

Contents list available at Directory of Open Access Journals (DOAJ) **Al-Ishlah: Jurnal Pendidikan** ISSN: 2087-9490 (Printed); <u>2597-940X</u> (Online)

Journal Homepage: http://www.journal.staihubbulwathan.id/index.php/alishlah



### A Corpus-Informed Materials Evaluation of EFL Textbooks and Teachers' Generated Materials in Indonesian Islamic Universities

Srifani Simbuka<sup>1</sup>, Andi Mukarramah Nagauleng<sup>2</sup>

DOI: 10.35445/alishlah.v13i3.1348

#### **Article Info** Abstract Keywords: The importance of vocabulary in the learning of EFL has been an indisputable Corpus-informed matter in ELT. The inclusion of vocabulary in ELT materials is often accompanied materials; by questions such as what type of vocabulary and how many words to teach Word list; students in diverse ELT contexts. Hence, evaluating the kinds of vocabulary in ELT materials is similarly important. This study demonstrates a considerably GSL: unconventional corpus-informed materials evaluation to assess the suitability of AWL: the vocabulary content of ELT coursebooks taught at the State Islamic Institute in IRSTV Manado with its ESP context. Four ELT coursebooks taught at the four schools of the State Islamic Institute were analysed using corpus-based methods. Special software designed for corpus linguistics studies called e Antwordprofiler was used to count the coverage/frequency of occurrence of three types of Nation' vocabulary classification. The results showed that the examined coursebooks contain sufficient English high-frequency words required by theory. They contain 84.14 % of high-frequency English words enlisted in the General Service List (GSL). However, in terms of academic and technical vocabulary coverage, these coursebooks content were still considered under the theoretically acceptable coverage of at least 12,4 % and 5 %, respectively. For ELT coursebooks used in ESP teaching, such as in field-specific Islamic studies programs, the course books should fulfil the minimum coverage threshold of high frequency, academic and technical vocabulary. Abstrak Kata kunci: Pentingnya kosakata dalam pembelajaran EFL telah menjadi hal yang tak terbantahkan dalam ELT. Pencantuman kosakata dalam materi ELT sering Materi informasi disertai dengan pertanyaan seperti apa jenis kosakata dan berapa banyak kata korpus:

Materi informas korpus; Daftar kata; GSL; AWL; IRSTV Pentingnya kosakata dalam pembelajaran EFL telah menjadi hal yang tak terbantahkan dalam ELT. Pencantuman kosakata dalam materi ELT sering disertai dengan pertanyaan seperti apa jenis kosakata dan berapa banyak kata yang harus diajarkan kepada siswa dalam konteks ELT yang beragam. Oleh karena itu, mengevaluasi jenis kosakata dalam materi ELT juga sama pentingnya. Studi ini menunjukkan evaluasi materi informasi korpus yang sangat tidak konvensional untuk menilai kesesuaian isi kosakata buku pelajaran ELT yang diajarkan di Institut Agama Islam Negeri Manado dengan konteks ESP-nya. Empat buku pelajaran ELT yang diajarkan di empat sekolah Institut Agama Islam Negeri dianalisis menggunakan metode berbasis korpus. Sebuah perangkat lunak khusus yang dirancang untuk studi linguistik korpus yang disebut e Antwordprofiler digunakan untuk menghitung cakupan/frekuensi kemunculan tiga jenis klasifikasi kosakata Bangsa. Hasil penelitian menunjukkan bahwa buku

<sup>1</sup> Institut Agama Islam Negeri Manado, Indonesia

Email: srifani.simbuka@iain-manado.ac.id

2 Institut Agama Islam Negeri Manado, Indonesia

Email: andi.mukarramah@iain-manado.ac.id

Vol.13 (3) December, 2021

Received: November 19, 2021; Received in revised form: December 3, 2021; Accepted: December 26, 2021; Available online: December 31, 2021. This is an open access article under a <u>Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License</u>.

pelajaran yang diperiksa berisi cakupan yang cukup dari kata-kata bahasa Inggris frekuensi tinggi seperti yang disyaratkan oleh teori yang berisi cakupan total 84,14% dari kata bahasa Inggris frekuensi tinggi yang terdaftar dalam Daftar Layanan Umum (GSL). Namun, dalam hal cakupan kosakata akademik dan teknis, isi buku pelajaran ini masih dianggap di bawah cakupan yang dapat diterima secara teoritis masing-masing minimal 12,4% dan 5%. Direkomendasikan bahwa untuk buku-buku pelajaran ELT yang digunakan dalam pengajaran ESP seperti dalam program studi Islam khusus bidang, buku-buku pelajaran harus memenuhi ambang batas cakupan minimum kosakata frekuensi tinggi, akademik dan teknis.

### **INTRODUCTION**

Efforts have been made to ensure that students are exposed to the proper materials in order to enhance their learning in English as a Second Language. There are various methods for determining if materials are suitable for use in ELT programs. It is common practice in English language teaching (ELT) to evaluate materials based on their level of difficulty, compatibility with the program's goals, and cost (Richards, 2001; Tomlinson, 2008, 2012). In most cases, the evaluation of materials is based on linguistic, ELT, and general learning theories. The coursebook evaluation checklist, for instance, was drafted in part by (Mukundan, Nimehchisalem, & Hajimohammadi, 2011).

With the advent of technology and applied linguistics theory, corpus linguistics emerged and made contact with language teaching, specifically ELT. Corpus studies developed its pedagogical wing, and more and more research has shown how corpora contributed to ELT by showing its connection and benefits for the development of ELT syllabus, materials and testing (Campoy, Cubillo, Belles-Fortuno, & Gea-Valor, 2010; I. Nation, 2001; P. Nation, 2016). The contribution of corpora and word lists created based on them in ELT materials development is apparent in their role as the provider of empirically backed vocabulary input for ELT materials. This contribution is a game changer that distinguish traditional materials development that mostly base their selection of vocabulary input for materials by employing more intuitive and subjective measures such as "expert judgement" selection of what vocabulary should go in ELT materials (course books). The vocabulary lists resulting from corpus-based research provide "real language," i.e. "real words/vocabulary" that occur in natural English texts supported by statistical evidence.

Referring to the concept of corpora as the provider of a valuable resource for ELT materials development, this particular study attempted to propose a non-traditional way to assess the suitability of vocabulary content of ELT course books used in an Indonesian Islamic university/college. This study is set to check the coverage of high frequency, academic and technical vocabulary in these ELT coursebooks based on their conformity to the corpus of English as used in Islamic studies discipline (academic and professional context). The purposes stated as follows: (1) To identify and evaluate the properties of vocabulary presented in EFL materials (course books) used in Indonesian Islamic universities.; (2) To evaluate these EFL materials in terms of the extent of their conformity (or disconformity) to the General Service List /GSL (West, 1953), Academic Word List/AWL (Coxhead, 2000), and the Islamic Religious Studies Textbooks Vocabulary/IRSTV List (Simbuka, Hamied, Sundayana, & Kwary, 2019); and (3) To examine whether the analyzed ELT textbooks content has met the properties of 'good materials', in terms of the vocabulary load or the coverage of the categories in (I. Nation, 2001) vocabulary classification as suggested by theory.

To begin, there are "High-Frequency Words"/HFW (Nation, 2001: 15), which are the words that appear most frequently in all kinds of written material. Although content words may also fall into this category in English, the most common function words are articles and prepositions, which are the most common. The General Service Lists (GSL) is the most widely used high-frequency list (1953). The roughly 2000-word families on this list are considered to be the most common general English words (Nation, 2001). The most commonly used list of high-frequency terms in corpus studies, despite its dated nature, is this one (Lessard-Clouston, 2012). The first 1000 and second 1000 word lists from West's GSL were used in this study to identify the most frequently occurring words in the examined

target corpus of English language textbooks used by the English language instructors at the IAIN Manado.

#### **METHODS**

The data of this study came from one primary and 3 secondary data sources due to the nature of the research questions that specifically target the use of a specific word/vocabulary list in a specific domain, i.e. ELT. The data used in this study comprised four ELT textbooks used in most of the departments of the four schools/faculties at the Manado State Institute for Islamic studies or *Institut Agama Islam Negeri* (IAIN Manado). Listed below were the four English language textbooks used as the data of this study:

- 1) Islamic Studies: Readings for the New Millennium was published in 2006 by Pabelan Cerdas Nusantara under the direction of Drs. Giyoto, M.Hum. In the Faculty of Ushuluddin and Dakwah (FUAD), one of the IAIN Manado's English instructors taught it in the Science of Qur'an and Interpretation (IAT) and the Dakwah Management (Dakwah Management). There were 159 pages in the book, but only 155 pages of that were used as the basis for this investigation. There were a total of 33.525 tokens in the book. Data 001 became the official designation for this textbook.
- 2) Three English textbooks previously used in Asnidar's (2021) study on a related topic served as secondary data sources. Thus, in addition to the researchers' primary data, these English textbooks' data was re-used: a) In 2018, UM Press published Economic English Instructional Material Based on Shariah Economy System, a textbook based on secondary data compiled by Syamsul Huda, et al. b) the second English textbook from the secondary data was Oxford University Press' 2005 Ready to use Person to Person Student Book 1; c) English textbook book titled *Business English* composed by E.B. Nikolaenko and published by the Tomsk Polytechnic University in 2008.

The data collection began with selecting textbooks that meet the purpose of this specific study under the purposive sampling frame. Hence, the data in mostly hardcopy English textbooks elaborated in the previous sub-chapter on the data source. The textbooks collected served as the raw data of this present study, therefore they should be processed further by scanning them into pdf formats and further converting them into the text format using a pdf converter program called the Antfile converter. Obeying to the corpus linguistic method, the data was stored as the corpus of English Subject textbooks used at the IAIN Manado that contained 4 sub-corpora that were built from the four English subject textbooks that were obtained from the data collection procedure. Once the hardcopy books had been turned into txt files, then they were ready for analysis.

The data analysis was conducted under the principles of the corpus linguistic method that was mainly conducted under the quantitative tradition. The analysis, however, did not involve any elaborate statistical hypothesis testing using statistical tests like most studies that employ the quantitative research method. Following many studies in corpus linguistics the statistical analysis centred on the descriptive statistic to determine the frequency of occurrence of certain vocabulary categories in the analysed corpus and its sub-corpora. The data analysis procedure was as follows:

#### Data analysis phase I- Data validity check

This phase was not an actual analysis to obtain the answers for any of the three research questions posed in this study. Rather, the first analysis was conducted in order to check the validity of the data, i.e. to check whether or not there were still words that were not recognized by the Wordprofile software. The process of checking the errors in the data also included editing the data in the txt formats to overcome the errors. Hence the analysis could yield more accurate results. The Antconc software was used to generate a word list of all the words that were contained in each of the sub-corpora. The process of data error checking/data validity checking was conducted by entering the scrutinized data set into

the AntConc software and it was then analyzed/checked using the software's wordlist tool. This is shown by the following illustration:

AntConc 3.5.8 (Windov	vs) 2019			[	
ile Global Settings To	ol Preferen	ces Help			
Corpus Files	Conco	rdance Co	ncordan	lot File View Clusters/N-Grams Collocates Word List Keyword List	it .
14 Juli DATA research	Word T	ypes: 416	1	ord Tokens: 33950 Search Hits: 0	
	Rank	Freq	Word	Lemma Word Form(s)	^
	1	2574	the		
	2	1205	to		
	3	1064	of		
	4	767	and		
	5	677	in		
	6	628	а		
	7	626	is		
	8	337	he		
	9	287	that		
	10	286	i i		
	11	286	it		
	12	277	his		
	13	246	what		
		> < >	<	> <	> •
	Search	Term 🖂 🕅	Nords [	Case Regex Hit Location	
				Advanced Search Only 0	
< >				Lemma List Loaded	
lotal No.	Sta	irt	Stop	Sort Word List Loaded	

It is important that the "token settings" are set to the right measurement as the research questions require. The token setting is a tool that translates the definition of what count as "words" or "token" based on specific research questions. In this specific study, the token definition was set to include all combinations of letters that form English words plus the apostrophe (') and hyphen that were often used in words inspired by Arabic language. These settings were suitable based on the research questions that seeks for any "Arabic inspired" English words.

This process resulted in a txt format that lists all words in a sub-corpora, and each was stored as ready to be analysed data.

Global Settings						
Category Character Encoding Colors Files Fonts Tage	Token Definition Settings         Letter Token Classes         Letter [       Uppercase         Number Token Classes         Number Token Classes					
Token Definition Wildcards	Replace Numbers (with %)       Punctuation Token Classes       Punctuation       Connector       Dash       Open					
	Symbol Token Classes Symbol [ Math Currency Modifier Other ] Mark Token Classes Mark [ Non Spacing Spacing Enclosing ]					
	User-Defined Token Class Use Following Definition bcdefghijklmnopqrstuvwxyzABCDEFGHUKLMNOPQRS TUVWXYZ					

### Data Analysis phase II- Word profile of the corpus

In this phase, the actual data analysis was conducted to answer all three research questions in this particular study. This analysis aimed to profile or describe the types of words/vocabulary in a given corpus. The analysis was conducted by uploading all the data onto the Antwordprofiler program and analysed using the GSL, AWL and the IRSTV List as the "filter" or the tools for categorizing the words/vocabulary in the corpus that was being analysed. This is shown in the figure below:



The program was operated by uploading the data under question into the "user file" tab, whilst the "filter" files or the vocabulary list that are used to categorize the words in the examined corpus were uploaded onto the "level list(s)" tab. In the "output settings" tab a researcher could set up the types of information that would be displayed on the result of the word profiler analysis files: the "statistics" button showed the statistical information or frequency of each of the vocabulary types in the analysed corpus. The checked "word types" button provided the result with the list of words types or individual words and their frequencies. Similarly, the "word groups (families)" provided the word groups or families in the corpus and their frequencies. In this present study, however, this feature was not necessary because the focus of this study was not on word families but on words tokens and word types. The result of the analysis would contain a complete frequent list of each of the category of the vocabulary in the analysed corpus, when the "include complete frequency list, which was crucial in this phase of analysis since it describes the detail statistics of the corpus. The result of the analysis would contain full information on which words of the level lists/vocabulary category lists did not existed in the analysed corpus and which did not when the two buttons of "Include words in user file(s) but not in level list(s)" and " Include words in the level list(s) but not in user file(s)" are checked.

Figure 3.4. Antconc WordProfiler analysis for finding the IRSTV						
<ul> <li>AntWordProfiler 1.4.0w (Windows) 2013</li> <li>File Edit Settings Help</li> </ul>		X				
User File(s) Choose View Clear	Results Clear Progress					
ANGGAPLAH READY-DATA research 2020-edit giyoto5.txt	Finished loading user files Creating lexical profile Finished creating lexical profile File name: C:/Users/IAIN 001/Documents/FANI pe/research 2020-IRSTV IN Number of lines: 7881 Number of types: 4128 Number of tokens: 33525 Level list 1: 1_gsl_1st_1000.txt Number of groups (families): 998 Level list 2: 2_gsl_2nd_1000.txt Number of types: 3708 Number of groups (families): 988	•				
Level List(s)         Choose         View         Clear           1_gsl_1st_1000.txt         2 and 1000.txt         3 and 1000.t	Level list 3: 3_awl_570.txt Number of types: 3082 Number of groups (families): 569					
3_awL570.bd IRSTV-RFK-Simbuka et-al.bd	Level list 4: IRSTV-RFK-Simbuka et-al.txt	>				
	Output Settings Sort Settings					

#### Data analysis phase III: -IRSTV checking

The next phase was the analysis for finding the percentage of IRSTV in the data. Again, the Antwordprofiler was used the same way as it was in the previous phase. Following the same procedure, the data was uploaded to the program and analysed using the IRSTV txt file as the additional filter in the "level list" of the program. This process was shown in the figure below:

AntWordProfiler 1.4.0w (Windows) 2013     File Edit Settings Help     User File(s) Choose View Clear     Results Clear Progress     READY-DATA 01 research 2020-gyoto.bd     DAAT 020-DUSMEDS ElsCLEICLEIALd     DAAT 020-DUSMEDS ElsCLEICLEIALd     DAAT 02-DUSMEDS ElsCLEICLEIALd     DAAT 04-PERSON TO PERSON bd     DATA 04-PERSON TO PERSON TO	Figure 3.5. Antconc WordProfiler	r analysis for finding the IRST	V
User File(s)     Choose     View     Clear     Progress       Image: Choose     View     Clear     Clear       Image: Level List(s)     Choose     View     Clear       Image: Choose     View     Clear     Clear       Image: Choose	<ul> <li>AntWordProfiler 1.4.0w (Windows) 2013</li> <li>File Edit Settings Help</li> </ul>		
Image: Control of the search 2020-gryoto bit DATA 04-PERSON LSH bit Data 02-economic English.bit DATA 04-PERSON TO PERSON bit Number of lines: 7881         Number of lines: 7881         Number of tokens: 33525         Level List(s)         Choose       View         Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         Level List(s)       Choose         View       Clear         View	User File(s) Choose View Clear	Results Clear	Progress
	READY-DATA 01 research 2020-glyoto txt         DAAT 03- BUSNESS ENGLISH txt         Data 02-economic English txt         DATA 04-PERSON TO PERSON txt             Level List(s)       Choose         View       Clear         1_gsl_1st_1000.txt       3_avi_570.txt         RSTV-RFK-Simbuka et-al.txt       RSTV-RFK-Simbuka et-al.txt	File name: C:/Users/IAIN 001/Documents/FANI per Number of lines: 781 Number of types: 4128 Number of types: 4128 Number of types: 4128 Number of types: 4114 Number of groups (families): 998 Level list 2: 2_gal_2nd_1000.txt Number of groups (families): 988 Level list 3: 3_avl_570.txt Number of types: 3082 Number of types: 3082 Level list 3: 3_avl_570.txt Number of types: 569 Level list 4: IRSTV-RFK-Simbuka et-al.txt Number of types: 263 Number of groups (families): 263	v/research 2020-IRSTV IN Sort Settings Sort Level 1 Sort Level 2 Batch Process ( No Yes Exit

One important setting that had to be applied in this particular analysis was the "token setting" in the Antwordprof program. In the 'token definition setting", the token must be defined as "user defined class" and the the aposhtrophe mark (') and the dash mark (-) had to be added to the existing

"token definition", so that the program could "read" all the words containing these marks, such as in the words and names in Arabic language that existed in the data. Figure 3.6 below showed this setting:

Figure 3.6. Tok	ten settings in the Wordprofiler Program	
- Global Settings		
Category Color Settings File Settings Tag Settings Token Definition Settings	Token Definition Settings <ul> <li>Default Class (Regular Expression): (?<!--\\p(N)\p(L)]\\p(L)+[\p(L)\p(N)]*</li--> <li>User Defined Class (Regular Expression)</li> <li>[a-zA-Z'-]+</li> </li></ul>	

Once the procedures in the various phases of analysis have been carefully done, the result are ready to be displayed and interpreted. All the results of this analysis are presented and discussed in the following chapter of this report.

#### FINDINGS AND DISCUSSION

The analysis of profiling the types of vocabulary contained in the corpus of English language textbooks used at IAIN Manado that were conducted with the aid of corpus analysis computer programs and tools could yield a number of results depending on the research questions.

#### The vocabulary categories in the Corpus of English Language Textbooks used

Addressing to the first research question of this particular study that is "What types of vocabulary group are covered in the EFL textbooks/coursebooks?" an analysis using the Antwordprofiler program was used in order to profile or to describe the kinds of vocabulary in the entire corpus of English Textbooks at the IAIN Manado and next to profile the kinds of vocabulary that exist in each of the sub-corpora or each of the individual English textbooks data files. The result of the first-world profile step analysis resulted in descriptions of words/vocabulary types as shown in table 1.

## Table. 1 The Result of Wordprofiler Of All 4 English Textbooks/the corpus of Englishlanguage Textbooks

LEVEL	FILE	TOKEN	TOKEN%	TYPE	TYPE%
1	1_gsl_1st_1000.txt	80122	77.97	2481	30.40
2	2_gsl_2nd_1000.txt	6343	6.17	1163	14.25
3	3_awl_570.txt	5203	5.06	1015	12.44
	IRSTV-RFK-Simbuka et-				
4	al.txt	2314	2.25	136	1.67
0	-	8780	8.54	3365	41.24
TOTAL:		102762		8160	

The Table above contains several information including "the level" which represent the kinds of vocabulary that exist in the examined corpus. "The level" or kinds of vocabulary entered at the Antwordprofiler program as the filter to categorize the vocabulary in the corpus were "level 1" was the first 1000 words of the GSL, "level 2" was the second 1000 words of the GSL, "level 3" was the AWL and

the last level "level 3" was the IRSTV compiled by Simbuka et.al (2019). Other information that existed in the Table were token, token percentage, types and types percentage. "Token" means the number of total words that exist in a sub-corpora or simply in a text, "token% or token percentage" was the percentage of the tokens/total number of words that were categorized according to a specific "level" or kinds of vocabulary in the analysed corpus/sub-corpora. "Types" referred to the number of individual words, and "types% or percentage" referred to the percentage of types or individual words that exist in the examined corpus/sub-corpora.

The analysis resulted in a description of the vocabulary of the corpus of English language textbooks at the IAIN Manado with the size of 102.762 tokens (or a total number of words) as follows:

- a) The examined corpus contained 86.465 tokens of the words that belong to the general service list or the list of most frequent English words.
- b) The corpus also contained 5.203 tokens of words that belong to the academic word list or the most frequent English academic words.
- c) There are 2314 words that belong to the special word list called the Islamic Religious Studies Vocabulary (IRSTV), as the name suggests, this is a list of words that are specifically high frequency in Islamic Religious studies textbooks.

In detail, the distribution of the vocabulary categories found in each of the four sub-corpora of the corpus of English language Textbooks is summarized by the table 2.

### Table 2. The result of the vocabulary categories that existed in each of the sub-corporaof the corpus of English language Textbooks

				DATA 02		DATA 03			
		Data 01-				Ū		DATA 04	
LEVEL	FILE	TOKEN	TOKEN%	TOKEN	TOKEN%	TOKEN	TOKEN%	TOKEN	TOKEN%
	$1_gsl_1^{st}$								
1	_1000.txt	26019	77,6	7830	67,73	14589	78,28	31684	81,16
	$2\_gsl\_2^{nd}$								
2	_1000.txt	1929	5,8	661	5,72	1545	8,29	2208	5,66
		27948	83,4	8491	73,45	16134	86,57	33892	86,82
	3_awl								
3	_570.txt	1285	3.83	1116	9.65	514	2.76	2288	5.86
	IRSTV-								
	RFK-								
	Simbuka								
4	et-al.txt	1432	4.27	277	2.40	162	0.87	443	1.13
0	-	2860	8.53	1677	14.51	1826	9.80	2417	6.19
TOTAL:		33525		11561		18636		39040	

The table above indicated that the total number of tokens of each category of vocabulary in each of the sub-corpora differed in number/percentages. In the first sub-corpora there was 27.948 tokens of words fell into the GSL category (GSL 1st 1000 and 2nd 1000), whilst sub-corpora 02/data 02, sub corpora-03 sub-corpora-04, contained 8491, 16134, and 338902 tokens of GSL words, respectively. The second category of vocabulary, the AWL, also existed in the corpus of English Language Textbooks at the IAIN Manado. The first sub-corpora of this corpus, the sub-corpora/Data 01 contained 1285 tokens of the AWL, whilst sub-corpora/Data 02, sub-corpora/Data 03 and sub-corpora/Data 04 contained 1116 tokens, 514 tokens and 2288 tokens of AWL in each of them, respectively.

Based on the category of "types" that existed in the examined corpus of English Language Textbooks that are used in IAIN Manado, there were 3644 "types" or "individual kinds of words. The result also revealed that there were also different numbers of "types" in each of the sub-corpora. Table 4.3. below indicated the number of the "types" that existed in the sub-corpora examined in this particular study. Indicated below, there were 2320 types of the sub-corpora/Data 01, 986 types of GSL in the sub-corpora/Data 02, 1344 types of GSL in sub-corpora/Data 03 and 2556 types of GSL existed in the fourth sub-corpora/Data 04.

			DATA 02		DATA o	A 03			
		Data 01						DATA	
LEVEL	FILE	TYPE	TYPE%	TYPE	TYPE%	TYPE	TYPE%	TYPE	TYPE%
	1_gsl_1 <sup>st</sup>								
1	_1000.txt	1700	41,18	819	40,83	1001	48,06	1883	42,51
	$2_gsl_2^{nd}$								
2	_1000.txt	620	15,02	167	8,33	343	16,47	673	15,19
		2320	56,2	986	49,16	1344	64,53	2556	57,7
	3_awl								
3	_570.txt	466	11.29	329	16.40	83	3.98	671	15.15
	IRSTV-RFK								
	-Simbuka et-								
4	al.txt	108	2.62	37	1.84	11	0.53	42	0.95
0	-	1234	29.89	654	32.60	645	30.96	1161	26.21
TOTAL:		4128		2006		2083		4430	

#### Table 3. The GSL, AWL in the Sub-corpora

# The Islamic Religious Studies Textbooks Vocabulary (IRSTV) Existed in the Corpus of English Language Textbooks used

The second result that was yielded from the word profiler corpus program was the existence of a words special vocabulary list that was specifically developed to measure the number of words related to the specific field of Islamic religious, academic studies in any text. This was the second research question that was set as one of the purposes of this present study, i.e. "How are the IRSTV presented in ELT textbooks used in the Indonesian Islamic university in terms of frequency and their distribution on the text/context?". The result of the AntconcWordprofiler program indicated that overall there were 2314 tokens of IRSTV that existed in the examined corpus. This number was approximately 2,25% of all the words that built this corpus. In terms of word types. There are 136 word types of the IRSTV list that the examined corpus contains. In other words, there was only around 1,67 % word types of the examined corpus that belonged to the IRSTV list.

In detail, the IRSTV in the sub-corpora differs in which the first sub-corpora/data 01 contained 1432 tokens or 4,27 % of the total words in the sub-corpora. Sub-corpora/data 02 contained 277 tokens or 2,40% of the total words in the sub-corpora. Sub-corpora/Data-03 contained only 162 words of the IRSTV or only 0,87 % of its total tokens (number of total words). In the meantime, sub-corpora/data 04 contained 443 tokens of IRSTV or 1,13 % of its total tokens. This information is summarized in table 4.

#### Table 4. The Table of IRSTV Percentage in Each of The Sub-Corpora/Each Textbook

IRSTV-RFK-Simbuka et-al.txt						
	TOKEN	TOKEN%				
Data 01	1432	4.27				
DATA 02	277	2.40				
DATA 03	162	0.87				
DATA04	443	1.13				

The result presented in the previous section of this report is discussed in the following section. To begin the discussion, it is beneficial to revisit the concept of "coverage" of a vocabulary list in a certain corpus. Coverage is often defined as the total amount of a specific vocabulary category in a corpus. Coverage is often calculated and presented in terms of the percentage of the vocabulary category under question in an examined corpus. The statistic showing the percentage of each vocabulary category produces, a general picture of these categories' coverage to which IAIN Manado students' were exposed through the investigated ELT textbooks.

First to be discussed is the coverage of the GSL in the corpus of ELT textbooks taught to IAIN first-year students. As stated in the result section, the examined corpus of ELT textbooks contained 86.465 tokens or 84.14 % of GSL. This result fell in the acceptable coverage of GSL as suggested by major research in corpus linguistics literature that general English words that occurs in high frequency with the highest frequency should occupy the most space in any corpus, with at least 78,4% (Sutarsyah et al., 1994) to 90,6 % in a study by Hirsh (as cited in Nation & Waring, 1997). This also confirmed Zipf's (1949) law that corpora should comprise of a small number of words with a high frequency of occurrence, in which they are the majority of the corpora content (Weisser, 2016).

In a closer examination, each sub-corpora's content should also display the same or at least similar result with the main corpus explained above. The result showed that the first sub-corpora/Data 1 contained 83,36%, the second sub-corpora/Data 2 contained 73,45%, the third sub-corpora/Data 3 contained 86,57% and the last sub-corpora/Data 4 contained 86,82% of combined GSL 1<sup>st</sup> 1000 and 2<sup>nd</sup> 1000 lists. This result showed that almost all but the first sub-corpora/Data 1 confirmed the abovementioned theory. Each of the sub-corpora showed a confirmative percentage of GSL to the required coverage of GSL in theory. The result also showed that the sub-corpora/data 4 contained the highest coverage of GSL of all the sub-corpora examined in this particular study. Meanwhile, the second sub-corpora seemed to fall a little below the popular coverage range mentioned earlier, having only 73,45% tokens of GSL. These findings might seem unmatched with the book's title that suggests that it is meant for teaching ESP on business discipline.

In relation to ELT materials, this result means that most of the sub-corpora or ELT coursebooks examined in this study have met the requirement as good resources for teaching general EFL courses due to their theoretically sufficient GSL coverage. Therefore, the sub-corpora/Data 4, which is an ELT textbook entitled *"Business English"* is the best ELT coursebook to be used for teaching general English courses because of its highest coverage of GSL.

The next point of discussion is the coverage of AWL or the academic words, the one that is labelled as specialized vocabulary in Nation's (2001) classification. The results showed that the overall coverage of AWL in the examined corpus was 5.06%. Each of the sub-corpora contained different coverage: sub-corpora/Data 1 has a 3.83% of AWL, sub-corpora/Data 2 contained 9.65 % AWL, sub-corpora/Data 3 has an AWL coverage of 2.76%, and sub-corpora/Data 4 contained 5.86 % AWL. The coverage of AWL in the examined ELT coursebooks used as the sub-corpora of this particular study fall slightly under the usual coverage of AWL found in other corpus studies, which ranges from 12.04% (Kwary & Artha, 2017) to as much as 27.8% Shabani, and Tazik (2014). The reason for the higher coverage of AWL these corpora of these previous studies can be traced to the differences of the source of texts used in those studies that were solely academic research articles, whilst this current study examined texts from more general ones, not entirely academic texts such as research reports.

On the last filter or stop list, the IRSTV is the marker of technical vocabulary /the third in Nation's (2001) category of vocabulary. The results indicated that the number of technical words in the examined ELT coursebooks was only 2.25%. This coverage may be relatively small compared to the required minimum coverage of technical vocabulary in any text, which is at least 5% (Nation, 2001, 2016).

In general, the content of the corpus of ELT coursebooks used to teach English at IAIN Manado has confirmed the suggested content of "acceptable/normal" English texts, in that it contains acceptable coverage of each of the high-frequency English words measured by the coverage of the GSL, specialized vocabulary represented by the coverage of AWL and technical vocabulary represented by the coverage

of the IRSTV in the corpus and its sub-corpora. The discrepancy in the coverage of each of the vocabulary categories, especially the GSL and the IRSTV in the sub-corpora, is unavoidable due to the fact that the goals of each of these course books as the sub-corpora were also different. In terms of the learning aim and the topic as suggested by the title of each of the course books, one can easily understand that sub-corpora/Data4/coursebook 4 is by nature a coursebook to be used for teaching general English. The other three sub-corpora/data/course books are naturally suitable for ESP contexts. This follows the notion that when a corpus is compiled based on a generic source, it was expected that it has "normal" coverage high coverage of GSL, rather than other types of vocabulary. Similarly, the lower coverage of IRSTV in sub-corpora of general ELT coursebooks sounded 'normal' for this genre in to have very little coverage of technical vocabulary compared to the other sub-corpora.

Problems might arise when the attention is turned to the other sub-corpora/data/ELT coursebooks bearing ESP titles such as "Business English" or even more field-specific ones, i.e. the two ELT coursebooks that bear the title of specific ELT in Islamic university or Syariah Economy context. These two ELT coursebooks contained technical vocabulary content that are considerably lower than the theoretically suggested technical vocabulary content thresh-hold. This means that for these course books to be deemed less suitable for use as text sources for the learning of technical vocabulary in Islamic Religious Studies because it contained insufficient coverage of the IRSTV list word families

#### CONCLUSION

There are several points that can be concluded from the analysis of this study: First, to be considered acceptable texts of English, the examined ELT coursebooks should fulfil the theoretically required coverage of high frequency and academic vocabulary for the coursebooks to be deemed useful for general academic Programs (EAP). Furthermore, course books that are designed to be used in the context of field-specific, such as in the context of Islamic studies, the context of ELT in state Islamic universities where the examined ELT course books are used should fulfil the theoretically suggested coverage of technical vocabulary. Thus, to be considered suitable for ELT in Islamic studies major, these coursebooks should contain at least 5% of technical vocabulary in the Islamic studies field, represented by the content of the Islamic Religious Studies Textbook List.

#### REFERENCES

- Baker, P., Hardie, A., & Mcenery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Campoy, M. C., Cubillo, M. C. C., Belles-Fortuno, B., & Gea-Valor, M. L. (2010). *Corpus-Based Approaches to English Language Teaching*. (M. C. Campoy-Cubillo, B. Belles-Fortuno, & L. Gea-Valor, Eds.). London, New York: A & C Black.
- Coxhead, A. (2000). A New Academic Word List. TESOL Quarterly, 34(2), 213-238.
- Flowerdew, L. (2012). Corpora and Language Education. Palgrave Macmillan.
- Kwary, D. A., & Artha, A. F. (2017). The Academic Article Word List for Social Sciences 1. *Mextesol Journal*, 41(4), 1–11.
- Mukundan, J., Nimehchisalem, V., & Hajimohammadi, R. (2011). Developing an English Language Textbook Evaluation Checklist: A Focus Group Study. *International Journal of Humanities* and Social Science, 1(12), 100–106. https://doi.org/10.19030/cier.v4i6.4383
- Nation, I. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, P. (2016). *Making and Using Word Lists for Language Learning and Testing*. (I. S. P. Nation, Ed.) (e-Book). John Benjamins. https://doi.org/10.1075/z.208
- Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary : Description , Acquisition and Pedagogy* (1st ed., pp. 122–139). Cambridge: Cambridge University Press.

- Richards, J. C. (2001). *Curriculum Development in Language Learning*. Cambridge: Cambridge University Press.
- Shabani, M. B., & Tazik, K. (2014). Coxhead 's AWL across ESP and Asian EFL Journal Research Articles (RAs ): A Corpus-Based Lexical Study. *Procedia - Social and Behavioral Sciences*, 98, 1722–1728. https://doi.org/10.1016/j.sbspro.2014.03.599
- Simbuka, S., Hamied, F. A., Sundayana, W., & Kwary, D. A. (2019). A Corpus-Based Study on the Technical Vocabulary of Islamic Religious Studies. *TEFLIN Journal*, *30*(1), 47–71. https://doi.org/http://dx.doi.org/10.15639/teflinjournal.v30i1/47-71
- Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How Useful is EAP Vocabulary for ESP? A Corpus BAsed Case Study. *RELC Journal*, *25*(2), 34–50.
- Tomlinson, B. (2008). Language Acquisition and Language Learning Materials. English language learning material: A critical review.
- Tomlinson, B. (2012). Materials development for language learning and teaching. *Language Teaching Teach*, *45*(452), 143–179. https://doi.org/10.1017/S0261444811000528
- Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, 28(3), 170–182. https://doi.org/10.1016/j.esp.2009.04.001
- Weisser, M. (2016). *Practical Corpus Linguistics: An Introduction to Corpus-Based Language Analysis* (1st ed.). Malden: Wiley & Blackwell. https://doi.org/10.1002/9781119180180
- West, M. (1953). GSL-AWL word list.pdf. In *A general service list of English*. London: Longman, Green & Co.