



Comparing the use of lexical bundles in Indonesian-English translation by student translators and professional translators

Hilda Novita

Atmajaya Catholic University, Indonesia
hildinza@yahoo.com

Deny A. Kwary

Universitas Airlangga, Indonesia
d.a.kwary@fib.unair.ac.id

DOI: 10.12807/ti.110201.2018.a04

Abstract: Knowledge of lexical bundles is essential for translators, and the study of lexical bundles should be an integral part of translation training. This study investigated the lexical bundles made by student translators and professional translators in the Indonesian-English translation work. The student translators were sophomore students who had completed a translation course, while the professional translators were certified translators who were also members of Indonesian Translators Association. The findings of this study and the statistical calculations using Chi-square tests revealed that the group of professional translators produced a higher frequency and number of types of lexical bundles than the student translators did. This means that the frequency and the number of types of lexical bundles depended on the group of translators. This study also classifies the lexical bundles into the five functional categories described by Mahlberg (2007a). The classification shows a similarity in the number of functional categories the lexical bundles fitted into. However, the types of lexical bundles found within each category were significantly different. Consequently, we can conclude that the production of lexical bundles, even in translation, depends on the ESL proficiency level.

Keywords: lexical bundles, N-grams, professional translator, student translator, translation, translation training.

1. Introduction

Corpus linguistics has enabled linguists to observe language patterns in different types of texts. One of the language patterns which is quite commonly observed is that of lexical bundles. Biber, Johansson, Leech, Conrad, and Finnegan (1999, p.899) defined lexical bundles as recurrent sequences of words which appear across texts in the same register and help shape distinctiveness of the register. These sequences of words have been termed differently by different scholars. Pawley and Syder (1983, p.191) used the

term “lexicalized sentence stem”, Scott (2004, p.204) used the word “clusters”, and Stubbs (2005, p.5) termed them “N-grams”.

Many studies have been conducted to investigate and identify possible links between lexical bundles and language proficiency. Some studies found that a higher frequency of lexical bundles was positively correlated with a higher language proficiency. Vidacovic and Barker (2010, p.145) found that the larger the collection of lexical bundles the students had at their disposal, the higher their levels of proficiency were. Adel and Erman (2012, p.85) and Chen and Baker (2010, p.43) discovered that students who were native speakers of English had larger frequencies and types of lexical bundles at their disposal than students who were non-native speakers of English. Boers, Eyckmans, Kappel, Stengers, and Demecheleer, (2006, p.247) looked at lexical bundles (LBs) from a psycholinguistics perspective and further explained that LBs helped speed up language processing thereby assisting learners in achieving language accuracy in the second language.

Studies on lexical bundles have mostly occurred in the tertiary education context although they may have had a different research focus than the study reported on here. Wray (2000); Biber, Conrad and Cortes (2004); and Biber (2006) investigated lexical bundles to explore the use and usefulness of lexical bundles in university teaching and textbooks. Cortes (2004) and Hyland (2008a) looked at lexical bundles in student writings across different disciplines. Hyland (2008b) researched bundles in the writings of students at different academic levels and created a taxonomy or classification of lexical bundles according to their functions. Chen and Baker (2010), Salazar (2011), and Adel and Erman (2012) focused on the use of lexical bundles in English native and non-native student writings.

L2 language production, however, is different from translated (target) language production. The PACTE (2003) model suggests that translation includes declarative and procedural knowledge, the ability to carry out the transfer process, taking into account the purpose of the translation and the characteristics of the target text readers. This means that the production of lexical bundles in an L2 writing process may also be different from that which can be identified in the translation process.

There have only been a few studies conducted on lexical bundles that focused on literary texts and fewer still in the field of translation. Mahlberg (2007a; 2007b) used a corpus stylistics approach to investigate “clusters” in Dickens’ work. She discovered that lexical bundles in literary texts are different from those in academic texts and therefore should be classified differently. Lee (2012) used the lexical bundle approach to identify individual features of texts which have been translated from English into Korean language. She argued that research on lexical bundles was also applicable to the field of translation studies and that a corpus stylistics approach to lexical bundle investigation was more suitable for this kind of study.

The study described here employed a frequency-driven approach, as used by Biber and Jones (2009) to identify general patterns of lexical bundles, in order to identify the frequency of lexical bundles and the number of types of lexical bundles produced by student and professional translators in the Indonesian context. The Chi-square statistical test was used to determine whether or not the frequency and the number of types of lexical bundles produced were independent of the groups of translators (students or professionals). In addition, Mahlberg’s (2007a) classifications of 5-word lexical bundles in literary texts were used to compare the types of lexical

bundles produced by the student and professional translators. It is hoped that the findings can help translator educators to identify areas trainee translators need to improve in.

2. Identifying lexical bundles

Biber and Conrad (1999, pp.183-184) state that lexical bundles must co-occur together frequently (frequency) and uninterruptedly (sequence/fixedness). They must also be distributed across different texts within the same register (distribution). Since these three characteristics – frequency, distribution and length of sequences – are important, they were used for the identification of lexical bundles in this study. These three characteristics were taken into account in determining lexical bundles in this study, and this is further explained in the next section.

The word sequences chosen in this study started from 3-word sequences. This is because, according to Conrad and Biber (2005), sequences of two words mostly concern word associations which do not have functions at the discourse level. Conrad and Biber (2005) instead used 3- and 4-word lexical bundles. Furthermore, Hyland (2008b, p.44), who analysed student writings, stated that a certain length of lexical bundles was actually a part of a longer string. For example, a very common four-word string such as *due to the fact* was a part of a five-word string *due to the fact that* and also part of the six-word string *is due to the fact that*. In her research on lexical bundles in literary texts, Mahlberg (2007a, p.7) used 3-, 4- and 5-word lexical bundles, because she felt that the longer the lengths of lexical bundles were, the more restricted the bundles were to specific texts. Therefore, in this study, used 3-, 4- and 5-word lexical bundles were used.

Since lexical bundles in literary texts are different from those used in academic texts, Mahlberg (2007b, p.228), developed five functional categories of bundles for literary texts which she found relevant to the style of Charles Dickens as follows:

- Labels – contain the names of characters (e.g. *Mr. Pickwick and his friends*) and names of places (e.g. *a tavern of a dropsical appearance*).
- Speech clusters – contain a first or second person pronoun or possessive which is taken as an indication of interaction (e.g. *what do you mean by*).
- As If clusters – contain clusters that starts with *as if* (e.g. *as if he would have*).
- Body Part clusters – contain at least one noun referring to a part of the human body (e.g. *his hands in his pockets*).
- Time and Place clusters – contain a nominal time and place expression with or without a preposition (e.g. *on the top of his, the opposite side of the*).

In her study, Mahlberg (2007, p.9) added that shorter clusters were more flexible and more frequent, but also more difficult to characterize. Therefore, she preferred to use only the 5-word lexical bundles in her classification since the 5-word lexical bundles could occur across different texts, and their frequencies were sufficient to lend themselves to analysis. The five categories

established by Mahlberg (2007) were applied in the present study, since the data analysed were also taken from a literary text, as shown in Appendix 1.

3. Methodology

This study used a corpus linguistics approach. McEnery and Hardie (2012, p.1) characterize corpus linguistics as a methodology since it is an area which focuses on a set of procedures, or methods, for studying language. Corpus linguistics procedures were applied in this study to analyse the lexical bundles (henceforth LBs) obtained from the corpus of translations produced by both student and professional translators, as set out below.

3.1 Research participants

Two groups of translators were chosen to participate in the study: one group consisted of translation students while the second group consisted of professional translators. All of these research participants had native level proficiency in their first language (L1) Indonesian and a certain degree of proficiency in English as an additional language (EAL). The translation students were chosen to represent novice translators who, although equipped with knowledge of Indonesian and English, might not have had much translation competence. The translation students, who will be referred to as student translators (STs) in this study, were selected based on these three criteria: They were third-year students who studied at the same university in Jakarta, Indonesia, and who had achieved an average paper-based TOEFL score of 550; they had taken a translation class in the previous semester; and, they had all obtained similar scores in their previous translation class (namely a grade between A and B, on a scale from A to E, where A represented the highest grade possible).

The professional translators (PTs) were chosen to represent expert translators whose professional qualifications reflected the fact that they possessed both linguistic and translation competence in both English and Indonesian. The PTs selected had to meet the following three criteria: They were all members of the Indonesian Translators Association (HPI – *Himpunan Penerjemah Indonesia*) and all held translator certification; they all worked as translators on a freelance basis; and they all lived in Jakarta, the capital of Indonesia. Five translators were chosen from each group. This small number was due to the fact that only a handful of professional translators met the criteria and agreed to participate in the research. The Indonesian Translators Association started its certification exam in 2010, and at the time when this study was conducted in 2016 there were only 32 certified Indonesian-to-English translators (<http://www.hpi.or.id/sertifikasi/daftar-penerjemah-bersertifikat-hpi>).

3.2 Methods of data collection

For this study, one fictional text taken from a random untranslated and unpublished Indonesian short story was chosen as the source text. The number of words in this text was around 600 words. Care was taken to ensure that the text remained confidential to all translators until the day of the test (please see Appendix 1 for the source text). The following method was implemented to collect the translations:

- The groups of translators were asked to translate this text in separate venues and at different times. The PTs and STs did not know each other. The test for the STs was conducted in the morning, while the test for the PTs was conducted in the evening, at different venues. The test for the STs was conducted in the university where the STs were studying. The test for the PTs was conducted in a different university building.
- The translation test followed the same procedures as those set by the HPI in its translator certification test:
 - A maximum of 90 minutes was allocated to all the translators to complete the translation of the same 600-word Indonesian short story into English.
 - All the translators were expected to work individually. The researchers acted as invigilators to make sure that the research participants worked individually.
 - All the translators used Notepad software on their laptops/computers to prevent the use of autocorrect and any other help.
 - Dictionaries (Indonesian–English, English–English, and English thesaurus) were provided for the translators. Participants were not allowed to use online sources or electronic dictionaries.
 - All translations were collected at the end of the given time and each translation was given an identification number (1-10) and the type of group (Group A for STs and Group B for PTs).

3.3 Data analysis methods

The data was analysed in three phases. These phases refer to the aim of this study, i.e. analysing the frequency of LBs, the number of types of LBs, and the categories of LBs. The texts translated were classed as either: student or professional. All texts from each group were put into *AntConc* (Anthony, 2014), a concordance tool, to extract 3-, 4- and 5-word LBs. LBs which appeared with a minimum frequency of 3 and in a minimum of 3 different versions of the translation were selected for analysis. The frequencies and the types of the LBs from each group were recorded and sequenced from highest to the lowest.

In phase one, the frequency of LBs produced by the STs was compared with that produced by the PTs. We felt it was necessary to explore the frequency of the LBs to see if language users, in this case translators, preferred to use the same LBs. High frequencies might indicate that many translators use the same LBs or translators use certain LBs repeatedly.

A Chi-square Test of Independence was conducted on the frequency of the LBs to see whether the total frequency of LBs was independent of the groups of translators.

In phase two, the types of LBs were manually classified into two categories: the LBs produced by STs and the LBs produced by PTs. Another Chi-square Test of Independence was conducted on the number of types of the LBs to see whether the total number of types of LBs was independent of the groups of translators.

In phase three, the 5-word LBs were classified according to functions (Mahlberg, 2007a). The results were then analysed to see which functions had the highest number of types for each group of translators.

4. Frequency of LBs

The data in Table 1 shows the 3-, 4-, and 5-word LBs which were extracted using *AntConc* from the English translations done by the STs and the PTs.

Table 1: Detailed frequencies of LBs for each LB sequence

LB Sequence	3-word LBs	4-word LBs	5-word LBs	Total
Student (ST)	378	157	71	606
Professional (PT)	531	253	115	899

To display the differences in the data presented in Table 1, we used the numbers to produce histograms, as shown in Figure 1.

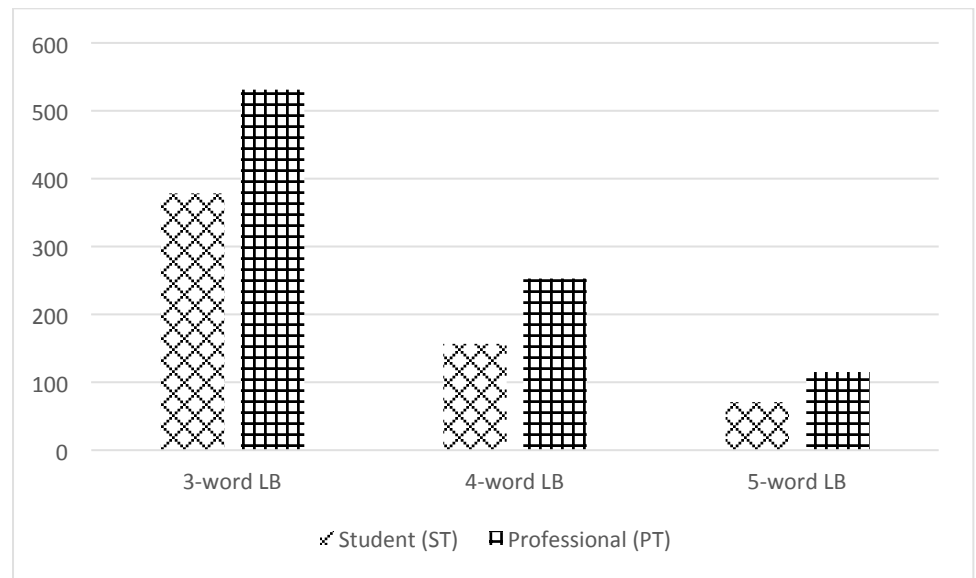


Figure 1. Frequencies of LBs for each LB sequence

Table 1 and Figure 1 both show that LBs can be found across all three LB sequences and that PTs produced higher frequencies of 3-, 4-, and 5-word LBs than STs did. As regards the 3-word LBs, the PTs produced 531 LBs while the students produced 378 LBs. In terms of the 4-word LBs, the PTs produced 253 LBs while the STs produced more than half this number (157 LBs). Lastly, with regard to the 5-word LBs, the PTs produced 115 LBs while the STs produced 71 LBs. This result is in line with the findings of Chen and Baker's (2010, p.43) study, which looked at the relationship between LBs and language proficiency. Chen and Baker stated that learners with higher levels of language proficiency were able to produce a higher frequency of LBs than learners with lower levels of proficiency. In this study, the PTs, whom we

assumed had a higher level of English language proficiency and translation competence, produced higher frequencies of LBs in all aspects, compared to the STs, who were expected to have a lower level of both English language proficiency and translation competence.

The lower frequency of LBs produced by the STs also corresponds with findings by Vidacovic and Barker (2010) and Feng (2014). Vidacovic and Barker (2010) found that learners with lower proficiency relied more on individual words than conventionalized multi-word sequences or LBs, while Feng (2014) found that translated business texts produced by English-to-Chinese translators tended to involve fewer bound collocations and idioms than comparable texts written by native speakers of English. In addition, Vidacovic and Barker (2010) also stated that although learners from a lower level of proficiency have started learning about LBs, these LBs could only be productively used when the learners had reached a higher level of proficiency.

Although LBs can be found across all three LB sequences, for both groups, the frequencies significantly decreased as the sequences got longer. Both groups only produced approximately half of the number of LBs they produced in the shorter sequence. The STs produced 378 LBs for the 3-word LBs, 157 LBs for the 4-word LBs, or less than half of the frequency of 3-word LBs, and 71 LBs for the 5-word LBs, or slightly less than half of the frequency of 4-word LBs, and only 1/5 of the frequency of 3-word LBs. Following the same trend, the PTs produced 531 LBs for the 3-word LBs, 253 LBs for the 4-word LBs, or slightly less than half of the frequency of 3-word LB, and 115 LBs for the 5-word LBs, or less than half of the frequency of 4-word LB, or only 1/5 of the frequency of 3-word LBs. This finding corresponds with Mahlberg's (2007a, p.7) findings that the frequency of occurrence of the LBs decreased as their length increased; that is, the larger the LBs were, the fewer there were. Mahlberg (2007a) found that in the literary text she studied the frequency of occurrence of the LBs decreased as their length increased. In other words, shorter LBs such as the 3-word LBs were more commonly found in the literary text than longer LBs such as the 5-word LBs. This finding is similar to the finding in this present study, in the case of the translations of the literary text we looked at.

This finding also suggests that among all three LB sequences, the 3-word LB was the most frequently encountered in both PT and ST translations and involved the most conventionalized word sequences. Biber, et al. (1999, p. 992) explained that 3-word LBs were extremely common because they are "a kind of extended collocational association", while longer bundles were "more phrasal in nature and correspondingly less common."

Table 2. The number of LBs according to the frequency and the number of sequences

Number of word sequences	Group of Translators	Number of LBs for each frequency				
		7	6	5	4	3
3-word LBs	STs	0	4	3	36	65
	PTs	2	1	23	36	84
4-word LBs	STs	0	0	0	16	31
	PTs	0	0	5	21	48
5-word LBs	STs	0	0	0	8	13
	PTs	0	0	0	13	21

The number 2 there means that there were 2 types of 3-word LBs (i.e. "me to go" and "my best friend" in Appendix 3) which were produced 7 times by the PTs. We refer to this as frequency category 7 because each of these 3-word LBs was produced 7 times by the PTs.

Table 2 shows a further comparison of the frequencies of the LBs. This table shows the number of LBs for each frequency and the number of sequences of the LBs. The details of the LBs can be found in the Appendix 2 and 3.

Table 2 shows that PTs consistently produced more LBs than the STs for each frequency, except for the frequency of 6, where the STs produced 4 LBs, while the PTs only produced 1 LB. As for frequency category 7 (see Appendices 2 and 3), the PTs were able to produce 2 LBs while the STs were not able to produce any LBs. As for the frequency category 5, the PTs were able to produce 23 LBs while the STs were able to produce only 3 LBs. Please see the cell in Table 2 which we highlighted in orange. Here the number 3 means that there were 3 types of 3-word LBs which were produced 5 times by the STs. Similarly, the number 23 means that there were 23 types of 3-word LBs which were produced 5 times by the PTs. We call this the frequency category 5 because each of these 3-word LBs was produced 5 times.

As to frequency category 4, both PTs and STs were able to produce 36 LBs and for the frequency category 3, the PTs were able to produce the most LBs with 84 LBs

If we refer again to the data in Table 2 (and Appendices 2 and 3), it can be seen that the 3-word LBs used by the PTs for each LB sequence was dominated by 3, 4, and 5 frequency categories. Meanwhile, for the STs the LB sequence was only dominated by the 3 and 4 frequencies. This result suggests that the PTs were more familiar with the same conventionalized word sequences than the STs.

In order to confirm statistically that the use of LBs by STs was significantly different from the use of LBs by PTs, we conducted a Chi-square test. It was hoped that the Chi-square test would determine whether or not the use of LBs depended on which groups the translators represented. The hypotheses were as follows:

H_0 = The frequency is independent of the group.

H_a = The frequency is not independent of the group.

With $\alpha = >0.05$ and the Chi-square value was 4.26489E-14 which is lower than α , so we rejected the null hypothesis and accepted the alternative hypothesis. This means the total frequency depended on the group. This confirmed the conclusion that the group of PTs was statistically proven to produce a higher frequency of LBs, while the group of STs produced a lower frequency of LBs. This was also in line with the findings by Vidacovic and Barker (2010, p.143) and Chen and Baker (2010, p.44), namely that the PTs, who could be described as more advanced users of English as a Second Language, produced higher frequencies of LBs than the STs, who could be described as less proficient learners of English as a Second Language in this study (see also Feng, 2014).

5. Types of LBs

After analysing the frequencies of the LBs, we go on to discuss the types of the LBs used by STs and PTs. We again used *AntConc* software to extract the types of the 3-, 4-, and 5-word LBs. The total number of types for each of the LB sequence is presented in Table 3 after identification according to minimum frequency (3) and minimum distribution (3).

Table 3. The total number of types of LBs for each of the LB sequence identified from the translations of the text

LB sequence	3-word LBs	4-word LBs	5-word LBs	Total
Student (ST)	108	47	21	196
Professional (PT)	146	74	35	285

To display the differences in the data presented in Table 3, we again used the numbers to generate histograms, shown in Figure 2.

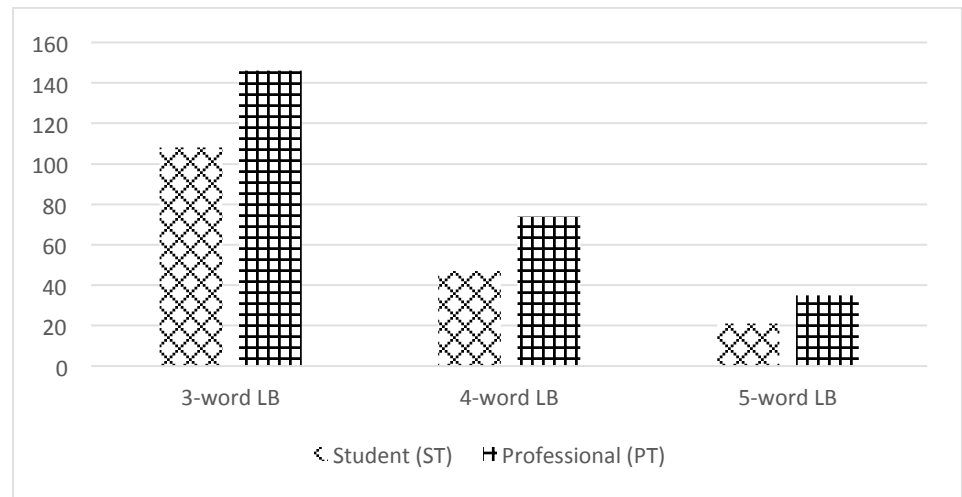


Figure 2: The number of types according to the LB sequence

Based on the data in Table 3 and Figure 2, we can see that the 3-word LBs showed the highest frequency of LB types for both STs and PTs. The PTs also produced more LB types compared to STs. In terms of these 3-word LB types, the STs produced 108 types while the PTs produced a higher frequency with 146 types. As for the 4-word LBs, the STs produced 47 types while the PTs produced almost double with 74 types. With regard to the 5-word LBs, the STs produced 21 types while the PTs produced 34 types.

As with the frequencies, the number of types also decreased by almost 50% as the sequences got longer. The PTs produced 146 LB types for the 3-word LBs. With regard to the 4-word LBs, the PTs produced slightly more than half (74) of all 3-word LB types. As for the 5-word LBs, the PTs produced slightly less than half (35) of the number of the 4-word LB types or about one quarter of the number of the 3-word LB types. Meanwhile, as to the 3-word LB, the STs produced 108 LB types. As to the 4-word LBs, the STs produced slightly less than half (47) of all 3-word LB types. For 5-word LBs, the STs produced more than half (21) of all 4-word LB types or less than ¼ of the number of the 3-word LB types.

A further calculation was generated in order to confirm statistically that the types of LB produced by the STs was significantly different from those produced by the PTs. In this case, we used the Chi-square test to determine whether or not the types of LBs depended on the group the translators were part of. The hypotheses were as follows:

H_0 = The number of types of LBs is independent of the group.

Ha₂ = The number of types of LBs is not independent of the group.

Since the calculation showed that $\alpha \geq 0.05$ and a Chi-square value of 4.9484E-05, we rejected the null hypothesis and accepted the alternative hypothesis. This means the number of types was not independent of the group. This result confirmed that the PTs produced a higher number of types of LBs than the STs. This result also confirmed findings by Vidacovic and Barker (2010, p.144) and Chen and Baker (2010, p.43) to the effect that the number of types of LBs increased as writing proficiency improved. The English language users with a higher English language proficiency, in this case are the PTs, produced a higher number of types of LBs than the STs, who could be referred to as less proficient English language learners.

6. Classification of types of LBs

As a further comparison of the types of LBs produced by STs and PTs in translating the literary text, this study also established a classification of the types of LBs. The classification follows the five functional categories established by Mahlberg (2007a, p.13). The classification of the 5-word LBs produced by the STs can be seen in Table 4.

Table 4: Classification of the 5-word LBs produced by the STs

Functional Category	Types of LBs
Label	<i>as a best friend he</i> <i>his father and mother also</i>
Speech	
As If	<i>if this is only about</i>
Body Parts	<i>face who opened the door</i> <i>from his hand then we</i> <i>the torch from his hand</i>
Time/Place	<i>a collection placed in a</i> <i>collection placed in a box</i> <i>placed in a box then</i> <i>than me that night I</i> <i>loyal than me that night</i> <i>me that night I was</i>

As we can see in Table 4, four out of five categories contained the LBs produced by the STs. The number of types of LBs that fitted into the categories or functions suggested by Mahlberg (2007a) was relatively small. This might be due to the fact that Mahlberg's functional classification was based on the keywords, or words which were especially high in frequency in Dickens's texts and lower in frequency in the reference corpus she used in her study. This means that the LBs which were included in the classification were specific to the texts and thus, the functions were also specific to the texts she analysed. The data in Table 4 also show that there were no LBs for the "Speech" category. This might be due to the fact that there were only a few dialogues in the source text and most dialogues were only one or two words which would not pass the minimum requirement for the length of sequence, i.e. a 5-word LB.

The results of the classification of the 5-word LBs produced by the STs and the PTs were quite similar in terms of category, but very different in terms

of types. The classification of the 5-word LBs produced by the PTs based on Mahlberg's (2007a) classification can be seen in Table 5.

Table 5. Classification of the 5-word LBs produced by the PTs

Functional Category	Types of LBs
Label	<i>his mother a woman who</i>
Speech	
As If	
Body Parts	<i>friend s waist I followed from his hand we then his hand we then walked the torch from his hand torch from his hand we</i>
Time/Place	<i>across the room it felt after your family returned and all since then so what at all since then so</i>

As shown in Table 5, only three out of five categories involved the LBs produced by the PTs. The PTs did not produce any LBs for the "Speech" and "As if" categories. The ST produced only one LB for the "As if" category, i.e. *if this is only about*, which involved the translation of the Indonesian string of words *Kalau ini hanya soal*. The PTs translated this string of words differently. The concordance lines in Table 6 confirm that each PT produced different translations of the string of words *Kalau ini hanya soal* differently. This may mean that the PTs used their individual creativity in the translation.

Table 6. The concordance lines showing the translation of *Kalau ini hanya soal* in **bold**

It if were about memories, there had been no need to wait for 10 years	P1
If this is just about a memory, no need to wait for 10 years after your	P2
If it's just about memories, you don't need to wait 10 damn years	P3
If it's all about memory, it doesn't have to take 10 years after your family	P4
If this is about memories, you don't have to wait 10 years after your family	P5

If we compare the 5-word LBs produced by STs and PTs within the categories shown in Table 4 and Table 5, we can see that there is a significant difference between the LBs created by STs and those produced by PTs. There is only one 5-word LB shared by both the STs and the PTs, i.e. *the torch from his hand*, which is in the category of "Body Parts". The significant differences in the types of 5-word LBs produced by STs and PTs confirm the results of the previous analyses that the production of LBs is indeed dependent on the groups of translators.

7. Conclusion

The findings of this study show that the 3-, 4- and 5-word lexical bundles (LBs) can be found in all translated texts produced by the two groups of translators. The study also shows that as the lexical bundle sequences got longer, the frequency and the number of types significantly decreased. This

means that both student translators (STs) and professional translators (PTs) were able to produce a larger collection of shorter LBs which in this case involved 3-word lexical bundles. The PTs managed to produce a higher frequency of LBs and on average a wider distribution of LBs compared to the STs. The PTs also managed to produce more types of LBs. The functional classification of the 5-word LBs shows a similarity between student translators and PTs in terms of the number of categories their LBs fitted into. However, the types of 5-word LBs they produced were significantly different. Consequently, the results confirm the hypothesis that the production of LBs is not independent on the groups of translators. In other words, we can say that in translation too, the production of LBs also depends on ESL proficiency level, and evidently the PTs in this study possessed a higher level of ESL proficiency than the STs.

This study emphasizes the needs to include LB training material in translator education. A large collection of useful and conventionalized LBs with their functions and usage should be included in the translation classroom as a part of translation education in order to help the students succeed in their future work as translators. This study has shown the significant role of LB production in translating literary texts. In the Indonesian setting, in addition to the literary texts, there is also a growing demand for translation of legal texts (e.g. company agreements, government legislation, and treaties) and medical texts (e.g. prescription labels, medication forms, and medical articles). Consequently, further studies on LBs involving these types of texts need to be conducted. In teaching LBs for student translators, lecturers could first ask students to identify lexical bundles in Indonesian texts, and then ask them to identify comparable lexical bundles in native English texts of the same genre, or on the same topic. Students could then discuss and compare these LBs in small groups, as a precursor to translation of texts containing lexical bundles, and this might help student translators become more aware of LBs.

References

- Adel, A. & Erman, B. (2012). Recurrent word combination in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes* 31, 81-92.
- Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net>
- Biber, D. (2006). *University language: a corpus-based study of spoken and written registers*. Amsterdam & Philadelphia: John Benjamin Publishing Company.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselard & S. Oksefjell (Eds.). *Out of corpora: Studies in honor of Stig Johansson* (pp.181-189). Amsterdam: Ropodi.
- Biber, D., Conrad, S. & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25, 371-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finnegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman
- Biber, D. & Jones, J. (2009). Quantitative methods in corpus linguistics. In A. Ludeling & M. Kytö (Eds). *Corpus Linguistics: An international handbook volume 2* (pp.1287-1304). Berlin: Walter de Gruyter.
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research* 10 (3), 245-261.

- Chen, Y. & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology* 14 (2), 30-49.
- Conrad, S. & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Applied Linguistics Faculty Publications and Presentations*, 56-71.
- Cortes, V. (2004). Lexical bundles in publishes and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23, 397-423
- Feng, H. (2014). *A corpus-driven study of Chinese translators' use of English collocations in commercial Chinese to English translation* (Doctoral dissertation, Auckland University of Technology).
- HPI (Himpunan Penerjemah Indonesia 'Indonesian Translators Association'). *List of HPI Certified Translators*. Retrieved 14 July 2017 from <http://www.hpi.or.id/sertifikasi/daftar-penerjemah-bersertifikat-hpi> .
- Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purpose* 27, 4-21.
- Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 18 (1), 41-62.
- Lee, C (2013). Using lexical bundles analysis as a discovery tool for corpus-based translation research. *Perspectives: Studies on Translatology* 21 (3), 378-395.
- Mahlberg, M. (2007a). Clusters, key clusters and local textual functions in Dickens. *Corpora* 2 (1), 1-31.
- Mahlberg, M. (2007b). Corpus stylistics: Bridging the gap between linguistic and literary study. In M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert (Eds). *Text, discourse and corpora: Theory and analysis* (pp.219-242). London & New York: Continuum.
- McEnery, T. & Hardie, A. (2012). *Corpus Linguistics*. Cambridge: Cambridge University Press.
- PACTE. (2003). Building a translation competence model. In F. Alves (ed.) *Triangulating translation: Perspectives in process oriented research* (pp.43-66). Amsterdam: John Benjamins.
- Pawley, A. & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Smith (Eds.) *Language and communication* (pp. 191-226). New York: Routledge.
- Salazar, D. (2011). Lexical bundles in scientific English: A corpus-based study of native and non-native writing (Phd Thesis, Universitat de Barcelona).
- Scott, M (1996). *Wordsmith tools 4*. Oxford: Oxford University Press.
- Stubbs, M. (2005). The most natural thing in the world: Quantitative data on multi-word sequences in English. In *Phraseology Conference*. Louvain-la-Neuve.
- Vidaković, I & Barker, F (2010) – Lexical development across second language proficiency level: A corpus-informed study. *Proceedings of the 42nd annual meeting of the British association for applied linguistics*. Retrieved from http://www.baal.org.uk/proc09/vidakovic_barker.pdf
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics* 12 (4). pp. 463-489

APPENDICES

Appendix 1 – Source text

An extract of an unpublished Indonesian short story taken from <https://cerpenkompas.wordpress.com/> titled Seragam). Total words: 660

Lelaki jangkung berwajah terang yang membukakan pintu terlihat takjub begitu mengenali saya. Pastinya dia sama sekali tidak menyangka akan kedatangan saya yang tiba-tiba.

Ketika kemudian dengan keramahan yang tidak dibuat-buat dipersilakannya saya untuk masuk, tanpa ragu-ragu saya memilih langsung menuju amben di seberang ruangan. Nikmat rasanya duduk di atas balai-balai bambu beralas tikar pandan itu. Dia pun lalu turut duduk, tapi pandangannya justru diarahkan ke luar jendela, pada pohon-pohon cengkeh yang berderet seperti barisan murid kelas kami dahulu saat mengikuti upacara bendera tiap Isnin. Saya paham, kejutan ini pastilah membuat hatinya diliputi keharuan yang tidak bisa diungkapkannya dengan kata-kata. Dia butuh untuk menetralsirnya sebentar.

Dia adalah sahabat masa kecil terbaik saya. Hampir 25 tahun lalu kami berpisah karena keluarga saya harus boyongan ke kota tempat kerja Ayah yang baru di luar pulau hingga kembali beberapa tahun kemudian untuk menetap di kota kabupaten. Itu saya ceritakan padanya, sekaligus mengucapkan maaf karena sama sekali belum pernah menyambangnya sejak itu.

”Jadi, apa yang membawamu kemari?”

”Kenangan.”

”Palsu! Kalau ini hanya soal kenangan, tidak perlu menunggu 10 tahun setelah keluargamu kembali dan menetap 30 kilometer saja dari sini.”

Saya tersenyum. Hanya sebentar kecanggungan di antara kami sebelum kata-kata obrolan meluncur seperti peluru-peluru yang berebutan keluar dari magasin.

Bertemu dengannya, mau tidak mau mengingatkan kembali pada pengalaman kami dahulu. Pengalaman yang menjadikan dia, walau tidak setiap waktu, selalu lekat di ingatan saya. Tentu dia mengingatnya pula, bahkan saya yakin rasa yang diidapnya lebih besar efeknya. Karena sebagai seorang sahabat, dia jelas jauh lebih tulus dan setia daripada saya.

Malam itu saya berada di sini, memperhatikannya belajar. Teplok yang menjadi penerang ruangan diletakkan di atas meja, hampir mendekat sama sekali dengan wajahnya jika dia menunduk untuk menulis. Di atas amben, ayahnya santai merokok. Sese kali menyalakan pemantik jika bara rokok lintingannya soak bertemu potongan besar cengkeh atau kemenyan yang tidak lembut diirisnya. Ibunya, seorang perempuan yang banyak tertawa, berada di sudut sembari bekerja memilin sabut-sabut kelapa menjadi tambang. Saat-saat seperti itu ditambah percakapan-percakapan apa saja yang mungkin berlaku di antara kami hampir setiap malam saya nikmati. Itu yang membuat perasaan saya semakin dekat dengan kesahajaan hidup keluarganya.

Selesai belajar, dia menyuruh saya pulang karena hendak pergi mencari jangkrik. Saya langsung menyatakan ingin ikut, tapi dia keberatan. Ayah dan ibunya pun melarang. Sering memang saya mendengar anak-anak beramai-ramai berangkat ke sawah selepas isya untuk mencari jangkrik. Jangkrik-

jangkrik yang diperoleh nantinya dapat dijual atau hanya sebagai koleksi, ditempatkan di sebuah kotak, lalu sesekali digelitik dengan lidi atau sehelai ijuk agar berderik lantang. Dari apa yang saya dengar itu, proses mencarinya sangat mengasyikkan. Sayang, Ayah tidak pernah membolehkan saya. Tapi malam itu toh saya nekat dan sahabat saya itu akhirnya tidak kuasa menolak.

”Tidak ganti baju?” tanya saya heran begitu dia langsung memimpin untuk berangkat. Itu hari Jumat. Seragam coklat Pramuka yang dikenakannya sejak pagi masih akan terpakai untuk bersekolah sehari lagi. Saya tahu, dia memang tidak memiliki banyak pakaian hingga seragam sekolah biasa dipakai kapan saja. Tapi memakainya untuk pergi ke sawah mencari jangkrik, rasanya sangat-sangat tidak elok.

”Tanggung,” jawabnya.

Sambil menggerutu tidak senang, saya mengambil alih obor dari tangannya. Kami lalu berjalan sepanjang galengan besar di areal persawahan beberapa puluh meter setelah melewati kebun dan kolam gurami di belakang rumahnya. Di kejauhan, terlihat beberapa titik cahaya obor milik para pencari jangkrik selain kami. Rasa hati jadi tenang. Musim kemarau, tanah persawahan yang pecah-pecah, gelap yang nyata ditambah angin bersiuran di areal terbuka memang memberikan sensasi aneh. Saya merasa tidak akan berani berada di sana sendirian.

Kami turun menyusuri petak-petak sawah hingga jauh ke barat. Hanya dalam beberapa menit, dua ekor jangkrik telah didapat dan dimasukkan ke dalam buntung yang terikat tali rafia di pinggang sahabat saya itu. Saya mengikuti dengan antusias, tapi sandal jepit menyulitkan saya karena tanah kering membuatnya berkali-kali terlepas, tersangkut, atau bahkan terjepit masuk di antara retakan-retakannya. Tunggak batang-batang padi yang tersisa pun bisa menelusup dan menyakiti telapak kaki. Tapi melihat dia tenang-tenang saja walaupun tak memakai alas kaki, saya tak mengeluh karena gengsi.

Appendix 2 – List of lexical bundles produced by student translators.

N o.	Fre q	3-word Lexical Bundles	Fre q	4-word Lexical Bundles	Fre q	5-word Lexical Bundles
1	6	he did not	4	a best friend he	4	a collection placed in a
2	6	me to go	4	a collection placed in	4	as a best friend he
3	6	that night i	4	as a best friend	4	collection placed in a box
4	6	you don t	4	asked me to go	4	face who opened the door
5	5	home because he	4	collection placed in a	4	he asked me to go
6	5	opened the door	4	face who opened the	4	studying he asked me to
7	5	to go to	4	he asked me to	4	than me that night i
8	4	a best friend	4	his father and mother	4	the simplicity of his family
9	4	a collection placed	4	me that night i	3	asked me to go home
10	4	after your family	4	placed in a box	3	from his hand then we
11	4	allowed me to	4	simplicity of his family	3	had to move to the
12	4	as a best	4	studying he asked me	3	his father and mother also
13	4	asked me to	4	than me that night	3	if this is only about
14	4	be sold or	4	the simplicity of his	3	loyal than me that night
15	4	best friend he	4	who opened the door	3	me that night i was
16	4	but that night	4	years after your family	3	me to go home because
17	4	clothes i asked	3	father and mother also	3	placed in a box then
18	4	collection placed in	3	father never allowed me	3	the torch from his hand
19	4	face who opened	3	from here i smiled	3	to go home because he
20	4	father and mother	3	from his hand then	3	to move to the city
21	4	from here i	3	go home because he	3	to stay in the regency
22	4	from his hand	3	had to move to		
23	4	from what i	3	heard the process of		
24	4	go to the	3	his face when he		
25	4	he asked me	3	his hand then we		
26	4	his father and	3	i often heard that		
27	4	his father was	3	if this is only		
28	4	i often heard	3	in a box then		
29	4	if this is	3	loyal than me that		
30	4	in a box	3	me to go home		
31	4	in the past	3	move to the city		
32	4	me that night	3	stay in the regency		

33	4	of his family	3	that night i was
34	4	on the table	3	the island where my
35	4	placed in a	3	the regency city i
36	4	simplicity of his	3	the torch from his
37	4	studying he asked	3	the window to the
38	4	than me that	3	this is only about
39	4	the garden and	3	to go home because
40	4	the simplicity of	3	to go to the
41	4	to the city	3	to move to the
42	4	who opened the	3	to stay in the
43	4	years after your	3	torch from his hand
44	3	a box then	3	unfortunately my father never
45	3	almost every night	3	years later to stay
46	3	and mother also	3	you don t change
47	3	as a collection	3	you don t need
48	3	because my family		
49	3	but wearing it		
50	3	ceremony every monday		
51	3	crickets the crickets		
52	3	directly go to		
53	3	don t change		
54	3	don t need		
55	3	face when he		
56	3	father never allowed		
57	3	friend almost years		
58	3	go home because		
59	3	had to move		
60	3	hand then we		
61	3	have many clothes		
62	3	he didn t		
63	3	he was my		
64	3	heard the process		
65	3	here i smiled		
66	3	his face when		
67	3	his hand then		
68	3	i heard the		
69	3	i know he		
70	3	in my mind		
71	3	in the regency		
72	3	is only about		

73	3	island where my
74	3	it was friday
75	3	later to stay
76	3	loyal than me
77	3	minutes two crickets
78	3	move to the
79	3	my father never
80	3	never allowed me
81	3	night i was
82	3	of course he
83	3	often heard that
84	3	outside the island
85	3	outside the window
86	3	regency city i
87	3	stay in the
88	3	tall man with
89	3	the corner while
90	3	the island where
91	3	the process of
92	3	the regency city
93	3	the torch from
94	3	the window to
95	3	this is only
96	3	to go home
97	3	to go it
98	3	to his face
99	3	to join him
100	3	to move to
101	3	to stay in
102	3	to the west
103	3	torch from his
104	3	unfortunately my father
105	3	wearing it to
106	3	where my father
107	3	window to the
108	3	years later to

Appendix 3 – List of lexical bundles produced by professional translators.

No.	Freq	3-word Lexical Bundles	Freq	4-word Lexical Bundles	Freq	5-word Lexical Bundles
1	7	me to go	5	his mother a woman	4	across the room it felt
2	7	my best friend	5	made me feel closer	4	friend s waist i followed
3	6	that night i	5	me to go home	4	from his hand we then
4	5	across the room	5	pond behind his house	4	hand we then walked along
5	5	after your family	5	years after your family	4	he told me to go
6	5	be sold or	4	across the room it	4	his hand we then walked
7	5	behind his house	4	and my best friend	4	his mother a woman who
8	5	he did not	4	friend s waist i	4	me to go home because
9	5	he was my	4	from his hand we	4	studying he told me to
10	5	his mother a	4	from what i heard	4	the torch from his hand
11	5	i told him	4	go home because he	4	to go home because he
12	5	i wanted to	4	hand we then walked	4	told me to go home
13	5	i was here	4	he told me to	4	torch from his hand we
14	5	made me feel	4	his hand we then	3	after your family returned and
15	5	me feel closer	4	mother a woman who	3	all since then so what
16	5	more sincere and	4	night i was here	3	amazed when he recognized me
17	5	mother a woman	4	s waist i followed	3	at all since then so
18	5	of clove or	4	since then so what	3	because he wanted to go
19	5	pond behind his	4	studying he told me	3	did not have many clothes
20	5	the room was	4	the room it felt	3	family had to move to
21	5	the torch from	4	the torch from his	3	go home because he wanted
22	5	there alone we	4	to go home because	3	he did not have many
23	5	to go home	4	to sit on the	3	he was my best childhood
24	5	to look for	4	told me to go	3	him at all since then
25	5	wanted to go	4	torch from his hand	3	home because he wanted to
26	5	years after your	4	we then walked along	3	i took over the torch
27	4	a while he	3	after your family returned	3	made me feel closer to
28	4	a woman who	3	all since then so	3	my family had to move
29	4	and my best	3	amazed when he recognized	3	pond behind his house from
30	4	family had to	3	as a best friend	3	s waist i followed him
31	4	friday the brown	3	at all since then	3	that night i was here
32	4	friend s waist	3	because he wanted to	3	took over the torch from
33	4	from here i	3	behind his house from	3	was my best childhood friend
34	4	from his hand	3	but that night i	3	years after your family returned
35	4	from what i	3	close to his face		

36	4	go home because	3	could be sold or
37	4	hand we then	3	did not have many
38	4	have many clothes	3	family had to move
39	4	he recognized me	3	from here i smiled
40	4	he told me	3	had to move to
41	4	he wanted to	3	have many clothes so
42	4	his hand we	3	he did not have
43	4	home because he	3	he finished studying he
44	4	many clothes so	3	he recognized me he
45	4	night i was	3	he wanted to go
46	4	room it felt	3	he was my best
47	4	s waist i	3	him at all since
48	4	since then so	3	him studying the oil
49	4	sit on the	3	home because he wanted
50	4	studying he told	3	i know that he
51	4	the room it	3	i took over the
52	4	then so what	3	me feel closer to
53	4	then walked along	3	minutes two crickets were
54	4	to go to	3	morning would still be
55	4	to sit on	3	my best childhood friend
56	4	told me to	3	my family had to
57	4	torch from his	3	not all the time
58	4	waist i followed	3	not have many clothes
59	4	was friday the	3	of clove or incense
60	4	we then walked	3	over the torch from
61	4	what i heard	3	so what brought you
62	4	you don t	3	students in our class
63	3	a best friend	3	that night i was
64	3	all since then	3	the corner twisting coconut
65	3	all the time	3	the door looked amazed
66	3	almost every night	3	to go to the
67	3	almost years ago	3	to write on the
68	3	amazed when he	3	took over the torch
69	3	as a best	3	waist i followed him
70	3	at all since	3	was friday the brown
71	3	awkwardness between us	3	was my best childhood
72	3	because he wanted	3	when he recognized me
73	3	best childhood friend	3	who opened the door
74	3	but that night	3	your family returned and
75	3	close to his		
76	3	clove or incense		
77	3	corner twisting coconut		

78	3	could be sold
79	3	did not have
80	3	door looked amazed
81	3	down to write
82	3	enthusiastically but my
83	3	family returned and
84	3	feel closer to
85	3	finished studying he
86	3	go to the
87	3	had to move
88	3	he finished studying
89	3	he then sat
90	3	here i smiled
91	3	him at all
92	3	him studying the
93	3	his family s
94	3	his father was
95	3	his house from
96	3	i followed him
97	3	i heard the
98	3	i know that
99	3	i took over
100	3	in a box
101	3	in our class
102	3	in the corner
103	3	know that he
104	3	minutes two crickets
105	3	morning would still
106	3	my best childhood
107	3	my family had
108	3	my friend s
109	3	need to wait
110	3	never allowed me
111	3	not all the
112	3	not have many
113	3	of his family
114	3	off my feet
115	3	on the table
116	3	opened the door

11 7	3	out of the
11 8	3	outside the window
11 9	3	over the torch
12 0	3	recognized me he
12 1	3	scout uniform he
12 2	3	since morning would
12 3	3	so what brought
12 4	3	students in our
12 5	3	studying the oil
12 6	3	the clove trees
12 7	3	the corner twisting
12 8	3	the door looked
12 9	3	the dry season
13 0	3	this surprise must
13 1	3	to his face
13 2	3	to move to
13 3	3	to neutralize it
13 4	3	to write on
13 5	3	took over the
13 6	3	two crickets were
13 7	3	walked along the
13 8	3	was my best
13 9	3	went to the
14 0	3	what brought you
14 1	3	when he recognized
14 2	3	who opened the
14 3	3	would still be
14 4	3	write on the
14 5	3	you here memories
14 6	3	your family returned