

## A CORPUS-BASED ANALYSIS OF LEXICAL COLLOCATION IN CONVERSATION TEXTS OF ENGLISH TEXTBOOKS

**Prisila Duana Intamanjana, Clary Sada**

Masters Study Program of English Language Education Teacher Training and  
Education Faculty Tanjungpura University Pontianak

Email: [tataduana@gmail.com](mailto:tataduana@gmail.com)

### *Abstract*

*This study revealed the analysis of lexical collocation in conversation texts of curriculum-based English textbook entitled “Bahasa Inggris” for tenth-grade students. In the textbook, it was found that there were 46 lexical collocations of VN, AdjN, NN, AdvAdj, and VAdv combination. Five out of seven types of lexical collocation combinations were presents in the conversation texts. The MI (Mutual Information) score of the collocation combinations were mostly above 3.00 which considered ‘strong collocation’. Semantically, the collocations were dominated by ‘free combination’ followed by ‘restricted collocation’ then figurative idiom’. The result also showed the lack of repetition of collocation combination in the conversation texts. Moreover, there was an absence of ‘pure idioms’ as one of the part that construct a conversation*

**Key Words:** *conversation, corpus, lexical collocation*

### INTRODUCTION

For senior high school EFL students, the primary source of materials for learning English in the classroom is textbooks. The textbook provide selected materials guided by the list of basic competence which based on the present curriculum. It is accessible to all students and teachers especially the books that are funded by the government. Presently, the government funded English textbook for senior high school used widely in Indonesia is titled ‘Bahasa Inggris’ based on Curriculum 2013.

Generally, the content of the English textbooks for EFL students in Indonesia mainly consist of texts, language rules, and exercises. The exercises in the book are designed to improve the students’ productive skills. It encourages the students to perform the language in written and spoken form individually and in group. These activities are placed in the final part of the chapter. It means that the first part of the chapter is to prepare the students.

To prepare the students, the textbook provides activities from reading materials, and the information of the language rules. The aim for these activities is to improve the students’ background knowledge on the topics so they can produce the language accurately and fluently. One of the knowledge that the students need to have in order to achieve fluency is the knowledge on collocation.

The definition of collocation according to Firth as stated in Hill (2000) “the way words combine in predictable ways” (p 48). Based on the definition, collocation is described as word combinations which habitually appear together making the combinations predictable. In other words, collocations are the word

combinations that are created because the native speakers of English constantly use the word combinations in their everyday interaction, written and spoken. Therefore, the knowledge on collocation can be a significant factor that can improve the students' fluency in English. Collocation combinations by nature are arbitrary. There are no specific rules that restrict the way the word combine. The record of the native speaker interactions in spoken and written form become reliable source for extracting collocation combinations.

Being described as arbitrary collocation combinations often created from the informal situation such as conversation. In "Bahasa Inggris" textbook, the conversation texts are mostly focus on informal everyday topics such as school life, social, and self (Oktavianti et al., 2020). From the informal situations of the conversations there are possibilities of collocation combination to be used. Moreover, the contexts of the conversation texts in the textbook are clear and contained the expressions that the students are supposed to acquire in the chapter. The previous research on lexical collocation in descriptive, recount, and narrative texts of "Bahasa Inggris" textbooks for tenth-grade students shows that narrative text contained the most collocation, followed by recount then descriptive text (Hutapea et al., 2019). The result of the research shows the possibility of collocation combination exist in conversation text because the characteristic of the narrative text which is consists of narrations accompanied by dialogue shows that collocation mostly used in informal situations similar to recount texts. Meanwhile a more technical text such as descriptive seems to contain less collocation.

There are four classification of collocation based on Cowie which listed by Nesselhauf, (2004), they are "free combination, restricted collocation, figurative idioms, and pure idioms" (pp. 14–15). This classification is based on the semantic features of the word combination. Further, Nesselhauf explained the criteria of each types of the collocation classification. Free combinations are the word combination where each word has literal meaning. Restricted collocation is the collocation combination where at least one of the words interpreted figuratively. Figurative idioms are the collocation where the words have figurative meaning however it still can be interpreted by referring it to the context of the conversation. Pure idiom are a combination of words that purely consist of figurative meaning which is cannot be deciphered without memorizing its meaning and the word combination in pure idioms cannot be modified. Another classification of collocations is based on their syntactic features. Wang & Good (2007) listed seven types of lexical collocation based on BBI Combinatory Dictionary they are; L1 (VN Creation Activation), L2 (VN Eradication Nullification), L3 (AdjN and NN), L4 (NV), L5 (NofN), L6 (AdvAdj), and L7 (VAdv)

Based on the content of the English textbook in Indonesia, it can be said that the EFL students in Indonesia are not familiar with collocation. Vocabulary section in most English textbook are consist of single word translation exercise which became redundant because there is already glossary section in the book. Moreover, the lack of awareness on collocation and the ignorance of the importance of collocation can cause a negative transfer from Indonesian to English (Said, 2011). Based on the importance of the knowledge in collocation, this research aimed to expose the lexical collocation combination in the conversation text of 'Bahasa Inggris' for tenth-grade students.

## METHODOLOGY

This study analyzed the conversation texts of Indonesian EFL textbook namely *Bahasa Inggris* for tenth grade students. There were nine conversation texts with seven topics (Self introduction, expressing congratulation, giving compliments and giving advice, future plan, interview, recount and admired person). In order to get an accurate data, these conversation texts were digitalized and compiled into a pedagogic corpus which then analyzed using concordancer tool AntConc version 3.5.9 (Anthony, 2020). The data were collected by identifying all possible collocation combination from the conversation texts. Two collocation dictionaries namely BBI Combinatory Dictionary of English (Benson et al., 2009) and Oxford Collocation Dictionary for Students of English (McIntosh et al., 2002) were used to verify whether the word combination was collocation or not.

The strength of the collocation combinations were measured by its Mutual Information (MI) score. The MI score of the collocation were determined by the number of co-occurrence in the reference corpus. Collocation combinations with the MI score that reach 3.00 were considered strong (Szudarski, 2017). The reference corpus for this study was British National Corpus (BNC). BNC is the British English reference corpus, consist of texts from newspaper, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, and many other texts. Moreover, there are also spoken language compiled from formal informal conversations (Burnard, 2009).

To analyze the collocations, this study classified the collocation combinations based on its semantic and syntactic characteristics. Cowie's free combination, restricted collocation, figurative idioms, and pure idioms (Nesselhauf, 2004) for its semantic features, and Benson et al.'s seven types of lexical collocations for its syntactic features (Wang & Good, 2007). Although collocation are not strictly constructed by two word combinations (Hill, 2000), this study listed only two-word collocations with the span of 4L:4R (Baker, 2006).

## FINDINGS AND DISCUSSION

This study listed 46 lexical collocations from conversation texts about self introduction, expressing congratulation, giving compliments and advice, future plan, interview, recount, and admired person. Five out of seven types of lexical collocation combinations, L1 (VN Creation Activation), L2 (VN Eradication Nullification), L3 (AdjN and NN), L6 (AdvAdj), and L7 (VAdv) were presents in the conversations texts. The collocations are in free combinations, restricted collocation, and figurative idioms categories with the absence of figurative idioms.

### Self Introduction

In self introduction conversation text there are two lexical collocation combinations which in the L3 (AdjN and NN) category. The strength of the collocation however, are weak with the score 1.26 for 'specific pattern' and 2.36 for 'tourism resort'. Both collocations classified as 'free combination'.

### Expressing Congratulation

There are three texts about expressing congratulation. In the first text there are ten lexical collocation combinations. The combination of L1 Creation

Activation (VN), L3 (AdjN and NN), L6 (AdvAdj) and L7 (VAdv) are present in the conversation text. Nine of the ten collocation combinations have high MI score except for the ‘food stall’ combination with the MI score 2.38. The word combination ‘very much’ is repeated three times in the text and it is classified as free combination. Other collocation combinations that are classified as free combination are ‘shake hands, having lunch, good luck, food stall, and very happy.’ Lastly, the collocation combination that categorized as restricted collocation such as ‘need help’ and ‘learn a lot’ are due to the meaning of one of the word in the word combinations are not literal. In the word combinations ‘need help’, the word ‘help’ can mean literal assistance or psychological assistance. As for the word combination ‘learn a lot’, the adverb ‘a lot’ can refer to many knowledge and skills that depends on the speakers who wanted to improve in. Therefore ‘learn a lot’ is categorized as restricted collocation.

**Table 1. Lexical Collocation in Expressing Congratulation  
Conversation Text 1**

Freq	Node	Collocate	Comb	MI	Classification
1	Need	Help	VN	3.01	Restricted Collocation
1	Shake	Hands	VN	6.44	Free Combination
1	Having	Lunch	VN	3.06	Free Combination
1	Good	Luck	AdjN	5.20	Free Combination
1	Food	Stall	NN	2.38	Free Combination
3	Very	Much	AdvAdj	3.57	Free Combination
1	Very	Happy	AdvAdj	3.22	Free Combination
1	Learn	A lot	VAdv	3.37	Restricted Collocation

In the second text, there are seven lexical collocation combinations. The combination of L3 (AdjN and NN) and L6 (AdvAdj) are present in the conversation texts. The word combination ‘story telling’ appears three times in the text and it is classified as free combinations. Other collocation classified as free combination are ‘good luck, and very good’. All the word combination in the second text have high MI score. Lexical collocation combinations that are categorized in restricted collocation are ‘best friend’ and ‘wish luck’. The term “best friend” refers to a person’s closest friend. The meaning of “best” is not literal which put this word combination in the “restricted collocation” category. The exact expressions that are used in the conversation text of the textbook are “wish me luck” it expresses hope for someone’s success. Success can refer to many forms which depend on the speakers’ standard.

**Table 2. Lexical Collocation in Expressing Congratulation  
Conversation Text 2**

Freq	Node	Collocate	Comb	MI	Classification
1	Best	Friend	AdjN	3.56	Restricted collocation
1	Good	Luck	AdjN	5.20	Free Combination
1	Wish	Luck	NN	5.23	Restricted collocation
3	Story	Telling	NN	5.31	Free Combination
1	Very	Good	AdvAdj	3.10	Free Combination

In the third text, there are three lexical collocation combination of L3 (AdjN) that categorized as free combination. The word combinations are ‘nice day’, ‘great feeling’ and ‘great job’. All three collocations have low MI scores.

**Table 3. Lexical Collocation in Expressing Congratulation Conversation Text 3**

Freq	Node	Collocate	Comb	MI	Classification
1	Nice	Day	AdjN	2.16	Free Combination
1	Great	feeling	AdjN	1.39	Free Combination
1	Great	Job	AdjN	0.07	Free Combination

**Giving Compliments and Giving Advice**

In conversation text about giving compliments and giving advice, there are five lexical collocation of L1 Creation Activation (VN), L3 (AdjN), and L6 (AdvAdj). Two word combinations ‘express admiration’, and ‘very much are strong collocations because of the high MI score. While the word combinations of ‘nice dress’, ‘wonderful performance’, and ‘great job’ are weak collocation. All five lexical collocations categorized as free combination.

**Future Plan**

In the conversation text about future plan there are six lexical collocation combinations of L1 Creation (VN), L3 (AdjN), L6 (AdvAdj), and L7 (VAdv). The word combination ‘baking cookies’ appear twice and has the highest MI score. The rest word combination with strong collocations are ‘good idea’ and ‘very good’. The word combination ‘last time’ and ‘newly design’ are weak collocation, and all the lexical collocation combination are categorized as free combination.

**Interview**

In the conversation text in the form of interview, there are six lexical collocation of L1 Eradication Nullification (VN), and L3 (AdjN and NN). The lexical collocation of ‘big thing’ and ‘newest version’ have low MI score, therefore they are weak collocation. The word combinations ‘soften the landing’ and ‘wish luck’ are classified as restricted collocation and ‘big thing’ is categorized as figurative idiom. The collocation ‘soften the landing’ refers to controlling the airplane to land in a specific area to reduce impact when the plane touches the ground to avoid destroying the airplane. The term “big thing” refers to something well known, or important. However, the speaker in the conversation actually makes a pun from the term. Since the topic of the conversation is about an airplane which is also big in appearance, the term “big thing” is used figuratively and literally. The rest of the word combination such as ‘good luck’, ‘newest version’, and ‘rubber band’ are classified as free combination.

**Table 4. Lexical Collocations in Interview Conversation Text**

Freq	Node	Collocate	Comb	MI	Classification
1	Soften	Landings	VN	4.24	Restricted Collocation
1	Big	Thing	AdjN	2.13	Figurative idiom
1	Good	Luck	AdjN	5.20	Free Combination
1	Newest	Version	AdjN	2.45	Free Combination
1	Wish	Luck	NN	5.23	Restricted collocation
1	Rubber	Bands	NN	6.87	Free Combination

**Recount**

In conversation text about past experience, there are three lexical collocation of L1 Creation Activation (VN), L3 (AdjN) and L6 (AdvAdj). All three word combinations are strong collocation. The word combination ‘unforgettable

experience' is categorized as restricted collocation. It refers to an event that is important, enjoyable, interesting, or unusual so the person who experiences it will remember for a long time. The meaning of 'unforgettable' is not literal which put this word combination in the 'restricted collocation' category. The other word combination 'planted tree' and 'very good' are classified as free combination.

#### **Admired Person**

Conversation text about admired person has four lexical collocation combinations of L1 Creation Activation (VN), L3 (AdjN), and L6 (AdvAdj). All collocation combinations in the admired person conversation text are classified as free combination. The MI score are high for three word combinations, 'enjoy retirement', 'good job', and 'very much'. Except for the word combination 'share knowledge', the MI score for the word combination is low.

Overall there are 32 lexical collocation combinations among 46 word combinations that are categorized as strong, and 14 of them are weak. The classifications are dominated by free combination. 39 of the listed lexical collocation are in free combination, followed by 6 restricted collocations and 1 figurative idiom, with the absence of pure idiom.

#### **Discussion**

The result shows that despite the limited amount of text, there are a considerable amount of lexical collocations found in the conversation texts of 'Bahasa Inggris' for tenth-grade students. The number of collocation in each conversation text are varies. It was due to the length of the text, shorter text tend to have small number of collocation. With the number of lexical collocations used in the conversation texts there are lack of repetition in the usage of the word combination. The research by Kim & Oh, (2020) also shows similar result in the Korean English textbook. Repetition on the usage of collocation combinations are needed for the students to be aware of the terms and familiar with it so they can use the collocation accurately. The result of the research also shows the absence of 'pure idiom.' This is not in accordance with Lewis & Gough, (1997) who stated that idiom is one of the parts that construct conversation text. The absence of 'pure idiom' in the conversation texts in "Bahasa Inggris" textbook avoids the chance for the students to learn idiomatic expressions in conversation texts. The number of collocation in the conversation text might overwhelm the students because they are not only need to be aware of the collocation but also need to be able to use the word combination for their productive activity. Because of their familiarity with the students, Woolard (2000) suggested the teachers to choose the collocation. He stated that the teacher can tell the students which collocation they can learn and encourage them to learn the collocation independently.

The data in this research describe a specific type of text in the textbook that is the conversation texts. Moreover, the amount of the conversation text does not dominate the content of the book. With the relatively small amount of texts, there is a considerable amount of collocation combinations used in those texts. There are possibly many more collocation combinations in different types of text in the textbook. British National Corpus (BNC), which is the sole reference corpus for this research reveals the strength of collocation from the British English. Perhaps other reference corpora can show a different result from these data.

## CONCLUSION AND SUGGESTION

The findings in this study shows that there are various types of lexical collocation contained in the conversation texts of 'Bahasa Inggris' textbook for tenth-grade students. The numbers of collocation in the texts are determined by the length of the texts. Between these word combinations, none of the lexical collocation is purely idiomatic. Most of the combinations have strong mutual information and some of them are weak. However, the weakness of the mutual information is not caused by syntactic matter, but semantically these word combinations are rarely used.

From the result of the study, the future researcher can consider to analyze lexical collocation in other types of texts and using different corpora in order to get a more varied result on the strength of the word combinations. As for material designer, it is advisable to put some idiomatic expressions especially in conversation text to fulfill the criteria of conversation texts. Material designer can also consider modifying the vocabulary exercise in the textbook to put multiple words instead of single word translation activity.

## REFERENCES

- Anthony, L. (2020). *AntConc* (Version 3.5.9) [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software>
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. Bloomsbury Academic.
- Benson, M., Benson, E., & Ilson, R. (2009). *The BBI combinatory dictionary of English* (3rd ed. rev). John Benjamins.
- Burnard, L. (2009). *[Bnc] What is the BNC? - About the British National Corpus*. <http://www.natcorp.ox.ac.uk/corpus/index.xml?ID=intro>
- Hill, J. (2000). Revising Priorities: From grammatical failure to collocational success. In *Teaching Collocation: Further Developments in the Lexical Approach* (pp. 47–67). Language Teaching Publications.
- Hutapea, S. U. P., Siregar, M., & Dewi, N. R. (2019). The Collocation of English Textbook for Grade X in Senior High School. *GENRE Journal of Applied Linguistics of FBS Unimed*, 7(3). <https://doi.org/10.24114/genre.v7i3.12439>
- Kim, Y. S., & Oh, S.-Y. (2020). A Corpus-based Analysis of Collocations in Korean Middle and High School English Textbooks. *Lanaguage Research*, 56(3), 437–461. <https://doi.org/10.30961/lr.2020.56.3.437>
- Lewis, M., & Gough, C. (1997). *Implementing the lexical approach: Putting theory into practice* (Nachdr.). Heinle Cengage Learning.
- McIntosh, C., Francis, B., & Poole, R. (Eds.). (2002). *Oxford Collocations Dictionary for Students of English* (1st ed.). Oxford University Press.

- Nesselhauf, N. (2004). *Collocations in a learner corpus*. J. Benjamins Pub. Co.
- Oktavianti, I. N., Prayogi, I., Amal B., M. A., & Pertiwi, R. S. (2020). An Analysis of Conversations in Curriculum-based EFL Textbooks for Senior High School in Indonesia and the Comparison with Corpus-based English Textbooks. *Universal Journal of Educational Research*, 8(9), 4151–4162. <https://doi.org/10.13189/ujer.2020.080941>
- Said, M. (2011). Negative Transfer of Indonesian Collocations into English and Implications for Teaching English as a Foreign Language. *LiNGUA: Jurnal Ilmu Bahasa Dan Sastra*, 6(2), 164–167.
- Szudarski, P. (2017). *Corpus Linguistics for Vocabulary: A Guide for Research* (1st ed.). Routledge. <https://doi.org/10.4324/9781315107769>
- Wang, J. T., & Good, R. L. (2007). The Repetition of Collocations in EFL Textbooks: A Corpus Study. *The Sixteenth International Symposium and Book Fair on English Teaching in the Republic of China*, 1–13.
- Woolard, G. (2000). Collocation—Encouraging learner independence. In *Teaching Collocation: Further Developments in the Lexical Approach* (pp. 28–46). Language Teaching Publications.