

## **THE WORD FREQUENCY LIST OF STUDENTS' MINI THESES COLLECTION**

**Paulina Erawati Paramita**

**Program Studi Bahasa Inggris Akademi Bahasa Asing Widya Dharma Pontianak  
Jalan H.O.S. Cokroaminoto Nomor 445 Telp. 0561-743303 / 0561 - 731966 (Ext. 802 / 805)  
e-mail: paulina\_paramita@yahoo.com**

### **Abstract**

The study of vocabulary is an effective way to help the learners to increase their vocabulary in the target language. Word frequency list helps learners determine the most useful words to study. To systematically study vocabulary, learners may use those words to help them to write a better mini-theses enriched with huge vocabulary knowledge, and then apply the appropriate words on their writings. The collection of twenty-one students' mini-theses from English study program was processed by TextStat to list the frequency word vocabulary. Finally, 1000 frequently used words were presented to see the words selection used by the students in writing their mini-theses. Most of the words found in mini-theses are functional words. This corpus of 1000 words can be utilized as a reference for the students to improve their vocabulary and writing.

**Keywords:** words, vocabulary, frequency, corpus, textstat

### **Abstrak**

Studi kosakata adalah cara yang efektif untuk membantu para siswa meningkatkan kosakata mereka dalam bahasa target. Daftar frekuensi kata membantu mahasiswa menentukan kata-kata yang paling berguna untuk dipelajari. Untuk belajar kosakata secara sistematis, Mahasiswa dapat menggunakan kata-kata itu untuk membantu mereka menulis tugas akhir yang lebih baik yang diperkaya dengan pengetahuan kosakata yang sangat besar, dan menggunakan kata-kata yang tepat pada tulisan-tulisan mereka. Koleksi dua puluh satu tulisan tugas akhir mahasiswa program studi Bahasa Inggris diproses oleh TextStat untuk mendaftarkan frekuensi kata. Akhirnya, 1000 kata yang sering digunakan ditampilkan untuk melihat pemilihan kata yang digunakan oleh mahasiswa dalam menulis tugas akhir mereka. Sebagian besar kata yang ditemukan dalam tulisan tugas akhir mahasiswa adalah kata-kata fungsional. Korpus 1000 kata ini dapat digunakan sebagai referensi bagi siswa untuk meningkatkan kosakata dan tulisan mereka.

**Kata kunci :** kata, kosakata, frekuensi, korpus, textstat

## **INTRODUCTION**

One of the most significant components for students in writing is vocabulary. It implies that students must know the vocabulary they use in their writings. There are some reasons why vocabulary is significant. Lessard-Clouston (2013) and Alqahtani (2015) expressed that vocabulary is central to English language teaching because without sufficient vocabulary students cannot understand others or express their own ideas. Vocabulary mastery is needed to express our ideas and to be able to understand other people's sayings. Nation (2001) realised that the acquisition of vocabulary is essential for successful second language use and plays a vital role in the formation of complete spoken and written texts. In English as a second language (ESL) and English as a foreign language (EFL) learning vocabulary items plays a vital role in all language skills (i.e. listening, speaking, reading, and writing). As Schmitt (2010) noted, "learners carry around dictionaries and not grammar books". Teaching vocabulary helps students understand and communicate with others in English. Vocabulary acquisition affects language skill's development. Vocabulary is the critical element in the second language skills. Vocabulary expansion is essential for learners to reach proficiency in English. In language learning and teaching, vocabulary knowledge is central to communicative competence and the acquisition of a language. Students learn vocabulary through course books, and the size of their vocabulary determines their language performance. Vocabulary is the most critical element to build their English knowledge and has a substantial effect on their future language learning progress.

Another area of study relating to vocabulary coverage is the issue of word frequency. It is often stated that learner should focus on learning the most common 2,000 or so words of a language since these account for 80% or so of most contexts (Nation, 2002). Nation also calculated that learner need 3,000-word families to feel even somewhat comfortable reading, and 5,000 to be comfortable in most situations. This is based on the assumption that the texts should contain 98% known words. It is considered as the standard of comfortable reading for students to achieve to make them able to maximize their vocabulary learning.

Words can be difficult because of factors like frequency (Chen & Truscott, 2010). Therefore, it is necessary to discover the frequency of specific words occurs in the students' mini-theses to identify the factors that make words difficult. It is important as it will analyze the difficulty level of an individual word for effective learning of vocabulary acquisition.

The word frequency lists and many others are available online and can be used by material writers, researchers, and students to drill the most useful words. Word frequency list is the foundation on which intentional study of vocabulary lie. Thus, this study used frequency list vocabulary generated from the collection of students' mini-theses of Foreign Language Academy of Widya Dharma Academic Year 2011 to see the frequency of words in their writings.

The collection of students' mini-theses is in the form of corpus. Desagulier (2017) stated that a corpus is a digital text or collection of texts. A corpus may contain texts in a single language or in multiple languages. Browne (2013) created a new list of important high-frequency vocabulary words for second language learners of English. The New General Service List (NGSL) is a list of approximately 2,800 core vocabulary words published. First published in early 2013, the NGSL provides over 92% coverage for most general English texts (the highest of any corpus-derived general English word list to date). Word frequency lists, the focus of these mini-theses, are built from corpora.

In 2000, Avril Coxhead of Massey University created Academic Word Lists (AWL) which began after the 2,000th word on the BNC. In particular, these academic words lists, which contained 570 word families, are meant for students aiming to study at an English speaking tertiary institution. While a great number of corpora presently exist from a variety of sources, the British National Corpus (BNC) is one of the most comprehensive and most utilized for vocabulary research. The BNC contains 100 million samples of both written and spoken language; 75% of the written texts were chosen from informative writings, i.e., fields of applied sciences, arts, finance, while 25% of the written texts were imaginative, that is, literary and creative works (Leech et al., 2001, p. 1). Accordingly, a word frequency list of the BNC is a useful tool for researchers of

English, in particular, British English. As with other word frequency lists, the first 2,000 words on the BNC are considered useful for general language purposes, such as reading simple texts and for everyday conversations. Nation (2001) believes that language learners need a minimum vocabulary size of 2000 word families and a good

The New Academic Vocabulary List (AVL) (Gardner & Davies 2014) was compiled on the basis of 120 million words from nine academic disciplines. With the publication of the 2013 NGSL, which is based on the concept of modified lexemes, a new list of important high frequency academic words that would fit tightly together with the NGSL was created and therefore published the New Academic Word List (NAWL), which was based on a 288 million word academic corpus. Like the NGSL, the NAWL provides a bit better coverage than the original AVL.

The word frequency list and many others are available online and can be used by material writers, researchers, and students to drill the most useful words. Word frequency list is the foundation on which intentional study of vocabulary lies. Thus, this study used frequency list vocabulary generated from the collection of students' mini-theses of Foreign Language Academy of Widya Dharma Academic Year 2011 to see the frequency of words in their writings.

There were some previous researches on words frequency list conducted by researchers. Vongpumivitch, Huang, and Chang (2009) explored the use of words in Coxhead's (2000) Academic Word List (AWL) in journal articles in the field of applied linguistics. A 1.5 million-word corpus called the Applied Linguistics Research Articles Corpus (ALC) was created for this study. Most of the AWL word forms belong to the word families included in Coxhead's first and second sublists. The non-AWL word forms are specialized terms in applied linguistics, terms related to language education and research methodology, and adjectives/nouns indicating countries/languages. Brysbaert, Mandera, and Keuleers (2018) observed that high-frequency words are processed more efficiently than low-frequency words. Matsuoka and Hirsh (2010) also researched analysis of an ELT course book. Their Research findings suggested that it only provided minimal

opportunities for students to develop vocabulary knowledge beyond frequency and academic words.

Taking advantage of large modern corpora, the current study explores the word frequency of the collection of students' mini-theses academic year 2011 in terms of different word classes. The word frequency was investigated at the same time to describe the variety of vocabulary used by the students to write their mini-theses. It is expected the students would increase their vocabulary in the target language. The words choice and variety were shown to help the students in writing their mini-theses.

## **METHODOLOGY**

This research used corpus as the only data because it aims to look the number of words in the mini-theses of the 21 Foreign Language Academy of Widya Dharma students. Conrad (1999) stated that there are three important characteristics of corpus-based research. First, corpus-based research uses a principled collection of naturally occurring texts. The second characteristic is that corpus-based research uses a computer to analyze. In this kind of research, it will not be feasible to analyze the corpora or many complex features without a computer. However, in some cases, it needs human judgments when the features are ambiguous. The third characteristic is that corpus-based research includes both quantitative analyses and functional interpretation of language use.

In this research, the corpus was in written form. Document analysis was used as data collection instrument since the corpus was obtained from the students' mini-theses. Qualitative research was used in this paper to describe the word frequency list the 21 students' mini-theses of Foreign Language Academy of Widya Dharma academic year 2011. Computer technology was required to build the corpora of the 21 students' mini-theses. TextSTAT was used in listing the words frequency list of the 21 students' mini-theses.

A corpus-based study is carried on in this paper to investigate the word frequency lists of Foreign Language Academy of Widya Dharma student's mini-theses in Pontianak. The first step is to collect the mini-theses written by twenty-

one students of Foreign Language Academy of Widya Dharma academic year 2011. The mini-theses were in the form of PDF format. To obtain the corpus, the PDF file was converted into a TXT file. It was done to make it easier to analyze. After that, each file was analyzed using software. The software used as the data analysis instrument is TextSTAT for retrieval analysis. TextSTAT is a simple programme for the analysis of texts. It reads plain text files (in different encodings) and HTML files (directly from the internet) and it produces word frequency lists and concordances from these files. It is a freeware available at <https://sourceforge.net/projects/textstat/>. The software was used to count the type frequency of each corpus. After getting the list of the words, the researcher sorted 1000 the most frequent words occurred in the mini-theses. The analysis result was used by interpreting the analysis result of the word frequency and the context of each repeated word. In brief, the data analysis process was explained in the following table.

**Table 1. Data Analysis Process**

<b>Phase</b>	<b>Process</b>	<b>Result</b>
1.	Collecting data. (Foreign Language Academy of students' mini-theses ). The data were scanned first, to make it into softcopy.	Data (PDF file)
2.	Converting the PDF file to a TXT file.	Corpus
3	Checking txt files word by word to make sure that all of the words in the mini-theses are copied entirely and readable.	Corpus
4.	Analyzing the corpus of the file using the TextStat program.	Processed data
5.	Interpreting the processed data.	Description the processed data

## **FINDINGS AND DISCUSSION**

### **Findings**

The table shows 1000 most frequent words. The words are found in twenty-one students' mini-theses of Foreign Language Academy of Widya Dharma

academic year 2011. The words are ranked from the most frequent to the least frequent.

**Table 2. List of 1000 words frequently found in twenty-one students' mini-theses of Foreign Language Academy of Widya Dharma academic year 2011**

No	words	Frequency										
1	the	13387	56	she	398	112	where	195	168	its	143	
2	of	5900	57	used	388	113	into	194	169	method	143	
3	to	4589	58	also	382	114	characters	193	170	me	142	
4	in	4389	59	harry	381	115	were	191	171	problem	142	
5	and	4376	60	perfect	380	116	him	187	172	part	141	
6	a	3371	61	when	375	117	know	186	173	population	141	
7	is	3219	62	all	347	118	action	185	174	fact	140	
8	that	1874	63	widya	347	119	find	185	175	main	140	
9	are	1424	64	dharma	344	120	games	184	176	Foreign	138	
10	this	1320	65	hadk	322	121	tenses	184	177	answer	138	
11	for	1152	66	some	321	122	very	183	178	than	137	
12	as	1116	67	analysis	320	123	first	182	179	learn	136	
13	students	1102	68	more	318	124	may	181	180	evan	135	
14	it	1079	69	simple	315	125	plot	181	181	voice	134	
15	by	1050	70	paper	306	126	mean	180	182	verb	133	
16	be	930	71	continuous	304	127	literature	179	183	each	132	
17	on	930	72	use	298	128	new	179	184	must	132	
18	tense	914	73	what	298	129	total	179	185	fourth	131	
19	or	868	74	data	295	130	excellent	177	186	subject	131	
20	research	821	75	reading	294	131	teaching	177	187	technique	129	
21	i	816	76	been	293	132	based	176	188	theme	129	
22	not	801	77	them	275	133	most	176	189	point	128	
23	with	788	78	other	272	134	elements	174	190	teacher	128	
24	from	769	79	would	271	135	correlation	173	191	student	127	
			80	story	270	136	jakarta	173	192	mastery	126	
25	writers	768	81	because	268	137	work	170	193	against	125	
26	english	726	82	novel	265	138	after	169	194	still	125	
27	he	712	83	pontianak	262	139	our	169	195	then	125	
28	have	681	84	vocabulary	259	140	character	168	196	want	125	
29	you	681	85	foreign	257	141	writing	167	197	sample	124	
30	can	643	86	do	252	142	number	166	198	school	123	
31	which	603	87	so	252	143	person	166	199	online	122	
32	they	597	88	clauses	251	144	second	165	200	study	122	
33	past	582	89	mr	248	145	up	165	201	knowledge	120	
34	at	580	90	average	244	146	hypothesis	164	202	help	119	
35	an	548	91	using	244	147	should	164	203	opinion	119	
36	movie	546	92	people	243	148	form	163	204	following	118	
37	score	544	93	chapter	240	149	academic	162	205	post	118	
38	his	523	94	between	237	150	conflict	160	206	take	118	
39	their	523	95	how	237	151	make	159	207	us	118	
40	was	514	96	table	237	152	only	158	208	just	117	
41	who	490	97	semester	233	153	place	158	209	could	116	
42	about	476	98	year	230	154	well	158	210	through	116	
43	said	476	99	poor	228	155	any	156	211	way	116	
44	time	474	100	academy	227	156	words	154	212	world	115	
45	there	464	101	if	224	157	give	153	213	made	114	
46	language	463	102	two	223	158	such	153	214	see	114	
47	we	461	103	children	220	159	section	151	215	passive	112	
48	test	458	104	many	218	160	according	149	216	these	112	
49	will	452	105	no	217	161	class	149	217	those	112	
50	has	440	106	my	216	162	important	146	218	dobby	111	
51	but	438	107	get	215	163	setting	146	219	information	111	
52	her	414	108	out	208	164	ability	145	220	life	111	
53	good	412	109	learning	205	165	your	145	221	parents	110	
54	present	406	110	future	198	166	result	144	222	did	108	
55	one	398	111	like	195	167	book	143	223	different	107	

224	government	107	304	happy	77	384	six	62	464	holy	51
225	correct	105	305	often	77	385	skills	62	465	hugh	51
226	great	104	306	play	77	386	both	61	466	rhymes	51
227	family	103	307	say	77	387	events	61	467	takes	51
228	articles	102	308	analyzing	76	388	high	61	468	wizard	51
229	before	102	309	consists	76	389	name	61	469	already	50
230	doing	102	310	kate	76	390	parts	61	470	film	50
231	example	102	311	however	75	391	situation	61	471	human	50
232	father	102	312	seen	75	392	things	61	472	kindergarten	50
233	given	102	313	above	74	393	yudhoyono	61	473	lecturers	50
234	over	102	314	don	74	394	department	60	474	libya	50
235	especially	101	315	go	74	395	harmon	60	475	meeting	50
236	here	101	316	understand	74	396	literary	60	476	moral	50
237	individual	101	317	always	73	397	playing	60	477	parties	50
238	intrinsic	101	318	mrs	73	398	problems	60	478	strong	50
239	therefore	100	319	word	73	399	v2	60	479	suggestions	50
240	verbs	100	320	adjectives	72	400	able	59	480	away	49
241	coalition	99	321	come	72	401	definition	59	481	better	49
242	conclusion	99	322	qaddafi	72	402	dialogue	59	482	education	49
243	lyla	99	323	variable	72	403	every	59	483	malfoy	49
244	researchers	98	324	york	72	404	kids	59	484	night	49
245	shows	98	325	press	71	405	sister	59	485	performance	49
246	game	97	326	review	71	406	support	59	486	results	49
247	conversation	96	327	structure	71	407	anna	58	487	teach	49
248	does	96	328	Tell	71	408	business	58	488	try	49
249	indonesia	96	329	third	71	409	four	58	489	classroom	48
250	items	96	330	describe	70	410	group	58	490	contains	48
251	analyze	95	331	edition	70	411	love	58	491	express	48
252	another	95	332	farmer	70	412	scope	58	492	keeper	48
253	idea	95	333	old	70	413	around	57	493	long	48
254	party	94	334	accessed	69	414	certain	57	494	march	48
255	view	94	335	down	69	415	criteria	57	495	speaking	48
256	years	94	336	errors	69	416	got	57	496	yeo	48
257	hope	93	337	fifth	69	417	happened	57	497	young	48
258	purpose	93	338	independent	69	418	mastering	57	498	advantages	47
259	university	93	339	list	69	419	rowling	57	499	developing	47
260	gpa	92	340	nursery	69	420	son	57	500	hagrid	47
261	order	92	341	related	69	421	condition	56	501	including	47
262	scores	92	342	stated	69	422	indonesian	56	502	james	47
263	since	92	343	why	69	423	materials	56	503	june	47
264	too	92	344	am	68	424	next	56	504	kind	47
265	discussion	91	345	follows	68	425	role	56	505	negative	47
266	grammar	91	346	noun	68	426	usually	56	506	public	47
267	introduction	91	347	renisenb	68	427	alternative	55	507	ragnarok	47
268	while	91	348	until	68	428	beginning	55	508	society	47
269	gamers	90	349	descriptive	67	429	shown	55	509	sounds	47
270	three	90	350	holman	67	430	speech	55	510	whole	47
271	understandin	90	351	irregular	67	431	william	55	511	climax	46
272	adjective	88	352	material	67	432	adjective-	54	512	contents	46
273	article	88	353	own	67	433	called	54	513	etc	46
274	best	88	354	without	67	434	forces	54	514	let	46
275	end	88	355	country	66	435	found	54	515	set	46
276	questions	88	356	nim	66	436	house	54	516	several	46
277	weasley	88	357	president	66	437	period	54	517	techniques	46
278	daily	87	358	ron	66	438	sentence	54	518	thought	46
279	readers	87	359	during	65	439	state	54	519	whether	46
280	statements	87	360	international	65	440	taylor	54	520	effect	45
281	teachers	86	361	man	65	441	tolson	54	521	finally	45
282	last	85	362	uncle	65	442	although	53	522	goes	45
283	now	85	363	adverbs	64	443	background	53	523	members	45
284	told	85	364	appendix	64	444	college	53	524	null	45
285	being	84	365	improve	64	445	death	53	525	overcoming	45
286	books	84	366	process	64	446	done	53	526	relationship	45
287	never	84	367	taken	64	447	l	53	527	social	45
288	sentences	83	368	author	63	448	letter	53	528	adverb	44
289	clause	82	369	entitled	63	449	positive	53	529	august	44
290	collecting	82	370	home	63	450	tom	53	530	central	44
291	music	82	371	lot	63	451	asked	52	531	ideas	44
292	read	82	372	much	63	452	case	52	532	looking	44
293	right	82	373	percentage	63	453	debate	52	533	louis	44
294	suggestion	82	374	statement	63	454	having	52	534	pronunciatio	44
295	even	81	375	think	63	455	mother	52	535	really	44
296	need	81	376	v3	63	456	newspaper	52	536	s.pd	44
297	same	81	377	vernon	63	457	object	52	537	activity	43
298	back	80	378	comes	62	458	skill	52	538	american	43
299	level	79	379	day	62	459	tells	52	539	defined	43
300	means	79	380	going	62	460	accepted	51	540	dictionary	43
301	written	79	381	meaning	62	461	differences	51	541	little	43
302	something	78	382	potter	62	462	director	51	542	minister	43
303	achievement	77	383	show	62	463	february	51	543	net	43

544	activities	42	624	making	38	704	political	33	784	later	29
545	again	42	625	moment	38	705	practice	33	785	methodology	29
546	attention	42	626	published	38	706	samantha	33	786	might	29
547	body	42	627	side	38	707	terms	33	787	participle	29
548	boy	42	628	washington	38	708	town	33	788	re	29
549	chengcai	42	629	writer	38	709	white	33	789	rejected	29
550	control	42	630	bad	37	710	women	33	790	religious	29
551	face	42	631	came	37	711	bring	32	791	suggest	29
552	hand	42	632	chamber	37	712	care	32	792	suster	29
553	local	42	633	development	37	713	choose	32	793	system	29
554	makes	42	634	fifa	37	714	christie	32	794	testing	29
555	message	42	635	finding	37	715	early	32	795	various	29
556	page	42	636	findings	37	716	easily	32	796	whose	29
557	protagonist	42	637	general	37	717	feeling	32	797	write	29
558	provide	42	638	learners	37	718	festival	32	798	anything	28
559	tools	42	639	oxford	37	719	groups	32	799	art	28
560	approach	41	640	bambang	36	720	implementati	32	800	ask	28
561	become	41	641	bibliography	36	721	intended	32	801	award	28
562	billy	41	642	common	36	722	interesting	32	802	black	28
563	fitzgerald	41	643	countries	36	723	job	32	803	completion	28
564	friends	41	644	fiction	36	724	knows	32	804	create	28
565	grade	41	645	golkar	36	725	mistakes	32	805	dan	28
566	hermione	41	646	jr	36	726	possible	32	806	denouement	28
567	major	41	647	rising	36	727	saying	32	807	gave	28
568	measurement	41	648	talk	36	728	says	32	808	imhotep	28
569	particular	41	649	topic	36	729	secrets	32	809	meet	28
570	percent	41	650	types	36	730	small	32	810	middle	28
571	pks	41	651	within	36	731	sound	32	811	needed	28
572	please	41	652	asean	35	732	uses	32	812	others	28
573	sara	41	653	child	35	733	age	31	813	pictures	28
574	source	41	654	debaters	35	734	ahmadiyah	31	814	places	28
575	took	41	655	drama	35	735	communicati	31	815	police	28
576	wants	41	656	gives	35	736	content	31	816	pssi	28
577	abstract	40	657	ha	35	737	far	31	817	rather	28
578	basic	40	658	happen	35	738	front	31	818	reader	28
579	change	40	659	hard	35	739	ho	31	819	rights	28
580	easy	40	660	himself	35	740	java	31	820	soon	28
581	event	40	661	interested	35	741	money	31	821	sunday	28
582	feel	40	662	jazz	35	742	novacek	31	822	wiley	28
583	george	40	663	kinds	35	743	power	31	823	begins	27
584	head	40	664	known	35	744	real	31	824	car	27
585	highest	40	665	lockhart	35	745	reshuffle	31	825	clear	27
586	internet	40	666	look	35	746	stupid	31	826	days	27
587	least	40	667	states	35	747	subjects	31	827	democratic	27
588	off	40	668	stop	35	748	supervisor	31	828	examples	27
589	room	40	669	thank	35	749	violence	31	829	explain	27
590	speak	40	670	thing	35	750	watching	31	830	follow	27
591	taking	40	671	v1	35	751	aspects	30	831	formed	27
592	text	40	672	yahmose	35	752	belief	30	832	fred	27
593	today	40	673	yang	35	753	cabinet	30	833	friday	27
594	ways	40	674	below	34	754	cannot	30	834	god	27
595	yes	40	675	city	34	755	degree	30	835	hospital	27
596	cambridge	39	676	classified	34	756	didn	30	836	low	27
597	consist	39	677	complete	34	757	divided	30	837	maria	27
598	east	39	678	drs	34	758	food	30	838	master	27
599	itself	39	679	falling	34	759	fu	30	839	media	27
600	languages	39	680	final	34	760	hori	30	840	methods	27
601	nick	39	681	five	34	761	library	30	841	outside	27
602	penelitian	39	682	free	34	762	live	30	842	preparation	27
603	phonetic	39	683	giving	34	763	mind	30	843	sources	27
604	question	39	684	hogwarts	34	764	morning	30	844	struggle	27
605	rush	39	685	lowe	34	765	once	30	845	term	27
606	sir	39	686	mastered	34	766	position	30	846	themes	27
607	someone	39	687	nofret	34	767	procedure	30	847	themselves	27
608	sometimes	39	688	put	34	768	rebels	30	848	thursday	27
609	taught	39	689	reference	34	769	scene	30	849	tuesday	27
610	team	39	690	short	34	770	tool	30	850	validity	27
611	under	39	691	steps	34	771	week	30	851	ve	27
612	useful	39	692	theory	34	772	went	30	852	wanted	27
613	variables	39	693	united	34	773	xile	30	853	working	27
614	water	39	694	works	34	774	air	29	854	yet	27
615	worked	39	695	buy	33	775	among	29	855	big	26
616	wrong	39	696	cassavetes	33	776	baby	29	856	considered	26
617	conclude	38	697	dudley	33	777	becomes	29	857	ever	26
618	decision	38	698	getting	33	778	bed	29	858	guide	26
619	develop	38	699	interest	33	779	collected	29	859	happens	26
620	elementary	38	700	john	33	780	daughter	29	860	joined	26
621	eyes	38	701	keep	33	781	don't	29	861	key	26
622	formula	38	702	news	33	782	gets	29	862	letters	26
623	hear	38	703	picture	33	783	inside	29	863	lim	26

864	lives	26	899	stage	25	934	thomas	24	969	sure	23
865	lowest	26	900	symbols	25	935	together	24	970	terminology	23
866	meets	26	901	theories	25	936	agatha	23	971	thanks	23
867	offer	26	902	type	25	937	alive	23	972	times	23
868	pearson	26	903	waiting	25	938	available	23	973	transcription	23
869	richard	26	904	believe	24	939	category	23	974	trying	23
870	special	26	905	border	24	940	company	23	975	viewers	23
871	wijaya	26	906	characterizati	24	941	conclusions	23	976	won	23
872	woman	26	907	coefficient	24	942	conducted	23	977	ago	22
873	actions	25	908	conduct	24	943	denzel	23	978	apply	22
874	actually	25	909	declaration	24	944	design	23	979	avoid	22
875	adalah	25	910	diploma	24	945	difference	23	980	chairman	22
876	advanced	25	911	effective	24	946	ending	23	981	close	22
877	answers	25	912	essential	24	947	enough	23	982	conditions	22
878	categorized	25	913	everything	24	948	except	23	983	developed	22
879	choice	25	914	fell	24	949	fall	23	984	dialogues	22
880	classes	25	915	forms	24	950	feels	23	985	element	22
881	collection	25	916	further	24	951	fuel	23	986	experience	22
882	described	25	917	girl	24	952	heard	23	987	explanation	22
883	difficult	25	918	include	24	953	hold	23	988	friend	22
884	due	25	919	involved	24	954	hours	23	989	fun	22
885	fighting	25	920	judge	24	955	inc	23	990	gerindra	22
886	focus	25	921	killed	24	956	jerry	23	991	helps	22
887	handbook	25	922	large	24	957	leader	23	992	meyer	22
888	held	25	923	named	24	958	measure	23	993	movies	22
889	increase	25	924	oil	24	959	messages	23	994	narrator	22
890	learned	25	925	pakpahan	24	960	month	23	995	needs	22
891	left	25	926	pay	24	961	narrative	23	996	permission	22
892	listening	25	927	provides	24	962	near	23	997	physical	22
893	loves	25	928	purposes	24	963	price	23	998	players	22
894	motivation	25	929	quickly	24	964	principal	23	999	points	22
895	office	25	930	run	24	965	standard	23	1000	previously	22
896	qualification	25	931	santa	24	966	start	23			
897	ryhmes	25	932	saturday	24	967	started	23			
898	solve	25	933	sixth	24	968	studying	23			

The word *the* is the most frequent type occurred in the course book. It occurred 13387 times. The word *of* is the second most frequent type with its occurrence 5900 times. The word *to* is the third most frequent type occurred in the course book which occurred 4589 times. The students must know the use of those words very well as they occurred the most in the students' mini-theses.

The finding of this research shows that most of the words found in mini-theses are functional words. Functional words might be prepositions, pronouns, auxiliary verbs, conjunctions, grammatical articles, or particles. As stated by Thomas, Pfister, and Peterson (2004), high-frequency words include functional/structural words. Functional words determine the meaning of a phrase or a chunk. Although functional words do not carry information, they contribute to meaning in some ways. Moreover, they are essential for understanding because they define the relationship between two words. It might lead to the misconception if someone does not know the different meaning and use. Mastering functional words, students can understand the word used in a sentence.

## Discussion

The word *the* in the students' mini-theses occurred 13387 times. *The* is part of articles. It is a functional word. Nation and Waring (1997: 8-9) said that the word *the* has

important value in English since it is useful and commonly used. At daily used, the word *the* occurs so frequently about seven percents of the words on a page in written English and the same proportion of the words in a conversation are repetitions of the word *the*. It showed that the word *the* in the students' mini-theses is important and useful as it the frequently occurs in the text.

The words *the, of, to, in, and, a, is, that, are, and this* are the top ten of highest frequency in the students' mini-theses. Their frequency of occurrences should bring positive effect for students although there is no exact number of frequencies which contributes vocabulary acquisition. The words repeated ten times in minimum would be enough to have an effect on students' vocabulary acquisition.

Technical words refer to the topic of the students' research. The words used in the students' mini-theses were very common and limited. The use of technical terms related to the field of study was very poor as they should use the terms related to English language learning, education, linguistics, etc. The variety of words was not many.

## **CONCLUSION**

The finding of this research shows that most of the words found in mini-theses are functional words. Functional words might be prepositions, pronouns, auxiliary verbs, conjunctions, grammatical articles, or particles. The words *the, of, to, in, and, a, is, that, are, and this* is the top ten of highest frequency in the students' mini-theses. The word *the* in the students' mini-theses occurred 13387 times.

The use of technical terms related to the field of study to make their writing more accurate is strongly suggested. The choice and variety of words should also be taken into consideration to create qualified writing. The students can use this corpus as a reference for them to select and determine the proper words to be used in their mini-theses writing.

The author should let the learners to always broaden their vocabulary in order to follow their English need. If they could not improve their vocabulary, they would not be able to follow their English lesson well.

The last recommendation is for future researchers. This research analyzed the word frequency. However, the limitation of time and cost led the researcher to have this research only for twenty-one students' mini-theses from English study program. Research involves

more significant numbers of mini-theses from many academic years, and various study programs is essential to conduct for a better result.

## REFERENCES

- Alqahtani, M. 2015. *The importance of vocabulary in language learning and how to be taught*. *International Journal of Teaching and Education*, 3(3), 21-34.
- Browne, C. 2013. "The New General Service List: Celebrating 60 years of Vocabulary Learning". *The Language Teacher*. 4. 37: 13–16
- Brysaert, M., Mandera, P., & Keuleers, E. 2018. *The word frequency effect in word processing: An updated review*. *Current Directions in Psychological Science*, 27(1), 45-50.
- Chen, C., & Truscott, J. 2010. *The effects of repetition and L1 lexicalization on incidental vocabulary acquisition*. *Applied Linguistics*, 31, 693-713.
- Desagulier, G. 2017. *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*. Cham: Springer International Publishing.
- Gardner, D & Davies, M. 2013. *A New Academic Vocabulary List*. *Applied Linguistics*. 35. 305-327. 10.1093/applin/amt015.
- Leech G., Rayson, P., & Wilson, A. .2001. *Word frequencies in written and spoken English*. London:Pearson Education Limited.
- Lessard-Clouston, M. .2013. *Teaching vocabulary*. Alexandria: Tesol International Association.
- Matsuoka, W. & Hirsh, D. 2010. *Vocabulary Learning through Reading: Does an ELT Course Book Provide Good Opportunities?* *Reading in a Foreign Language*, 22 (1), 56-70.
- Nation, I. S. P. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P. & Waring, P. 1997. *Vocabulary Size, Text Coverage and Word Lists*. In N. Schmitt and M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.
- Schmitt, N. 2000. *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

- Thomas, L., Pfister, H. P., & Peterson, P. 2004. *Issues Related to the Construction of a Purpose-Built Domain-Specific Word Corpus*. *Australian Journal of Educational & Developmental Psychology*, 4, 13-28.
- Vongpumivitch, V., Huang, J. Y., & Chang, Y. C. 2009. *Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers*. *English for Specific Purposes*, 28(1), 33-41.