CHAPTER III
RESEARCH METHODOLOGY

This chapter describes the methodology used in the present study, including the research design, data collection, and data analysis procedure.

3.1 Research Design

This study employed a comparative approach using a combination of descriptive qualitative and descriptive quantitative designs to arrive at its intended purposes, which are to analyze and compare the manifestation of LBs between accepted research article introductions (hereafter ARAIs) and rejected research article introductions (hereafter RRAIs). According to Lambert and Lambert (2012), a descriptive qualitative study tends to draw from naturalistic inquiry in which something is analyzed in its natural state to obtain a comprehensive summarization of specific events experienced by individuals or groups of individuals. Hence, in this study, the descriptive qualitative design was employed to determine and code the LBs according to their structures and functions to result in the comprehension of the natural use of language in applied linguistic RAIs.

At the same time, the descriptive quantitative design was employed in the calculation processes. First of all, the utilization of computer software to generate the lists of LBs and their frequencies in this study belonged to the quantitative corpus-based method with the computational technique mentioned in Baker (2010). According to Baker, as computers are able to calculate frequencies and perform statistical tests more quickly and accurately, this technique is able to give researchers access to see linguistic patterns and trends, such as collocational information. Furthermore, the particular aspect of descriptive quantitative design in this study was the calculation of the proportion differences between the two groups using statistical analysis, namely Z-test. This approach has been applied by several previous comparative studies (e.g. Kanafani et al., 2022; Nurcik et al., 2022; Qurratu'aini et al., 2022). Z-test calculation is helpful to point out the exact proportion and determine whether the difference in the proportion of data

Zahra Fadillah Haerunisa, 2022
A COMPARATIVE STUDY OF LEXICAL BUNDLES IN ACCEPTED AND REJECTED APPLIED LINGUISTIC RESEARCH ARTICLE INTRODUCTIONS
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

occurrences in the two compared groups could be considered significant (Kanafani et al., 2022; Qurratu'aini et al., 2022). For this study, Z-test calculation was performed under the following hypotheses:

- H_0 = There is no significant difference in the proportion of LBs occurrences in the two corpora.
- H_1 = There is a significant difference in the proportion of LBs occurrences in the two corpora.

Furthermore, the alpha for the Z-test was set at 0.05 (p-value > 0.05 = H_0 is accepted). The results of analyses in this study would be presented in the form of figures, tables, explanations, and excerpts to enhance the comprehension of the results.

3.2 Data Collection

The data source of this study derived from 30 RAs that were grouped into two corpora comprising 15 introduction sections for each group (table 3.1). The RAs chosen for this study were retrieved from Indonesian Journal of Applied Linguistics (IJAL), an Indonesian Scopus-indexed journal, and were accessed with permission from the journal administrator. The selection of the RAs was done by limiting the submission and/or publication year of 2018-2021. All of the selected RAs were written in English by Indonesian authors to make a fair comparison in revealing the difference between the use of LBs in the RAIs which was hypothesized to influence the RAs to be accepted or rejected.

Table 3.1 Description of the corpora

Corpus	Number of RAIs	Word Counts
Accepted	15	23,211
Rejected	15	21,141
Total	**30**	**44,352**

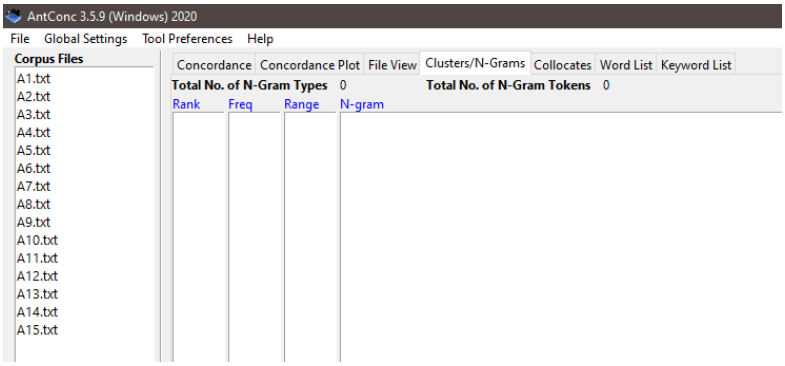
3.3 Data Analysis Procedure

3.3.1 LBs lists identification

LBs identification in this study was conducted using AntConc 3.5.9, computer software developed by Anthony (2020a). The first step began with

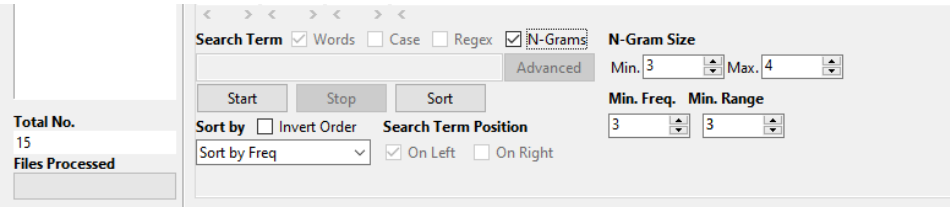
separating each RAI from the selected RAs to be copy-pasted into a notepad and saved in .txt format due to the program demand. The ARAIs files were coded by A1-A15, while the RRAIs by R1-R15. Each corpus was then respectively inserted into the software, to be further identified using the N-Grams feature.

Figure 3.1 AntConc's N-Grams feature



This tool allows the users to find frequently used expressions in the corpus by scanning the determined 'N'-word clusters (e.g. 2/3/4-word clusters), along with the minimum frequency and range of clusters dispersion. As the corpora in this study can be considered small—since this study is focused on specific purposes (i.e. examining LBs in a specific RA section and a specific discipline)—the software was set to generate a list of LBs from each corpus by following the cut-off criteria proposed by Biber and Barbieri (2007), i.e. the word combinations should have a minimum frequency of three and must be distributed across a minimum of three different texts to avoid the authors' idiosyncrasy. Furthermore, this study was focused on the combinations of three- and four-word bundles as they had been tested manageable and able to display relevant expressions (Lee, 2020). The options to set these criteria onto the software can be found at its bottom area.

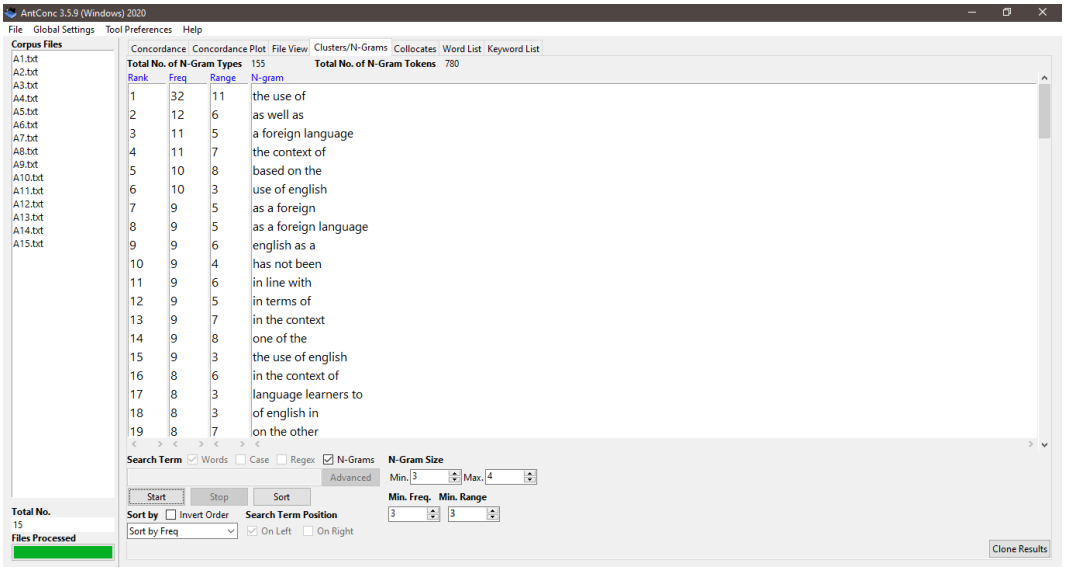
Figure 3.2 Configuration of LBs criteria



As displayed in figure 3.2, the N-Gram size was set at a minimum of 3 and a maximum of 4 as this study was aimed to identify three- and four-word LBs. Both minimum frequency and range were set at 3, in accordance with the cut-off criteria

mentioned earlier. After these cut-off criteria were inserted, the start button was clicked for the software to generate the list of LBs.

Figure 3.3 LBs list identification by AntConc



After the lists of LBs were generated, they were put into Microsoft Excel sheets for further codification and analyses.

3.3.2 Exclusion step

Before the structural and functional analyses were started, the initial lists of LBs were checked to eliminate identical bundles which might create repetitive and exhausting lists. Some of the identical bundles were manually eliminated by following the exclusion criteria proposed by Salazar (2014). The exclusion criteria adopted in this study included:

a) Fragments of other bundles. This criterion eliminates short bundles that are incorporated into longer bundles with the same or similar frequency. For example, a three-word bundle *on the other* and a four-word bundle *on the other hand* occur 8 times in the corpus. The concordance lines show that in all instances, *on the other* occurs as a fragment of *on the other hand*. Therefore, *on the other* is excluded.

b) Bundles ending in articles. This criterion excludes longer bundles ending in articles, if they are already part of shorter bundles. For instance, four-word bundles *in accordance with the* is an extended bundle of *in accordance with* and both have the same frequency of occurrence. Since

the article '*the*' does not provide additional information to the bundle, *in accordance with the* is disregarded.

c) Bundles composed exclusively of function words that have no textual evidence or semantic function, such as *has not been*.

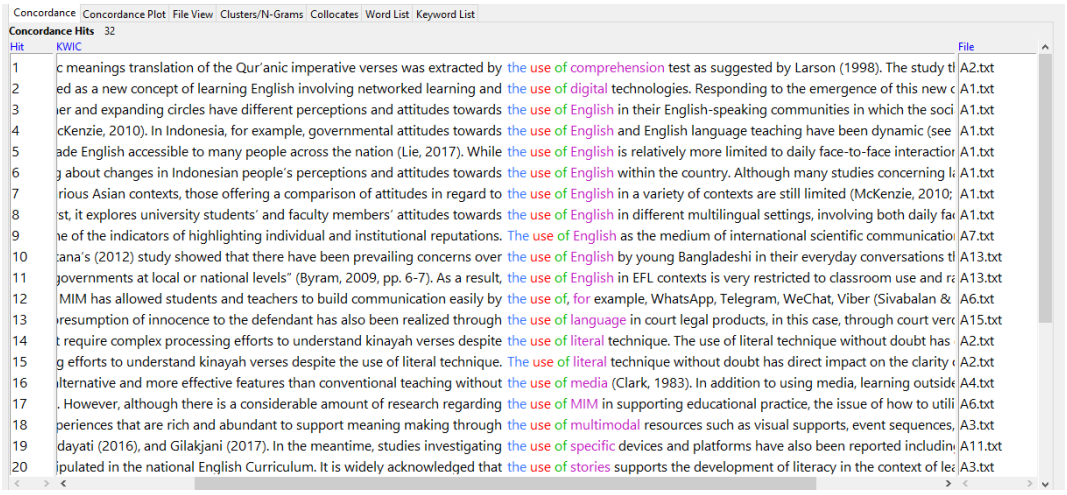
d) Bundles with random numbers, such as *two or more*.

e) Meaningless bundles, such as *et al in*.

3.3.3 LBs classification

The final lists of LBs were then continued to be classified structurally according to Biber et al.'s framework (table 2.1) and functionally according to Hyland's framework (table 2.2) to guide the comparison. To classify the LBs, Antconc's concordance tool was employed. This tool shows search results in a 'KWIC' (KeyWord in Context) format to see how the word combinations are commonly used in a corpus of texts (Anthony, 2020b). The step was just to click on the selected bundles on the list which would automatically lead to the display of the concordance tool feature.

Figure 3.4 Example of AntConc's concordance line display

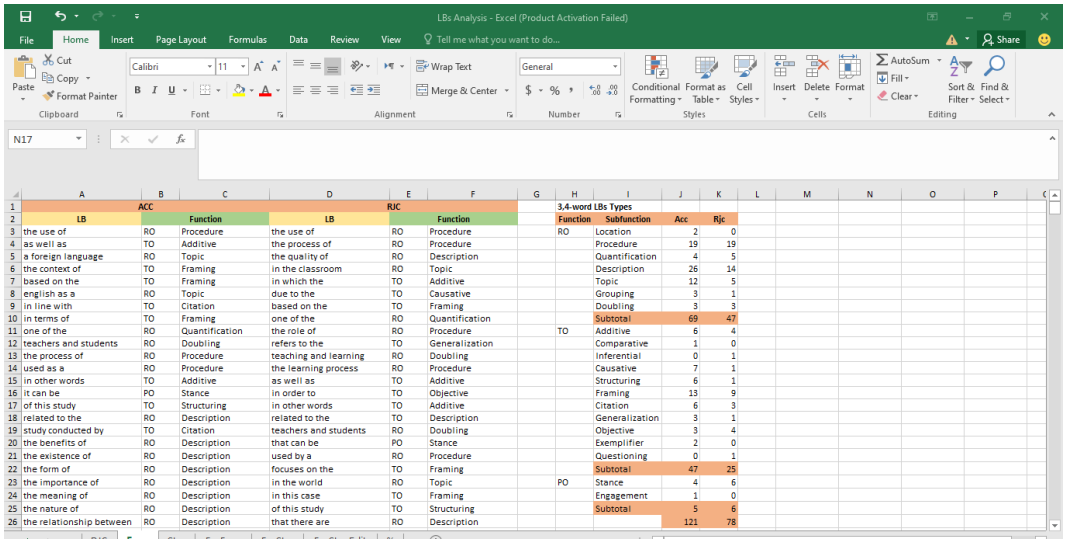


As the structures of bundles were mostly able to be directly retrieved even in their three- or four-word forms, this tool was a great helper to identify LBs functions in which they needed to be seen from how it was used within a context.

3.3.4 Data trustworthiness

As an attempt to prevent a high tendency of subjectivity, this study further adopted the inter-coder reliability assessment in analyzing the data. Author and a chosen inter-coder, who had an ongoing similar study, worked independently to check and code the LBs in the two corpora. The results achieved the inter-coder reliability rate of 95.48% agreement. The remaining disagreements were then discussed and aligned for the refinement of the coding results.

Figure 3.5 Example of LBs analysis in Microsoft Excel



	ACC			RJC			3,4-word LBs Types			
	LB	Function	LB	Function		Function	Subfunction	Acc	Rjc
3	the use of	RO	Procedure	the use of	RO	Procedure	RO	Location	2	0
4	as well as	TO	Additive	the process of	RO	Procedure		Procedure	19	19
5	a foreign language	RO	Topic	the quality of	RO	Description		Quantification	4	5
6	the context of	TO	Framing	in the classroom	RO	Topic		Description	26	14
7	based on the	TO	Framing	in which the	TO	Additive		Topic	12	5
8	english as a	RO	Topic	due to the	TO	Causative		Grouping	3	1
9	in line with	TO	Citation	based on the	TO	Framing		Doubling	3	3
10	in terms of	TO	Framing	one of the	RO	Quantification		Subtotal	69	47
11	one of the	RO	Quantification	the role of	RO	Procedure	TO	Additive	6	4
12	teachers and students	RO	Doubling	refers to the	TO	Generalization		Comparative	1	0
13	the process of	RO	Procedure	teaching and learning	RO	Doubling		Inferential	0	1
14	used as a	RO	Procedure	the learning process	RO	Procedure		Causative	7	1
15	in other words	TO	Additive	as well as	TO	Additive		Structuring	6	1
16	it can be	PO	Stance	in order to	TO	Objective		Framing	13	9
17	of this study	TO	Structuring	in other words	TO	Additive		Citation	6	3
18	related to the	RO	Description	related to the	TO	Description		Generalization	3	1
19	study conducted by	TO	Citation	teachers and students	RO	Doubling		Objective	3	4
20	the benefits of	RO	Description	that can be	PO	Stance		Exemplifier	2	0
21	the existence of	RO	Description	used by a	RO	Procedure		Questioning	0	1
22	the form of	RO	Description	focuses on the	TO	Framing		Subtotal	47	25
23	the importance of	RO	Description	in the world	RO	Topic	PO	Stance	4	6
24	the meaning of	RO	Description	in this case	TO	Framing		Engagement	1	0
25	the nature of	RO	Description	of this study	TO	Structuring		Subtotal	5	6
26	the relationship between	RO	Description	that there are	RO	Description			121	78