

**Teaching Tamil Language through the Standard
Spoken Tamil Corpus Data Bank: A Study (CRP6/04SL)
Dr Seetha Lakshmi, ALC & Dr Vanithamani Saravanan, CRPP
National Institute of Education, Singapore**

The Corpus Data Approach can be used for learning about a language and for teaching the language. Sinclair (COBUILD Project, 1987) and Huston (2002) highlight that working on a corpus data provides a challenge for language teachers to use a lexical, rather than a grammatical, syllabus for language teaching. A lexical syllabus could offer a corpus mirroring the distribution of structures, word frequencies and phraseology that learners require in language learning. In view of the recent emphasis on the use of Spoken Tamil in Singapore classrooms, a systematic data bank would assist the teachers by providing spoken corpus data as an additional link language in the teaching of Tamil.

Our Tamil Corpus Data Bank, we believe, is a pioneering effort with respect to the Tamil language in Singapore. The Data Bank contains a Tamil alphabetical listing of a total of one million words compiled through sit-in observations and recording-based transcriptions of exchanges in primary school, secondary school and junior college classrooms, as well as resources gathered from the mass media. This paper reports on our second study on the Standard Spoken Tamil Corpus Data Bank.

Schiffman (2005) and Nadaraja Pillai (2005) have identified the early use of the Standard Spoken Tamil variety in limited contexts within the classroom. In this paper, we present an analysis of the frequency of specialized terms representing Standard Spoken Tamil used in three specific domains, namely:

1. A secondary school classroom discussion about a Tamil written literature text
 - A total of 9454 words were recorded. No English words were used.
2. A short interview with a film director broadcast over the national Tamil radio station
 - A total of 158 words were transcribed. Of these, 6 were English words and 1 was a combination of English and Tamil used basically to aid in the flow of the conversation.
3. A portion of a children's drama telecast on the national Tamil television station
 - A total of 449 words were used in this section relating to a family situation. English words, like 'taxi', were used where their Tamil equivalents were not commonly used in Singapore. Words and phrases of English-Tamil combination were also used but to a relatively small extent.

The aforementioned data were examined for the following aspects of language use:

- Letters of the Tamil alphabet used to begin the words, and
- The use of the 1st, 2nd and 3rd person pronouns.

On the whole, the letters used at the beginning of the Tamil words in all three data conformed to the phonetic rules governing word structure in the Tamil language. Tables 1 to 3 provide a listing and frequency of the letters used and not used at the onset of the words in the data examined. We also noted that when Written Tamil words were engaged in the oral language, certain letters which begin words underwent a phonic change. For example, in the

classroom data, 'mu' was pronounced as 'mo' in the word 'mudhal' (first). However, there were instances when the same phoneme, in similarly related words like 'mudhalil' (at first), did not undergo any phonic change.

The 1st, 2nd and 3rd person pronouns were generally used in the three contexts examined. However, whilst all three types of pronouns occurred more freely in the exchanges in the first and third data, a rather restricted usage of the 1st person pronoun was observed in the second data. (See Tables 4, 5 and 6.) This was probably due to the specific conventions governing an interview and the restrictions relating to radio broadcasting.

It is deducible from the analysis of our present study that the Singapore Tamil corpus data could be used as an additional resource for developing curriculum materials to promote the use of Spoken Tamil among Tamil students and later develop the literary Tamil for their educational success. As already noted, a systematic data bank would also provide spoken corpus data for use as an additional link language for students learning to use Tamil both in the classroom and in social contexts. This would, in turn, ensure that the Tamil language, especially the Spoken variety, continues to thrive in Singapore.

Reference books:

Bao Zhiming and Hong Huaqing. (2006). Diglossia and register variation in Singapore English. *World Englishes*, Vol 25, No 1, pp.105-114.

David Deterding and Low Ed Ling, (2005). The NIE corpus of spoken Singapore English. In *English in Singapore: Phonetic research on a Corpus*. (eds) David Deterding, Adam Brown and Low Ed Ling. Singapore: Mc Graw Hill. pp. 1-6.

Hunston. (2002). *Corpora in Applied Linguistics*., Cambridge Applied Linguistics.

Luke, A., Freebody, P., Cazden, C., and Lin, A., (2004). *Singapore pedagogy coding manual*. (Unpublished technical report). Singapore: National Institute of Education, Centre for Research in Pedagogy and Practice.

Ministry of Education. *Tamil Language Curriculum and Pedagogy Review Committee*. Singapore.

Norbert Schmitt and Bruce Dunham. (1999). Second Exploring native and non-native intuitions of word frequency *Language Research*, 15, 4(1999), pp. 389-411.

Seetha Lakshmi, Viniti Vaish, Gopinathan S., and Vanithamani Saravanan. (2005). *A Critical Review of the Tamil Language Syllabus and Recommendations for Syllabus Revision (CRP36/03SL)*.

Sinclair, John. (1987). *Collin COBUILD English Language Dictionary*.

Sinclair. John. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Soili Nokkonen(2006). The semantic variation of NEED TO in four recent British English corpora, International Journal of Corpus Linguistics, 11:1(2006), John Benjamins Publishing Company. pp. 29-71.

Vanithamani Saravanan, Seetha Lakshmi.(2004). An Examination of the use of Standard Spoken Tamil in the school and media domains in Singapore in order to establish SST as an additional resource for the teaching and learning of Tamil (CRPP Project): Code no. CRP 06/04SL.