

## Vocabulary in Primary School Tamil Textbooks (A Corpus Based Analysis)

Prem Kumar LR\*

Central Institute of Indian Languages, Mysore, Karantaka, India

### Abstract

This paper mainly deals with vocabulary frequency in the light of findings regarding depth of vocabulary learning, the cyclical nature of vocabulary learning, and vocabulary size and reading comprehension and concentrates on the number of Words and their cumulative Percentage of Tamil Textbooks for all the standards, namely, from first standard to fifth standard. A comparative study of Number of Words and Their Cumulative Percentage from 1<sup>st</sup> to 5<sup>th</sup> Standard Tamil Textbooks is also developed to find out what exactly is the area where concentration should be given for developing vocabulary in the students.

**Keywords:** Vocabulary; Tamil textbooks; Language; Literacy

### Introduction

Vocabulary in curriculum is always a matter of discussion throughout the history of language teaching that a question arises as to how much vocabulary do learners need to know? This was never been a concluding debate. It always goes with the need based syllabus one has. However, it was to some extent accepted that any basic course, which teaches a language as a second language, should teach approximately one thousand five hundred basic words [1]. This is still a matter of question to be answered. On the other hand if it is a first language teaching course as in the State of Tamilnadu, for example, emphasizes about 250 words in the first standard and about 300 in the second standard. This is always refuted by the fact that the number of words is considerably differs as the number of known words which the child brings to the school is not measureable. This study, however, brings out a different story as far as this point is concerned.

While designing a language course or planning our own course of study, it is useful to be able to set learning goals that will allow us to use the language in the ways the study want to. When the study plan the vocabulary goals of a long-term course of study, the study can look at three kinds of information to help decide how much vocabulary needs to be learned: the number of words in the language, the number of words known by native speakers and the number of words needed to use the language.

In the same way, the question that how many words are there in a language cannot be answered fruitfully. The most ambitious goal is to know all of the language. However, even native speakers do not know all the vocabulary of the language. There are numerous specialist technical vocabularies, such as those of nuclear physics or computational linguistics, which are known only by the small groups who specialize in those areas. Still, it is interesting to have some idea of how many words are there in the language. This is not an easy question to resolve because there are numerous other questions which affect the way the study answer for it, including the following.

### Review of Literature

A less ambitious way of setting vocabulary learning goals is to look at what native speakers know in their language. Unfortunately, research on measuring vocabulary size has generally been poorly done (I.S.P. Nation, 1993), and the results of the studies stretching back to the late nineteenth century are often wildly incorrect. This research may not discuss these points here.

The reliable studies [2] suggest that educated native speakers of English know around 20,000 word families. These estimates are rather low because the counting unit is word families which have several derived family members and proper nouns are not included in the count. A very rough rule of thumb would be that for each year of their early life, native speaker add on average 1,000 word families to their vocabulary. These goals are manageable for non-native speakers of English, especially those learning English as a second rather than foreign language, but they are way beyond what learners of English as another language can realistically hope to achieve. There is no frequency based vocabulary analysis for Tamil.

### Discussions and Analysis

In this paper, a very important distinction has been made between high-frequency words and low-frequency words. This distinction has been made on the basis of the frequency, coverage and quantity of these words. The distinction is important because teachers need to deal with these two kinds of words in quite different ways, and teachers and learners need to ensure that the high-frequency words of the language are well known.

It is, therefore, important that teachers and learners know whether the high-frequency words have been learnt. This paper contains a vocabulary test that can be used to measure whether the high-frequency words have been learnt, and the progress of the learner in the learning of low-frequency vocabulary which could be based on the test in two different versions.

The test is designed to be quick to take, easy to mark and easy to interpret. It gives credit for partial knowledge of words. Its main purpose is to let teachers quickly find out whether learners need to be working on high-frequency or low-frequency words, and roughly how much work needs to be done on these words.

There is much more to vocabulary testing than simply testing if

\*Corresponding author: Prem Kumar LR, Central Institute of Indian Languages, Mysore, Karantaka, India, Tel: 8095047362; E-mail: [lrprem90@gmail.com](mailto:lrprem90@gmail.com)

Received May 11, 2015; Accepted May 22, 2015; Published May 28, 2015

Citation: Prem Kumar LR (2015) Vocabulary in Primary School Tamil Textbooks (A Corpus Based Analysis). Arts Social Sci J 6: 103. doi:10.4172/2151-6200.1000103

Copyright: © 2015 Prem Kumar LR. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

a learner can choose an appropriate meaning for a given word form, and the study should look closely at testing in a later. However, for the purpose of helping a teacher decide what kind of vocabulary work learners need to do, the levels test is reliable, valid and very practical.

“When the students complete the fifth standard, they should be able to read and speak for themselves without mistake. since the students come from different regional and social dialects, importance should be given to vocabulary exercise. Only, then this objective can be achieved”.

It is explicit from the syllabus that vocabulary exercise is very important at the beginning level.

Moreover, there are two more important reasons as to why the listening and speaking skills in the primary schools should be given importance:

- 1) The need for developing communicative competence in ‘Standard Spoken Form’
- 2) The existence of diglossia Tamil.

The students of the primary school level have to be trained to use the ‘Standard Spoken Form’ for their wider communication.

Due to the existence of diglossic situation, children who come to the school know only the ‘low variety’. Since the entire Tamil literature, prose or poetry, function or science uses only the ‘high variety’; literacy in Tamil could almost be equated with familiarity with this variety. One who is not familiar with this variety will never be considered literate in Tamil. So, student who come to primary school have to learn a ‘Standard Spoken Form’ for their day to day communication and along with this skill they have to acquire the ‘high variety’ in order to learn the text book material. The ‘Standard Spoken Form’ differs in many respects from their regional/social form and still more difference is found between ‘high variety’ and ‘Standard Spoken Form’.

Due to these two reasons the two skills (Listening and Speaking) are to be given importance. By this concept there seems to be not much difference between second language teaching and mother tongue teaching at least in the primary school level in Tamilnadu.

How do count a word? A number of questions arise as far as counting of words in a textbook or a language is concerned. Does it count /puttakam/ ‘book’ and /puttakaṅka/ ‘books’ as the same word? Does it count /paccai/ ‘green color’ and /paccai/ ‘a large grassed area’ in /paccai kuzhantai/ where the meaning is not ‘green’ but ‘very young’ as the same word? Do count people’s names? Does it count the names of products like /spiraiṭ/, /pepsi/, /kēṅāṅ/, /jāṅcaṅ and jāṅcaṅ/ as words of a language? The few brave or astonishing attempts to answer these questions and the major question ‘How many words are there in Tamil?’ have counted the number of words in very large dictionaries. When deals with following dictionaries, ‘Cre-A: Dictionary of Contemporary Tamil’ [3], are taken for the count of number of words for Tamil. It contains around 21,000 words. This is a very large number and is well beyond the goals of most first and second language learners. There are several ways of deciding what words will be counted.

For example, the second language teaching materials in Tamil, An Intensive course in Tamil [1] contains about one thousand five hundred words as per the syllabus. In fact the number of words used in the book is more than that. A semantically classified Vocabulary’ [4] contains about six thousand words. The actual number is not given. Hindi-Tamil Common Vocabulary [1] contains about two thousand words. But the actual number of words given in the book is not

known. These are some of the works published by Central Institute of Indian Languages, Mysore. There are plenty of materials available for teaching Tamil as second language, like ‘Spoken Tamil for foreigners’ by Arokiyanathan [5]. They are also to be subjected for such a study.

However, the present study does not go into the counting of words in the second language materials.

## Tokens

One way is simply to count every word form in a spoken or written text and if the same word form occurs more than once, then each occurrence of it is also counted. So the sentence /oru ūril oru rājā iruntār/ ‘There was a king in a place’ would contain five words, even though, two of them are the same word form /oru/ ‘one’. Words which are counted in this way are called ‘tokens’, and sometimes ‘running words’. The study tries to answer questions like ‘How many words are there in a page or in a line?’ ‘How long is this book?’ ‘How fast can you read?’ ‘How many words does the average person speak per minute?’, then our unit of counting will be the token.

## Types

We can count the words in the sentence in a different way, the same sentence as discussed above, /oru ūril oru rājā iruntār/ ‘There was a king in a place’. If we see the same word, we do not count it again. So the sentence of five tokens consists of four different words or ‘types’. We count words in this way if we want to answer questions like ‘How large was Bharathiar’s vocabulary?’ ‘How many words do you need to know to read his book?’ ‘How many words does his book contain?’ To get all these answers, one need to do a little bit of statistics based on the corpus or on the textbooks which would be the main target in this paper.

## High-frequency words

In the ‘CIIL Tamil corpus’, these words are marked as the most frequently used ten words: /oru/ ‘one’, /eṅru/ ‘said as’, /vēṅṅum/ ‘want’, /inta/ ‘this’, /eṅra/ ‘said as’, /allatu/ ‘or’, /koṅṅu/ ‘with’, /pala/ ‘many’, /itu/ ‘it’, /āṅāl/ ‘but’. The high-frequency words also include many content words also. The classic list of high-frequency words is given by Uma Maheswar Rao and Thennarasu [6]. A General Service List of Tamil Words which contains around 1 lakh word families. Almost 85% of the running words in the text are high-frequency words. There is a small group of high-frequency words which are very important because these words cover a very large proportion of the running words in spoken and written texts and occur in all kinds of uses of the language.

How large is this group of words? The usual way of deciding how many words should be considered as high-frequency words is to look at the text coverage provided by successive frequency-ranked groups of words. The teacher or course designer then has to decide where the coverage gained by spending teaching time on these words is no longer worthwhile.

The above Word list-I (Table 1) shows coverage figures for successive 50 word form from the Tamil Textbook have 2.29 percent – a collection of various 2,000 words forms of Tamil textbook just over three 39.47 percent. It is to be noted here that a word normally will have a boundary of its own. However, one can observe here that the word boundary here most often is a grammatical one also. Take for example, /payaṅpaṅṅuttik/ ‘to be used’, /kēṅṅuk/ ‘asked’, etc have the doubling of the next stop consonants which are not the part of the word. Secondly, the word counting goes with the case markers also if the word has that grammatical category, as in, /tolaikkāṅṅiyil/ ‘in television’, /coṅṅalaṅṅai/

S. No.	CF	Freq.	Words
1	0.31%	93	en
2	0.60%	91	vēṇṭum
3	0.90%	91	nī
4	0.93%	9	kīṭṭkāṇum
5	0.96%	9	kuḷukkaḷākap
6	0.99%	9	payaṇṇapaḷuttuka
7	1.02%	9	uruvākkuka
8	1.05%	9	kuppucāmi
9	1.08%	9	virumpum
10	1.11%	9	aḷaitṭuc
11	1.14%	9	iṭaṅkaḷil
12	1.17%	9	uyirmey
13	1.20%	9	uruvākkū
14	1.23%	9	eḷuttuc
15	1.26%	9	ellārum
16	1.29%	9	kuṟippatu
17	1.32%	9	coṟkaḷaik
18	1.35%	9	naḷittuk
19	1.38%	9	nikaḷcci
20	1.41%	9	varicaiyil
21	1.43%	9	viṭutalaip
22	1.46%	9	avaṟṟiṇ
23	1.49%	9	eṇakkum
24	1.52%	9	eṇṇiṭam
25	1.55%	9	vaṭṭamiṭu
26	1.58%	9	varukīratu
27	1.61%	9	aṟivurai
28	1.64%	9	eṇṇāl
29	1.67%	9	eḷuttai
30	1.70%	9	evvaḷavu
31	1.73%	9	kataiyait
32	1.76%	9	kāṭṭuka
33	1.79%	9	cūriyaṇ
34	1.82%	9	tēvaiyāṇa
35	1.85%	9	ponkal
36	1.88%	9	vellam
37	1.91%	9	amaikka
38	1.94%	9	eṇṇatai
39	1.97%	9	kaḷiṭam
40	2.00%	9	takunta
41	2.03%	9	naṇṇaṇ
42	2.06%	9	nēṟṟu
43	2.08%	9	paṭṭam
44	2.11%	9	pirivu
45	2.14%	9	porulai
46	2.17%	9	vāyil
47	2.20%	9	appu
48	2.23%	9	ulakam
49	2.26%	9	eḷuti
50	2.29%	9	kalvi

Table 1: Word List – I.

'words' etc. This also leads to the conclusion that the word count is on the basis of #----#, where /#/ means the juncture.

Usually, the 3,855 words have been set as the most suitable limit for high-frequency words. The study presents evidence that counting the 3,855 most frequent words of Tamil as the high-frequency words is still the best decision for learners going on to academic study because they cover up to 60 percentage of in the overall vocabulary that used in the textbooks.

It has to be understood that what would be the words in this group first of all, before going into the details of the frequency of words. As has been noted, the classic list of high-frequency words is General Words List which contains 3,855 words families. In the most frequently occurring list of vocabularies, most of them are function words such as /oru/ 'one', /eṇṇu/ 'said as', /eṇṇa/ 'said as', /pala/ 'many', /āṇāl/ 'or'. The rest are a content word that is nouns, verbs, adjectives and adverbs. Older series of graded readers are based on this list.

One more question arises as to how stable are the high-frequency words. In other words, does one properly researched list of high-frequency words differ greatly from another? Frequency lists may disagree with each other about the frequency rank order of particular words but if the research is based on a well-designed corpus there is general agreement about what particular words should be included. The research on the General word list showed quite large overlap between it and more recent frequency counts. Replacing some of the words in the General Corpus with other words resulted in only 1% increase in coverage. It is important to remember that the 3,855 high-frequency words of Tamil consist of some words that have very high frequencies and some words that are only slightly more frequent than others are not in the list. The first 1,000 words cover about 25% and the second 1,000 about 15.55% of the running words in academic texts. When making a list of high-frequency words, both frequency and range must be considered. Range is measured by seeing how many different texts each particular word occurs in. A word with wide range occurs in many different texts.

The high-frequency words of the language are clearly so important that considerable time should be spent on them by teachers and learners. The words are a small enough group to enable most of them to get attention over the span of a long-term Tamil programme. This attention can be in the form of direct teaching, direct learning, incidental learning, and planned meetings with the words. The time spent on them is well justified by their frequency, coverage and range. In general, high-frequency words are so important that anything that teachers and learners can do to make sure they are learned is worth doing.

### Low-frequency words

The second group is the low-frequency words. Here, this group includes words like /tacai/ 'muscle', /taṭu/ 'prevent', /taṭai/ 'obstruction', /taṇi/ 'seperate', /talā/ 'each', /tavaṟa/ 'to slip', /tāva/ 'to jump', /tiya/ 'bad', /tira/ 'thoroughly', /tēta/ 'to find', /toḷa/ 'loss', /naṭi/ 'act', /narai/ 'grey hair', /nila/ 'good', /nila/ 'blue', /nēca/ 'lovely', /paṭu/ 'sleep', /paṇ/ 'song', /paṇi/ 'cloud', /parata/ 'national', /paṟi/ 'pick', /puṟa/ 'outside', /pūca/ 'to paint', /pēṇa/ 'to maintain', /maṟai/ 'hide', /māṟa/ 'to change', /muka/ 'face', /ravi/ 'person name', /vaṭai/ 'cutlet like snack', /vāṇa/ 'sky', /vāra/ 'weekly', /viṭa/ 'than', /vēta/ 'make up', /vaira/ 'diamond', /aṭa/ 'exclamatory word', /iṭa/ 'palce', /ōṭa/ 'to ran', /ṭai/ 'dye', /tara/ 'to give', /nā/ 'tongue', /paṟa/ 'to fly', /vaṇa/ 'sky', /vala/ 'richness'. They make up over 1-5% of the words in an academic text. There are thousands of them in the language, by far the biggest group of words. They include all the words that are not high-frequency words, not academic words and not technical words for a particular subject. They consist of technical words for other subject areas, proper nouns, words that almost got into the high-frequency list, and words that the vocabulary rarely meet in our use of the language.

Let us now look into literature in English at a longer text and a large collection of texts. Sutarsyah [7], Nation and Kennedy looked at a single

economics textbook to see what vocabulary would be needed to read the text. The textbook was 295,294 words long. The study shows the result that the academic word list used in the study was the University Word List [8]. What should be clear from this study and from the text looked at earlier that a reasonably small number of words cover a lot of text. The study used an academic textbook corpus made up of a balance of science and arts texts totaling 11,479 running words in this corpus. There is a very large group of words that occur very infrequently and cover only a small proportion of any text.

Some low-frequency words are words of moderate frequency that did not manage to get into the high-frequency list. It is important to remember that the boundary between high-frequency and low-frequency vocabulary is an arbitrary one. Any of several thousand low-frequency words could be candidates for inclusion within the high-frequency list simply because their position on a rank frequency list which takes account of range is dependent on the nature of the corpus the list is based on. A different corpus would lead to a different ranking particularly among words on the boundary. This, however, should not be seen as a justification for large amounts of teaching time being spent on low-frequency words that in the seven thousand one hundred and ninety two (7192) words in all the level. Many low-frequency words are proper names. Approximately 4% of the running words in the Tamil textbooks are words like /kaṇṭupitikkappaṭuvataṅku/ 'to find out', /ceyttittākaḷiliruntu/ 'from newspaper', /terivittukkoḷkiṛōm/ 'we are informing', /tērtetukkappaṭṭār/ 'he have been elected', /tērtetukkappaṭṭāl/ 'if he have been elected', /niṛuttappaṭṭirukkumē/ 'to be stopped', /aṭukkappaṭṭirukkum/ 'to be arranged', /amaikkappaṭṭirukkum/ 'to be set up', /innaṇceyalukkellām/ 'these all good work, /eṭuttukkāṭṭukaḷuṭaṅ/ 'with example'.

As novels and newspapers, proper nouns are like technical words – they are of high-frequency in particular texts but not in others, their meaning is closely related to the message of the text, and they could not be sensibly pre-taught because their use in the text reveals their meaning. Before one reads a story, he/she do not need to learn the characters' names.

'One person's technical vocabulary is another person's low-frequency word.' This ancient vocabulary proverb makes the point that, beyond the high-frequency words of the language, people's vocabulary grows partly as a result of their jobs, interests and specializations. The technical vocabulary of our personal interests is important to us. To others, however, it is not important and from their point of view is just a collection of low-frequency words. This is true of second language learners also. If they belong to the same family of language, their understanding and sharing of cognate words will help them in learning, whereas, the non-cognate language speakers may not share the words and hence the difficulty level of learning increases.

Some low-frequency words are simply low-frequency words. That is, they are words that almost every language user rarely uses, for example: /vāra/ 'weekly', /viṭa/ 'more than', /vēṭa/ 'make up', /vaira/ 'Diamond', /aṭa/ 'exclamatory word', /iṭa/ 'Place', /ōṭa/ 'run', /ṭai/ 'dye', /tara/ 'to give', /nā/ 'tongue', /paṛa/ 'to fly', /vaṅa/ 'sky', /vaḷa/ 'richness' They may represent a rarely expressed idea; they may be similar in meaning to a much more frequent word; they may be marked as being old-fashioned, very formal, belonging to a particular dialect words.

## Methodology

The paper follows computational linguistic model with corpus for collection and analysis of the data. A computer assisted statistical

method of frequency analysis with a qualitative and quantitative approach is followed in the research.

The first step in an analysis of the vocabulary in a textbook was to create a corpus consisting of the contents of the textbook. For the present study, selected contents of the primary school Tamil textbooks prescribed by the Government of Tamil Nadu published in 2009 were converted into machine-readable Unicode format doc files so that they could be "read" by the computer program used to analyze them. The entire primary textbooks typed the contents into files using word processing software (Microsoft Word 2010).

Omitted from the textbook corpus were the credits, acknowledgements, introduction, Resources page explaining support materials, two pages of extra vocabulary consisting of days, months, number in the textbooks. The omitted pages were considered to contain supplementary material that not all teachers would be likely to use. And also the poems verses are omitted from the textbook corpus.

The study have used PERL programme to obtain the frequency and cumulative percentage of words in the corpus.

The following section gives the detailed study of vocabulary in each standard. It analyzes the importance of vocabulary learning in the primary classes and the load of vocabulary learning and accordingly the difficulty level.

Considering the size of the 1<sup>st</sup> standard textbook and representative of the Table 2 and Figure 1 reveals that the number of words required to cover a particular percentage in the text. In the first standard, it takes 22 words to reach 10% and 63 words to reach 20%, 120 words to reach 30%, 201 words to reach 40%, 299 words to reach 50%, 441 words to reach 60%, 636 words to reach 70%, 832 words to reach 80% and so

Sl. No.	No. of words	(%)
1	22	10%
2	63	20%
3	120	30%
4	201	40%
5	299	50%
6	441	60%
7	636	70%
8	832	80%
9	1028	90%
10	1223	100%

Table 2: Number of Words and their Cumulative Percentage of 1<sup>st</sup> Standard Tamil Textbook.

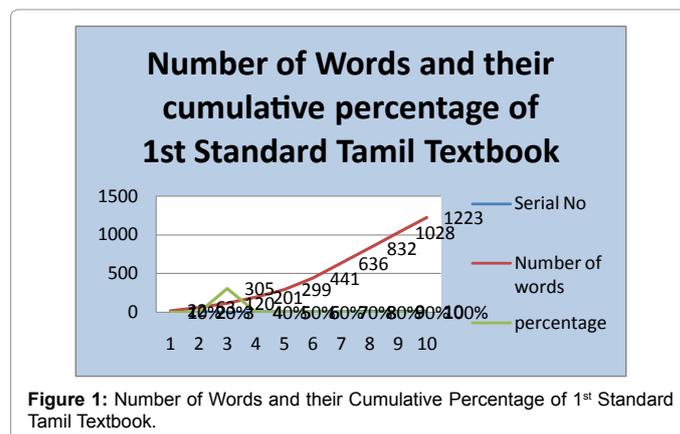


Figure 1: Number of Words and their Cumulative Percentage of 1<sup>st</sup> Standard Tamil Textbook.

on. The Table 2 shows that little less than 1k words cover 80% of the language. As one goes down the table one may observe that to reach a particular percentage it needs more and more distinct words in the textbook. The study show that the number of words increases more rapidly to reach higher percentages in the textbook.

If the study considers the second standard textbook and representative of the Table 3 and Figure 2 disclose that the number of words required to cover a particular percentage in the text. Here, unlike the first standard text one needs to learn 25 words to reach 10% and 70 words to reach 20%, 144 words to reach 30%, 255 words to reach 40%, 410 words to reach 50%, 600 words to reach 60%, 922 words to reach 70%, 1303 words to reach 80% and so on. The Table 3 shows that little less than 2k words cover 80% of the text. As one go down the table are may observe that to reach a particular percentage it needs more and more distinct words in the textbook. The study shows the

Sl. No.	No. of words	(%)
1	25	10%
2	70	20%
3	144	30%
4	255	40%
5	410	50%
6	600	60%
7	922	70%
8	1303	80%
9	1684	90%
10	2065	100%

Table 3: Number of Words and their Cumulative Percentage of 2<sup>nd</sup> Standard Tamil Textbook.

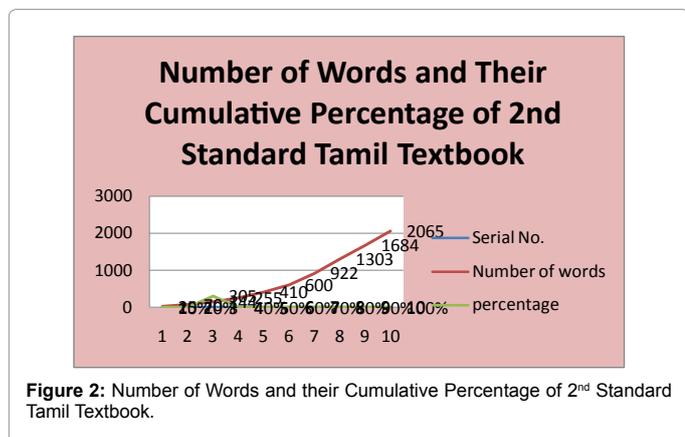


Figure 2: Number of Words and their Cumulative Percentage of 2<sup>nd</sup> Standard Tamil Textbook.

Sl. No.	No. of words	(%)
1	28	10%
2	88	20%
3	185	30%
4	320	40%
5	527	50%
6	809	60%
7	1243	70%
8	1806	80%
9	2370	90%
10	2934	100%

Table 4: Number of Words and their Cumulative Percentage of 3<sup>rd</sup> Standard Tamil Textbook.

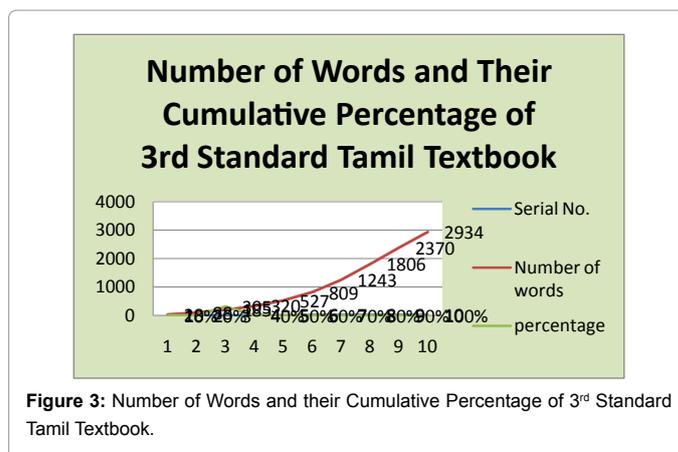


Figure 3: Number of Words and their Cumulative Percentage of 3<sup>rd</sup> Standard Tamil Textbook.

Sl. No.	No. of words	(%)
1	28	10%
2	95	20%
3	208	30%
4	377	40%
5	620	50%
6	922	60%
7	1442	70%
8	2046	80%
9	2650	90%
10	3253	100%

Table 5: Number of Words and their Cumulative Percentage of 4<sup>th</sup> Standard Tamil Textbook.

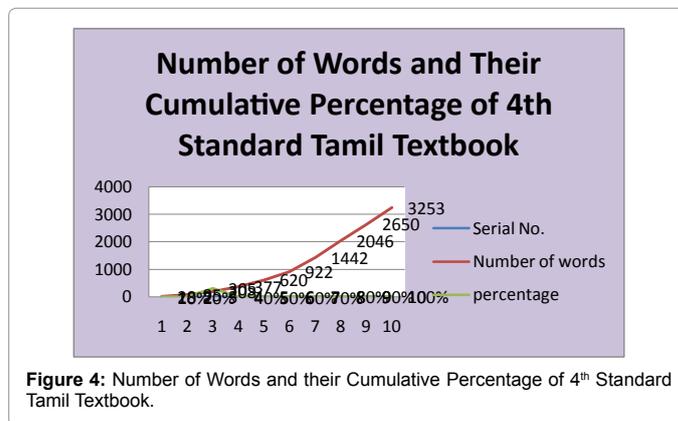


Figure 4: Number of Words and their Cumulative Percentage of 4<sup>th</sup> Standard Tamil Textbook.

number of words increase more rapidly to reach higher percentages in the textbook. If the study compare between first standard and second standard textbook one need to know little more than 1k words to reach 100% but whereas in second standard textbook one needs to learn little more than 2k word to reach 100%.

The Table 4 and Figure 3 represent the vocabularies of third standard textbook. In the third standard one does not need to learn more 28 words to reach 10% and 88 words to reach 20%, 185 words to reach 30%, 320 words to reach 40%, 527 words to reach 50%, 809 words to reach 60%, more than 1k words i.e. 1243 words to reach 70%, little less than 2k i.e. 1806 words to reach 80% and so on.

The Table 5 and Figure 4 represent the vocabulary of fourth standard textbook. In the fourth standard one needs to learn 28 words

to reach 10% and 95 words to reach 20%, 208 words to reach 30%, 377 words to reach 40%, 620 words to reach 50%, 922 words to reach 60%, 1442 words to reach 70%, 2046 words to reach 80% and so on. As one go down from the beginning to end of the Table 5 the study also observe that to reach 30 percentages it requires more distinct words in the textbook. The study focuses on number of words increase more rapidly to reach higher percentages in the textbook compared to 10% and 20% in the same table.

The above Table 6 and Figure 5 represents the vocabulary that is used in the fifth standard textbook. Here in this text, to master 10% of vocabulary one needs to know 46 words and to reach 20% one needs to learn 145 words, to reach 30% one requires 319 words, to reach 40% needed 596 words, to reach 50% one might learn 998 words, to reach 60% one might learn 1599 words, to reach 70% 2355 words one needs to know, to reach 80% one needs to know 3625 words so on. To be master to the whole vocabulary one need to know 6166 word. The

Sl. No.	No. of words	(%)
1	46	10%
2	145	20%
3	319	30%
4	596	40%
5	998	50%
6	1599	60%
7	2355	70%
8	3625	80%
9	4896	90%
10	6166	100%

Table 6: Number of Words and their Cumulative Percentage of 5<sup>th</sup> Standard Tamil Textbook.

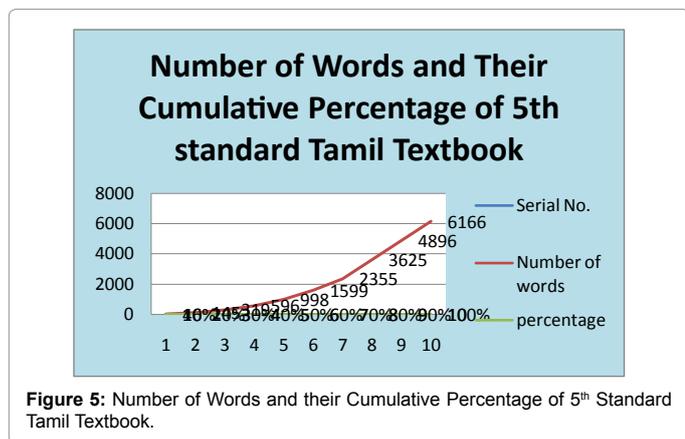


Figure 5: Number of Words and their Cumulative Percentage of 5<sup>th</sup> Standard Tamil Textbook.

study says that in every standard the number of words increase to reach higher percentages.

The above Table 7 and Figure 6 shows that the comparison of vocabulary from 1<sup>st</sup> standard to 5<sup>th</sup> standard textbook. One can observe that to reach 10% from 1<sup>st</sup> standard to 4<sup>th</sup> standard it takes only 22-28 words but in the 5<sup>th</sup> class it requires 46 words. Followed by that, to reach 20% 63-95 word from 1<sup>st</sup> standard to 4<sup>th</sup> standard but in 5<sup>th</sup> standard it requires 145 words. In the same way, if the study take 30% 1-4<sup>th</sup> standard takes 120-208 words, whereas, 5<sup>th</sup> standard textbook takes 319 words. Likewise, if the study go down one can see that the number of words increases more rapidly to reach higher percentages in 5<sup>th</sup> standard textbook not in the 1-4<sup>th</sup> standard textbooks. The table shows that the 5<sup>th</sup> standard students need to learn more words (to reach particular percentage) which are really extra burden for them. That is why syllabus for the textbooks needs to consider the corpus based syllabus creation for all the standards based on the frequency. It should also be taken into account that the teaching and learning of not only the high frequency words but the others also is very difficult since the vocabulary load is too much for the primary standards. The analysis shows that in the primary standards from 1<sup>st</sup> to 5<sup>th</sup> the learners are expected to acquire about 15,000 words, out of which known vocabulary or already learnt vocabulary is to be studied. Whatever be the reason these textbooks, in fact, increase the learning load altogether.

## Findings

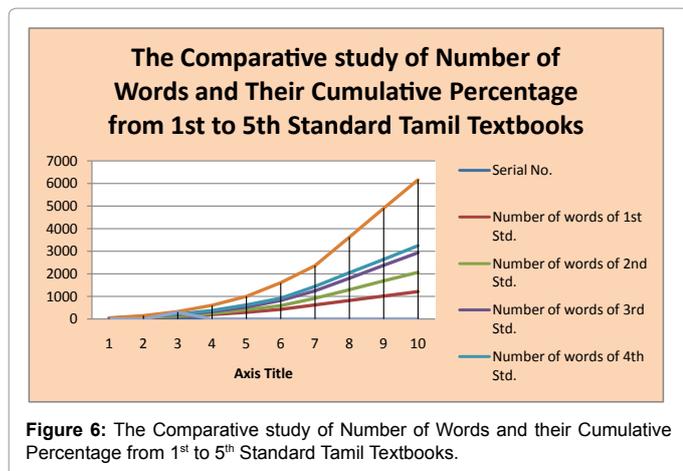
This observation is reinforced in the revised syllabus for primary school of Tamilnadu. This sort of quantitative information must be relatable to the complexity of inflectional morphology of the language. Since, Tamil is morphologically rich; it confirms the facts of the existing knowledge about Tamil language. Thus, the fact that the language has more word types, hence less word frequency is proved. The study finds in the above table that the lesser frequency word types are more in the language.

If the stock of vocabulary to be introduced in the first standard to help a child learn formal variety of language in his studies is to be roughly estimated it should be the size of about 500 words. The syllabus committee also had prescribed the word limit as 500 words only at first standard level. And there after a level wise increase of 350 new words. But in real increase is almost three- fold the consequences of which on the comprehensibility of the books by the students can be estimated.

An analysis of the textbook can show if vocabulary occurs frequently enough and is given enough repetitions over time to provide optimum vocabulary-learning conditions. The results can guide teachers in deciding how best to textbook the text with activities that will give learners exposure to target vocabulary that is not sufficiently presented in the textbook.

Sl. No.	No. of Words of 1 <sup>st</sup> Std.	No. of Words of 2 <sup>nd</sup> Std.	No. of Words of 3 <sup>rd</sup> Std.	No. of Words of 4 <sup>th</sup> Std.	No. of Words of 5 <sup>th</sup> Std.	Cumu. %
1	22	25	28	28	46	10%
2	63	70	88	95	145	20%
3	120	144	185	208	319	30%
4	201	255	320	377	596	40%
5	299	410	527	620	998	50%
6	441	600	809	922	1599	60%
7	636	922	1243	1442	2355	70%
8	832	1303	1806	2046	3625	80%
9	1028	1684	2370	2650	4896	90%
10	1223	2065	2934	3253	6166	100%

Table 7: The Comparative study of Number of Words and Their Cumulative Percentage from 1<sup>st</sup> to 5<sup>th</sup> Standard Tamil Textbooks.



This paper concentrated on the frequency of vocabulary in the textbooks. The vocabulary frequency is that aspect of vocabulary that has to do with the number of times, and therefore the duration of contact with a word, which facilitates its acquisition by enhancing the chances to draw learners' attention to it. The number of times the learners listens, speaks, reads, and writes a word has a lot of bearing on learning and developing a lexicon of his/her own. Therefore the frequency of the vocabulary is very important from the point of view of both teaching and learning.

The analysis shows that there is no gradual increment in the vocabulary across the standards through tables and graphs.

Furthermore the study has brought out an important factor that a root can be categorized by the functional endings attached to. Typologically, functional categories of each language are limited, which gives a clue to the justification of the lexical status in terms of categories. That is, to be precise, if an aspectual marker goes with a root, the composite morphological form is a verb and, similarly, if a plural marker goes with a root, the outcome is a noun. So it can posit the following hypothesis: Categorical Determination: Determine the categorical status of the root by the functional categories it goes with. It

is expected that there will be some reduction average in the ambiguity rate without considering other functionality in the local domain in which it appears. It also can be studied when it does the frequency study based on the part-of-speech tagged corpus looking into the context.

The study has concentrated on the number of Words and their Cumulative Percentage of Tamil Textbooks for all the standards, namely, from first standard to fifth standard. A comparative study of Number of Words and Their Cumulative Percentage from 1st to 5th Standard Tamil Textbooks is also developed to find out what exactly is the area where concentration should be given for developing vocabulary in the students.

## Conclusion

It is also very important to remember that the boundary between high-frequency and low-frequency vocabulary is an arbitrary one. Many of the low-frequency words could be included within the high-frequency list. A different corpus would lead to a different ranking particularly among words on the boundary. The study is, therefore, very important for the materials production from the point of view of selection of vocabulary for each of the lessons including the exercises.

## References

1. Rajaram S (1979) An Intensive Course in Tamil: Dialogues, Drills, Exercises, Vocabulary, Grammar, and Word Index. Central Institute of Indian Languages, Mysore.
2. Goulden R, Nation P, Read J (1990) How Large can a Receptive Vocabulary be? Applied Linguistics. Oxford University Press 11: 341-363.
3. (2008) Cre-A: Dictionary of Contemporary Tamil. Cre-A Publications, Chennai.
4. Nadaraja PN (1986) A Guide for Advanced Learners of Tamil. Central Institute of Indian Languages, Mysore.
5. Arokianathan S (2012) Spoken Tamil for Foreigners. A-Team Info Media Publishers Pvt. Limited, Chennai.
6. Uma Maheshwar Rao G, Thennarasu S (2007) A Course Material-book on Corpus Linguistics. University of Hyderabad, Hyderabad.
7. Sutarsyah C, Nation ISP, Kennedy G (1994) How Useful is EAP Vocabulary for ESP? A Corpus based Case Study. RELC Journal 25: 34-50.
8. Xue G, Nation ISP (1984) A University Word List. Language Learning and Communication 3: 215-229.