

Word Frequency in Language Teaching – A Case Study of Tamil Textbooks of Tamilnadu

L.R. Prem Kumar, M.A. Linguistics, M.A. Tamil, Ph.D. Linguistics

Lecturer-cum-Junior Research Officer

Central Institute of Indian Languages, Mysuru 570006, Karnataka, India

lrprem90@gmail.com

Abstract

This study on word frequencies deals with some words and word forms used in the Primary School Tamil Textbooks produced and prescribed by the Tamil Nadu Government published in 2009. Though the topic is an important one from the point of view of language teaching and learning, it is unfortunate that such a study on the most frequent words used in any textbook in Tamil has not been undertaken so far. As the learners go up in the levels of their primary education, the frequency of words learnt will find more uses in those classes to expand their vocabulary stock. The objective of the present study is to find out the effect of word frequency in teaching and learning. The primary data as a corpus for the present study was collected from the Tamil Textbooks produced in 2009. A computer assisted statistical method of frequency analysis with a qualitative and quantitative approach is followed in the study.

Keywords: Tamil, Language Teaching, word, vocabulary, Word Frequency, Textbook, Corpus Linguistics

1. Introduction

Frequency is one of the many important properties of a word. It is not an inherent property because it cannot be measured directly on the word using some operational definitions. It can be determined by counting occurrences of the word in a finite sample of text. Its relative value changes from text to text and its population value cannot be stated at all because there are no true populations in language (Orlov, Boroda and Nadarejšvili, 1982). Hence, the concept of word probability in language as propagated by Herdan (1960) has no empirical correlation. It is also true that a word has no definite frequency in all the texts. For example, a word like 'god' will have more occurrences in the subject like Philosophy and not in Physics. It would mean that the frequency of a word has to be defined in a particular way to suit the study. However, in a word analysis of textbooks definitely varies from the general frequency study as in the case of Linguistic Data Consortium for Indian Languages, CIIL, Mysuru, which has been creating a corpus of each of the twenty-two scheduled languages of India. Hence, frequency in the case of particular studies is to be well defined and demarcated.

The words can be divided into two, namely, lexical and functional. The most frequent words also fall into these two categories. There should be a conscious effort to make a distinction between these two; the question of focusing on either one of these two is also important. A word like ‘iTam’ in Tamil has two uses: as a lexical one it has the meaning of ‘place’ and as a functional one it is a locative case marker or post position used with animate nouns meaning ‘to’ or ‘with’. At this point of discussion, there arises a question as to which of the meaning or category to look into for teaching and practice. A solution to this kind of problems is also to be found out using the corpus.

2. Corpus for Language Teaching and Learning

The corpus-based approach to language education has gained importance over the past few decades, particularly since the mid of 1980s (Tony A. McEnery and R. Xiao, 2011). This is because corpus analysis can be enlightening ‘in virtually all branches of linguistics or language learning’ (Leech, 1997). Therefore, the teaching and learning of a language in the modern period, mainly, concentrate on corpus driven language teaching. This helps the teacher, materials producer and the learners as well in enhancing the teaching learning process. One of the strengths of corpus data lies in its empirical nature, which pools together the intuitions of a great number of speakers and makes linguistic analysis more objective (McEnery & Wilson, 2001). Obviously, corpora have been used extensively in nearly all branches of linguistics including lexicographic and lexical studies, grammar, language variation, contrastive and translation studies, diachronic studies, semantics, pragmatics, stylistics, sociolinguistics, discourse analysis, forensic linguistics, language pedagogy, etc.

3. Use of Frequency of Words in Language Teaching

Nation (1900) suggests that most frequent words can be so useful; each word requires a specific focus. Since words are the focus point in communication, frequently used words can find a place in the content to be taught or learnt as well as in communication practices. They are otherwise called ‘active vocabulary’. They need to be identified in developing all the basic skills, namely, listening, speaking, reading and writing and in developing higher order skills, like coherence and cohesion, creativity, etc., as well, where the words play a very important role. These words are also need to be directly studied. Most frequently used words can be taught through audio-lingual or Communicative methods, or a mixture of both. The key is to make learners understand the importance of these words and reinforce them at every opportunity.

4. Teaching Most Frequent Words

Like any other words for teaching and practice, the most frequent words can be drilled and practiced: pair-work; word games; repetition and substitution drills, etc. It is a good idea to get the learners to build their own list of most frequent words by adding words to a wall chart or making a list in the back of their notebooks, which they refer to regularly. This may be called a list of ‘words for active usage’.

The main problem with most frequent words is that it is difficult to interconnect the meaning because they do not have, sometimes, any concrete object. (e.g. the quotative particle /enru/) has a high frequency value but has no concrete meaning.) Direct translation or using in sentence is,

therefore, often needed to teach such most frequent words in a second language teaching situation. Whereas, in a first language or mother tongue teaching class such a situation may not arise where the meaning or meanings are to be explained by means of their spoken dialects, as opposed to ‘written Tamil’ or ‘standard spoken Tamil’, since Tamil is a diglossic language. As far as Tamil is concerned, in the schools language education is imparted in the written Tamil and outside the classroom the learners use the spoken Tamil. In such a case, the use of the most frequent words requires to be spoken forms used in active spoken communication to start with. Information is often best remembered, when we can use it within a visual context. For example, we may like to introduce the most frequent word /oru/. We may have to tell the learners what it means and drill it as an individual vocabulary item. When the learners know it well (i.e. they can understand to say and write it confidently), introduce it into simple sentences such as:

1. avan oru māṇavan
 he one student
 ‘He is a student.’

Teachers can help the learners get the meaning of this type of sentences by using flashcards, mimes or drawings, or they can look them up in a dictionary in the higher classes. The learners can then make up their own sentences. This technique uses most frequent word as a starting point and builds the context around them. The second approach is to use them in the conversation. Introducing different contexts and then teaching the most frequent words as an individual vocabulary item is very important. For example, if the most frequent word /paṭi/ ‘to study’ is to be taught, the following steps are to be taken up and introduced.

2. nī naṅrāka paṭi
 you well study
 ‘You study well’
3. nī paṭiyil ninru payanam ceyyātē
 you step stand travel do not do
 ‘You do not stand on the steps and travel’
4. avan conṇapatik kēl
 he tell listen
 ‘Listen as he tells’

In sentence (2), the word is used as a verb; in sentence (3), it is a noun and in (4), it is a post position. Therefore, it is very much necessary and important to teach the meaning of the word in different contexts so that the learners understand the usage and try to use them in their own sentences. The learners guess the meaning of /paṭi/ through the contexts. /paṭi/ is then drilled and practiced as an individual vocabulary item (e.g. using methods as suggested above). By repeating new most frequent words in different contexts, they are reinforced and learners are more likely to recognize them in future language uses also. This is an excellent way to increase confidence - an important factor for motivation. Because the learners immediately recognize the new most frequent

words and are able to work out most of the sentences through their knowledge of common most frequent word. These are two different approaches to teaching most frequent words.

5. Exercises and Tests

Giving exercises to practice most frequent words regularly is crucial. After full practice of the words, a test is to be conducted to understand their achievement level. The test, however, should allow the learners the opportunity to produce a sentence in context, using the most frequent words that they learnt. For example, if the new most frequent words are:

5. oru	‘one’
eṇru	‘said as’
vēṇṭum	‘want’
inta	‘this’
eṇra	‘said as’
allatu	‘or’
koṇṭu	‘with’
pala	‘many’
itu	‘it’

The learners must be tested on the meaning of the words individually and then be required to produce a sentence, such as

6. avan	kuṭai	oṇRaik	kaṭaiyil	vāṅkiṇāṇ.
He	umbrella	one	shop	brought
‘I brought one umbrella in a shop’.				

These exercises must be done both orally and written, so that they would be internalized easily. Less able learners could make smaller sentences.

Example

7. enakku	oru	paḷam	tā.
Me	one	fruit	give
‘Give me a fruit’.			

6. Use of Frequent Words for Syllabus Design and Materials Development

Corpora have been used extensively to provide more accurate descriptions of language use in this paper. Mindt (1996) finds that the use of grammatical structures in textbooks for teaching English differs considerably from the use of these structures in L1 English. He, further, observes that one common failure of English textbooks is that they teach ‘a kind of school English which does not seem to exist outside the foreign language classroom’. It is true of teaching Tamil as a second language, as well. Since a standard spoken Tamil is being used in many courses, the learners often find it difficult to communicate successfully with native speakers.

A simple yet important role of corpora in language education is to provide more realistic examples of language usage that reflect the complexities and shades of meanings of natural language. Absence of this has created a lacuna in the materials. In addition to the above, however, corpora may provide data, especially frequency data, which may further alter what is to be taught. However, the use of corpus for syllabus framing, materials production and language teaching or testing in the context of Indian languages in general and Tamil in particular has not been made use of. In this era of information communication technology (ICT), the use of corpus for the educational purposes presupposes that the language speaking community is in hard-hitting with the modern methods and technology.

For example, on the basis of a comparison of the frequencies of modal verbs, future time expressions and conditional clauses in the native English corpora and their grading in textbooks used widely in Germany, Mindt (1996) concludes that one problem with non-corpus-based syllabuses is that the order in which those items are taught in syllabuses ‘very often does not correspond to what one might reasonably expect from corpus data of spoken and written English’, arguing that syllabuses should be based on empirical evidence rather than tradition and intuition, with frequency of usage as a guide to priority for teaching (Mindt, 1996). While frequency is certainly not the only determinant of what to teach and in what order, it can indeed help to make learning more effective.

7. Frequencies of Part-of-Speech (POS) Tagged Words from Texts

List of words in the textbooks are not at all given in any form to help the learners or the teachers. Such a list of words to be used in the textbooks so prepared is also not given to the textbook writers. In such a context the writers depend on their own knowledge of the words to be included. If a corpus is used to prepare a textbook, the focus can be given where exactly the learners require words and the most frequent words to be taught in each of the classes in the research area can be well listed and the teaching will go smoothly and to reach the exact objective proposed.

The word in a text can be arranged in order of their frequency. This is most interesting when texts are compared in terms of their frequency lists along with their category. The following list of words cover the top 10% from the primary level of Tamil text-books:

The following table 1 contains 10% of vocabularies with POS tagged words for both first standard and second standard in the primary textbooks.

1 st Standard			2 nd Standard		
Cumul. %	Freq.	Tagged word	Cumul. %	Freq.	Tagged word
0.99%	19	e_lutu\VF	1.01%	39	e_lutu\VF
1.93%	18	ceyka\VF	1.92%	35	oru\QTF
2.76%	16	vāl\NN	2.63%	27	eṇṇa\PRP
3.49%	14	pār\VF	3.33%	27	eṇṇu\CONJ
4.22%	14	vā\VF	3.80%	18	kūru\VF
4.79%	11	pomma\NN	4.19%	15	irukkum\VF
5.26%	9	eṇṇa\PRP	4.58%	15	tāmarai\NN

5.73%	9	e <u>l</u> uttu\NN	4.94%	14	ce <u>n</u> ratu\VF
6.20%	9	pa <u>l</u> am\NN	5.31%	14	ā <u>c</u> ai\NN
6.61%	8	e <u>t</u> uttu\VNF	5.67%	14	ku <u>ṭṭ</u> i\ADJ
7.03%	8	mu <u>t</u> al\QTF	6.01%	13	vā <u>r</u> u <u>ṅ</u> ka <u>l</u> \VF
7.45%	8	u <u>ṅ</u> \PRP	6.35%	13	pā <u>ṭṭ</u> iyammā\NN
7.86%	8	co <u>l</u> lai\NN	6.68%	13	ku <u>r</u> uvi\NN
8.28%	8	pa <u>l</u> li\NN	7.02%	13	mu <u>y</u> al\NN
8.65%	7	kū <u>r</u> ac\VINF	7.33%	12	pā <u>r</u> \VF
9.01%	7	kū <u>r</u> u\VF	7.65%	12	va <u>r</u> ai\VF
9.38%	7	o <u>r</u> u\QTF	7.96%	12	vā\VF
9.74%	7	co <u>ṅ</u> ka <u>l</u> ai\NN	8.27%	12	it <u>u</u> \PRP
10.10%	7	pū <u>n</u> ai\NN	8.58%	12	e <u>ṅ</u> \PRP
			8.89%	12	nī\PRP
			9.21%	12	mā <u>m</u> pa <u>l</u> am\NN
			9.52%	12	nā <u>l</u> \NN
			9.83%	12	pu <u>ḷ</u> u\NN
			10.12%	11	ko <u>ṅ</u> ṭu\VNF

Table 1

The following table 2 contains 10% of vocabularies with Part-of-Speech (POS) tagged words for both second standard and third standard primary textbooks.

2 nd Standard			3 rd Standard		
Cumul. %	Freq.	Tagged word	Cumul. %	Freq.	Tagged word
1.01%	39	e <u>l</u> utu\VF	1.13%	64	o <u>r</u> u\QTF
1.92%	35	o <u>r</u> u\QTF	2.04%	52	e <u>ṅ</u> ru\CONJ
2.63%	27	e <u>ṅ</u> na\PRP	2.92%	50	e <u>ṅ</u> na\PRP
3.33%	27	e <u>ṅ</u> ru\CONJ	3.60%	39	e <u>l</u> utu\VF
3.80%	18	kū <u>r</u> u\VF	4.03%	24	e <u>ṅ</u> \PRP
4.19%	15	i <u>r</u> ukkum\VF	4.45%	24	ta <u>ṅ</u> \PRP
4.58%	15	tā <u>m</u> arai\NN	4.85%	23	va <u>n</u> atatu\VF
4.94%	14	ce <u>n</u> ratu\VF	5.26%	23	ka <u>t</u> ai\NN
5.31%	14	ā <u>c</u> ai\NN	5.64%	22	ē <u>ṅ</u> \PRP
5.67%	14	ku <u>ṭṭ</u> i\ADJ	6.03%	22	na <u>r</u> i\NN
6.01%	13	vā <u>r</u> u <u>ṅ</u> ka <u>l</u> \VF	6.40%	21	at <u>u</u> \PRP
6.35%	13	pā <u>ṭṭ</u> iyammā\NN	6.77%	21	nī\PRP
6.68%	13	ku <u>r</u> uvi\NN	7.10%	19	kū <u>r</u> u\VF
7.02%	13	mu <u>y</u> al\NN	7.42%	18	aru <u>k</u> il\PSP
7.33%	12	pā <u>r</u> \VF	7.72%	17	at <u>il</u> \PRP
7.65%	12	va <u>r</u> ai\VF	8.02%	17	mu <u>t</u> alai\NN

7.96%	12	vā\VF	8.30%	16	onru\QTF
8.27%	12	itu\PRP	8.58%	16	pōnra\PSP
8.58%	12	en\PRP	8.86%	16	nān\PRP
8.89%	12	nī\PRP	9.14%	16	āttukkuṭṭi\NN
9.21%	12	māmpaḷam\NN	9.43%	16	amutā\NN
9.52%	12	nā\NN	9.71%	16	palli\NN
9.83%	12	puḷu\NN	9.97%	15	koṇṭē\VNF
10.12%	11	koṇṭu\VNF	10.23%	15	irukkum\VF

Table 2

The following table 3 contains 10% of vocabularies with Part-of-Speech (POS) tagged words for both third standard and fourth standard primary textbooks.

3 rd Standard			4 th Standard		
Cumul. %	Freq.	Tagged word	Cumul. %	Freq.	Tagged word
1.13%	64	oru\QTF	1.19%	73	eḷutuka\VF
2.04%	52	enru\CONJ	2.34%	70	enṇa\PRP
2.92%	50	enṇa\PRP	3.14%	49	oru\QTF
3.60%	39	eḷutu\VF	3.83%	42	enru\CONJ
4.03%	24	en\PRP	4.38%	34	nān\PRP
4.45%	24	taṇ\PRP	4.86%	29	onru\QTF
4.85%	23	vantatu\VF	5.25%	24	valli\NN
5.26%	23	katai\NN	5.62%	23	kantaṇ\NN
5.64%	22	ēn\PRP	5.98%	22	kuruvi\NN
6.03%	22	nari\NN	6.33%	21	vēṇṭum\VF
6.40%	21	atu\PRP	6.67%	21	ēn\PRP
6.77%	21	nī\PRP	7.01%	21	nī\PRP
7.10%	19	kūru\VF	7.34%	20	poru\NN
7.42%	18	arukil\PSP	7.65%	19	viṭai\NN
7.72%	17	atil\PRP	7.95%	18	vā\VF
8.02%	17	mutalai\NN	8.22%	17	illai\VF
8.30%	16	onru\QTF	8.50%	17	payircikaḷ\NN
8.58%	16	pōnra\PSP	8.76%	16	pārttuṇṇu\VNF
8.86%	16	nān\PRP	9.02%	16	maram\NN
9.14%	16	āttukkuṭṭi\NN	9.27%	15	taruka\VF
9.43%	16	amutā\NN	9.51%	15	yār\PRP
9.71%	16	palli\NN	9.76%	15	en\PRP
9.97%	15	koṇṭē\VNF	10.00%	15	akarātiyaip\NN
10.23%	15	irukkum\VF			

Table 3

The following table 4 contains 10% of vocabularies with Part-of-Speech (POS) tagged words for both third standard and fourth standard primary textbooks.

4 th Standard			5 th Standard		
Cumul. %	Freq.	Tagged word	Cumul. %	Freq.	Tagged word
1.19%	73	eḷutuka\VF	0.91%	104	eḷutuka\VF
2.34%	70	eṇṇa\PRP	1.55%	73	eṇru\CONJ
3.14%	49	oru\QTF	2.09%	62	oru\QTF
3.83%	42	eṇru\CONJ	2.57%	55	kūruka\VF
4.38%	34	nāṇ\PRP	2.93%	41	vēṇṭum\VF
4.86%	29	oṇru\QTF	3.29%	41	eṇṇa\PRP
5.25%	24	vaḷḷi\NN	3.57%	33	makkaḷ\NN
5.62%	23	kantaṇ\NN	3.85%	32	nāṇ\PRP
5.98%	22	kuruvi\NN	4.13%	32	poruḷ\NN
6.33%	21	vēṇṭum\VF	4.40%	30	eṇ\PRP
6.67%	21	ēṇ\PRP	4.65%	29	nī\PRP
7.01%	21	nī\PRP	4.89%	27	nām\PRP
7.34%	20	poruḷ\NN	5.12%	27	nam\PRP
7.65%	19	viṭai\NN	5.34%	25	ceṇru\VNF
7.95%	18	vā\VF	5.56%	25	uḷḷa\VNF
8.22%	17	illai\VF	5.78%	25	pōṇra\PSP
8.50%	17	payiṛcikaḷ\NN	5.99%	24	eṇra\VNF
8.76%	16	pārttup\VNF	6.19%	23	uyir\NN
9.02%	16	maram\NN	6.39%	23	evvāru\ADV
9.27%	15	taruka\VF	6.59%	23	pala\ADJ
9.51%	15	yār\PRP	6.78%	22	uṇ\PRP
9.76%	15	eṇ\PRP	6.97%	22	col\NN
10.00%	15	akarātiyaip\NN	7.16%	21	atu\PRP
			7.34%	21	nāḷ\NN
			7.52%	20	yāvai\PRP
			7.69%	20	eḷuttukaḷ\NN
			7.87%	20	aṇiviyal\NN
			8.04%	20	ceyalkaḷ\NN
			8.22%	20	tamiḷ\NN
			8.38%	19	pārttup\VNF
			8.55%	19	eṇappaṭum\VF
			8.71%	19	avai\PRP
			8.88%	19	ivai\PRP
			9.05%	19	eḷuttu\NN
			9.21%	19	kaṇiṇi\NN

			9.37%	18	pirittu\VNF
			9.53%	18	vantu\VNF
			9.68%	18	viṅākkaḷukku\NN
			9.84%	18	ācīriyar\NN
			10.00%	18	ciṅkam\NN

Table 4

The above tables of the high frequency words among the 10% of the total vocabulary in the primary classes had taken for analysis give very important information that how much the materials producers and the teachers should concentrate on the inclusion of the most frequent vocabulary in the lessons and teaching them. Interestingly, the word **oru** ‘one adjectival form’ is the most frequent one whereas **ōr** ‘adjectival form of one’ which occurs before vowel beginning nouns is not that frequent. This would mean that the teaching of the variations in the use of **oru** and **ōr** need not be insisted upon in the use of them. Further, in the spoken also **ōr** is not in use, except in an educated and well planned platform speech.

The high frequent grammatical form of the verb is the verbal participle form, which is used abundantly in Tamil to join verbs to make complex sentences. Since this is very much in use both in the spoken and written Tamil variations, this had to be focused on in teaching grammar, rather than the teaching of grammar for sake of teaching it. Ultimately, this analysis brings out lots of information as to what should be the modifications necessary in the teaching and learning process.

Yet another grammatical form used very frequently is the hortative form of the verbs, for example, the form **eḷutuka** ‘write’, **taruka** ‘give’, etc. these are used in the exercises and in questions. However, these are not used in the common communication either in speech or in written. There are disadvantages also as explained above.

The tables given above for the high frequency of 1-10% can be analyzed to get what kinds of words are being taught in the first two standards. It is interesting and valid to discuss the information provided in this table.

7.1 High Frequency Words of 1% to 10%

7.1.1 Nouns

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word
16	Vāl	15	Tāmarai	23	katai	24	vaḷḷi	33	makkaḷ
11	pomma i	14	Ācai	22	nari	23	kantaṅ	32	poruḷ
9	eḷuttu	13	pāṭṭiyamm ā	17	mutalai	22	kuruvi	23	uyir

9	paḷam	13	Kuruvi	16	āṭṭukkutt i	20	poruḷ	22	col
8	collai	13	Muyal	16	amutā	19	viṭai	21	nāḷ
8	paḷli	12	māmpaḷam	16	paḷli	17	payiṛcikaḷ	20	eḷuttukaḷ
7	coṛkaḷa i	12	nāḷ			16	maram	20	aṛiviyal
7	pūṇai	12	puḷu			15	akarāṭiyai p	20	ceyalkaḷ
								20	tamiḷ
								19	eḷuttu
								19	kaṇiṇi
								18	viṇākkalukk u
								18	āciryar
								18	ciṅkam
75		104		110		156		303	

Table 5

A comparison of the nouns introduced in the five standards show a view that the most common animals like pacu ‘cow’, kālai ‘ox’, or birds like kākam ‘crow’ etc. do not find a place in the first 1-10% frequent vocabulary. This indicates that efforts to introduce words which are nearer to the learner are not taken.

7.1.2. Finite Verbs

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
19	eḷutu	39	eḷutu	39	eḷutu	73	eḷutuka	104	eḷutuka
18	ceyka	18	kūru	23	vantatu	21	vēṇṭum	55	kūruka
14	pār	15	irukkum	19	kūru	18	vā	41	vēṇṭum
14	vā	14	ceṇṛatu	15	irukkum	17	illai	19	eṇappaṭum
7	kūru	13	vāruṅkaḷ			15	taruka		
		12	pār						
		12	varai						
		12	vā						
72		135		96		144		219	

Table 6

In the same way if finite verbs are taken into consideration, familiar and most used verbs like pō ‘go’ col ‘say’ are replaced by cel ‘go and kūru ‘say respectively. This study shows that the principle of known to unknown was not taken into consideration for the preparation of textbooks.

Furthermore, the learners, while coming to the class, bring the spoken variety of Tamil with verbs and nouns so close to their speech. This fact is also not taken care of while preparing the materials.

7.1.3 Non-finite Verbs

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
8	eḷuttu	11	koṇṭu	15	koṇṭē	16	pārttup	25	ceṅru
								25	uḷla
								24	eṅra
								19	pārttup
								18	pirittu
								18	vantu
8		11		15		16		129	

Table 7

As far as the finite verbs are concerned, the verbal participle forms such as eḷuttu, ‘having taken’ koṇṭu, ‘having had’, pārttu ‘having seen’, ceṅru ‘having gone’, pirittu ‘having separated’, vandu ‘having come’ and two relative participle forms, namely, uḷla ‘which is there’ and eṅra ‘which/who is called as’ only are used in the textbooks. As said earlier, the verb cel ‘to go’ is not a familiar one though it is introduced in the 5th standard. Most frequent one should be pōy the verbal participle form of the verb pō ‘to go’. This is true of the relative participle forms used in the 5th std. Textbook.

They are also uncommon as well as not known to the learners of these classes. A decision then has to be taken by the syllabus framers whether those familiar and known verbs and their other grammatical forms should be used in the textbook or not. The grammatical knowledge and its development also will be hampered.

7.1.4 Infinitive Verbs

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
7	kūra								

Table 14

It is a captivating truth that this infinitive form of the verb also does not find a place in the most frequent words list. This definitely raises a question as to why this much of difference in the use of grammatical forms which are necessary for story telling or narration and conversation in general. The grammatical forms to be introduced in the primary classes are to be revisited in the light of the tables and statement made thereof.

7.1.5 Pronouns

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
9	eṇṇa	27	eṇṇa	50	eṇṇa	70	eṇṇa	41	eṇṇa
8	uṇ	12	itu	24	eṇ	34	nāṇ	32	nāṇ
		12	eṇ	24	taṇ	21	ēṇ	30	eṇ
		12	nī	22	ēṇ	21	nī	29	nī
				21	Atu	15	yār	27	nām
				21	Nī	15	eṇ	27	nam
				17	Atil			22	uṇ
				16	nāṇ			21	atu
								20	yāvai
								19	avai
								19	ivai
17		63		195		176		287	

Table 9

A glance through the table 9 showing the pronouns introduced, one will come to the conclusion that first person singular pronoun nāṇ ‘I’, nī ‘you’ etc. do not find a place in the teaching or learning. It is also captivating that all the eleven pronouns, namely nāṇ ‘I’, nām ‘We (Inclusive)’, nāṅkaḷ ‘We (Exclusive)’, nii ‘You’, avaṇ ‘He’, avaḷ ‘She’, avar ‘he-singular honorific’, avarkaḷ ‘they - human’, atu ‘it’, avai ‘They’, avaiḷ ‘They (Plural)’ do not find a place in the most 10% of frequency words. One can even say that the back bone of vocabulary for speech which is pronouns does not find proper attention.

7.1.6 Conjunctions

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
		27	eṇru	52	eṇru	42	eṇru	73	eṇru

Table 8

It is interesting that eṇru ‘that- quotative particle’ is the only conjunction word used in the textbooks except the first standard. This form has a very rich functional value as far as narration or story telling is concerned. But unfortunately this is not introduced in the first standard textbook. This shows that the learners will be deprived of the knowledge of this form used in conversation also.

7.1.7 Quantifiers

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
8	mutal	35	Oru	64	oru	49	oru	62	oru
7	oru			16	onru	29	onru		
15		35		80		78		62	

Table 10

The quantifiers oru ‘adjective of one’ and onru ‘one’ only find place in these textbooks. mutal ‘first’ is also not a very commonly used form. The more nearer form for the learners would be mutalaavatu ‘first. This grammatical form which is more productive is not used would mean that such a preliminary grammatical form necessary for the learners has not been taken into consideration while preparing the textbooks.

7.1.8 Post Positions

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
				18	arukil			25	pōnra
				16	pōnra				

Table 11

Only two postpositions, namely, arukil ‘near’ and pōnra ‘like that’ are used in the third and fifth standard textbooks. As discussed earlier, words like arukil ‘near, is purely a written variety high order word. This will not find a place in the spoken or written variety of the primary school learners. This also helps to take a new look at the vocabulary to be taught to the learners.

7.10.1.9 Adverbs

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
								23	evvāru

Table 12

No adverb has found a place in the 1-10% most frequent word list, though evvāru ‘how’ is also an adverb. As discussed this word does not find a use in the standard spoken Tamil also. More adverbs will be used in the common conversation or speech when the learners tell a story.

7.1.10 Adjectives

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word

		14	kuṭṭi					23	pala
--	--	----	-------	--	--	--	--	----	------

Table 13

The same kind of description and comment can be extended to this grammatical form too.

When the study goes into the details of the frequency from 11% to 20% the facts to some extent remains same. But there is some interesting finding which is very much essential for further studies.

7.2 High frequency Words of 11% to 20%

7.2.1 Nouns

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
6	kuñcu	11	coṛkaḷai	15	pāpu	15	eḷuttu	18	aṛivu
6	kiḷi	11	kutirai	13	kiḷi	15	iṭam	18	nampi
6	nāy	11	kiḷi	13	Nīr	15	nari	17	payiṛcikaḷ
5	viṇākkaḷ	11	Nūl	13	viṭai	14	āṇ	17	ilakkaṇam
5	eṛumpu	10	viṭai	12	Vikkal	13	ceyalpātukaḷ	17	añcal
5	eḷuttai	9	Kuppucāmi	12	Māmā	13	coṛkaḷai	16	akarātiyaip
5	veñku	9	kuṭṭi	11	coṛkaḷai	12	kāṭci	16	kutirai
5	ammā	9	vaṇakkam	11	paṛavaikaḷ	12	appā	15	ṭāktar
5	paṭṭu	9	aṇil	11	niṛam	10	kurañku	15	nāṭu
5	valli	9	Māmā	10	varikkutirai	10	tāttā	13	maṇṇaṇ
5	iṭam	8	kuḷuvil	10	Puli	10	nēram	13	aruḷ
5	kāl	8	Pāppā	9	kuḷuvil	10	kīrai	13	āṇṭu
5	kuḷu	8	Pañcu	9	kaṭal	9	pērunthu	13	āṭai
5	kōḷi	8	yāṇai	9	Col	8	teṇāḷirāmaṇ	12	amaiccarum
5	yāṇai	8	maṇi	9	Kalā	8	kutiraiyai	12	māṇavarkaḷ
5	eli	7	āṛraik	8	kuḷantaikaḷ	8	paṭattil	12	aṇaivarum
4	turattiyatu	7	Oruvar	8	Ācīriyar	8	vīṭṭil	12	kappal
4	vaṭivaṅkaḷai	7	kaṇṇaṇ	8	Ellām	8	kaṇṇaṇ	12	kumaṇṇaṇ
4	carikkaś	7	piṛanta	8	kurañku	8	pēccu	12	ammā
4	māmpaḷam	7	viṇāk	8	appu	8	tampi	12	aracē
4	kurañku	7	paḷam	8	naḷiṇi	8	col	12	mey
4	paṭaṅkaḷai	7	pūṇai	8	Mēkam	8	nīr	12	viḷā
		7	mīṇā	8	paḷam	8	pūṇai	11	kaṭikāram
		7	nari	8	vēru	8	eli	11	celvam
						8	kaṭai	11	pulavar

						7	kiṇarriṇ	11	akkā
						7	valliṇ	11	peyar
						7	aṇaivarum	11	kuṛi
						7	ciṛuvaṇ	11	mālai
						7	kāraṇam	10	irātākirusṇaṇ
						7	talaivar	10	paḷḷiyil
						7	māmaram	10	palkalaik
								10	viḷakkam
								10	kōṇaṅki
								10	tīpā
								10	nīr
								10	eṇ
								9	uyirmey
								9	viṇākkal
								9	nāṭṭiṇ
107		202		239		303		493	

Table 15

Out of these nouns if one analyses the animals and birds separately, the following facts will emerge.

7.2.1.1 Animals

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
6	nāy	11	kutirai	10	varikkutirai	15	nari	16	kutirai
5	eṇṇuppu	9	aṇil	10	Puli	10	kuraṅku	10	kōṇaṅki
5	yāṇai	8	yāṇai	8	kuraṅku	8	kutiraiyai		
5	eli	7	pūṇai			8	pūṇai		
4	kuraṅku	7	nari			8	eli		
25		42		28		49		26	

Table 16

The animal nari ‘fox’ which is available in almost all children’s stories is taught only in second and fourth standards. Likewise kuraṅku ‘monkey’ comes in first and third standards. yāṇai ‘elephant’ occurs in the first two standards. A cursory look at the table 16 shows the animals taught occurs only in the 11 to 20 % of the frequency.

7.2.1.2 Birds

As far as birds are concerned, the scene is not a pleasant one. Only two birds occur in the 20% frequency. It is not known why birds do not become a frequent word in the primary classes.

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
6	kili	11	kili	13	kili				
5	kōli								
11		11		13					

Table 17

The study shows that the importance of animals and birds is to be understood from the point of view of the age group as well as the use of them in the conversation or language use.

On the other hand, the verbal nouns used both in the 10% and the 20% frequency is also not good, in the sense that they are not used frequently in the common language use. However, the words like eḷuttu ‘writing letter’, payirci ‘exercise’, etc. seem to occur only in the exercises.

The general nouns are simple and are commonly used ones. But, unfortunately, the kinship terms which are necessary for the children are not fully introduced in the primary classes. Only ammā ‘mother’, appā ‘father’, māmā ‘uncle’, tāttā ‘grandfather’, and akkā ‘elder sister’ are introduced. It shows that the vocabulary selected is not confirmed with the other materials producers or the previous class textbooks, which would have helped to decide the vocabulary for the next classes.

7.2.2 Finite Verbs

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
5	celvōm	11	eṅṅratu	12	teriyumā	13	ceṅṅratu	17	viṭaiyaḷikka
5	vaṭṭamiṭu	10	vantatu	12	vēṅṅtum	12	pār	15	paḷakuka
5	vēṅṅtum	9	vēṅṅtum	11	eṅṅratu	11	nirappuka	14	paṭikka
4	eḷuttuvōm	9	ceyka	11	pār	10	kēṭṭatu	14	ākum
4	paṭippōm	8	paṭi	10	uḷḷatu	10	varai	12	nirappuka
4	vāṅkalām	7	cey	9	eṅṅrār	9	eṅṅrār	11	uṅṅtu
4	Pārāy			9	ceṅṅratu	9	vantatu	10	pāruṅkaḷ
				9	nirappu	9	ākum	10	ceṅṅraṅar
				8	eṅṅṅiṅāḷ	7	payanpaṭuttuka	10	eṅṅratu
				8	pāruṅkaḷ	7	pārttatu	10	illai
				8	iruntatu	7	ceṅṅraṅar	10	ceyka
				8	kūṅṅiyatu	7	muṭiyātu	9	kūṅṅiṅār
				8	ceytatu	7	uṅṅtu	9	aṅṅika
				7	koṅṅiruntatu			9	vā

				7	uruvāḱku				
				7	ōṭiyatu				
31		54		144		118		160	

Table 18

When the study looks into the use of finite verbs the following observations are made. In the first standard itself, the verb celvōm ‘let us go’ is used the commonly used verb in the normal use as well for the age group is poo ‘go’. In the same way, the following verbs which are not used in ordinary conversations need not have been used in the primary classes.

enṛatu	‘it said’
kūrīyatu	‘it said’
uḷḷatu	‘it is there’
enṛār	‘he said’
cenṛaṇar	‘they went’
viṭaiyalikka	‘give answer’

7.2.3 Nonfinite Verbs

There are other options for such verbs, which are not used in the common usages. The use of nonfinite verbs in all the four standards in the 11-20% frequency and even till 10% show that the verbal participle form of the verb in question is used. It is one of the techniques used to make longer and complex sentences in Tamil specially and in Dravidian in general. The following table 19 shows the nonfinite verbs used in the 20% frequency.

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
4	paṛantu	7	eṭuttu	10	ceṇru	13	uḷḷa	16	varum
4	varum	7	paṛantu	8	vanta	12	koṇṭu	15	paṭittup
				7	pārttu	9	pārttu	14	kōṭiṭṭa
				7	eṭuttu	9	pirittu	13	paṭittu
				7	paṛantu	9	kōṭiṭṭa	12	piṇvarum
						7	toṭarntu	12	koṇṭu
								11	cērttu
								11	ōṭi
								10	eṇṇum
								9	aṇintu
								9	eṭuttu
								9	kaṇṭu
								9	kūrī
								9	vanta

8		14		39		59		159	
---	--	----	--	----	--	----	--	-----	--

Table 19

The other point to be discussed is the use of relative participle for of the verbs. Only vanta ‘that which came’ is used in the texts, whereas the relative participle form kōṭiṭṭa the underlined is used in the fourth and fifth standard textbooks in the exercises portions only. This means that this grammatical form is not looked into while writing texts for the classes. As such the grammar teaching is to be revised for better performance of the learners.

7.2.4 Infinitive Verbs

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
-	-	-	-	-	-	-	-	-	-

Table 20

The textbooks did not care about the use of infinitive forms of verbs. In the 20% frequency there is no infinitive form of the verb used, whereas in the 10% frequency only one infinitive form of a single verb is used in the first standard, namely, kūra ‘to tell’. As discussed earlier the textbooks are, to a certain extent, defective in imparting use of different grammatical forms, which are essential for the good communication.

7.2.5 Pronouns

Yet another grammatical form which is not looked into properly for language education in the primary classes is the use of or introduction of pronouns. See the table 21 given below.

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
6	eṇakku	11	etu	12	itu	12	atu	16	eṇṇai
6	eṇ	10	nāṇ	9	yār	10	nām	15	atil
6	ēṇ	10	yār	8	nāṅkaḷ	10	itu	15	ataṇ
4	nāṇ	7	uṇakku	8	uṇ	8	eṇakku	14	itu
4	yār	7	etaṇku	7	eṅkaḷ	7	eṅkaḷ	13	ēṇ
		7	ataṇ			7	uṇ	12	nīṅkaḷ
						7	taṇ	12	avar
								12	taṇ
								12	tam
								10	nāṅkaḷ
								10	ivar
								10	yār
								9	uṅkaḷ

								9	ettaṇai
								9	avaṇ
								9	yātu
26		52		44		61		187	

Table 21

Very importantly, the use of pronouns, which is very much necessary for communication is not looked into properly. This also exemplifies the fact that the texts produced did not care for grammatical forms to be taught in the primary classes.

First person pronouns nāṇ ‘I’ which should be the basic one does not appear in the 10% frequency in the first standard but occurs in the 20% frequency only. This should have been prioritized. nām ‘we (inclusive)’ and nānkaḷ ‘We (exclusive)’ appear in the 20% frequency.

nī ‘you’ occurs in the 10% frequency and nīnkaḷ ‘you –honorific singular and plural’ occurs only in the 20% frequency.

avaṇ ‘he’, avar ‘he-singular honorific’, atu ‘it’, avai ‘they- neuter’

It is interesting to know that this list of pronouns up to the stand V does not include two very important pronouns, namely, aval ‘she’ and avarkaḷ ‘they- human’. There is no valid reason for the absence of these two pronouns in the primary school textbooks.

7.2.6 Conjunctions

The other grammatical form concentrated on in the textbooks is the conjunctions or the sentence connectors. The three forms used are āṇāl ‘but’, eṇṇum ‘always’ and eṇṇa ‘said as’. Out of these three āṇāl ‘but’ is the sentence connector and the other two are phrase connectors. The absence of this conjunctive in the first four standards is a surplice to both materials producers and the teachers.

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
								15	āṇāl
								14	eṇṇum
								12	eṇṇa
								41	

Table 22

7.2.7 Adverbs and Adjectives

The study considers the frequency and use of adverbs and adjectives the wide gap in imparting knowledge of these two is not at all done in the primary classes. These are required to describe the nouns.

7.10.2.7.1 Adverbs

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
		7	eṅkē	11	aṅku	11	aṅkē	17	uṭaṅē
				8	mīṇṭum	11	eṅkē	12	iṅru
								10	eṅkum
								10	aṅku
								10	aṅkē
		7		19		22		59	

Table 23

It is to be noted that no proper adverb is used in these classes. The derivation of adverbs using the marker - āka is not at all taught, which is a mystery. Adverb of place aṅku, aṅkē ‘there’ only is used. In the case of adjectives also there are no adjectives used in the fifth standard textbook in the 20% frequency.

7.2.7.2 Adjectives

1 st STD		2 nd STD		3 rd STD		4 th STD		5 th STD	
Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word
6	civappu	10	viruppamāṇa	13	putiya	10	poruttamāṇa		
5	putiya			10	periya				
				8	pala				
11		10		31		10			

Table 24

A quick look at the table 24 shows that the derivation of adjectives with the marker - āṇa is not appropriately used to impart the grammatical feature of the adjective. viruppamāṇa ‘which you like’ in the 2nd standard and poruttamāṇa ‘appropriate’ in the 4th standard are the only derived adjectival forms used. The lessons should have got some more adjective used so that the learners would get more competence in the use of these forms.

8. Use of Corpora

Any discussion on the use of corpora in language pedagogy should address appropriately some objections to the use of corpora in language learning and teaching. While frequency and authenticity are often considered two of the most important advantages of using corpora, they are also the focus of criticism from language pedagogy researchers. For example, Cook (1998) argues

that corpus data impoverishes language learning by giving undue prominence to what is simply frequent at the expense of rarer but more effective or salient expressions. Widdowson (1990) argues that corpus data is authentic only in a very limited sense in that it is de-contextualized and must be re-contextualized in language teaching.

Widdowson (2003) makes a distinction between “genuineness” and “authenticity”, which are claimed to be the features of text as a product and discourse as a process respectively: corpora are genuine in that they comprise attested language use, but they are not authentic for language teaching because their contexts, as opposed to co-texts, have been deprived.

Examples from the study are

cel	‘go’
arukil	‘near’
yāvai	‘what (plural)’
uḷḷa	‘which is (there)’
kūru	‘tell’
yātu	‘what’
enratu	‘it said’

These are genuine ones but highly literary ones.

It is not to engage in the debate here but would like to draw the attention that the study of frequency of words definitely helps in the process of language teaching and learning.

9. Conclusion

Frequency of words and their use in Primary school Tamil Textbooks are studied elaborately in this article. The standard-wise frequency of most frequent 10% and 20% are analyzed. The frequency of words of different grammatical categories, such as nouns, verbs, conjunctions, adverbs, adjectives, pronouns, quantifiers, etc., are also analyzed and presented with their frequency. In addition to these, there are some verb forms, which are essential for instructions in the textbook but not used in common communication are also discussed in this chapter.

From the point of view of the teaching learning process the study on frequency is very important. However, this is not fully utilized in the present scenario of textbook production.

Corpus-based learning activities are nearly absent in teaching of Tamil as a first language in the primary classes. Of the various causes for this absence mentioned earlier, perhaps the most important are the access to appropriate corpus resources and the necessary training of teachers, which is one of the main priorities for future tasks of corpus linguists, if corpora are to be popularized to more general language teaching context.

It is recommended here that an integration of corpus studies in language teacher training is the only solution available for us to face this problem. It will at least take more time and perhaps a new generation of teachers, for corpora to find their way into the language classroom.

This study brings to light some grammatical forms and imparting the same indirectly has not been taken care of seriously. In the light of the above analysis, it may also be said that the textbooks do not cater to the needs of the minority children whose mother tongues are Telugu, Kannada, Malayalam, Urdu, Marathi and Sourashtri which are the six major minority languages and a number of tribal children who speak different tribal languages in the State of Tamil Nadu.

If these two tasks are accomplished, the corpora will not only revolutionize the teaching of grammar in the coming years, they will also fundamentally change the modus operandi of language education.

=====
Colophon: I am very much indebted to Prof. N. Nadaraja Pillai, former Professor-cum-Deputy Director, CIIL, Mysuru, India for his valuable suggestions and guidance in writing this article.
=====

References

Cook, G. 1998. 'The Uses of Reality: A Reply to Ronald Carter'. *ELT Journal*, Vol. 52, pp. 57-63.
Herdan, G. 1960. *Type, Token Mathematics: A Textbook of Mathematical Linguistics*. Den Haag: Mouton.
Leech, G. 1997. 'Teaching and language corpora: a convergence'. in A. Wichmann, S. Fligelstone, A. McEnery and G. Knowles (eds.) *Teaching and Language Corpora*, London: Longman. pp. 1-23.
Leech, G. 1997. *Teaching and Language Corpora*. New York: Addison Wesley Longman.
Mindt, D. 1996. 'English corpus linguistics and the foreign language teaching syllabus' in J. Thomas and M. Short (eds.) *Using Corpora for Language Research*. London: Longman. pp. 232-247
Nation, I.S.P. 1990. *Teaching and Learning Vocabulary*. New York: Newbury House.
Orlov, J.K., Boroda, M.G., & Nadarejvili, I.S. 1982. *Sprache, Text, Kunst. Quantitative Analysen*. Bochum: Brockmeyer.
Widdowson, H. 1990. *Aspects of Language Teaching*. London: Oxford University Press.
..... 2003. *Defining Issues in English Language Teaching*. London: Oxford University Press.

Notes

Abbreviations:

ADJ.	-	Adjective
ADV.	-	Adverb
CONJ.	-	Conjunction
Cumul.	-	Cumulative Frequency
DMD.	-	Demonstrative
e.g.	-	for example
Freq.	-	Frequency
INT.	-	Interrogative Marker

NN.	-	Noun
PAR.	-	Particle
PRP.	-	Pronoun
PSP.	-	Post Position
QTF.	-	Quantifier
V.	-	Verb
VF.	-	Finite Verb
VINF.	-	Infinitive Verb
VNF.	-	Non-finite Verb



Dr. L.R. Prem Kumar
Lecturer-cum-Junior Research Officer
Central Institute of Indian Languages, Mysuru, India
lrprem90@gmail.com