



VOCABULARY INPUT IN ESL TEXTBOOKS: A CORPUS-BASED ANALYSIS

Muthyala Udayaⁱ

Department of ESL Studies,
The English and Foreign Languages University,
India

Abstract:

Although textbooks are a major source in language learning, few research studies explored to what extent vocabulary input in ESL textbooks supports English language learning. The present study compares the vocabulary input of two English textbooks prescribed for Indian university undergraduate students. A corpus is constructed using online software based on the words in the textbooks. Comparison of frequency against different corpora reveals the differences in the order of words. The results are discussed both from a quantitative and a qualitative perspective. The analysis reveals that many words in the textbooks occasionally occur in common everyday language use and vocabulary selection in these textbooks demonstrates great variation in the number and selection of vocabulary. Therefore, the study confirms that the communication needs of learners are afforded greater weight than frequency criteria.

Keywords: corpus, ESL, frequency, textbook, vocabulary, wordlist

1. Introduction

Vocabulary frequency information provides a rational basis for ensuring that learners get the best return for their vocabulary learning effort by ensuring the words studied will often be met. Vocabulary frequency lists that take account of the range have a significant role in curriculum design and setting learning goals. The essential thing that can be learned from studying the language contained in a corpus is how frequently any particular word occurs. A frequency list records the number of times each word occurs in the text. Therefore, it can provide interesting information about the words that appear (and do not appear) in a text. A word list can be arranged in order of the first occurrence, alphabetically, or in frequency order. First occurrence order serves as a quick guide to the distribution of words in a text, an alphabetic listing is built mainly for indexing purposes, but a frequency-ordered listing highlights the most commonly-occurring words in the text. Word frequency counting dates back to Hellenistic times and made word lists. Word

ⁱ Correspondence: email udaya@efluniversity.ac.in

frequency count in English was started for making word lists. Since the initiation of word counting made by Kaeding (mentioned by Engels 1968), there have been several developments and innovations in the study of word frequency. A few important works include Thorndike's *Teacher's Word Book* in 1921 (Bright and McGregor 1970), Michael West's *A General Service List of English Words* (1953), the work of Kučera and Francis (1967). development of computerized corpora, word frequency study has been "revolutionized" (Read 2004: 156); and very large size corpora have been established in this way, from which more reliable, authentic, and up-to-date word lists based on frequency counts have been prepared. English word lists based on the British National Corpus, COBUILD Corpus, and Cambridge International Corpus are worth mentioning.

2. Review of literature

To fully know a word means knowledge of meaning and form, register, collocation, and association, among other things (Nation, 2013). According to Tyler (2012), to gain this knowledge and enable the automatization of word knowledge and recognition, a word must be encountered in various contexts. According to (Aitchinson, 2012), words are not stored in isolation in our brains but are connected in semantic networks. When learners encounters the same word repeatedly and in different contexts, those connections are likely to be internalized and thus remembered (Nation, 2008). As Cameron (2001) emphasizes, language learning materials should be constructed to make it possible for learners to meet new words repeatedly and in different contexts and how often a word should recur in a textbook to enable learning is unclear. The figures vary among researchers, with estimates ranging between 5–6 (Cameron, 2001; Nation, 1990) and twenty (Waring & Takaki, 2003) occurrences. Repeated exposure has been shown as 3–4 times more important for beginners than for advanced learners (Zahar, Cobb & Spada, 2001). Studies suggest that knowledge of high-frequency words is closely related to text coverage and successful language learning (Nation, 2006; Nation & Beglar, 2007).

Relevance of word frequency in language learning, it seems natural that a learner can pick the more frequently occurred items with ease compared to the rarely occurred ones in the course of his learning career. This implies the need for making provision for the occurrence of words time and again in language learners' course materials. Moreover, particularly in the initial stages of language learning, it seems reasonable to emphasize those vocabulary items that occur very frequently in the concerned language. Learning such items will be more helpful for learners in general than learning those items that are rarely used in the language. Word frequency is highly important, and despite some other criteria, it has been considered an essential criterion for vocabulary selection in language pedagogy (Honeyfield, 1987; Verghese, 1989). The primary purpose of depending on frequency is to assist the learners in developing "*a survival-level repertoire for comprehension and production in language*" (McCarthy 1990: 79). This sort of exploration will help us in predicting how far the learners' exposure to the target language can facilitate the comfortable mastery over vocabulary and can thus contribute towards language

development in such a way that they can function well once they encounter the situation whereby the use of authentic, real-world English is demanded.

The present study investigates the vocabulary input contained in selecting textbooks for ESL teaching in an Indian University: Degree first and second-year undergraduate program. It pursued the following specific objectives:

- 1) finding out the number of words in the textbooks at two levels;
- 2) to identify the different words contained in the textbooks;
- 3) to draw a profile of the distribution of word categories in the whole textbook corpus;
- 4) to determine whether there is an increase in vocabulary input through the first and second-year texts.

3. Method

A corpus consisting of two English textbooks were created for the study. An analysis is done to what extent the vocabulary in the textbooks corresponds to common everyday language use and compares lexical occurrences, adjectives, verbs, and nouns. The university curriculum at the advanced level states that learners need to develop the ability to use language in everyday life and focus particular attention on the words they use and how they use them. The textbooks also mention that learners need to develop everyday verbal communications to be accurate and effective with exposure and study of the language. These textbooks are designed to develop command over the English language with specimens of modern English, ordinary conversations, personal letters, formal speech, and reflective writing, fiction, and drama activities. The objectives of teaching the second-year learners are to develop the LSRW skills, and the exercises and the classroom activities are specifically designed to train the students in compositional and communication skills. The English prose selections serve as models for content and style and develop writing and communication skills through the exercises and activities.

The two-word lists of undergraduate English textbooks of Degree first and the second-year are prepared using Web Frequency Indexer v1.3. In the present study, the unit of analysis is the word as Carter (1998:4) defined "*any sequence of letters (and a limited number of other characteristics such as hyphen and apostrophe) bounded on either side by a space or punctuation mark*" For Lemmatization as Read (2000:18) proposes as the grouping under the same heading of the base and inflected forms of a word. However, we are aware that in most corpus studies, words are lemmatized. The study has aimed to see how far the frequency patterns of words found in a large corpus with a wide range of sources match with the selected ESL textbooks corpus. From a pedagogical view, it is essential to know which words forms are included in the textbooks and which are not. Researchers like (Laufer 1991: Nation 1990: Richards 1976: Wallace 1992) have claimed that knowing a word means knowing its different word forms. The present study is quantitative, and it attempts to count the number of words comprising the vocabulary input in a sample of textbooks. It establishes ratios and percentages to compare the vocabulary load of the textbooks.

4. Results

The number of words contained in the textbooks corresponding to the levels is presented. The two levels' analysis of the textbooks with the following objectives is addressed.

- 1) The number of words contained in each textbook;
- 2) The top fifty words (both grammatical and content word types);
- 3) The top hundred content words;
- 4) Common words found in textbooks at different levels.

4.1. Number of words in the textbooks

The results indicate a significant difference in the number of types analyzed: 2424 in Degree first-year textbook compared to 3777 in the second year textbook. The difference (1353) in the number of types in each textbook: considering the number tokens (total number of running words) in each textbook: 10071 in Degree first-year textbook versus 16827 in Degree second-year textbook and the difference (6756). The lexical variation between the two textbooks is 22.46 and 24.06 and not as high as it might appear by looking at the raw figures. An analysis of word types and tokens can be summarized as follows:

Table 1: Number of types and tokens

	Types (distinct words)	Tokens (absolute number)
Text 1	2424	10071
Text 2	3777	16827

4.2. Vocabulary size and text coverage

One of the most important points of interest in word frequency studies has been to see how many words cover what percentage of the entire corpus. There have been attempts to establish several frequency bands (levels) in terms of the most frequent words (e.g., the most frequent 100 words, 500 words or 1000 words, etc.) and calculate in percentage the entire corpus's size that the items falling within a given frequency band occupy. The result gives us an estimation of the coverage of a given frequency band in the given corpus. With this overview, an attempt has been made to establish different bands of word frequency as given in the table below and calculate the coverage of the items falling into them, based on both corpora separately.

The frequency analysis informs us that the crucial part of the learner's vocabulary lies in the first 1000 words, which have 80-90% coverage in the whole text. The fifty most frequent function words in these words list contribute coverage of 44.2% & 44.7% of the whole text. For the following 1000 words, there is an increase of 5% in coverage for every 500 words. The remaining vocabulary (roughly 1000) in the text contributes only 10% or less than 10%, and the distribution of function words and content words across the two textbooks maintains the same range (i.e., coverage and contribution). The percentage coverage of words in the whole textbook corpus is presented in Table 2, along with the graphs. Table 2 compares the frequency of the textbooks with well-established corpora.

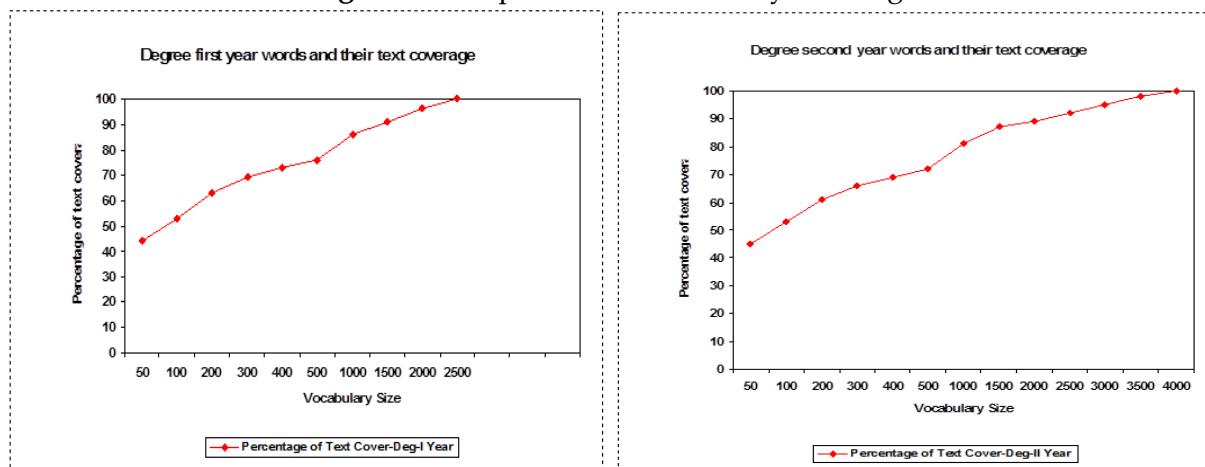
The top fifty words for each textbook alongside their occurrences with respect to the total number of tokens in the text are shown in decreasing order.

Table 2: Vocabulary size and text coverage

Vocabulary size	Text Coverage	
	Textbook 1	Textbook 2
50	44.2%	44.7%
100	53.3%	53.0%
200	63.1%	61%
300	68.6%	65.7%
400	72.6%	69.0%
500	75.7%	71.8%
1000	85.9%	80.8%
1500	90.8%	86.5%
2000	95.8%	89.4%
2500	100% (2424)	92.4%
3000	--	95.3%
3500	--	98.3%
4000	--	99.9% (3777)

The information given in the table above is depicted in the following figure (G1, G2) below, highlighting the coverage pattern of words and the two corpora across different levels. When compared, the Degree second-year textbook has greater coverage than the first-year textbook.

Figure 1: Comparison of Vocabulary Coverage



Note: G1: vocabulary coverage in first-year textbook G2: vocabulary coverage in second- year textbook

As the figure depicts, there is a slight disparity between the percentage values derived from the two in the first frequency band (50 words level), which shows the tendency of a gradual increase at each next upper band until the curve reaches the five hundred band. After that, the increase or decrease in the disparity is negligible, and we can notice more or less a parallel increase in the coverage of the given frequency band in the two corpora up to the 300 words level. Despite the difference between the two corpora regarding the

percentage value seen in all the frequency bands, a general principle established by several word frequency studies in quite large corpora that a small number of items account for a very high proportion of general use (Branford 1967) seems to be applicable in the case of both corpora.

4.3. Comparison of the fifty most frequent words contained in the textbooks

Table 3 illustrates the frequency lists from three different corpora, and a comparison is made against Cambridge International Corpus (CIC), The Cambridge and Nottingham Corpus of Discourse in English (CANCODE), the British National Corpus (BNC). The fourth and fifth lists are Degree first and second-year textbook. Function words are predominant in the two textbooks, including pronouns, determiners, prepositions, modal verbs, auxiliary verbs, conjunctions. Several lexical words such as *know, well, got, think, and right* are the most frequent words in the first fifty words. Function words dominate the highest frequencies of five lists, and indeed one of the defining criteria of function words is their high frequency. As we go down the frequency list, there is no absolute cut-off between function words and lexical words of high frequency such as *thing, young, science, and day*. On closer examination, some of the 'lexical' words which appear in the high-frequency function word list prove to be elements of interpersonal markers (e.g., *you know, I think*) or single-word organizational markers (*well, right*). The ranks of most of the words in corpus conform closely to those obtained in the Brown corpus (Kucera & Francis, 1967), BNC, CIC, and CANCODE. The relative frequencies of words, ranging from the most frequent (the) to the least frequent (been), are the same. Some words have changed their order that can be ascribed to the differences in the size of the corpus. The derived words contribute only 0.61% in Degree first-year text, 0.56% in Degree second-year textbook. The first-fifty most frequent function words of Degree first-year text contribute 44.2% of the whole vocabulary, and second-year degree text is 44.7%. The first 100 most frequent content words contribute only 9.6% in Degree first-year text and 8.01% in Degree second-year text in terms of coverage of the whole text.

Table 3: Comparison of first fifty words

CIC corpus	CANCODE corpus	BNC corpus	Degree first year textbook	Frequency coverage	Degree second year textbook	Frequency coverage
the	the	the	the	625	the	915
of	i	of	and	304	of	572
to	you	and	to	269	to	539
and	and	a	of	245	and	524
a	to	in	a	236	a	429
in	it	to	i	226	i	392
that	a	is	in	193	in	328
is	yeah	to	it	122	that	234
for	that	for	is	117	it	174
it	of	it	he	114	is	173
was	in	be	that	100	you	169
he	was	was	you	100	for	152
on	is	on	was	91	my	144
with	its	with	his	82	was	136
i	know	that	not	79	be	132
as	no	by	on	79	as	125
at	oh	are	for	76	he	121
be	so	's	have	68	have	115
by	but	this	at	61	on	108
his	on	from	with	61	but	103
but	they	which	but	60	not	101
have	well	at	had	58	with	100
from	what	not	my	58	had	90
are	yes	or	be	55	are	88
said	have	an	they	54	me	88
not	we	he	said	53	his	86
they	he	but	we	53	should	81
you	do	has	are	52	which	81
this	got	will	as	51	at	77
an	that's	i	no	42	if	76
had	for	they	me	40	an	75
has	this	his	man	38	this	75
or	just	were	what	38	would	71
one	all	their	him	37	one	65
which	there	have	lady	36	by	61
will	like	you	all	35	see	61
were	one	we	an	35	all	56
their	be	been	has	35	can	54
who	right	had	this	35	from	54
we	not	its	been	34	will	52
would	don't	can	do	33	or	51
all	she	all	from	32	so	51
she	think	there	one	32	we	48
her	if	who	our	32	what	47
more	with	would	young	32	your	47
been	then	as	if	29	has	41
about	at	if	by	28	their	41
there	about	that	her	28	day	40
when	are	some	science	28	do	40
its	as	one	out	27	only	47
Total number of words				4448		7523
Coverage				44.2%		44.7%

4.4. The top hundred content words contained in the textbooks

The content words are grouped into nouns, verbs, adjectives, and adverbs. There is similar distribution in each textbook. It was easy to find out which word among the synonyms is the most frequent from the above list.

Table 4: The top hundred most frequent words contained in the textbooks

Degree first year textbook				Degree second year textbook			
Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency
Well	22	Let	8	Knowledge	38	Friend	11
Valentine	21	Long	8	Elephant	34	Information	11
Dentist	20	Next	8	People	32	Keep	11
Cop	18	Night	8	World	31	Old	11
Dolly	18	Nothing	8	Eyes	30	Taxi	11
English	17	Road	8	Life	28	Freedom	10
Smiles	17	Room	8	New	28	Hydrogen	10
Years	17	Short	8	Time	24	Last	10
People	16	Snake	8	War	24	Latin	10
Philip	16	Take	8	Man	22	Made	10
Time	16	Under	8	Sight	22	Particular	10
Anil	15	Whole	8	Society	22	Went	10
Before	15	Without	8	Human	21	Work	10
Magda	15	Work	8	India	21	Years	10
World	15	Bomb	7	First	19	Beauty	9
Arthur	14	Close	7	Things	18	Bombs	9
Five	14	Cobra	7	Three	18	Found	9
Think	14	Done	7	Way	18	History	9
Day	13	Fear	7	Write	18	Important	9
Eve	13	Found	7	Good	17	Interest	9
Future	13	Gate	7	Little	17	Live	9
First	12	Home	7	Never	17	London	9
Hand	12	Husband	7	Face	16	Means	9
House	12	Language	7	Writing	16	Moment	9
Mind	12	Last	7	Days	15	Myself	9
New	12	Look	7	English	15	Own	9
Life	11	Might	7	Great	15	Play	9
Months	11	Mother	7	Home	15	Point	9
Way	11	Notes	7	Living	15	Present	9
Century	10	Oxford	7	Long	15	Second	9
Down	10	Pot	7	Reader	15	Style	9
Eyes	10	Stories	7	Two	15	Subject	9
Love	10	Thought	7	Hammar skjold	14	Thousand	9
Two	10	Told	7	Hand	14	Well	9
War	10	Walked	7	Men	14	White	9
Car	9	Woman	7	Want	14	York	9
Dasa	9	Written	7	Crowd	13	Yourself	9
Face	9	Ago	6	Driver	13	Best	8
Going	9	Atomic	6	Rifle	13	Better	8
Good	9	Darwin	6	Language	12	Children	8
Great	9	England	6	Mind	12	Education	8
Money	9	Evening	6	Read	12	England	8
Never	9	Family	6	Same	12	Expression	8
Own	9	Feel	6	Shoot	12	Family	8
Quite	9	Forward	6	Blind	11	Feel	8
Today	9	Friends	6	Born	11	Future	8
Want	9	Generation	6	Cribbs	11	General	8
Write	9	Grass	6	Edward	11	Interested	8
Back	8	Hard	6	End	11	Learn	8
Total number of words		976		Total number of words		1348	
Coverage		9.69%		Coverage		8.01%	

The content words are grouped into nouns, verbs, adjectives, and adverbs. There is similar distribution in each textbook. It was easy to find out which word among the synonyms is the most frequent from the above list. For example, if we take **start (0),(0)**, **begin(0),(1)** and **commence(0),(0)**, begin is more frequent than start, and commence is the least frequent without any citation as against (1) occurrence for begin. The words that commonly go together are not included in the textbooks, such as knife, fork, spoon, but loan words such as teacher, educator, and tutor are given, learners are familiar with and occur in their mother tongue.

Many new words are introduced at the advanced level, such as *punctiliousness* and *yanked*, which occur as less frequently occurring content words, where the learners need to spend more time learning them. Idioms are also introduced, such as *scrambled out*, *throttle down*, *on the wane*, *leaps*, and *bounds* where the complete meaning cannot be deduced from the sum of its parts. These words include inflected forms of nouns such as plural possessive, *neighbour – neighbours*, *beggar–beggars*, *person–persons*, *child– children*. Word lists/vocabulary extracted from these texts does not include all the inflected forms of the relevant categories.

<i>long</i>	-	<i>longer</i>	-	<i>longest</i>
<i>early</i>	-	<i>earlier</i>	-	<i>earliest</i>
<i>large</i>	-	<i>larger</i>	-	<i>largest</i>
<i>hard</i>	-	<i>harder</i>	-	<i>hardly</i>
<i>clever</i>	-	<i>cleverer</i>	-	<i>cleverest</i>
<i>young</i>	-	<i>younger</i>	-	<i>youngest</i>
<i>great</i>	-	<i>greater</i>	-	<i>greatest</i>

The words in Italics do not occur in the word lists, however, included here for comparison.

4.5. Commonly occurring prefixes

Many irregular verb forms, mainly the past tense forms, occur, but the present tense forms do not occur, e.g., *took*, *cut*, and *went*. The comparative inflections -er and -est (sooner, soonest; quicker, quickest) have been found. The content list contains many derived words such as nouns, verbs, adjectives, adverbs, and only specific derivational vocabulary is frequent and significant in learning and teaching. Derived words here mainly consist of words derived by adding derivational prefixes and suffixes to the stem. These affixes often change parts of the speech of the existing word. Some of the derivational affixes listed below are very productive. The most commonly occurring prefixes in these textbooks are *in-*, *dis-*, *de-*, *un-*, *over-*, *under-*, *fore-*, *re-* and *ir-*. Table 5 displays the distribution of prefixes throughout the textbooks. These two textbooks have 0.56% and 0.61% coverage in the whole corpus. The prefix *un-* had the highest frequency and topped the list. Each prefix was used an equal number of times in each textbook, and few prefixes included more than others.

Table 5: Commonly occurring prefixes

Degree first year textbook				Degree second year textbook			
unconscionable	1	improve	2	immeasurably	1	unbalanced	1
unconsciously	1	impulse	1	impartial	1	unbearable	1
undercurrent	1	improved	1	impersonal	1	unconscious	1
understand	1	incapable	1	impossible	4	undiscoverable	1
undertaking	2	indiscretions	1	imprinting	1	uneasy	1
unexpectedly	2	inflexible	2	imprisoned	1	unendurable	1
unfolds	1	inheritance	1	improbable	1	unfolding	2
unhappily	1	inmates	1	inaccurately	1	unforeseen	1
unknown	1	disappeared	2	inadequate	1	unimaginable	1
unobstructed	1	discomfort	1	incapable	1	unless	6
uplifted	1	distress	1	interwoven	1	unlike	1
uppermost	1	distressed	1	indefinitely	1	unlikely	1
upstream	1	overbalanced	1	inescapable	3	unnerving	1
undergraduate	3	overcome	1	inevitably	1	unnoticed	1
undetected	1	overnight	1	inexact	1	unorthodox	1
unfailing	1	undercurrent	1	infatuation	2	unrelieved	1
unfortunately	1	undergraduate	3	ingenuity	1	unrivalled	1
unimproved	1	undertaking	1	inhabitants	3	unseen	1
united	2	reconstruction	1	innumerable	1	unsuitable	1
unlined	1	remaking	1	international	5	overseas	1
unspectacular	1	resolve	1	interplay	1	overtones	1
upset	2	reunion	1	irrational	2	undertook	1
impossible	1	reunited	1	irreconcilable	1	undervaluing	1
imprisonment	1			irrelevant	2	forecast	1
Total number of words	58			irresistibly	1	forecasting	1
Coverage	0.3%			irresponsible	2	forego	1
				disagreement	1	reappeared	1
				disappearance	1	rediscover	1
				disappointed	1	rediscovery	1
				disarmament	1	irrational	2
				disintegration	1	irreconcilable	1
				disobedience	3	irrelevant	2
				disreputable	1	irresistibly	1
				distasteful	1	irresponsible	1
				unarmed	1	irresponsible	1
				Total number of words	95		
				Coverage	0.9%		

For example, non- de- anti are equally shared in the textbooks. There are many affixes that are still productive and used to generate new words. Among the most common prefixes that are used to form new words.

unfriendly - unhelpful or harmful

enrain - not concerned about or harmful to the environment

unwaged - unemployed

unleaded - not containing lead

Although it is sometimes difficult to process the meaning of words that have been formed by affixation, new ones are generally transparent. Once the learner knows the meanings of the root and that of the affix, the combination can be easily predicted.

The prefix *anti-* can also be used in the sense of preventing or neutralizing. The prefix *de-* means to remove or reverse and only two occurrences are contained in the text. The prefix *non-* means *negation, exclusion or refusal*.

Table 6: Negative prefixes

Word form	Frequency	Word form	Frequency	Word form	Frequency
anti-communism	2	degraded	1	non-cooperation	1
anti-communist	1	delighted	1	non-violence	1
anticolonial	1			non-violent	1
antidote	1				

The second language learner has to learn the most productive affixes and also their patterns of distribution, e.g., *teach – teacher – teaching*. This particular affix *-ing* does not change the part of speech as both are nouns, rather the grammatical meaning of the word is changed. A teacher is one who teaches, teaching is a classroom activity of a teacher.

4.6. Commonly occurring suffixes

Table 6 shows the distribution of types, tokens, and the odds of encounter with suffixes in the textbooks. Compared with the distribution of prefixes shown in Table 7,8 many types of suffixes are included. A suffix is a group of letters which is added to the end of a word that changes how a word is used as a part of speech. Suffixes can carry grammatical information (inflectional suffixes), or lexical information (derivational suffixes) The most commonly used inflectional suffixes are *-est, -ies, -ed, -ing,* and derivational suffixes are *-tion, -ly, -ness, -ful*.

Table 7: Frequency of suffixes

Suffixes	Frequency	
-able	27	Commonly shared in the textbooks
-dom	04	
-er	48	
-ise	4	
-ism	12	

Throughout the textbooks a majority of suffixes are included in the textbooks. The combination of a base and suffix may decrease readers' opportunities to encounter other variants. Compared with prefixes, suffixes exhibited substantially more types and tokens in the textbooks. Though suffixes with various types and many tokens (e.g., *-er* and *-ly*) are attached to certain bases, which may prohibit expansion of the reader's suffix knowledge, they can be useful tools for teaching suffixes and are relatively easy to enhance using other materials such as graded readers and supplemental reading. Using these suffixes in explicit instructions is also effective in helping learners increase their

morphological knowledge (cf. Bowers, Kirby, & Deacon, 2010). The suffix *-able* occurs 27 times and are shared commonly in both the tests. Examples extracted from degree first and second-year frequency lists. The suffix *-dom* occurs four times in the second year degree textbook. Most of these words are abstract nouns. The suffix *-er* has the highest frequency and coverage.

Table 8: Suffixes (-able, -er, -dom)

Word form	Frequency	Word form	Frequency	Word form	Frequency
lamentable	1	younger	1	longer	2
respectable	2	cutter	2	longer	7
delectable	1	harder	1	sooner	2
comparable	1	idler	1	owner	2
fashionable	1	larger	1	wider	1
miserable	1	sewer	1	trigger	1
unbearable	1	greater	4	thinner	1
unendurable	1	lower	2	stranger	1
unsuitable	1	wiser	1	shopper	1
valuable	2	upper	1	nearer	1
respectable	2	traveller	1	glitter	1
remarkable	2	taller	1	bluster	1
unconscionable	1	simpler	1		
acceptable	1	sharper	1	kingdom	1
disreputable	1	mover	1	seldom	1
irreconcilable	1	gainer	1	wisdom	1
sizeable	1	earlier	2	freedom	1
undiscoverable	1	lover	1		
unimaginable	1	fuller	1		
remarkable	2	higher	1		
inescapable	2	inner	1		

It has been observed that when more than one affix is involved in derivation, the resulting words are more complex and hence present difficulties in the acquisition of the word. Often these words have different meanings and different restrictions for word collocations from those of their bases. This makes language learning very difficult. In such a case, each word has to be acquired as a separate one. However, if frequencies of words are correlated with their usage patterns, it seems prefixation in derivation is less frequent hence present difficulties in acquisition when compared to suffixation.

Table 9: Suffixes (-ism, -ise)

Word Form	Frequency	Word form	Frequency	Word form	Frequency
realism	1	euphemism	1	capitalise	1
socialism	1	totalitarianism	1	economise	1
secularism	1	vegetarianism	1	utilise	1
journalism	1	anti –communism	1	emphasize	1
rheumatism	1	communism	1		
		imperialism	1		

4.7. Contracted forms and their frequencies

Some of the contracted forms are acquired later than the full forms. These examples suggest that learners begin lexical development by using features that vary along perceptual dimensions.

Table 10: Contracted forms

Abbreviated form / Full form	Frequencies	
	Degree first-year	Degree second-year
n't/not	45/79	5/101
'll/will	5/18	-/52
'm/am	6/7	1/28
's/is	13/113	15/173
're/were	2/23	-/32
'd / would, had	2/23/53	-/76/90

The forms in the left most columns are spelling variants (abbreviated) of the corresponding forms in the right of the table. However, they occur in fewer frequencies. In formal learning or teaching one may avoid use of such variants. These examples suggest that learners begin lexical development by using features that vary along perceptual dimensions. Learners acquire nouns earlier than verbs, but the learnability criteria explain how the learners establish word classes that are equivalent to noun or verb.

4.8. Compound words

Few of the compound words are difficult to process and ESL learners take time in understanding the semantic relations hence find these forms difficult. Learners are not sure whether compounding is possible. A second problem for learners is to relate the order in which the parts of a compound appear. There are pie apples and apple pies, there is gum chewing (often used as an adjective) and chewing gum. This, undoubtedly, is confusing to an L2 learner as shown in the word order errors from Hatch & Brown (1995:194).

All time I have my *book map* in my bag. Soldiers have *glasses shoes*.
We eat in *room banquet*. Prince married *shoes girl*.

Such errors suggest that learners do not learn all compounds as new items, rather put the parts together to create the compounds. In doing so, they use the typical word order of their first language. Another possibility is that they place the words such that the meaning on which they want to focus comes first and then the modifying part is added. Some of the interesting compounds presented in the textbooks are as follows.

<i>cock - eyed</i>	<i>river - bed</i>
<i>half - opened</i>	<i>heart - boiled</i>
<i>cold - blooded</i>	<i>noble - hearted</i>
<i>many - horned</i>	<i>well - oiled</i>

<i>many - getting</i>	<i>well - educated</i>
<i>over - shadowed</i>	<i>easy - going</i>
<i>fast - changing</i>	<i>price - winning</i>
<i>tele - education</i>	<i>self - satisfaction</i>

5. Conclusion

To conclude, the corpus study provided a basis for analyzing the textbooks has an important role to play setting learning goals and curriculum design. In the present study, quantitative analysis on the vocabulary input in two textbooks from two educational levels was carried out. Depending on the prescribed textbooks learners may be exposed to different words and learners may be provided wordlists as the major source of vocabulary learning activity. Course designers can refer to wordlists when they consider the vocabulary component of a language course and the teachers need to have reference lists to judge whether a particular word deserves attention as frequency provides a key indication of their importance. Percentage coverage was the main criterion for selecting what has to be taught to the degree level students. Frequency information allows teachers to focus appropriately on the most common words, ensuring that learners know and actively use them. The less frequent words are topic-specific and can be acquired when needed, e.g., *benefactor, despotic, flutter*. The common words need less learning effort as the frequency of exposure helps in easy understanding. Concerning language textbook designers, this study provided an analysis of two textbooks, which, hopefully, should make teaching material designers reflect on the need to follow common objectives. Textbooks for different levels should contain words depending on age appropriateness. Regarding language teachers, the results of this study highlight the nature of the input that textbooks should include. As textbooks provide students with different kinds of input, and that this difference may have an effect on language learning, so teachers need to find out what criteria are used in vocabulary selection. From a research perspective, this study contributes to the field of second language acquisition and teaching with a description of the type of vocabulary contained in English language textbooks as representative of an important educational genre. It provides the evidence of vocabulary input contained in small sample of books with empirical evidence. However, a further analysis with comparison to different ESL textbooks might help in arriving at definite conclusions. Finally, word frequency is an important dimension in ESL textbooks along with other aspects such as vocabulary input, lexical density, lexical variation that require investigation such as the influence of the number of word encounters in word learning, or the effect of word length and word class in the acquisition of vocabulary. Other aspects that need further investigation are the relationship between the vocabulary input provided by textbooks and learners' vocabulary output in each level. It is essential to control certain variables that may be having an effect on this relationship, such as teachers' treatment of the vocabulary contained in the textbook, or learners' strategies to retain and recall the unfamiliar words.

Conflict of Interest Statement

The author declares no conflicts of interests.

About the Author

Dr. Muthyala Udaya is an Assistant Professor in the Department of ESL Studies, School of ELE, The English and Foreign Languages University, India. Her research interests include Corpus and Language Teaching, Bilingualism, Second Language Acquisition, and Learning.

References

- Aitchinson, J. (2012). *Words in the Mind: An Introduction to the Mental Lexicon*. (4th ed.). Oxford: Oxford University Press.
- Branford, W. (1967). *The Elements of English*. London. Routledge and Kegan Paul.
- Bright, J. A. and McGregor. G. P. (1970). *Teaching English as a Second Language*. London: Longman. BNC-British National Corpus.
- Cameron, L. (2001). *Teaching languages to Young learners*. Cambridge: Cambridge University Press.
- Carter, R. (1998). *Vocabulary: Applied Linguistic Perspectives*. London: Routledge.
- CANCODE - The Cambridge and Nottingham Corpus of Discourse in English
CIC- Cambridge-International-Corpus
- Engels, L. K. (1968). The Fallacy of Word Counts. *IRAL* 6(3). PP.213-32.
- Hatch, E., & Brown. S. (1995). *Vocabulary, Semantics, and Language Education*. New York: Cambridge University Press.
- Honeyfield, John G. (1987). Word Frequency and the Importance of Context in Vocabulary Learning. In *Methodology in TESOL: A Book of Readings* (eds) Michael H. Long and Jack C. Richards. New York: Newbury House/Harper and Row. PP. 318-24.
- Kučera, H. and Francis. W. N. (1967). *Computational Analysis of Present-Day American English*. USA: Brown University Press.
- Laufer, B. (1991). Knowing a word: What is so difficult about it? *English Teachers' Journal* 42: 82-88.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House Publishers.
- Nation, I. S. P. (2006). How large a Vocabulary is needed for Reading and Listening? *Canadian Modern Language Review*, 63(1). PP. 59–82.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7). PP. 9–13.
- Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston, MA: Heinle Cengage Learning.

- Nation, I. S. P. (2013). *Learning Vocabulary in another Language (2nd ed.)*. Cambridge: Cambridge University Press.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Read, J. (2004). Research in Teaching Vocabulary, *Annual Review of Applied Linguistics* 24. PP.146-61.
- Richards, J. (1976). The Role of Vocabulary Teaching. *TESOL Quarterly* 10 (1): PP.77-89.
- Tyler, A. (2012). *Cognitive linguistics and Second language learning: Theoretical Basics and Experimental evidence*. London: Routledge.
- Udaya, M. (2005). Learning English vocabulary at the primary level- A frequency count of the vocabulary input in the textbooks of Andhrapradesh. *Language in India*. 5(10)
- Vergheese, C. Paul. (1989). *Teaching English as a Second Language*. New Delhi: Sterling Publishers.
- Wallace, C. (1992). Critical Literacy Awareness in the EFL Classroom. In N. Fairclough (Ed.), *Critical Language Awareness* (pp. 59-92). London: Longman.
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader. *Reading in a Foreign Language*, 15(2), PP.130–163.
- West, M. (1953). *A General Service List of English Words*. London: Longman
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring Vocabulary through Reading: Effects of Frequency and Contextual Richness. *Canadian Modern Language Review*, 57(4), 541–572.

Creative Commons licensing terms

Author(s) will retain the copyright of their published articles agreeing that a Creative Commons Attribution 4.0 International License (CC BY 4.0) terms will be applied to their work. Under the terms of this license, no permission is required from the author(s) or publisher for members of the community to copy, distribute, transmit or adapt the article content, providing a proper, prominent and unambiguous attribution to the authors in a manner that makes clear that the materials are being reused under permission of a Creative Commons License. Views, opinions and conclusions expressed in this research article are views, opinions and conclusions of the author(s). Open Access Publishing Group and European Journal of Education Studies shall not be responsible or answerable for any loss, damage or liability caused in relation to/arising out of conflicts of interest, copyright violations and inappropriate or inaccurate use of any kind content related or integrated into the research work. All the published works are meeting the Open Access Publishing requirements and can be freely accessed, shared, modified, distributed and used in educational, commercial and non-commercial purposes under a [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).