high level of proficiency in language learning by a great number of theoreticians (Boers & Lindstromberg, 2008; Zu, 2009; Schmitt, 2000). It is believed that having large and varied vocabulary knowledge is the indicator of communicative competence and learning vocabulary should be considered as an integral part of learning a foreign or second language since it leads the way to communication (McCrostie, 2007). It’s a widely accepted idea that collocations are very important part of knowledge of second language acquisition and they are essential to non-native speakers of English in order to speak or write fluently and accurately (Jaén, 2007). Skrzypek (2009) indicates the importance of collocation by stating that one of the criteria for knowing a word is being aware of other words with which it keeps company. Shin and Nation (2008), also explain one of the reasons as to why teachers and learners should be interested in collocations is that collocations improve learners’ language fluency and ensure native-like selection. The term collocation has been labeled in a variety of ways, e.g., prefabs, multi-wordunits, and has been defined in many ways (Moon, 1998). Some of these definitions are as follows.

Collocations are words that occur together with high frequency and refer to the combination of words that have a certain mutual expectancy. “The combination is not a fixed expression but there is a greater than chance likelihood that the words will co-occur” (Jackson, 1988, p. 96). Also McCarten (2007, p. 5) states that the way in which two or more words are typically used is generally called collocation. Stubbs (2002, p. 215) defines collocation as the habitual co-occurrence of two unordered content words, or of a content word and a lexical set.

Although collocations seem to be vital in daily communications, most EFL/ESL learners, even at advanced levels, have various problems in their oral and written products. This is in spite of the fact that they apparently have sufficient access to lexical or grammatical knowledge. Erroneous utterances like ‘the manager of university’, heavy tea’, ‘to be bad in something’ are not due to poor lexical or grammatical knowledge and arise partly from the lack of knowledge about the companies that words keep. Bahns and Eldaw (1999), based on the results of their own studies, argue that EFL/ESL learners face relatively greater difficulty with lexical collocations rather than grammatical ones. This difficulty has been discussed from different aspects and according to researchers...
like, Tajalli (1994) and Gitaski (1999), lack of awareness of collocational patterns and insufficient exposure to these patterns account for learners' difficulties in processing and producing appropriate collocational patterns.

**Statement of problem**

Collocation runs through the whole of the English language. No piece of natural spoken or written English is totally free of collocation. For the student, choosing the right collocation will make his or her speech and writing sound much more natural, more native-like, even when basic intelligibility does not seem to be at stake. A student who talks about strong rain may make himself or herself understood, but it requires more effort on the part of the listener and ultimately creates a barrier to communication. But, perhaps even more importantly than this, language that is collocationally rich is also more precise. This is because most single words in the English language embrace a whole range of meanings, some quite distinct, and some that shade into each other by degrees. A student who chooses the best collocation will express him or herself much more clearly and be able to convey not just a general meaning, but something more precise (Oxford Collocations Dictionary, 2009).

Taking insufficient exposure to the problematic patterns as an effective reason for ESL/EFL learners' difficulties in gaining collocational competence (Wray, 1999), and considering the role of ELT textbooks as the main source of teachers output and learners input in classrooms present study aims to examine three series of general English textbooks with the focus on implication of Lexical collocations.

**Objectives**

The current study by investigating, face to face and telephone conversations from three series of general English textbooks seeks to give comprehensible profile over implication of Lexical collocations by these series. So the present study aims to seek the following objectives:

1. Study the types and frequency of the usage of lexical collocations which are presented in the target ELT textbooks series under study.
2. Look for the similarities and differences in the representation of lexical collocations in the selected coursebooks.
3. Compare the three selected ELT textbook series compare with American National Corpus in terms of Lexical Collocations.

**Research questions**

The current study attempted to address the following questions:

1. What types of lexical collocations are presented in the target ELT textbooks series under study?
2. Are the three selected ELT textbook series similar or different in the representation of Lexical collocations?
3. How frequent are different Lexical collocations used in three ELT textbooks series?
4. How do the three selected ELT textbook series compare to American National Corpus in terms of Lexical Collocations?

**Review of Related Literature**

Zahedi et al. (2010) have emphasized the importance of drawing second language learners’ attention to standard multiword expressions such as collocations and idioms. In other studies done in the field of lexical collocation (Hsu, 2010) and Mounya, (2010) point to the impact of lexical collocation awareness on other specific aspects of general English proficiency. Nakata (2006) mentioned that learners have to acquire a large number of collocations to be able to produce and comprehend ideas accurately, fluently and naturally. Another study that showed the difficulties language learners encounter learning English collocations is Koosha and Jafarpour’s (2006) study The research was done on 200 Iranian university students (in three universities in Shahrekord) majoring in English. The subjects were then randomly divided into two groups. One group underwent the conventional treatment on prepositions and their patterns in which preposition and their collocational patterns were explicitly thought to the participants in English or Farsi. The second group (experimental group) received a data driven based instruction (treatment) that was based on concordancing lines presented in KWIC format. Two completion tasks on collocation of prepositions were administered as yielded the following results. First, DDL approach proved to be highly effective in teaching and learning of collocation of preposition. Second learners’ performance on collocation of preposition was shown to be positively related to their level of proficiency. Third, the analysis of errors of collocations used in EFL textbooks. *Latin American Journal of Content and Language Integrated Learning*, 7(1), 66-81. doi:10.5294/laclil.2014.7.1.4 eISSN 2322-9721
indicated that Iranian EFL learners tended to carry over their L1 collocational patterns to their L2 production. Generally results revealed that the Iranian students lacked collocation knowledge. Therefore it seems that collocations are L2 learners’ serious problem and this problem tends to be solved.

In another study Koya (2005), compared verb + noun collocations among six English I textbooks compiled as all the same level for Japanese upper secondary schools and examine how collocations from those which are frequently used in CoBuild. Koya (2004) also compared verb-noun collocations extracted from history textbooks for secondary school pupils in the UK to collocations from textbooks for 10th graders in Japan. Chamnan (2004) examined high frequency content words and their collocations in two civil engineering journals by using a corpus.

Nesselhauf (2003) explored the use of verb–noun collocations such as take a break and shake one’s head in essays written by advanced German-speaking learners of English. Kennedy (1998) in his book An Introduction to Corpus Linguistics defines corpus as: "a body of written text or transcribed speech which can serves as a basis for linguistic analysis and description" (, p. 21).


Zimmerman (1993) did a study in which she asked 12 native-speaking ESL teachers to identify, categorize and explain lexical anomalies in each of 14 sentences. Nattinger and Decarrico (1992) set up three criteria to distinguish between idioms, collocations and free-combinations. He considers them as a continuum with completely invariant clusters at one end, freely combining morphemes at the other end with all degrees of combinational flexibility in between. They accept semantic criteria of compositionality which means the meaning of the combination is predictable from each constituent.

According to John Sinclair (1991), a man seen by many as a pioneer in the field of corpus linguistics thirty years ago when corpus emerged it was impossible to process texts of several million words twenty years ago it was considered marginally possible but lunatic ten years ago it was considered quite
possible but still lunatic while today it is very popular. Benson (1990) defines collocation as the occurrence of two or more words within a short span of each other, and at the same time, McCarthy (1990) highlights the importance of collocation in language teaching and argues that collocation is a marriage contract between words and it forms an important organizing principle in the vocabulary of any language. Palmer (1983) reported that the term “collocation” was already used as a technical but rather vague term in linguistics field. He defined collocation as “a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts”.

In the 1970s and 1980s, studies on the acquisition and use of different types of frequent word combinations have drawn on a number of research methods: ethnography (Fillmore, 1979; Hakuta, 1974; Peters, 1983), conversational analysis (Manes & Wolfson, 1981; Tannen, 1987) and quantitative research on the use of multi-word expressions (Altenberg, 1993; Kjellmer, 1991; Sinclair, 1991), among others.

**METHODOLOGY**

This study focused on the analysis of the occurrence of Lexical collocation entries in textbooks described above. The process of data collection in present study includes counting, grouping, and further recording collocational units from the selected textbooks. This was done manually through the following steps:

1. The transcripts in each series were examined manually in order to identify the Lexical collocations.

2. After going through all transcripts in each series and underlining all possible lexical collocations, we confirmed the collocational status by consulting two dictionaries of collocation: *The Longman Contemporary 5th Edition DVD-Rom with 75000 collocations* and *Oxford collocations dictionary for students of English* with its 170,000 collocation entries mainly selected from the 100-million-word British National Corpus. Most importantly, these two collocational dictionaries exclude free combinations and fixed idioms, which is essential to keep the focus on collocations rather than other kinds of multiple-word items. Only collocations found at least in one of the two dictionaries were counted for this study.

3. Each identified collocational pattern was sorted into their corresponding categories. To obtain the frequency of collocational patterns in ELT
textbooks and compare them to (OANC) corpus the following steps were carried out:

a. Collected transcripts from each series were scanned and converted into electronic files.

b. The AntConc 3.2.1 was applied to figure out the frequency of occurrence of those legitimate lexical collocations in textbook series and spoken register of (OANC). The collocational patterns consist of all variations of the verb, adjective, adverb and the noun ignoring determiners, such as a, or the, and possessives, such as his, or my.

Data analysis

As was discussed, the present study tried to investigate the frequency of collocational patterns in different textbooks selected as the materials of the research. So in order to address the formulated questions, AntConc 3.2.1 computer concordance program was employed. This free text analysis computer program can identify word combinations of different lengths and frequencies in small or large corpora, so collected collocations from each series were examined in related electronic texts.

The collected data from the selected course books were ranked based on their frequency in corresponding categories and then occurrence of collocations with the same frequency was treated in percentage to show how different or similar are series in repetition of collocations. To make a comparison between textbooks corpora and reference corpus possible, the actual frequency of collocations was normalized in one thousand and the results presented in tabular form for clarity and ease of access. Furthermore descriptive analysis used to interpret the results. The results of categorizing collocations from series, as seen in table 1, revealed that series have similar emphasis on particular patterns such as (verb + noun and adjective + noun) and paying less attention to patterns such as n1 + of + n2, verb + adverb, noun + verb and adverb + adjective).
Table 1. Types and number of Lexical Collocations in three ELT textbook series.

<table>
<thead>
<tr>
<th>Books</th>
<th>V + N</th>
<th>Adj + N</th>
<th>N + V</th>
<th>N1 + of + N2</th>
<th>Adv + Adj</th>
<th>V + Adv</th>
<th>Total</th>
<th>Volume means</th>
<th>Units means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interchange 2 &amp; 3</td>
<td>89</td>
<td>34</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>137</td>
<td>68.5</td>
<td>4.2</td>
</tr>
<tr>
<td>American Headway 2 &amp; 3</td>
<td>75</td>
<td>32</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>116</td>
<td>58</td>
<td>4.3</td>
</tr>
<tr>
<td>American File 2 &amp; 3</td>
<td>71</td>
<td>26</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>109</td>
<td>54.5</td>
<td>6.8</td>
</tr>
</tbody>
</table>

As table 1 above shows, Interchange series not only includes the highest number of collocations 137 among three series it also included all six types of lexical collocations although the number of collocations in two types, (verb + adverb) and (N1 + of + N2), is very low 2 and 1 respectively. According to above table the total number of collocations in two other series, American Headway and American File, is 116 and 109 respectively and in both series two types of collocations, (verb + adverb) and (noun + verb) are neglected. Among six categories of Lexical collocations (verb + noun) and (adjective + noun) in all three series are respectively the first and second type in terms of number. The volume mean and units mean scores in the table show that Interchange series included the highest number of collocations per book and American File series takes the highest number of collocations per unit. In addition, distribution of different types of lexical collocations in series was demonstrated in figure 1.
On the other hand, just eight collocations were common in all three series (get job, get married, make decision, make friends, spend time, give lesson, good idea and wrong number) and most collocations such as chewing gum, take care, take photo just appeared in one series, indicating that there is no high agreement over presentation of specific collocations. Table 2 shows comparison of common collocations to reference corpus in terms of frequency. Although Interchange series includes more collocations compared with American Headway and American File series, an obvious harmony can be found among them while all of them put clear emphasis on including two types of collocations, (verb + noun) and (adjective + noun), much more than the other types.

The results of categorizing collocations from series, as seen in table 1, revealed that the series have similar emphasize on particular patterns such as (verb + noun and adjective + noun) and paying less attention to patterns such as n1 + of + n2, verb + adverb, noun + verb and adverb + adjective). On the other hand, just eight collocations were common in all three series (get job, get married, make decision, make friends, spend time, give lesson, good idea and wrong number) and most collocations such as chewing gum, take care, take photo just appeared in one series, indicating that there is no high agreement.
over presentation of specific collocations. Table 2 shows comparison of common collocations to reference corpus in terms of frequency.

**Table 2. frequency of common collocations in textbook corpora and reference corpus (No. per one thousand)**

<table>
<thead>
<tr>
<th>Collocations</th>
<th>Reference corpus</th>
<th>Interchange</th>
<th>American headway</th>
<th>American File</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good idea</td>
<td>237 (1.1)</td>
<td>9 (0.4)</td>
<td>5 (0.1)</td>
<td>4 (0.1)</td>
</tr>
<tr>
<td>Get married</td>
<td>217 (1)</td>
<td>6 (0.2)</td>
<td>9 (0.3)</td>
<td>6 (0.2)</td>
</tr>
<tr>
<td>Spend time</td>
<td>79 (0.4)</td>
<td>4 (0.1)</td>
<td>1 (0.03)</td>
<td>2 (0.08)</td>
</tr>
<tr>
<td>Get a job</td>
<td>62 (0.3)</td>
<td>6 (0.2)</td>
<td>3 (0.1)</td>
<td>5 (0.2)</td>
</tr>
<tr>
<td>Make decision</td>
<td>21 (0.1)</td>
<td>3 (0.1)</td>
<td>2 (0.07)</td>
<td>2 (0.08)</td>
</tr>
<tr>
<td>Make friends</td>
<td>5 (0.02)</td>
<td>3 (0.1)</td>
<td>2 (0.07)</td>
<td>1 (0.04)</td>
</tr>
<tr>
<td>Wrong number</td>
<td>1 (0)</td>
<td>2 (0.09)</td>
<td>1 (0.03)</td>
<td>1 (0.04)</td>
</tr>
<tr>
<td>Give lesson</td>
<td>0 (0)</td>
<td>1 (0.04)</td>
<td>1 (0.03)</td>
<td>1 (0.04)</td>
</tr>
<tr>
<td>Total</td>
<td>622 (2.9)</td>
<td>34 (1.2)</td>
<td>24 (0.7)</td>
<td>22 (0.7)</td>
</tr>
</tbody>
</table>

As table 2 shows, most frequent common collocations among textbook corpora (good idea, get married), also have a high frequency in reference corpus and lowest frequent collocation (give lesson) could not be found in reference corpus at all. Table 3 represents collocations repetition in series calculated by percentage.

**Table 3. Descriptive statistic of collocations repetition in textbooks.**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Interchange</th>
<th>American Headway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Once</td>
<td>76 [55.4%]</td>
<td>76 [65.5%]</td>
</tr>
<tr>
<td>Twice to five times</td>
<td>55 [40%]</td>
<td>46 [42.2%]</td>
</tr>
<tr>
<td>More than five times</td>
<td>6 [4.3%]</td>
<td>4 [3.6%]</td>
</tr>
</tbody>
</table>

**DISCUSSION**

To address the first question in this study textbooks were investigated manually and collected collocations were assigned into appropriate categories. The results revealed that two types of lexical collocations, *verb + noun and adjective + noun*, were much more common in series than other types. The findings are in tune with those from Lewis (2000), Gitsaki (1999), and Bahns & Eldaw (1993) which introduced *verb + noun* collocations as the most problematic type to ESL/EFL learners. Present study also confirmed findings of the study by Koya (2004) on Japanese students collocation learning that claimed typical collocations should be
presented from the beginning of second language acquisition especially the \((\text{verb} + \text{noun})\) type which can be learned by students with intermediate level of proficiency. In terms of collocations frequency, AntConc 3.2.1 computer program was employed to investigate the frequency of individual collocations in series. What the analysis of data shows is that over 50% of collocations are repeated just once, 31% - 42% occurs twice to five times and 2.5% - 4.3% with over five time occurrences. Actually the results of this study is comparable to those of single-word vocabulary counts done by Chang (2002) for locally produced English textbooks in Taiwan, and Nation (1990) for ELT textbooks worldwide and against strong recommendations of scholars like Conzett (2000), Gitsaki (1999), Harwood (2002), and Lewis (1997) which said collocations should be recycled regularly and systematically in teaching materials for recognition and production purposes to help learners integrate them into their mental lexicon. What this low frequency of collocations in selected course books implies is that textbook writers do not pay much attention to the recycling of collocations and the studied course books in this research are not reliable sources in directing students' attention toward collocational patterns and cannot have effective influence on increasing learners' collocational knowledge, while a substantial number of linguists have proven that applying corpus linguistic techniques has a significant role in teaching ESL/EFL and in materials development (Hunston, 2005; O’Keefe, McCarthy, & Carter, 2007).

The present study also compared textbooks corpora to reference corpus (OANC) in terms of lexical collocations and findings show that although high frequent collocations in series have high frequency in reference corpus, according to normalized frequencies of collocations, textbooks do not refer to actual use of language by native speakers in choosing and presenting collocations. The findings of this comparison were in tune with Biber & Reppen, (2002) and Gilmore (2004), who explored the difference between the information presented in ESL/EFL teaching materials and the information regarding actual language use, their corpus-based frequency findings showed a serious discrepancy between six popular ESL/EFL textbooks and real world language while Maddalena (2001) argued use of authentic and real-life examples with L2 learners is more beneficial than examples that are made up by the teacher and do not simulate real life use of language.
To conclude it can be mentioned that while textbooks play the role as the main source of language input in classrooms, the results of this study, that is low number of frequent collocations and low referral to real use of language by native speakers show that these series cannot play an effective role in making collocations part of learner actual competence. Furthermore, this research aimed at investigating three popular English textbooks series in terms of presentation and distribution of lexical collocations and their reference to real use of lexical collocations by native speakers in reference corpus (OANC). The findings of present research suggested that especial attention has been paid to specific types of lexical collocations noun + verb and adjective + noun while the frequency of collocations in series could not be affective on learners collocations learning and on the other hand comparing textbooks collocations to reference corpus (OANC) showed choice of collocations in these series did not have big refers to real use of language by native speakers. Therefore, it can also be concluded that improving learners' collocational knowledge just by focusing on these textbooks is impossible and use of supplementary materials to compensate textbooks deficiency and learners' difficulties is necessary.

REFERENCES


**BIODATA**

Ahmad MOLAVI is a Ph.D student of TEFL in Islamic Azad University, Khorasgan branch, Isfahan, Iran. He is currently teaching in Azad universities and Islamic Propagation Office in Isfahan, Iran. He is interested in teaching methodology and second language acquisition (SLA) studies, psycholinguistics and sociolinguistics as well.

Dr. Mansour KOOSHA is a faculty member of English Language Teaching Department, Islamic Azad University, Khorasgan (Isfahan) Branch. He is an associate professor at Khorasgan branch. He holds a Ph.D in applied linguistics. Dr. Koosha’s interests are language testing, research methodology, linguistics, material development and teaching methodology. He has published a lot of articles and books. Dr. Koosha has participated in several seminars and conferences and is an experienced professor of English.

Hossein HOSSEINI is a Ph.D student of TEFL in Islamic Azad University, Khorasgan branch, Isfahan, Iran. He is currently teaching in Azad university of Khorasgan and Language Institutes in Isfahan, Iran. He is interested in teaching methodology, second language acquisition (SLA) studies, and Discourse analysis as well.